Nearest Neighbour Continuum Percolation Model with Random Number of Connections

OMAR KHALIL

Master of Science

Department of Mathematics and Statistics

McGill University

Montreal,Quebec

2014-07-02

In Partial Fulfillment of the Requirements for the Degree of Master of Science

DEDICATION

This Thesis is dedicated to my mother and father

ACKNOWLEDGEMENTS

I would like to thank my parents for working tediously to make this possible, my sister Salma, my partner Dulcie and my supervisor Louigi Addario Berry for continued support

ABSTRACT

Percolation theory is the study of the behaviour of connected clusters in a random graph. Due not only to its contribution in the last five decades to numerous other fields, most notably materials science, but also to its standing as an interesting topic in theoretical probability in its own right, it continues to be a thriving area of mathematics. In this thesis, we discuss one model under the percolation framework, where each point connects to a random number of neighbours, and ask the usual questions of existence and uniqueness of infinite clusters. More specifically, we consider a Poisson process X in $\mathbb{R}^d$ with density 1. We connect each point in X to its B nearest neighbours, where B is a non-negative integer valued random variable. The parameter in the model is the distribution function of B. We introduce sufficient conditions for B such that for d large enough percolation occurs almost surely. Furthermore the infinite cluster will almost surely be unique.

.

ABRÉGÉ

La théorie de la percolation est l' étude du comportement de groupements reliés dans un graphe aléatoire. On peut attribuer son essor continu dans le monde des mathématiques non seulement à sa contribution des cinq dernières décennies à de nombreux domaines externes -notamment à la science des matériaux -mais également en tant que sujet d'intérêt de son propre chef. Dans cette thèse, nous discutons un modèle dans le cadre de la percolation, où chaque point est connecté à un nombre aléatoire de voisins, et posons les questions habituelles de l'existence de l'unicité de groupes infinis. Plus précisément, nous considérons un processus de Poisson X dans $\mathbb{R}^d$ avec densité 1. Nous connectons chaque point de X à ses B voisins immédiats, où B est une variable aléatoire à valeur entière et positive. Le paramètre dans le modèle est la fonction de répartition de B. Nous introduisons des conditions suffisantes sur B tel que pour D une percolation assez importante survient presque sûrement. En outre, l'amas infini sera presque sûrement unique

TABLE OF CONTENTS

CHAPTER 1
Introduction

1.1 Introduction

Beginning in the second half of the twentieth century and continuing strongly to this day, percolation theory has received considerable interest from mathematicians and physicists alike. As an area of theoretical probability, it is concerned with the formation of infinite clusters in random graphs under a certain rule that connect points to each other. In physics, chemistry and materials science, however, the term percolation is used to describe the study the movement and filtering of fluids through porous materials. The source of the name is a representative question dating back to a 1957 paper by Broadbent and Hammersley [1]. The question, stated informally, was the following: If a liquid is poured on top of some porous material, will it make its way to the bottom?

The mathematical model was a graph with vertex set $V = \{0, 1, \ldots, n\}^3$, representing points on the integer lattice in three dimensions. For each pair (x, y) such that $x, y \in V$ and $|x - y| = 1$ an edge is included between x and y with probability p, independently of any other pair of vertices in V. The question then becomes one of the presence of an open path from top to bottom. This is now called *bond* percolation with parameter p. A closely related model is that of *site* percolation with parameter p where, instead of edges, vertices are activated randomly and independently with probability p and only two vertices $x, y \in V$ which are both activated and are such

that $|x - y| = 1$ are joined by an edge. One could also consider *oriented* bond/site percolation where the edges in the graph are all oriented edges. The behaviour for large n was of primary interest, and mathematicians typically study the behaviour of infinite networks to gain insight into the behaviour of large but finite ones. In this setting the analogue of the initial question of the existence of a top to bottom path is the existence of a path of infinite length. The event that such a path exists is equivalent to the existence of an infinite connected component of points, since the former is implied by the latter and vice versa. If such a component exists we say that percolation occurs.

More precisely, fix an infinite graph $G = (V, E)$, and write $G_p = (V, E_p)$ for a bond percolation on G with parameter p. One basic object of study is the critical probability $p_c(G) = \inf\{p : P(G_p \text{ contains an infinite component}) > 0\}$. We say G has a *non-trivial percolation phase transition* if $0 < p_c(G) < 1$. Since the event that an infinite component exists is independent of any finite subset of G_p, it follows by the Kolmogorov zero-one law that, for any fixed $p \in [0, 1]$, the probability of percolation is either 0 or 1. This, combined with the fact that the probability percolation occurs is increasing as a function of p, ensures that the critical probability p_c is well defined for any graph G.

Naturally, one may pose the same questions for random graphs other than the one presented in the introductory example above, and also in dimensions other than 3. In this thesis we study sufficient conditions for percolation on *geometric* graphs. A geometric graph is a graph whose points live in some metric space and a random geometric graph is a geometric graph in which either the set of vertices, the set of

edges or both are random. When the graphs of interest live in a discrete geometric structure such as the lattice $\mathbb{Z}^d$, the study of the infinite clusters is referred to as discrete percolation, and indeed the first percolation model studied, described above, falls into this category. This thesis however is concerned with a model which falls under the second category of percolation settings, where the random graph lives in a continuous space. The analysis of the infinite clusters in this category is referred to as continuum percolation. Knowledge of the former is critical to the treatment of the latter, since many of the ideas carry over, and many of the results in continuum models are analogues of the results in their discrete counterparts. For a thorough treatment of both percolation models the reader is encouraged to refer to Grimmett [2] for an in depth discussion of discrete percolation and to Roy and Meester [3] for one on continuum percolation.

To give the reader an idea of the type of questions that continuum percolation could be applied to, we consider a useful explanatory example from Meester and Roy [3], which is that of rain falling on dry ground. When a raindrop falls on the ground, a wet circular patch is formed. As rain continues to fall the picture changes from wet islands on dry regions to dry islands inside wet regions. When such a drastic change takes place in the spatial structure this is referred to as a phase transition. The parameter in such a model is typically the density of the rain drops rather than time. As opposed to our introductory example, the dynamics cannot be studied in a discrete setting without assuming entirely unrealistic behaviour for the rain drops. Furthermore, though many of the results in continuum models have analogues that

hold in the discrete setting, many more have analogues which are either false or unknown in the discrete setting, and multiple results have no discrete analogues at all and vice versa. In addition, even for the first class of results whose analogues hold, extra technical complications arise when working in a continuum usually due to the topological setting or the dependency structure of the models.

Examples in the latter class include the non-triviality of phase transitions and the uniqueness of unbounded components. Examples in the second class of results, whose analogues are false in the discrete setting, include possible non-uniqueness of unbounded components, and some limit results in continuum fractal percolation. The third class of results in which no discrete analogue exists is large and includes high density results like compression and rarefaction, scaling properties and complete coverage results. Thus continuum models prove to be more than a simple extension of the discrete case, and many of them, including the one we study in this thesis, provide new insights into the possible behaviour of random graphs.

The RCM and Boolean models

To provide some context to the model discussed in the thesis it is relevant to introduce two of the continuum percolation models which have received much attention: The RCM and the Boolean models. For a comprehensive coverage of those models we refer to Meester and Roy [3]. In both of those models we have a Poisson point process X with positive density λ in $\mathbb{R}^d$, where $d \geq 1$. By this we mean that X is a random subset of $\mathbb{R}^d$ with the following two properties: first, the number of points of X in a bounded measurable set A follows a Poisson distribution with

mean $\lambda V(A)$ where V is the d-dimensional Lebesgue measure; second, for disjoint Borel sets $A, B \in \mathbb{R}^d$, $|A \cap X|$ and $|B \cap X|$ are independent. In the Boolean model we consider a sequence of independent identically distributed (i.i.d) non-negative random variables $r_1, r_2, \ldots$ which is independent of X and order the Poisson points according to some rule. We then place a closed ball of radius r_i around the i'th point of X. In an RCM we choose a function $g : \mathbb{R} \to [0, 1]$ and we connect points $x, y \in X$ with probability $g(|x - y|)$ independently of all other pairs of points. One of the important properties of both those models is the existence, for $d \geq 2$ and under some additional conditions, of a non-trivial phase transition, a concept which we tried to make intuitive with the example above.

This means that in the Boolean model there exists a critical density $0 < \lambda_c(r_1) < \infty$ which depends on the distribution function of r_1, such that for $\lambda < \lambda_c(r_1)$ the part of the space covered by balls consists almost surely of bounded connected components only, while for $\lambda > \lambda_c(r_1)$ the covered section consists of at least one unbounded component almost surely. The RCM model displays similar behavior with the critical density being a function of g. It also turns out that when percolation does occur, the unbounded component is unique in both the RCM and the Boolean case. Although there are bounds available for the critical densities (see, for example [3]) the exact values are not known in any of these models.

1.2 The $NN(d, k)$ and $NN(d, B)$ models

There are other natural ways to connect the Poisson points. A simple and intuitive method to connect them is given by the $NN(d, k)$ model which has been studied extensively and provides the motivation for the $NN(d, B)$ model that is the subject

of this thesis. NN stands for nearest neighbour, d describes the dimension and k is the parameter of the model. The connection rule in this model is to connect every point to its k nearest neighbours. In other words, the vertices of the $NN(d,k)$ model are the points X of a Poisson process in $\mathbb{R}^d$ and there is a directed edge from each point to its k nearest neighbours; the graph $NN(d,k)$ is almost surely, well defined since with probability 1, there are no points $x, y, z \in X$ for which $d(x,z) = d(y,z)$, where d is the standard Euclidean metric. Note that the density of the Poisson process plays no role in this model, since the connection rule is based on relative distance. It was shown by Häggström and Meester [4] that for d large enough, percolation occurs almost surely in the $NN(d,2)$ model.

