Cluster Detection in Networks using Percolation

Ery Arias-Castro* and Geoffrey R. Grimmett†

Abstract

We consider the task of detecting a salient cluster in a sensor network, i.e., an undirected graph with a random variable attached to each node. Motivated by recent research in environmental statistics and the drive to compete with the reigning scan statistic, we explore alternatives based on the percolative properties of the network. The first method is based on the size of the largest connected component after removing the nodes in the network whose value is lower than a given threshold. The second one is the upper level set scan test introduced by Patil and Taillie (2003). We establish their performance in an asymptotic decision theoretic framework where the network size increases. We make abundant use of percolation theory to derive our theoretical results and our theory is complemented with some numerical experiments.

Keywords: cluster detection; surveillance; multiple hypothesis testing; scan statistic; largest open cluster within a box; upper level set scan statistic; connected components; percolation.

1 Introduction

We consider the problem of cluster detection in a network. The network is modeled as a graph and we assume that a random variable is observed at each node. The task is then to detect a cluster, i.e., a connected subset of nodes with values that are larger than usual or unusual in some other way. There is a multitude of applications for which this model is relevant. Examples include the detection of hazardous materials (Hills, 2001) and target tracking (Li et al., 2002) in sensor networks (Culler et al., 2004), and disease outbreak detection (Heffernan et al., 2004; Rotz and Hughes, 2004; Wagner et al., 2001). Pixels in digital images are also sensors so that many other examples can be found in the rich literature on image processing, with examples such as road tracking (Geman and Jedynak, 1996) and fire prevention using satellite imagery (Pozo et al., 1997), and the detection of tumors in medical imaging (McInerney and Terzopoulos, 1996).

After specifying a distributional model for the observations at the nodes and a class of clusters to be detected, the generalized likelihood ratio test is the first method to come to mind. And, indeed, it is by far the most popular method in practice, and as such, is given different names in different fields. The likelihood ratio is known as the scan statistic in spatial statistics (Kulldorff, 1997, 2001) and the corresponding test as the method of matched filters in engineering (Jain et al., 1998; McInerney and Terzopoulos, 1996). We will use the former, where scanning a given cluster K means to compute the likelihood ratio for the simple alternative where K is the anomalous cluster. Various forms of scan statistic have been proposed, mostly differing by the assumptions made on the shape of the clusters. Most methods assume that the clusters are in some parametric family, e.g., circular (Kulldorff and Nagarwalla, 1995), elliptical (Hobolth et al., 2002; Kulldorff et al., 2006) or, more generally, deformable templates (Jain et al., 1998). Sometimes no explicit shape is

*Department of Mathematics, University of California, San Diego {math.ucsd.edu/∼eariasca}
†Statistical Laboratory, Centre for Mathematical Sciences, Cambridge University {www.statslab.cam.ac.uk/∼grg}

assumed (Duczmal and Assunção, 2004; Kulldorff et al., 2003; Tango and Takahashi, 2005), leading to nonparametric models.

We consider nonparametric methods based on the percolative properties of the network. The most basic approach is based on the size of the largest connected component of the graph after removing the nodes whose values fall under a given threshold. If the graph is a one-dimensional lattice, after thresholding this corresponds to the test based on the longest run (Balakrishnan and Koutras, 2002), which Chen and Huo (2006) adapt for path detection in a thin band. This test is studied in a series of papers (Davies et al., 2010; Langovoy and Wittich, 2011) and references therein, under the name of maximum cluster test. More sophisticated is the upper level set scan statistic of Patil and Taillie (2003), subsequently developed in the context of ecological and environmental applications (Patil et al., 2004, 2010, 2006; Patil and Taillie, 2004). In its basic form, it scans over the connected components of the graph after thresholding. Clearly, both methods are nonparametric and they depend on a single parameter, the threshold. Patil and Taillie (2003) originally proposed to scan over all the clusters that are connected components of the graph after thresholding at some threshold, leading to a parameter-free method.

We compare both methods to the scan statistic in a standard asymptotic decision theoretic framework (Arias-Castro et al., 2011, 2005; Perone Pacifico et al., 2004; Walther, 2010) where the network is a square lattice of growing size and the variables at the nodes are assumed independent and identically distributed for nodes inside (resp. outside) the anomalous cluster. Our theory generalizes to other settings as discussed in Section 7. When there is no anomalous cluster, the variables are therefore i.i.d., and after thresholding, we have a standard (site) percolation model (Grimmett, 1999). Percolation theory is central to this paper in the analysis of the two methods mentioned above as they both rely on the connected components after thresholding.

The rest of the paper is organized as follows. We formally introduce the framework in Section 2 and state some fundamental detection bounds. In Section 3, we describe the standard scan statistic and state some existing results on its performance showing that it is essentially optimal. In Section 4, we consider the size of the largest connected component after thresholding. In Section 5, we consider the upper level set scan statistic. We briefly discuss implementation issues and present some numerical experiments in Section 6. Finally, Section 7 is a discussion section where, in particular, we mention extensions. The proofs are postponed to the Appendix.

2 Mathematical framework and fundamental detection bounds

For concreteness, and also its relevance in signal and image processing, we model the network as a finite subgrid of the regular square lattice in dimension d, denoted $\mathbb{V}_m := \{1, \dots, m\}^d$. Our analysis is asymptotic in the sense that the network is assumed to be large, i.e., $m \to \infty$. To each node $v \in \mathbb{V}_m$, we attach a random variable X_v. For example, in the context of a sensor network, the nodes represent the sensors and the variables represent the information they transmit. The random variables $\{X_v : v \in \mathbb{V}_m\}$ are assumed to be independent with common distribution in a certain one-parameter exponential family $\{F_\theta : \theta \in [0, \theta_\infty)\}$ defined as follows. Let $\theta_\infty > 0$, let F_0 be a distribution function with finite non-zero variance σ_0^2, and assume the moment generating function $\varphi(\theta) := \int e^{x\theta}\, dF_0(x)$ is finite for $\theta \in [0, \theta_\infty)$. Then F_θ is the distribution function having density $f_\theta(x) = \exp(\theta x - \log \varphi(\theta))$ with respect to F_0. We will assume further regularity of F_0 at later points in this paper. Note that our results apply to other distributional models as discussed in Section 7.

Examples of such a family $\{F_\theta : \theta \in [0, \theta_\infty)\}$ include the following:

- *Bernoulli model:* $F_\theta = \text{Ber}(p_\theta)$, $p_\theta := \text{logit}^{-1}(\theta + \theta_0)$, relevant in sensor arrays where each sensor transmits one bit (i.e., makes a binary decision).

- *Poisson model:* $F_\theta = \text{Poi}(\theta + \theta_0)$, popular with count data, for example arising in infectious disease surveillance systems.
- *Exponential model:* $F_\theta = \text{Exp}(\theta_0 - \theta)$, e.g., to model response times.
- *Normal location model:* $F_\theta = \mathcal{N}(\theta + \theta_0, 1)$, standard in signal and image processing, where noise is often assumed to be Gaussian.

Let $\mathcal{K}_m$ be a class of clusters, where a cluster is defined as a subset of nodes which is connected in the graph. Under the null hypothesis, all the variables at the nodes have distribution F_0, i.e.

$$\mathbb{H}_0^m \ : X_v \sim F_0, \ \forall v \in \mathbb{V}_m.$$

Under the particular alternative where $K \in \mathcal{K}_m$ is anomalous, the variables indexed by K have distribution F_{θ_m} for some $\theta_m > 0$, i.e.

$$\mathbb{H}_{1,K}^m \ : X_v \sim F_{\theta_m}, \ \forall v \in K; \quad X_v \sim F_0, \ \forall v \notin K.$$

We are interested in the situation where the anomalous cluster K is unknown, namely in testing $\mathbb{H}_0^m$ against $\mathbb{H}_1^m := \bigcup_{K \in \mathcal{K}_m} \mathbb{H}_{1,K}^m$. We illustrate the setting in Figure 1 in the context of the two-dimensional square grid.

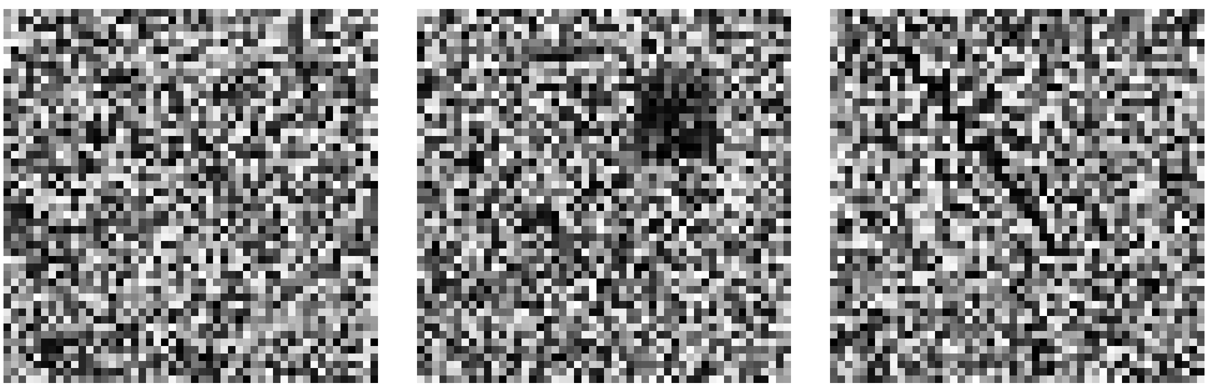

Figure 1: This figure illustrates the setting in dimension $d = 2$ for a beta model where $F_0 = \text{Unif}(0,1)$ and $F_\theta = \text{Beta}(\theta + 1, 1)$, $\theta \geq 0$. Left: An instance of the null hypothesis. Middle: An instance of an alternative with a square cluster. Right: An instance of an alternative with a path.