The model we consider here we can call $NN(d,B)$, where again NN stands for nearest neighbour, d is the dimension of the space and B is a random variable taking values in the non-negative integers. In this model, we work in the same framework as the $NN(d,k)$ model but instead of connecting each point to its k nearest neighbours for some fixed k, the number of nearest neighbours each point connects to is random and drawn from the distribution of B independently of the sample of B used to determine the number of neighbours for any other point. More precisely we will start with a sequence $(B_i : i \geq 1)$ of i.i.d copies of B, and order the Poisson process points in some way $x_1, x_2, \ldots$. Each point x_i will then connect to its B_i nearest neighbours under standard Euclidean distance in $\mathbb{R}^d$. We will provide conditions on B under which percolation occurs almost surely for large enough dimension. Furthermore it

will turn out that under the same conditions the infinite cluster formed is almost surely unique.

CHAPTER 2
Infinite Cluster in High Dimensions

Definition 2.0.1. *From here on, unless otherwise stated, B is an almost surely bounded random variable whose support is a subset of the non-negative integers.*

Our main result is the following theorem.

Theorem 2.0.2. *If $E[B] > 2$ then there exists a positive integer $d_0 < \infty$ such that for all $d \in \mathbb{N}$, $d \geq d_0$ percolation occurs almost surely in the $NN(d, B)$ model.*

It suffices to prove the theorem under the additional assumption that B is almost surely bounded since $E[B] > 2$ if and only if $E[B\chi(B < n)] > 2$ for some $n \in \mathbb{N}$ where χ denotes the indicator random variable of an event. We provide a short outline of the proof to give the reader an intuitive idea of how the result is obtained. We take d large and assume there is a point at the origin. This assumption is easy to remove but simplifies the exposition. We then provide a sequential construction of a subset of the cluster containing the origin via a modified breadth-first search exploration. Furthermore we will consider only directed paths starting at the origin for simplicity. The breadth-first search exploration allows us to locally approximate the behavior of the sequential process using a branching process which, for a given B with $E[B] > 2$ and for d large enough, proves supercritical. The approximation relies on projecting the branching process on $\mathbb{R}^2$ and then making a coupling based comparison with oriented site percolation on the lattice $\mathcal{L} = \{(i, j) \in \mathbb{Z}^2 : i \geq 0, |j| \leq i, (i+j)/2 \in \mathbb{Z}\}$ with oriented edges from (i, j) to $(i + 1, j \pm 1)$. For d sufficiently large the oriented

site percolation will be supercritical which will imply the existence of an infinite connected component in $NN(d, B)$ via the coupling.

2.1 The Sequential Search and the Branching Process

We begin by mentioning and proving two basic but important lemmas, whose statements can be found in [4], and which are crucial for our arguments. Let $\{X_d : d \geq 1\}$ be a sequence of homogeneous Poisson processes such that for all $d \geq 1$, X_d is a Poisson process in $\mathbb{R}^d$ with rate $\lambda(d) = \frac{\Gamma(d/2+1)}{\pi^{d/2}}$, where Γ is the gamma function. In other words, the rate is chosen such that the expected number of points in a ball of radius 1 is 1. Fix some arbitrary positive integer k, and, for $d \geq 1$, let $Y_k(X_d)$ be the position in $\mathbb{R}^d$ of the k'th nearest neighbour to the origin in X_d.

Lemma 2.1.1. *As $d \to \infty$ we have*

$$|Y_k(X_d)| \to 1$$

in probability. For any d we have furthermore that, in polar co-ordinates writing $Y_k(X_d) = (R, \Theta)$, then the conditional distribution of Θ given that $R \in H$ where H is any Borel subset of $[0, \infty)$, is uniform over the $d-1$-sphere. That is to say, for a Borel subset A of the (d-1)-sphere S^{d-1}, $P(\Theta \in A | R \in H) = \int_A d_{S^{d-1}}V$ where $d_{S^{d-1}}V$ is the volume element of the (d-1)-sphere.

Proof. The first part of the lemma is seen by noting that for any $0 < \epsilon < 1$,

$$\begin{aligned} P(|Y_k - 1| > \epsilon) &= P(Y_k > 1 + \epsilon) + P(Y_k < 1 - \epsilon) \\ &= P(Q((1+\epsilon)^d)) < k) + P(Q((1-\epsilon)^d)) \geq k) \end{aligned}$$

where $Q(\lambda)$ is a Poisson process with rate λ, and this expression goes to 0 as d approaches infinity. For the second part of the lemma, consider a random element O chosen uniformly from S^{d-1}. By this we mean that if μ is the surface measure on S^{d-1} then O is distributed according to $\mu* = \mu/\mu(S^{d-1})$. Now let O' be the associated rotation operator. $O'(X_d)$ is still a Poisson process since the two conditions for a point process to be Poisson are still satisfied. Hence the distribution of Θ is unchanged by the rotation. However, note that the distribution of Θ in $O(X_d)$ is now uniform over the (d-1)-sphere since the space underwent a uniformly chosen rotation. We conclude that the distribution of Θ in X_d is thus uniform over the (d-1)-sphere. $\square$

Lemma 2.1.2. *Let S_1^d and S_2^d be balls in $\mathbb{R}^d$ of radii r_1 and $r_2 \in [0.9, 1.1]$. Suppose furthermore that for all d, S_1^d and S_2^d are centered at least 0.9 units apart. Then*

$$\frac{V(S_1^d \cap S_2^d)}{V(S_1^d)} \to 0$$

as $d \to \infty$.

Proof. Note that under the given conditions we can bound the fraction given above by the worst case scenario which is when the balls are centered 0.9 units apart, S_1^d has radius 0.9 and S_2^d has radius 1.1. This can be seen by noting that without loss of generality S_1^d can be assumed to have radius 0.9 by symmetry and since a uniform scaling of space leaves the ratio $\frac{V(S_1^d \cap S_2^d)}{V(S_1^d)}$ unchanged. Now S_2^d having the closest possible distance to S_1^d and having the largest possible radius maximizes the volume of the intersection of the two balls. Then the volume of the intersection is the sum of the volumes of two spherical caps.

Using the Pythagorean theorem and basic Euclidean geometry we obtain that the

height of the spherical cap arising from S_1^d is bounded below by 0.672 and the height of the spherical cap arising from S_2^d is bounded below by 0.4278. We utilize a formula for the volume of spherical caps in n dimensions as a function of the radius r and height h, found in many multi-variable calculus textbooks, for example in [8]: $V_n^{cap}(r,h) = \frac{\pi^{(n-1)/2}}{\Gamma(n/2+1/2)} \int_0^{cos^{-1}((r-h)/r)} sin^n(t)dt$. We also note that the volume of the n-sphere as a function of the radius r is $V_n(r) = \pi^{n/2}/\Gamma(n/2+1)r^n$. This yields after simplification,

$$\begin{aligned}
\frac{V(S_1^d \cap S_2^d)}{V(S_1^d)} &\leq \frac{V_d^{cap}(0.9, 0.672) + V_d^{cap}(1.1, 0.4278)}{V_d(0.9)} \\
&\leq \frac{\Gamma(d/2+1)}{\sqrt{\pi}\Gamma(d/2+1/2)} \left(\int_0^{1.32} \sin^d(t)dt + (11/9))^d \int_0^{0.92} \sin^d(t)dt \right) \\
&\leq \frac{\Gamma(d/2+1)}{\sqrt{\pi}\Gamma(d/2+1/2)} \left((1.32)\sin^d(1.32) + (0.92)(11/9\sin(0.92))^d \right)
\end{aligned}$$

where the second inequality follows from $\cos^{-1}((0.9-0.672)/0.9) < 1.32$ and $\cos^{-1}((1.1-0.4278)/1.1) < 0.92$. The final expression goes to 0 as n approaches infinity, since $\sin(1.32)$ and $11/9\sin(0.92)$ are both strictly less than 1. □

We now formally describe the sequential construction of our cluster containing 0 in dimension d. We consider the Poisson process X_d and analyse $X_d' = X_d \cup v^{\emptyset}$ where $v^{\emptyset}$ is the origin. We will need a short lemma.

For d large enough, under the connection rule introduced in the $NN(d, B)$ section of chapter 1, where each point x in the Poisson process is assigned an independent copy of B, B^x. Consider the graph $G(X_d)$ with vertex set X_d, and an edge between two points $x, y \in X_d$, if y is one of the B^x neighbours of x or x is one of the B^y nearest neighbours of y.

Definition 2.1.3. *Let X be a Poisson process in $\mathbb{R}^d$, for some positive integer d. Then by $G(X)$ we mean the graph whose vertices are X and whose edge set is determined by the rule described above. Furthermore, assign $v^\emptyset$ another independent copy of B, which we call $B^\emptyset$. Then by $C(X)$ we mean the largest connected subset of the vertices of $G(X \cup v^\emptyset)$ containing $v^\emptyset$.*

Lemma 2.1.4.

$$\begin{aligned} & P(\textit{There is an infinite component in } G(X_d)) \\ = \; & P(\textit{There is an infinite component in } G(X'_d)) \end{aligned}$$

Proof. We will see that the degree of any vertex in the graph is almost surely finite in both $G(X_d)$ and $G(X'_d)$, and so the addition or removal finitely many vertices cannot affect the existence of an infinite component. If, for positive integers R we let A_R be the region enclosed between the balls of radii R and $R-1$, and consider the random objects X_R^d, defined as the restriction of X_d to A_R, we can see that the X_R^d's are independent, by the independence property of disjoint sets in a Poisson process. Furthermore the event that an infinite component exists in $G(X_d)$ is a tail event of the X'_Rs. Using the Kolomogorov zero-one law, we can deduce that the probability on the left hand side is either 0 or 1, and the same argument can be applied to X'_d. Therefore the only way which the two sides of the equation can be unequal is when one of the sides is 1 and the other is 0. Adding a point cannot delete an infinite component if it exists, hence we need only consider the possibility that the RHS is 1 and the LHS is 0. Consider the set of points ω of the underlying probability space such that $G(X_d(\omega))$ has no infinite component but $G(X'_d(\omega))$ has an infinite

component. Such a set must have the property that in $G(X'_d)$ the origin is connected to infinitely many points. Since $E[B]$ is finite, the directed edges coming out of the origin are finite almost surely.