Let $\mathcal{K}_m$ denote a cluster class for $\mathbb{V}_m$. For a test T, we define its worst-case risk as

$$\gamma_m(T) = \mathbb{P}(T = 1 | \mathbb{H}_0^m) + \max_{K \in \mathcal{K}_m} \mathbb{P}(T = 0 | \mathbb{H}_{1,K}^m).$$

A method is formally defined as a sequence of tests (T_m) for testing $\mathbb{H}_0^m$ versus $\mathbb{H}_1^m$. We say that a method (T_m) is asymptotically powerless if

$$\liminf_{m \to \infty} \gamma_m(T_m) \geq 1.$$

This amounts to saying that, as the size of the network increases, the method (T_m) is not substantially better than random guessing. Conversely, a method (T_m) is asymptotically powerful if

$$\lim_{m \to \infty} \gamma_m(T_m) = 0.$$

We focus on situations where the clusters in the class $\mathcal{K}_m$ are of same size, increasing with m but negligible compared to the size of the entire network. We do so for the sake of simplicity, as more general results could be obtained as in (Arias-Castro et al., 2011, 2005; Perone Pacifico et al., 2004; Walther, 2010) without additional difficulty. Assuming that the size of the anomalous cluster is large allows us to state general results applying to a wide range of one-parameter exponential families (via the Central Limit Theorem). Also, note the following. On the one hand, reliably detecting a cluster of bounded size is impossible in the Bernoulli model or any other model where F_0 has finite support. On the other hand, detecting a cluster of size comparable to that of the entire network is in some sense trivial as the simple test based on the total sum $\sum_{v \in \mathbb{V}_m} X_v$ is optimal up to a multiplicative constant.

We consider two emblematic classes of clusters, in some sense at the opposite extremes:

- *Hypercube detection.* Let $\mathcal{K}_m$ denote the class of *hypercubes* within $\mathbb{V}_m$ of sidelength $[m^\alpha]$ with $0 < \alpha < 1$. This class is parametric, with the location of the hypercube being the only parameter

- *Path detection.* Let $\mathcal{K}_m$ denote the class of *loopless paths* within $\mathbb{V}_m$ of length $[m^\alpha]$ with $0 < \alpha < 1$. This class is nonparametric, in the sense that its complexity (in the information theoretic sense) is exponential in the length of the paths.

See Figure 1 for an illustration. (Note that a hypercube of sidelength k may be seen as a loopless path of length k^d.) Though we obtain results for both, our main focus is in the setting of hypercube detection, which is relevant to a wider range of applications, in fact any situation where the task is to detect a shape that is not filamentary. The situation exemplified in the setting of path detection may be relevant in target tracking from video, or the detection of cracks in materials in non-destructive testing. Note that the two settings coincide in dimension one.

We first state fundamental detection bounds for each setting. The following result is standard; see e.g., (Arias-Castro et al., 2011, 2005). Remember that σ_0^2 denotes the variance of F_0.

Lemma 1. *In hypercube detection, all methods are asymptotically powerless if*

$$\limsup_{m \to \infty} (\log m)^{-1/2} m^{d\alpha/2} \theta_m < \sigma_0 \sqrt{2d(1-\alpha)}.$$

In fact, the conclusions of Lemma 1 apply for a wide variety of parametric classes such as discs, a popular model in disease outbreak detection (Kulldorff and Nagarwalla, 1995), and also to nonparametric classes of blob-like clusters; see (Arias-Castro et al., 2011, 2005).

The following result is taken from (Arias-Castro et al., 2008).

Lemma 2. *In path detection, in dimension $d = 2$, all methods are asymptotically powerless if $\lim_{m \to \infty} \theta_m (\log m)(\log \log m)^{1/2} = 0$; and the same is true in dimension $d \geq 3$ if $\limsup_{m \to \infty} \theta_m < \theta_*$, where $\theta_* > 0$ depends only on d.*

In dimension $d \geq 4$, θ_* may be taken to be the unique solution to

$$\rho\, \varphi(2\theta) - \varphi(\theta)^2 = 0,$$

where ρ is the return probability of a symmetric random walk in dimension d.