Hence, we must conclude that in this situation there exist infinitely many points v such that the origin is amongst the B^v nearest neighbours of v, where B^v is the associated integer random variable for v which determines how many of its nearest neighbours v connects to. Let V be the number of such points v and let $O_{x,y}$ denote the open ball centered at x with radius y. Let $|A|$ denote the cardinality of a set A, Furthermore, let $\chi(E)$ denote the indicator random variable of an event E and $B^{\emptyset}$ denote the copy of B associated with the origin.

$$\begin{aligned} E[V] &= E\left[\sum_{x\in X_d} \chi(B^x > |O_{x,||x||} \cap X_d|)\right] \\ &= \int_{R^d} P(B^x > |O_{x,||x||} \cap X_d|)\lambda(d)dx \\ &= \int_{R^d} P(B^{\emptyset} > |O_{0,||x||} \cap X_d|)\lambda(d)dx \\ &= E\left[\sum_{x\in X_d} \chi(B^{\emptyset} > |O_{0,||x||} \cap X_d|)\right] \\ &< E[B] < \infty \end{aligned}$$

where the third equality follows by translation invariance, and the second to last inequality follows since $E\left[\sum_{x\in K} \chi(B^{\emptyset} > |OB_{x,||x||} \cap K|)\right]$ is the expected number of points which are amongst the $B^{\emptyset}$ nearest neighbours of the origin, which is bounded above by $E[B]$. Hence V is finite almost surely. □

2.1.1 Sequential Search Process

We now analyze X'_d. Recall that for each point in X'_d, we have an associated copy of B independent of the rest of the copies, starting with $B^\emptyset$ associated with the origin. We view the $B^\emptyset$ nearest neighbours $v^1, v^2, \ldots, v^{B^\emptyset}$ of $v^\emptyset$ sorted in increasing order of modulus as its ***children*** and $v^\emptyset$ as their ***parent.*** In general when we use the terms parent and child we mean that the children are the nearest neighbours to the parent (edges are oriented from parent to child). We then relabel the children's random variables as $B^1, \ldots, B^{B^\emptyset}$. These points are said to belong to generation 1. Then for each $1 \leq i \leq B^\emptyset$ in turn, we label the $B^i - C_i$ nearest neighbours of v^i, $v^{i1}, v^{i2}, \ldots, v^{i(B^i - C_i)}$, again in order of increasing modulus (where C_i is the number of points amongst v^i's first B^i neighbours which have been encountered previously in the search process). We then relabel the random variables associated with v^i's children $B^{i1}, B^{i2}, \ldots, B^{i(B^i - C_i)}$. These vertices are then said to belong to generation 2.

In general, if construction up to generation n is complete then each vertex in generation n has a label L which is a string of n positive integers which is also common to its random variable B^L. We order the strings in lexicographic order. For each string L we search amongst the B^L nearest neighbours of v^L for points which have not been previously encountered, and label them $v^{L1}, \ldots, v^{L(B_L - C_L)}$ in increasing order of modulus, (where C_L is the number of points amongst v^L's nearest B_L neighbours which have been encountered previously in the search process). Completing this process for every string L in generation n yields generation $n+1$. By drawing edges between a point and its children as defined above, we obtain a graph whose vertex

set is a subset of the Poisson process.

Definition 2.1.5. *Let P be a Poisson Process in $\mathbb{R}^d$. SSP(P) is defined as the vertex set of the graph obtained by applying the sequential search procedure described above to $P \cup v^{\emptyset}$, where $v^{\emptyset}$ is the origin.*

Note that $SSP(X'_d)$ is a subset of the largest connected component containing the origin, $C(X_d)$. The dependence of $SSP(P)$ on the marks B_v is implicit. The next corollary follows directly.

Corollary 2.1.6.

$$P(|C(X_d)| = \infty) \geq P(|SSP(X_d)| = \infty)$$

For the remainder of our analysis we will thus focus on proving that $SSP(X'_d)$ is infinite with positive probability. For fixed L, the distribution of the rank of the parent of v^L among the B^L nearest neighbours depends sensitively on the distance between v^L and its parent, and this effect does not vanish as $d \to \infty$.

For this reason we study the distribution of the scaled distance as $d \to \infty$. For any point $x \in \mathbb{R}^d$ we consider the true Euclidean distance to the k'th nearest neighbour of x in X_d, which we will denote by $Y_k(x, X_d)$. Hence $Y_k(x, X_d) = \inf\{r; |B(x, r) \cap X_d| \geq k\}$ where $B(x, r)$ is the open ball of radius r centered at x. For $d \geq 1$, let $Z_k(x, X_d) = (Y_k(x, X_d) - 1) \cdot d$ where d is the dimension. We call $Z(x, X_d)$ the ***scaled distance*** from x to its k'th nearest neighbour in X_d. We will also define Z_k as a random variable whose distribution is the limit as d approaches

infinity of the distribution of $Z_k(x, X_d)$, which we will see exists.

Lemma 2.1.7. *Let $x \in \mathbb{R}$. Then as $d \to \infty$*

$$\{Z_k(x, X_d)\}_{k\geq 1} \to \{P_k\}_{k\geq 1}$$

in distribution, where $\{P_k\}_{k\geq 1}$ is an inhomogeneous Poisson process on $\mathbb{R}$ with rate $\lambda(t) = e^t$

Proof. Let $x \in \mathbb{R}$. We prove the convergence of the finite-dimensional distributions $P(Z_1(x, X_d) \leq z_1, Z_2(x, X_d) \leq z_2, \ldots, Z_n \leq z_n)$, where $n \in \mathbb{N}$ and $z_1 < z_2 < \cdots < z_n \in \mathbb{R}$. Note that if $l < m$ and $x > y$ then $P(Z_l(x, X_d) \leq x, Z_m(x, X_d) \leq y) = P(Z_m(x, X^d) \leq y)$. Thus the condition $z_1 < z_2 < \cdots < z_n$ is without loss of generality. Let $Q(\lambda)$ denote a Poisson distribution with parameter λ. Whenever $Q(\lambda)$ is mentioned it is referring to the same random variable. Furthermore, let $B_d(x, r)$ denote the ball in $\mathbb{R}^d$ with centre x and radius r.

$$\begin{aligned}
& P(Z_1(x, X_d) \leq z_1, Z_2(x, X_d) \leq z_2, \ldots, Z_n \leq z_n) \\
= \; & P(Y_1(x, X_d) \leq z_1/d + 1, Y_2(x, X_d) \leq z_2/d + 1, \ldots, Y_n(x, X_d) \leq z_n/d + 1) \\
= \; & \prod_{k=1}^{n} P\left(Y_k(x, X_d) \leq z_k/d + 1 \middle| \bigcap_{j=1}^{k-1} Y_j(x, X_d) \leq z_j/d + 1 \right)
\end{aligned}$$

by the chain rule for conditional probability. Now for any fixed k we have

$$\begin{aligned} & P\left(Y_k(x, X_d) \leq z_k/d+1 \middle| \bigcap_{j=1}^{k-1} Y_j(x, X_d) \leq z_j/d+1\right) \\ = \quad & P\left(|B_d(x, z_k/d+1) \cap X_d| \geq k \middle| |B_d(x, z_{k-1}/d+1) \cap X_d| \geq k-1\right) \end{aligned}$$

since the k nearest neighbour of x having distance less than $z_k/d+1$ from x, is precisely the event that the ball of radius $z_k/d+1$ centered at x contains k or more points of the Poisson process. Furthermore, this event is conditionally independent of the smaller balls given the number of points in $B_d(x, z_{k-1}/d+1) \cap X_d$. We can further condition on the number of points in $B_d(x, z_{k-1}/d+1)$ being exactly $k-1$, to obtain that the above expression is equal to:

$$\begin{aligned} & \Bigg[P\left(|B_d(x, z_k/d+1) \cap X_d| \geq k \middle| |B_d(x, z_{k-1}/d+1) \cap X_d| = k-1\right) \\ \times \quad & P(B_d(x, z_{k-1}/d+1) \cap X_d| = k-1 | B_d(x, z_{k-1}/d+1) \cap X_d| \geq k-1)\Bigg] \\ + \quad & P(B_d(x, z_{k-1}/d+1) \cap X_d| \geq k | B_d(x, z_{k-1}/d+1) \cap X_d| \geq k-1) \end{aligned}$$

Since if $B_d(x, z_{k-1}/d+1)$ contains k or more points then so does $B_d(x, z_k/d+1)$. The number of points in the region enclosed between the two balls is Poisson with rate $(z_k/d+1)^d - (z_{k-1}/d+1)^d$, and the number of points in $B_d(x, z_k/d+1)$ is Poisson

with rate $(z_{k-1}/d+1)^d$. Therefore we obtain that the above expression equals

$$\begin{aligned}
& \Big[P(Q((z_k/d+1)^d - (z_{k-1}/d+1)^d) \geq 1) \\
\times \quad & P(Q((z_{k-1}/d+1)^d) = k-1 | Q((z_{k-1}/d+1)^d) \geq k-1) \Big] \\
+ \quad & P(Q((z_{k-1}/d+1)^d) \geq k | Q((z_{k-1}/d+1)^d) \geq k-1) \\
\to \quad & P(Q(e^{z_k} - e^{z_{k-1}}) \geq 1) P(Q(e^{z_{k-1}}) = k-1 | Q(e^{z_{k-1}}) \geq k-1) \\
+ \quad & P(Q((e^{z_{k-1}}) \geq k | Q((e^{z_{k-1}}) \geq k-1)
\end{aligned}$$

as $d \to \infty$. Thus we have that

$$\begin{aligned}
& P(Z_1(x, X_d) \leq z_1, Z_2(x, X_d) \leq z_2, \ldots, Z_n(x, X_d) \leq z_n) \\
\to \quad & \prod_{k=1}^{n} \Bigg(P(Q(e^{z_k} - e^{z_{k-1}}) \geq 1) P(Q(e^{z_{k-1}}) = k-1 | Q(e^{z_{k-1}}) \geq k-1) \\
+ \quad & P(Q((e^{z_{k-1}}) \geq k | Q((e^{z_{k-1}}) \geq k-1) \Bigg) \\
= \quad & \prod_{k=1}^{n} P\left(P_k \leq z_k \middle| \bigcap_{j=1}^{k-1} P_j \leq z_j \right) \\
= \quad & P(P_1 \leq z_1, \ldots, P_n \leq z_n)
\end{aligned}$$

and we are done. □

Definition 2.1.8. *Let $\{Z_k\}_{k\geq 1}$ be an inhomogeneous Poisson process on $\mathbb{R}$ with rate $\lambda(t) = e^t$, and the atoms listed in increasing order.*

By the preceding lemma, $\{Z_k(v^\emptyset, X_d)\}_{k\geq 1}$ converges in distribution to $\{Z_k\}_{k\geq 1}$ as d approaches infinity. Here $v^\emptyset$ indicates the origin in $\mathbb{R}^d$. The result, however, also holds with $v^\emptyset$ replaced with any set of points $\{x_d\}_{k\geq 1}$ with $x_d \in \mathbb{R}^d$.

2.1.2 Defining the Branching Process

We can can approximate the behavior of the true sequential process search process in $\mathbb{R}^d$ using a multi-type branching process which we will denote G_d.

Definition 2.1.9. *Let $\mathcal{T} = \bigcup_{n=1}^{\infty} \mathbb{N}^n$, where $\mathbb{N}$ denotes the set of natural numbers and $\mathbb{N}^0$ is the empty string $\emptyset$. This is the set of all finite strings of positive integers as well as the empty string $\emptyset$. The Ulam-Harris tree is the infinite rooted tree with labelled node set $\mathcal{T}$ where $\emptyset$ is taken as the root, and with an edge joining $i_1 \dots i_k$ and $i_1, \dots, i_{k+1}$ for any $k \geq 0$ and $i_1, \dots, i_{k+1} \in \mathbb{N}$.*

For a node $v = i_1 \dots i_k$ its ***children*** are the strings $v1, v2, v3, \dots$ and we order them accordingly so that U is an infinite ordered rooted tree. The ***parent*** of $i_1 \dots i_k (k > 0)$ is $i_1 \dots i_{k-1}$. By generation of a node $v = i_1 \dots i_k$, denoted $|v|$, we will mean the length of the string which is k. Let $\mathcal{S}_d = \mathbb{R}^d \times \mathbb{R}^+ \times \{0, 1\}$. $\mathcal{S}_d$ is called the type space and the three components of its elements correspond to position, distance, as well as a binary mark indicating whether the point is "alive" or "dead", in the sense of being amongst an appropriate number of its parent's nearest neighbours. The random object G_d is an $\mathcal{S}$-valued function on the nodes of $\mathcal{T}$ drawn from the sample space $\mathcal{S}_d^{\mathcal{T}}$. $G_d(v)$ is called the type of node $v \in \mathcal{T}$. An element of $\mathcal{S}_d$ has the form $s = (w, x, b)$ where $w \in \mathbb{R}^d$, $x \in \mathbb{R}^+$ and $b \in \{0, 1\}$. For a node with type s we call x the ***position*** of the node and y its unscaled ***distance*** from its parent. A child will be referred to as a ***ghost*** if the third co-ordinate is 0.

For $v \in \mathcal{T}$, the types of its children are determined based on an associated distribution which depends on the type of v. Let $\mathcal{F}_d = \mathcal{S}_d^{\mathbb{N}}$ and for $f \in \mathcal{F}$ write

$f = (f_1, f_2, \dots)$. The children of an individual have types which are described by an element $f \in \mathcal{F}$ where f_1 is the type of the first child, f_2 is the type of the second child and so on, and the $f_k's$ are sorted in ascending order of distance. A kernel $P_s(df)$, from $\mathcal{S}_d$ to $\mathcal{F}$ specifies the reproduction mechanism; we will describe the kernels shortly. Thus, the governing distribution for the children of a node of type s is P_s and the corresponding expectation is denoted by E_s.

The family tree is produced in the usual way from the distributions. Given the family history up to generation n, individuals in the generation reproduce independently of each other with family distributions for each parent's type. The law for the branching process is constructed from the kernel $P_s(df)$ by using the Ionescu Tulcea extension theorem, which allows us to construct infinite dimensional distributions from a family of finite dimensional distributions if those finite dimensional distributions can be defined recursively through applications of probability kernels. This theorem can be found, for example, in [9]. The law here begins from an initial root of type $(v^{\emptyset}, 1, 1)$. In other words, the root corresponds to an initial individual at the origin in $\mathbb{R}^d$, with parent at distance 1 (though the root has no parent, some "parent distance" must be chosen so that the root has type within the type space; the value chosen is unimportant for future arguments), and the last co-ordinate indicates that the root is alive. Thus, to describe G_d fully, we need only provide the transition kernels, which depend only on the type. We describe the evolution of G_d directly, leaving the kernels implicit but their definition easily derivable from what follows.

We now describe the transition kernels $\{P_s : s \in \mathcal{S}\}$. Every vertex v is assigned an independent copy B_v of B, and an independent copy $X_{d,v}$ of X_d, the d dimensional Poisson process. For $i \geq 1$ write D_{vi} for the position of the i'th nearest neighbour of the origin in $X_{d,v}$.

For a ghost vertex v, with type $G_d(v) = (w, x, 0)$, for $k \geq 1$ the child vk has type $(w + D_{vk}, |D_{vk}|, 0)$. Note that all the children of ghost vertices are also ghost. Given a vertex $v \in \mathcal{T}$ with a non-ghost type $G_d(v) = (w, x, 1)$ its children's types are determined as follows. If $|D_{vB_v}| < x$, let $G_d(vk) = (w + D_{vk}, |D_{vk}|, 1)$ for $1 \leq k \leq B_v$, and $(w + D_{vk}, |D_{vk}|, 0)$ for $k > B_v$. Otherwise $|D_{vB_v}| \geq x$, let $G_d(vk) = (w + D_{vk}, |D_{vk}|, 1)$ for $1 \leq k \leq B_v - 1$, and $(w + D_{vk}, |D_{vk}|, 0)$ for $k > B_v - 1$. Intuitively, a node v of type $(w, x, 1)$ has distance x from its parent, so if $D_{vB_v} > x$, one edge leaving v attaches to an already explored node. In this case we explore only $B_v - 1$ rather than B_v children. For $n \geq 0$ we can define $\mathcal{F}_n^d = \sigma(G^d(v), |v| \leq n + 1)$, where σ denotes the sigma algebra generated by the random objects of interest, and again $|v|$ denotes the length of the string or, in other words, the generation of v. Now, note that for all v with $|v| \leq n$, B_v as well as the offsets of v's children is measurable with respect to $\mathcal{F}_n$. We are now in a position to give a formal definition of G_d.

Definition 2.1.10. *G_d is a random $\mathcal{S}_d$-valued function on $\mathcal{T}$, i.e, a random element of $\mathcal{S}_d^{\mathcal{T}}$ whose law is determined based on the transition kernels described above.*

Note that since B is bounded, we may equally view G_d as an $\mathcal{S}_d$-valued function on $\bigcup_{n=0}^{\infty}\{1, \ldots, m\}^n$, where $m = \sup\{k : \mathbb{P}(B \geq k) > 0\}$. The ***survival*** of G_d is the event that there does not exist a generation n such that all vertices in generation

n are ghosts. ***Extinction*** is the complement of this event. We will use survival and non-extinction interchangeably. The probability of survival depends only on the number of non-ghost children each point gives birth to, which depends on the distance aspect of the type. Therefore we can, for the analysis purposes of bounding the extinction probability, consider the multi-type branching process G'_d obtained from G_d by projection onto the second "distance" co-ordinate and scaling it, so G'_d has type space $\mathbb{R} \times \{0, 1\}$ and tracks only scaled distances.

Definition 2.1.11. *For $v \in \mathcal{T}$, define G'_d as follows. If $G_d(v) = (w, x, b)$ where $w \in \mathbb{R}^d$, $x \in \mathbb{R}, b \in \{0, 1\}$ then $G'_d(v) = ((x - 1) \cdot d, b)$.*

2.1.3 Survival of G'_d

Survival for G'_d is defined the same way as it is for G_d, that there exists no generation such that for all v in that generation the binary mark is 0. We also refer to those vertices whose binary mark is 0 as ghosts. We will also refer to the first component of the type in G'_d as the distance, which is a slight abuse of notation, but from the context it becomes clear whether we are talking about G'_d or G_d. The probability of survival/extinction for both G_d and G'_d are identical. Hence, we will analyse G'_d and give a full mathematical description of the number of non-ghost children of each object given the distance aspect of its type. Let $x \in \mathbb{R}$ and recall from section 2.1.1 the definitions of the scaled and unscaled distances $Z_k(x)$ and $Y_k(x)$ respectively.

We can now define a branching process G', which can be viewed as the $d \to \infty$ distributional limit of the processes G'_d. The type space is $\mathbb{R} \times \{0, 1\}$ and the transition kernels are such that the probability a point of type $(x, 1)$ gives birth to more than k

non-ghost points of distance less than or equal to y for $y < x$ is $P(Z_k(x) < y)P(B \geq k)$. We have already determined, in Lemma 2.2.3, that the law of $Z_k(x)$'s is that of a Poisson process on $\mathbb{R}$ with rate $\lambda(t) = e^t$. Thus, if $y > x$ this probability is $P(Z_k(x) < x)P(B \geq k) + P(Z_k(x) > x \cap Z_{k+1}(x) < y)P(B \geq k+1)$ since the encountered parent does not count as one of the children. Furthermore G' maintains the usual Galton-Watson independence structure and the offsets are independent for different sibling groups. This completely describes our transition kernels, and we now need to ask if the branching process is supercritical or not. To this end we define a family of measures M_x on $\mathbb{R}$.

Definition 2.1.12. *Let $x \in \mathbb{R}$, and $A \subset \mathbb{R}$. Define $M(A) = |i \in \mathbb{N} : G'_d(i) \in A \times \{1\}|$. Then $M_x(A) = E_x[M(A)]$.*

Thus $M_x(A)$ is the expected number of alive children a node of type $(x, 1)$ gives birth to, whose distance s fall within A. Of interest in studying supercriticality is the operator T defined as follows.

Definition 2.1.13. *For a measurable function $f : \mathbb{R} \to \mathbb{R}$, $f \in L_1(M_x)$ for all x,*

$$Tf(x) = \int_{\mathbb{R}} f(y) dM_x(y)$$

The following result is due to a corollary to Theorem 14.1 in [5].

Lemma 2.1.14. *If the greatest eigenvalue of the operator T is strictly greater than 1, then the probability that G'_d survives is non-zero.*

To analyze the operator spectrum it will be useful to calculate the densities of the family of measures M_x with respect to Lebesgue measure. Let $M(\leq y)$ be

the number of children within distance y of the parent. Also let $Q(\lambda)$ be a Poisson random variable with parameter λ.

Proposition 2.1.15. *In formal differential notation, there exists $m(x,y)$ such that $dM_x(y) = m(x,y)dy$ and*

$$m(x,y) = e^y \begin{cases} \sum_{k=0}^{\infty} \frac{e^{ky}e^{-e^y}}{(k)!} P(B > k) & \text{if } y < x \\ \sum_{k=1}^{\infty} \frac{e^{-e^y}}{k!} P(B > k)(e^{ky} - e^{kx}) & \text{if } y \geq x \end{cases}$$

Proof. Case 1: $y < x$.

$$\begin{aligned} M_x(\leq y] &= \sum_{k=1}^{\infty} P_x(M(\leq y) \geq k) \\ &= \sum_{k=1}^{\infty} P(Z_k(x) \leq y) P(B \geq k) \\ &= \sum_{k=1}^{\infty} P(Q(e^y) \geq k) P(B \geq k) \\ &= \sum_{k=1}^{\infty} \left(1 - \frac{\int_{e^y}^{\infty} t^{k-1}e^{-t}dt}{(k-1)!}\right) P(B \geq k) \\ &= E[B] - \sum_{k=0}^{\infty} \frac{\int_{e^y}^{\infty} t^k e^{-t}dt}{(k)!} P(B \geq k+1) \end{aligned}$$

Differentiating with respect to y we get

$$e^y \sum_{k=0}^{\infty} \frac{e^{ky}e^{-e^y}}{(k)!} P(B \geq k+1)$$

Case 2: $y \geq x$. Similarly we have

$$\begin{aligned} M_x(\leq y) &= \sum_{k=1}^{\infty} P_x(M(\leq y) \geq k) \\ &= \sum_{k=1}^{\infty} P(Z_k(x) \leq x) P(B \geq k) + \sum_{k=1}^{\infty} P(Z_k(x) > x \cap Z_{k+1}(x) < y) P(B \geq k+1) \end{aligned}$$

The first sum's evaluation is identical to what was done above, and does not involve y and so does not play when we differentiate. As for the second sum we have.

$$\begin{aligned} & \sum_{k=1}^{\infty} P(Z_k(x) > x \cap Z_{k+1}(x) < y) P(B \geq k+1) \\ = & \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} P(Q(e^x = j)) P(Q(e^y - e^x) \geq k+1-j) P(B \geq k+1) \\ = & \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \frac{e^{jx} e^{-e^x}}{j!} \left(1 - \frac{\int_{e^y - e^x}^{\infty} t^{k-j} e^{-t} dt}{(k-j)!} \right) P(B \geq k+1) \end{aligned}$$

Differentiating we get

$$\begin{aligned} & \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \frac{e^{jx} e^{-e^x}}{j!} e^y \frac{(e^y - e^x)^{k-j} e^{e^x - e^y}}{(k-j)!} P(B \geq k+1) \\ = & \sum_{k=1}^{\infty} \frac{e^y e^{-e^y}}{k!} P(B \geq k+1) \sum_{j=0}^{k-1} \binom{k}{j} e^{jx} (e^y - e^x)^{k-j} \\ = & \sum_{k=1}^{\infty} \frac{e^y e^{-e^y}}{k!} P(B \geq k+1)(e^{ky} - e^{kx}) \end{aligned}$$

Hence in formal differential notation $dM_x(y) = m(x,y)dy$ where

$$m(x,y) = e^y \begin{cases} \sum_{k=0}^{\infty} \frac{e^{ky}e^{-e^y}}{(k)!} P(B > k) & \text{if } y < x \\ \sum_{k=0}^{\infty} \frac{e^{-e^y}}{k!} P(B > k)(e^{ky} - e^{kx}) & \text{if } y \geq x \end{cases}$$

□

Proposition 2.1.16. *If $E[B] > 2$ then G' has a positive probability of survival.*

Proof. Using the function $f : y \mapsto 1$ we can verify that $Tf(x) > f(x)$ for all $x \in \mathbb{R}$ and thus the maximal eigenvalue for the operator is strictly greater than 1. G' thus has a non-zero probability of never going extinct. To see this note that for any fixed $x \in \mathbb{R}^d$, and rewriting $m(x,y)$ appropriately we can see that

$$\begin{aligned} \int m(x,y)dy &= \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{e^{(k+1)y}e^{-e^y}}{(k)!} P(B > k)dy - \int_x^{\infty} \sum_{k=0}^{\infty} e^y \frac{e^{-e^y}}{k!} P(B > k)e^{kx}dy \\ &= \sum_{k=0}^{\infty} \frac{P(B > k)}{k!} \int_0^{\infty} t^k e^{-t}dt - \sum_{k=0}^{\infty} \frac{P(B > k)e^{kx}}{k!} \int_{e^x}^{\infty} e^{-t}dt \\ &= \sum_{k=0}^{\infty} P(B > k) - \sum_{k=0}^{\infty} P(B > k) \frac{e^{kx}e^{-e^x}}{k!} \\ &= E[B] - \sum_{k=0}^{\infty} P(B > k) \frac{e^{kx}e^{-e^x}}{k!} \end{aligned}$$

which is greater than 1 if $2 < E[B] < \infty$. □

2.2 Existence of an Infinite Component

Definition 2.2.1. *Define the linear mapping*

$$L : \mathbb{R}^d \to \mathbb{R}^2$$

$$L(x_1, \ldots, x_d) = \sqrt{d}(x_1, x_2)$$

Recall the definition of the sequences of displacements $\{D_{vi}\}_{i\geq 1}$ for $v \in \mathcal{T}$ introduced in page 19, which we used to define the process G_d.

Definition 2.2.2. *Define a new process G_d^* on $\mathcal{T}$ with type space $\mathbb{R}^2 \times \mathbb{R}^+ \times \{0,1\}$ as follows. If $G_d(vi) = (w, x, b)$ and $D_{vi} = (a_1, \ldots, a_d)$ then let $G_d^*(vi) = (L(w), \sqrt{d}||(a_1, a_2)||_2, b)$.*

In other words, given the process G_d we can define a new process G_d^* whose type space is $\mathbb{R}^2 \times \mathbb{R}^+ \times \{0,1\}$ by mapping the position of the points in G_d to $\mathbb{R}^2$ using L and calculating the $\mathbb{R}^2$ Euclidean distance accordingly to obtain the distance aspect of the type. The binary mark is identical to what it is in G_d. We will now require the use of a result from Penrose [6].

Lemma 2.2.3. *Suppose $U = (U_1, U_2, \ldots, U_d)$ is uniformly distributed on $\{x \in \mathbb{R}||x| = 1\}$ according to the scaled spherical measure. Then, as d approaches infinity, the random vector $L(U)$ converges in distribution to the bivariate normal distribution $N(0, I)$ with zero mean and covariance matrix I.*

Corollary 2.2.4. *The offsets of G_d^* converge weakly to i.i.d with distribution $N(0, I)$ as d approaches infinity.*

This follows directly from lemmas 2.1.1 and 2.2.3. We will write G^* for the process started at the origin with initial type $((0,0), 1, 1)$ and G_y^* for the limiting process started at $y \in \mathbb{R}^2$, with initial type $(y, 1, 1)$. By limiting process we mean that the transition kernels of G^* are the limit as d approaches infinity of the transition kernels of G_d^*. In particular, writing (L^v, D^v, b^v) for the type of a node v, the offsets D^v are 2 dimensional $N(0, I)$ random variables. By Proposition 2.1.15, the process G' has a positive probability of non-extinction and hence there must exist d large enough such that G'_d (and thus G_d) has positive probability of survival as well. Since

the subtree of $\mathcal{T}$ induced by the set of alive nods is the same in G_d and in G_d^*, G_d^* must also have a uniformly positive probability of survival for d large enough. Since the transition kernels converge to those of G^*, G^* must have positive survival probability. Let q be the probability of extinction for G^*. The following lemma will be required. For a positive number M let $S_{ij}(M) = [M(i-\frac{1}{2}), M(i+\frac{1}{2})] \times [M(j-\frac{1}{2}), M(j+\frac{1}{2})]$. Then

Lemma 2.2.5. *Given $\epsilon > 0$ there exists a positive integer N_0, a positive number M, and a positive integer N such that given N points $x_1, x_2, \ldots, x_N \in S_{ij}(M)$, the probability that there exists $x_i, 1 \leq i \leq N$ such that generation N_0 of $G_{x_i}^*$ contains at least N points with position in $S_{i+1,j-1}$ and at least N points with position in $S_{i+1,j+1}$ exceeds $1-\epsilon$, where N is such that $1-q^N > \sqrt{1-\epsilon}$.*

Proof. For simplicity we give the proof for $i = j = 0$ although the argument is translation invariant. Conditional on non-extinction the (non-ghost) generation size in G^* tends to ∞ almost surely, due to, for example, theorem 11.2 in [5]. Pick N such that $1-q^N > \sqrt{1-\epsilon}$, and consider $x_1, x_2, \ldots x_N \in S_{0,0}$. With probability greater than $\sqrt{1-\epsilon}$ there exists $1 \leq i \leq N$ such that $G_{x_i}^*$ does not go extinct. Conditional on $G_{x_i}^*$'s survival we can pick N_1 such that the probability that generation N_1 has at least K individuals which are themselves root nodes of surviving subtrees exceeds $\sqrt{1-\epsilon}$, where K is to be specified shortly.

By the preceding sentence, $P(A_{N_1,K,M}) > \sqrt{1-\epsilon}$ for M sufficiently large. Let $A_{N_1,K,M}$ denote the event that at least K non-ghost individuals in generation N_1 are within distance M from x_i and are root nodes of surviving subtrees. Let $M' = \inf\{m : P(A_{N_1,K,m}) > \sqrt{1-\epsilon}\}$ and let $M = \lceil M' \rceil$. Let $N_2 = M^2$ and $N_0 = N_1 + N_2$.

For any vertex v in generation N_1 which is itself a root of a surviving subtree, the position of any of its descendants w with $|w| = N_0$ is the sum of N_2 $N(0, I)$ random variables. With such a Gaussian random walk, call it $(H_k, k \geq 0)$ starting from position $H_0 = y$ with $y \in \mathbb{R}^2$ and $|y - x_i| < M$, we can bound the probability that $H_{N_2} \in S_{1,-1}$ by noting that H_{N2} has distribution $N(y, N_2 I)$. Note that the maximum distance between any point in $S_{0,0}$ and any point in $S_{1,-1}$ is the distance between the 2 furthest corners at $M(-1/2, 1/2)$ and $M(3/2, -3/2) = 2M\sqrt{(2)}$. The furthest possible distance between $x \in S_{0,0}$ and y is M, hence, by the triangle inequality, $|z - y| \leq M(1 + 2\sqrt{(2)}) < 4M$. Let $||y - S_{0,0}||_2 = \inf_{x \in S_{0,0}} ||y - x||_2$. Then we have that

$$P(H_{N_2} \in S_{1,-1}) \geq \inf_{y:||y-S_{0,0}||_2<M} \int_{S_{1,-1}} \frac{1}{2\pi N_2} e^{-\frac{|z-y|^2}{2N_2}} dz \geq M^2 \frac{1}{2\pi N_2} e^{-\frac{4^2}{2}} = \frac{1}{2\pi} e^{-8}$$

The ancestors of different individuals of generation N_1 move independently. Hence, conditional on $A_{N_1,K,M}$ we can consider K individuals of interest in generation N_1 and say an individual is a success if it has at least one descendent in generation N_0 with position in $S_{1,-1}$. By the above calculation, for any individual at distance at most M from x the probability of success is at least $1 - \frac{1}{2\pi} e^{-8}$. Hence, we have that given $A_{N_1,K,M}$.

$$P(\text{less than } N \text{ individuals in generation } N_0 \text{ in } S_{1,-1} | A_{N_1,K,M}) \leq \sum_{l=0}^{N} \binom{K}{l} \left(1 - \frac{1}{2\pi} e^{-8}\right)^{K-l}$$

which can be made arbitrarily small for K large enough and fixed N and similarly for $S_{1,1}$. Hence,

$$P(\text{at least } N \text{ point in generation } N_0 \text{ in each of } S_{1,-1} \text{ and } S_{1,1} | A_{N_1,K,M}) \geq \sqrt{(1 - \epsilon)}$$

and we conclude that

$$\begin{aligned} P(\text{at least } N \text{ point in generation } N_0 \text{ in each of } S_{1,-1} \text{ and } S_{1,1}) &\geq \sqrt{1-\epsilon}\sqrt{1-\epsilon} \\ &= 1-\epsilon \end{aligned}$$

and we are done. □

The Approximation Algorithm

As a final step towards using G_d to approximate the behaviour of the sequential search procedure we will need to define an algorithm. The algorithm consists of steps (i, j) where $(i, j) \in \mathcal{L}$ the lattice defined at the beginning of chapter 2. We can order the elements of $\mathcal{L}$ lexicographically, i.e for (i, j) and $(i', j') \in \mathcal{L}$, $(i, j) < (i', j')$ if $i < i'$ or $i = i'$ and $j < j'$. The steps are carried out in ascending order according to this ordering. A step (i, j) is carried out if at least one of the steps $(i-1, j-1)$ and $(i-1, j+1)$ is carried out successfully. The algorithm is said to be successful if infinitely many steps are successful. Let p_c be the critical value for independent oriented site percolation on the lattice $\mathcal{L}$. It is known in the literature that $p_c < 1$, see for example [10]. Choose $\epsilon > 0$ such that $1 - 7\epsilon > p_c$. It will be useful to define functions for position and distance.

Definition 2.2.6. *Define*

$$\begin{aligned} pos : \bigcup_{d=1}^{\infty} \mathbb{R}^d \times \mathbb{R} \times \{0,1\} &\to \bigcup_{d=1}^{\infty} \mathbb{R}^d \\ ((x_1, \ldots, x_d), r, b) &\mapsto (x_1, \ldots, x_d) \end{aligned}$$

and

$$dist : \bigcup_{d=1}^{\infty} \mathbb{R}^d \times \mathbb{R} \times \{0,1\} \to \mathbb{R}$$
$$((x_1, \ldots, x_d), r, b) \mapsto r$$

Choose N_0 as in lemma 2.2.5 and R_0 large enough so that

$$P(\exists\, v \in \mathcal{T},\, |v| \leq N_0,\ \ v \text{ alive} | pos(G^*(v) - (0,0)| > R_0) < \epsilon$$

Let π_d be the volume of the d-dimensional unit ball, and define the family of events $\{E_r\}_{r>0}$ as $E_r = \{\exists\, v \in \mathcal{T}, 1 \leq |v| \leq N_0,\ \ v \text{ non-ghost}, pos(G_d(v)) \notin B_r(v)\}$, where $B_r(v)$ is the ball of volume $r\pi_d$ centered at the position of the parent of v. Then we define W as follows:

$$W = \inf\{r > 0 : P(E_r) < \epsilon)\}$$

W is independent of d and the radius of the ball in $\mathbb{R}^d$ of volume $W\pi_d$ is $W^{1/d}$ which goes to 1 as d approaches infinity. In the algorithm, when searching for out-neighbours of a vertex v we only search for points whose is position is within radius $W^{1/d}$ of v's position.

Predictable Subtrees

Predictable subtrees 1) If we write $T \subset \mathcal{T}$ we mean that T is a rooted subtree of $\mathcal{T}$. For any $S \subset \mathcal{T}$, write

$$\mathcal{F}_S = \sigma(G_d(v), v \in S).$$

2) If $S \subset \mathcal{T}$ is random, we say S is *predictable* if for any rooted subtree T of $\mathcal{T}$, we have $\{T \subseteq S\} \in \mathcal{F}_T$. If S is a predictable set then we let

$$\mathcal{F}_S = \sigma(\bigcup_{T \subset \mathcal{T}} \{E \cap \{T \subseteq S\}).$$

This is similar to the σ-algebra generated a random process stopped at a stopping time. All the random subsets of $\mathcal{T}$ considered here are predictable.

Seeds and explored Subtrees

Seeds and explored subtrees For each lattice site (i, j) we shall define a set $\Sigma_{i,j} \subset \mathcal{T}^*$ which is, informally, the subtree of $\mathcal{T}^*$ explored before step (i, j). We will also define a set $\sigma_{i,j} \subset \Sigma_{i,j}$ of *seeds* for step (i, j), which may be empty, with the property that v is alive and that $|v| = iN_0$ for all seeds $v \in \sigma_{i,j}$. The set $\sigma_{i,j}$ comprises the nodes from which step (i, j) of the search procedure begins.

We will also maintain the property that for each $(i, j) \in \mathcal{L}$, if (i', j') follows (i, j) in the total order of $\mathcal{L}$ then all nodes $v \in \Sigma_{i,j}$ satisfy that $iN_0 < |v| \leq (i + 1)N_0$. More strongly, $\Sigma_{i',j'} \setminus \Sigma_{i,j}$ induces a (possibly empty) forest, each tree of which is rooted at a node of $\sigma_{i,j}$ and has height at most N_0

The lexicographic order on $\mathcal{T}$ induces a total order of $\Sigma_{i',j'} \setminus \Sigma_{i,j}$; combined with the total order of $\mathcal{L}$, this yields a total order of $\mathcal{T}^* := \bigcup_{(i,j)\in\mathcal{L}} \Sigma_{i,j}$, and we list the elements of $\mathcal{T}^*$ in this order as $(v_k, k \geq 1)$. (In fact, in the search procedure it is possible that $\mathcal{T}^*$ turns out to be finite, but is notationally convenient to ignore this.) A key property of the search process is that for all $\ell \geq 1$, the nodes $v_1, \ldots, v_\ell$ form a rooted subtree of $\mathcal{T}$.

Explored regions

For any sequence of vertices $(v_k, k \geq 1)$ with $v_k \in \mathcal{T}$ for all k, and with $v_1 = \emptyset$, define the following quantities. For the root $v_1 = \emptyset \in \mathcal{T}$, let $A(v_1)$ be the ball centered at the origin in $\mathbb{R}^d$ with radius $W^{1/d}$. Let $I(v_1) = A(v_1)$. For $k \geq 2$,

- $r(v_k)$ is the minimum of $W^{1/d}$ and the distance to $I(v_{k-1})$.
- $A(v_k)$ is the ball centered at $pos(G_d(v_k))$ with radius $r(v_k)$
- $I(v_k)$ is the set $\bigcup_{j \leq k: v_j \neq p(v)} A(v_j)$

where $p(v)$ is the parent of v. It is possible that $r(v_k) = 0$, in which case $A(v_k)$ is a the single point $pos(G_d(v_k))$.

Nodes scanned in step (i, j)

If the set $\sigma_{i,j}$ of seeds has $|\sigma_{i,j}| < N$and $(i, j) \neq (0, 0)$ then step (i, j) is *null*. Otherwise, we define the set of nodes scanned in step (i, j) as follows. Inductively suppose the nodes $v_1, \ldots, v_k$ have already been defined. Then v_{k+1} is the lexicographically least alive node v of $\mathcal{T}$ satisfying the following properties.

1) v is a descendant of some node in $\sigma_{i,j}$, with $iN_0 \leq |v| \leq (i + 1)N_0$

2) p(v) $\in \{v_1, \ldots, v_k\}$

2) The distance from $pos(G_d(v))$ to $\bigcup_{j \leq k: v_j \neq p(v)} pos(G_d(v_j))$ is at least $W^{1/d}$.

3) $|pos(G_d^*(v) - (Mi, Mj)| < R_0$.

4) $|pos(G_d(v)) - pos(G_d(p(v)))| < W^{1/d}$

New Seeds

New seeds At the end of Step (i, j), let $\sigma_{i+1,j+1}$ be the set of nodes v scanned in Step (i, j) with$|v| = (i + 1)N_0$, and with $pos(G_d^*(v) \in S_{i+1,j+1}$; define $\sigma_{i+1,j-1}$ similarly.

Explored Regions and Points

The *region of space explored in step* k is $R_k = A(v_k) \setminus \bigcup_{j<k} A(v_j)$. By the definition of the search process, $A(v_j)$ is disjoint from $\bigcup_{j\leq k: v_j \neq p(v)} A(v_j)$, so $R_k = A(v_k) \setminus A(p(v_k))$. The set of *points discovered in step* k is $P_k = \{pos(G_d(v)) : p(v) = v_k, pos(G_d(v)) \in R_k\}$. We write $Q_k = \bigcup_{j\leq k} P_j$ and $U_k = \bigcup_{j\leq k} R_j = \bigcup_{j\leq k} A(v_j)$.

Definition 2.2.7. *Let* $k(i, j)$ *= time step* (i, j) *completes. We say step* (i, j) *is successful if*

- $|Q_{k(i,j)} \cap S_{i+1,j+1}| \geq N$
- $|Q_{k(i,j)} \cap S_{i+1,j-1}| \geq N$

Definition 2.2.8. *We also define the sigma algebras* $\mathcal{F}_{i,j}$ *generated by* $G_d(v)$ *for* v *with* $pos(G_d(v)) \in Q_{k(i,j)}$.

We now formally describe a coupling of the algorithm and the sequential search process. The true sequential search process described in section 2.1 defines a subset of the process X'_d, which we will call $SSP(X_d)$. We can define another point process on the same underlying probability space using our algorithm.

Definition 2.2.9. *Let* $U^\infty = \bigcup_{k\geq 1} U_k$ *and* $Q_\infty = \bigcup_{k\geq 1} Q_k$.

Using U_∞ we can define a new point process P on $\mathbb{R}^d$, such that on U_∞^c, $P = X_d$ and on U_∞, $P = Q_\infty$.

Proposition 2.2.10. P *is a Poisson point process and* $Q_\infty \subset SSP(P)$

Proof. We first show that P is a Poisson process. U_∞ is a countable union of balls $A(v_k)$'s where v_k is scanned. Let $\mathcal{G}_n = \sigma(U_n)$, and let $\mathcal{G}^\infty$ denote $\sigma(U_\infty)$.

Lemma 2.2.11. *Let S be a Borel set in $\mathbb{R}^d$, and define the event $E^n(S) = \{U_n \subset S\}$. Then $E^n(S)$ is measurable with respect to $P \cap S$.*

Proof. We prove this lemma by induction. For $n = 1$, $A_\emptyset$ is just the ball of radius $W^{1/d}$ centered at the origin and thus $\{U_1 \subset S\}$ is just the event that $\{x, |x| \leq W^{1/d}\} \subset S$, which is clearly $P \cap S$-measurable. Now assume that the hypothesis is true for some $n \in \mathbb{N}$. Note that $E^{n+1} = E^n \cap \{A(v_{n+1}) \in S\}$. E^n is $P \cap S$-measurable by the induction hypothesis. The order of exploration is $P \cap U_n$-measurable, since it is discernible from the distances of points from their parents. The $n+1$'th vertex to be scanned is thus $P \cap U_n$-measurable, and so is the radius of exploration since it depends only on U^n. Therefore if $E^n(S)$ does not hold then E^{n+1} clearly does not hold, otherwise if $E^n(S)$ holds then, $\{A(v_{n+1}) \in S\}$ is $P \cap U_n$-measurable and therefore $P \cap S$-measurable. □

We can thus conclude that for S Borel, $E(S) = \{U_\infty \subset S\}$ is $P \cap S$-measurable, since $\{U_\infty \subset S\} = \bigcap_{n=1}^\infty E^n(S)$, and since $E^n(S) \in \sigma(P \cap S)$ for all n, by the countable union property of σ-algebras we can conlude that $\{U_\infty \subset S\} \in \sigma(P \cap S)$.

Lemma 2.2.12. *$P \cap (\mathbb{R}^d \setminus U_\infty)$ is conditionally independent of $P \cap U_\infty$ given U_∞, and is distributed as a Poisson processes on $\mathbb{R}^d \setminus U_\infty$.*

Proof. For the rest of the proof $Q(m)$ denote a Poisson random variable with rate $\lambda(d)m$, recalling that $\lambda(d)$ is the rate of the Poisson process X_d. Let $D_1, \ldots, D_n$ be disjoint Borel sets in $\mathbb{R}^d$, and for $1 \leq i \leq n$ let $X_i = |(D_i \setminus U_\infty) \cap P|$. Let $\mathcal{C}$ be the

set of all finite unions of balls in $\mathbb{R}^d$ with rational radii and centres. Finally let *Leb* denote the Lebesgue measure on $\mathbb{R}^d$. We wish to show that almost surely that for non-negative integers $k_1, \ldots, k_n$,

$$P\left(\bigcap_{i=1}^{n} X_i \geq k_i | \mathcal{G}^\infty\right) = \prod_{i=1}^{n} P(Q(Leb(D_i \setminus U_\infty)) \geq k_i)$$

Note that for all $C \in \mathcal{C}$, and all $1 \leq i \leq n$,

$$\begin{aligned} X_i &= |(D_i \setminus U_\infty) \cap P| \\ &= \sup_{C \in \mathcal{C}} |(D_i \setminus C) \cap P| \chi(U_\infty \subset C) \\ &\geq |(D_i \setminus C) \cap P| \chi(U_\infty \subset C) \end{aligned}$$

Therefore we have that for all $C \in \mathcal{C}$, almost surely

$$\begin{aligned} P\left(\bigcap_{i=1}^{n} X_i \geq k_i | \mathcal{G}^\infty\right) &= E\left[\chi\left(\bigcap_{i=1}^{n} X_i \geq k_i\right) \middle| \mathcal{G}^\infty\right] \\ &\geq E\left[\chi\left(\bigcap_{i=1}^{n} |(D_i \setminus C) \cap P| \geq k_i\right) \chi(U_\infty \subset C) \middle| \mathcal{G}^\infty\right] \\ &= E\left[\chi\left(\bigcap_{i=1}^{n} |(D \setminus C) \cap X_d| \geq k_i\right) \chi(U_\infty \subset C) \middle| \mathcal{G}^\infty\right] \\ &= \prod_{i=1}^{n} P(Q(Leb(D_i \setminus C)) \geq k_i) \chi(U_\infty \subset C) \end{aligned}$$

since $D \setminus C$ is a deterministic set and X_d is a Poisson process. Since this holds for all C, we can conclude that

$$P(X \geq k | \mathcal{G}^\infty) \geq P(Q(Leb(D \setminus U_\infty)) \geq k)$$

almost surely. Similarly, an analogous argument considering the infimum over a union of increasing balls contained in U_∞ yields that

$$P(X \geq k|\mathcal{G}^\infty) \leq P(Q(Leb(D \setminus U_\infty)) \geq k)$$

and the result is shown. □

Lemma 2.2.13. *For all n, $P \cap U_n$ is a Poisson Process on U_n conditional on $\mathcal{G}_n$.*

Proof. We proceed by induction on n. $P \cap U_1$, is distributed as a Poisson process on U_1, since U_1 is just the ball around the origin of radius $W^{1/d}$ and $P \cap U_1$ is just the intersection of the U_1 with an independent Poisson process $X_\emptyset$, as the distribution of G_d entails. Assume the claim holds true for U_n. The radius and center of $A(v_{n+1})$ are both measurable with respect to $P \cap U_n$, and by lemma 2.2.12 $P \cap (A(v_{n+1}) \setminus U_n)$ is conditionally independent of $P \cap U_n$ given U_n. By the production mechanism of G_d, conditional on U_n, $P \cap (A_v \setminus U_n)$ is distributed according to the intersection of $(A(v_{n+1}) \setminus U_n)$ and an associated independent Poisson process, X_v and so $P \cap U_{n+1}$ is thus again a Poisson process on U_{n+1}. □

It follows directly from the above lemma that $P \cap U_\infty$ is distributed as a Poisson process on U_∞ given $\mathcal{G}^\infty$, since if $S \subset U_\infty$, then $S \subset U_n$ for some n, and therefore $P \cap S$ is distributed as a Poisson process on S.

Lemma 2.2.14. *Let $V_1, V_2, \ldots, V_n$ be disjoint Borel subsets of $\mathbb{R}^d$. Then for $k_1, \ldots, k_n$ non-negative integers, we have that*

$$P\left(\bigcap_{i=1}^{n} |V_i \cap P| = k_i\right) = \prod_{i=1}^{n} p_{k_i}(Leb(V_i))$$

where $m_i = Leb(V_i)$ *and* $p_k(m) = P(Q(m) = k)$.

Let $\mathcal{N} = \{(l_1, \ldots, l_n) \in (\mathbb{N} \cup 0)^n : 0 \leq l_i \leq k_i \text{ for } 0 \leq i \leq n\}$.

$$\begin{aligned} & P\left(\bigcap_{i=1}^{n} |V_i \cap P| = k_i\right) \\ = & \sum_{l \in \mathcal{L}} P\left(\bigcap_{i=1}^{n} \{|V_i \cap P \cap U_\infty| = l_i\} \cap \{|V_i \cap P \cap U_\infty^c| = k_i - l_i\}\right) \end{aligned}$$

For $l = (l_1, \ldots, l_n) \in \mathcal{N}$ let H^l be the event $\bigcap_{i=1}^{n} \{|V_i \cap P \cap U_\infty| = l_i\} \cap \{|V_i \cap P \cap U_\infty^c| = k_i - l_i\}$. Conditioning on $\mathcal{G}^\infty$ yields $P(H^l) = E[\chi(H^l)] = E[E[\chi(H^l)|\mathcal{G}^\infty]]$. Examining the internal conditional expectation yields,

$$E[\chi(H^l)|\mathcal{G}^\infty] = E[E[\chi(H^l)|\mathcal{G}^\infty, P|_{U_c^\infty}]\mathcal{G}^\infty] \tag{2.1}$$

but by the previous lemma and noting that for all i, $|P \cap V_i \cap U_\infty^c|$ is $P|_{U_\infty^c}$-measurable, we have that

$$\begin{aligned} E[\chi(H^l)|\mathcal{G}^\infty, P|_{U_\infty}^c] &= E\left[\prod_{i=1}^{n} \chi(|V_i \cap P \cap U_\infty| = l_i)\chi(|V_i \cap P \cap U_\infty^c| = k_i - l_i)\Big|\mathcal{G}^\infty, P|_{U_\infty^c}\right] \\ &= \prod_{i=1}^{n} p_{l_i}(Leb(V_i \cap U_\infty)) \prod_{i=1}^{n} \chi(|V_i \cap P \cap U_\infty^c| = k_i - l_i) \end{aligned}$$

Finally, returning to (2.1) we get

$$\begin{aligned} E[\chi(H^l)|\mathcal{G}^\infty] &= E\left[\prod_{i=1}^{n} p_{l_i}(Leb(V_i \cap U_\infty)) \prod_{i=1}^{n} \chi(|A_i \cap P \cap U_\infty^c| = k_i - l_i)\Big|\mathcal{G}^\infty\right] \\ &= \prod_{i=1}^{n} p_{l_i}(Leb(V_i \cap U_\infty)) \prod_{i=1}^{n} p_{k_i - l_i}(Leb(V_i \cap U_\infty^c)) \end{aligned}$$

Hence we have that by the above evaluation,

$$\begin{aligned} P\left(\bigcap_{i=1}^{n} |V_i \cap P| = k_i\right) &= E\left[\sum_{l\in\mathcal{N}} E[\chi(H^l)|\mathcal{G}^\infty]\right] \\ &= E\left[\sum_{l\in\mathcal{N}} \prod_{i=1}^{n} p_{l_i}(Leb(V_i \cap U_\infty)) \prod_{i=1}^{n} p_{k_i-l_i}(Leb(V_i \cap U_\infty^c))\right] \\ &= E\left[\prod_{i=1}^{n} p_{k_i}(Leb(V_i \cap U_\infty) + Leb(V_i \cap U_\infty^c))\right] \\ &= E\left[\prod_{i=1}^{n} p_{k_i}(Leb(V_i))\right] \\ &= \prod_{i=1}^{n} p_{k_i}(Leb(V_i)) \end{aligned}$$

That $Q_\infty \subset SSP(P)$ is now clear since a point is only in Q_∞ if it's part of a successful step, which in turn implies that it is part of the directed connected component of $G(P)$ rooted at the origin.

□

Corollary 2.2.15. *Recall from Def 2.1.3 the definition of $C(P)$.*

$$P(|C(P)| = \infty) \geq P(|Q_\infty| = \infty)$$

This follows from the previous proposition and corollary 2.1.6. We now direct our attention towards proving that the algorithm has a positive probability of being successful, using a comparison to percolation on the lattice $\mathcal{L}$.

Lemma 2.2.16.

$$P(\textit{Step } (i,j) \textit{ is successful } |\mathcal{F}_{i,j}) > p_c$$

where p_c is the critical probability for site percolation on $\mathcal{L}$.

Proof. Consider the following types of failures

- There exists a vertex v scanned in Step (i, j) such that v is not in the set of explored nodes because v was born in the previously explored region.
- There exists a vertex v scanned in Step (i, j) such that $pos(G_d(v))$ was within distance $W^{1/d}$ of a previously scanned node w such that $w \neq p(v)$.
- $|pos(G_d^*(v) - (Mi, Mj)| < R_0$.
- $|pos(G_d(v)) - pos(G_d(p(v)))| < W^{1/d}$

We recall that $1 - 7\epsilon > p_c$ by choice. The probability that the set of vertices scanned in step (i, j) contain less than N points in $S_{i+1,j+1}$ or less than N points in $S_{i+1,j-1}$ given that one of the steps $(i-1, j-1)$ and $(i-1, j+1)$ was successful, is upper bounded independently of $\mathcal{F}_{i,j}$ as a direct result of lemma 2.2.5 by ϵ for all sufficiently large d. The same holds for failures of type (3) by definition of R_0 and for failures of type (4) by definition of W. Thus to prove the result we need only bound the probabilities that a failure of types 1 or 2 occurs.

Next it remains to bound failures of types 1-2. First we consider a failure of types 1-2 due to step (i', j') with $M\sqrt{(i-i')^2 + (j-j')^2} < R$ where R is to be determined later. Here we're including the possibility that $(i', j') = (i, j)$. Both failures are subsets of the event that a point in step (i', j') is within distance $2W^{1/d}$ of a point in step (i, j). The number of individual points born in such steps of the algorithm is bounded as a function of R and the bound on the integer mark B. Hence we can use lemma 2.1.2 to deduce that we can make the probability of this happening less than ϵ by taking d large.

Now we consider failures of type 1-2 due to a step (i', j') with $M\sqrt{(i-i')^2+(j-j')^2} \geq R$.

Consider a failure of type 1 due to a step (i', j') with

$$\lfloor M\sqrt{(i-i')^2+(j-j')^2}\rfloor = Q$$

where $\lfloor x \rfloor$ denotes the smallest integer greater than or equal to x. The total volume scanned in step (i', j') is uniformly bounded by construction, since B is bounded and so are the radii of the M_v's. Now note that by condition 4, the projected distance between a point in step (i, j) and a point in step (i', j') around which scanning has taken place is at least $Q - 2R_0$. It follows from this and from Lemma 2.2.3 and the super-exponential decay of the normal distribution that the fraction of the volume scanned in step (i', j') whose projection falls into $S((Mi, Mj), R_0)$ is less than $\frac{2pi_d\epsilon}{Q^3}$ for all sufficiently large Q. The number of points (i', j') with

$$\lfloor M\sqrt{(i-i')^2+(j-j')^2}\rfloor = Q$$

is bounded by a constant times Q. Hence for any $\delta > 0$ we can make the total such volume from all points (i', j') with $M\sqrt{(i-i')^2+(j-j')^2} \geq R$ smaller than $\pi_d\delta$ by choosing R large enough. Choose R such that this volume is small enough such that each individual of the G_d in step (i, j) has probability at most ϵ/C of landing there, where C is the maximum possible number of individuals in step (i, j). C is dependent on the bound on B. The probability that there exists individual in step (i, j) which lands in such a volume is thus less than ϵ. The probability of a failure of types 2 due to steps (i', j') with $M\sqrt{(i-i')^2+(j-j')^2} \geq R$ can be bound by ϵ

using an analogous argument. We obtain the lemma by summing up the probabilities of failure.

□

Using a comparison to percolation on $\mathcal{L}$ we can conclude that the probability that the algorithm is successful is positive for d large enough. Thus for such d the probability that Q_∞ is infinite is positive, and the component containing the origin in C has positive probability of being infinite by corollary 2.2.15.

CHAPTER 3
Open Problems and Further Research Possibilities

Our analysis provides a condition under which there exists a dimension where percolation occur for all greater dimensions in the $NN(d, B)$ model. This adds to previous literature by incorporating a randomness factor in the number of points each point connects to, but leaves some questions open that arise naturally from our results. The first such question is whether or not the bound given is tight. Formally, the question can be stated as follows: Does there exist a constant $0 < M < 2$ such that, for all non-negative integer valued random variables B with $M < E[B] < \infty$, there exist d_0 such that for $d \geq d_0$ percolation occurs almost surely in the $NN(d, B)$ model?

A possible path to tackling this question could be an examination of the operator T defined in chapter 2 in an attempt to learn more about its eigenspace, and its dependence on the properties of B. If one could show that the eigenspace necessarily contains a value greater than 1 if $M < E[B] < \infty$ then this would relax the bound stated in the theorem. Furthermore, we could question if the expectation bound stated could be relaxed if further conditions on B are imposed, such as bounds on the variance or the decay of the distribution's tail. This would yield an interesting

perspective on the general $NN(d, k)$ model, by allowing us to think, albeit informally, of the critical k as not necessarily an integer value.

Another natural question that arises from our work is, whether or not we can provide any similar expectation conditions to ensure percolation occurs in specific finite dimensions. While the work we have done proves the existence of a dimension beyond which percolation occurs under certain conditions, the dimension remains unknown, and the behaviour of the model in low dimensions remains open to exploration.

Finally there exists the ambitious question of necessary conditions. Namely, if B is such that percolation occurs in the $NN(d, B)$ model for high enough dimensions, what is necessarily true of B? Answering this question gives a more comprehensive understanding of the model, as well as offering the prospects of shedding light on the more studied model with a deterministic number of connections per point.

CHAPTER 4
References

[1] Broadbent, S.R. and Hammersley, J.M. (1957). Percolation processes I: crystals and mazes. Proceedings of the Cambridge Philosophical Society, 53, 629 641.

[2] Grimmett, G. (1982). Probability and Random Processes. Oxford University Press.

[3] R. Meester and R. Roy (1996). Continuum Percolation. Cambridge University Press.

[4] Häggström and Meester, R. (1996). Nearest neighbor and hard sphere models in continuum percolation. Random Structures Algorithms, 9(3), 295315.

[5] Theodore E. Harris (1963). The Theory of Branching Processes. Springer-Verlag, Berlin.

[6] M. D. Penrose (1996). Continuum percolation and Euclidean minimal spanning trees in high dimensions. Applied Probability, 6(2), 528-544.

[7] G. W. Stewart (1980). The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators. SIAM Journal on Numerical Analysis, 17(3), 403-404.

[8] J. Stewart (2011). Multivariable Calculus 7th edition. Brooks Cole.

[9] Ethier, S. N., and Kurtz, T. G. (1986). Markov processes: characterization and convergence. New York: Wiley.

[10] Balister, P., Bollobas, B., and Stacey, A. (1993). Upper Bounds for the Critical Probability of Oriented Percolation in Two Dimensions. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 440(1908), 201-220.