3 The scan statistic

For a subset of nodes $K \subset \mathbb{V}$, let $|K|$ denote its size and define

$$\bar{X}_K = \frac{1}{|K|} \sum_{v \in K} X_v.$$

Given a cluster class $\mathcal{K}$, we define the (simple) scan statistic as

$$\max_{K \in \mathcal{K}} \sqrt{|K|}\, (\bar{X}_K - \mu_0), \tag{1}$$

where μ_0 is the mean of F_0. If μ_0 is not available, we may use the grand mean $\bar{X}_{\mathbb{V}_m}$ instead. In Appendix B, we derive this form of scan statistic as an approximation to the scan statistic of Kulldorff (1997), which is strictly speaking the generalized likelihood ratio and arguably the most popular version, particularly in spatial statistics. We use this simpler form to streamline our theoretical analysis.

The test based on the scan statistic, which we called the scan test, is near-optimal in a wide range of settings (Arias-Castro et al., 2011, 2005; Walther, 2010). In particular, in the context of a class of hypercubes, and in fact many other parametric classes, this test is asymptotically optimal to the exact multiplicative constant.

Lemma 3. *In hypercube detection, the scan test is asymptotically powerful if*

$$\liminf_{m \to \infty} (\log m)^{-1/2} m^{d\alpha/2} \theta_m > \sigma_0 \sqrt{2d(1-\alpha)}.$$

In the context of a class of paths, the following result states that the scan test detects if θ_m is bounded away from zero and sufficiently large. Note that this does not match the order of magnitude of the lower bound given in dimension $d = 2$. Let $\Lambda(\theta) = \log \varphi(\theta)$ and $\Lambda^*(x) = \sup_{\theta \geq 0} [\theta\, x - \Lambda(\theta)]$. ($\Lambda^*$ is the rate function of F_0 when $x \geq \mu_0$.) The following result is established in (Arias-Castro et al., 2008).

Lemma 4. *In path detection, the scan test is asymptotically powerful if*

$$\liminf_{m \to \infty} \theta_m > \theta_* := (\Lambda^* \circ \Lambda')^{-1}(\log(2d)).$$

4 The size of the largest open cluster

We study the test based on the size of the largest connected component after thresholding the values at the nodes. This test was independently[1] considered in a series of papers (Davies et al., 2010; Langovoy and Wittich, 2011). Our results are seen to sharpen and elaborate on these results. In particular, we study this test under all three regimes (subcritical, supercritical and critical).

Adapting terminology from percolation theory (Grimmett, 1999), for a threshold $t \in \mathbb{R}$, we say that a subset $K \subset \mathbb{V}$ is open (at threshold t) if $X_v > t$ for all $v \in K$. Let $S_m(t)$ (resp. $S_K(t)$) denote the size of the largest open cluster within $\mathbb{V}_m$ (resp. within K). The analysis of the test based on $S_m(t)$, which we call the largest open cluster (LOC) test, boils down to bounding the size of $S_m(t)$ from above, under $\mathbb{H}_0^m$, and, since $S_m(t) \geq S_K(t)$, bounding the size of $S_K(t)$ from below, under $\mathbb{H}_{1,K}^m$. Define $\xi_v(t) = \mathbf{I}\{X_v > t\}$, which is Bernoulli with parameter $p_\theta(t) := \mathbb{P}_\theta(X_v > t)$.

[1] The authors were not aware of this (unpublished) line of work until M. Langovoy contacted them in the final stages of this manuscript.

The process $(\xi_v(t) : v \in \mathbb{V}_m)$ is a site percolation model (Grimmett, 1999). In general, consider a process $(\xi_v : v \in \mathbb{V}_m)$ i.i.d. Bernoulli with parameter p, and let S_m denote the size of the largest open cluster within $\mathbb{V}_m$. In dimension $d = 1$, this process may be seen as a sequence of coin tosses and S_m as the longest heads run in that sequence. In this context, the Erdős–Rényi Law (Erdős and Rényi, 1970) says that

$$\frac{S_m}{\log m} \to \frac{1}{\log(1/p)}, \quad \text{almost surely.} \tag{2}$$

In higher dimensions $d \geq 2$, the situation is much more involved. Let p_c denote the critical probability for site percolation in $\mathbb{Z}^d$, defined as the supremum over all $p \in (0, 1)$ such that the size of the open cluster at the origin, denoted S, is finite with probability one. (The dependency in d is left implicit.) We consider the subcritical ($p_0(t) < p_c$), supercritical ($p_0(t) > p_c$) and near-critical ($p_0(t) \approx p_c$) cases separately.

4.1 Subcritical percolation

In the subcritical case, where t is such that $p_0(t) < p_c$, we are able to obtain precise, rigorous results on the performance of the test based on $S_m(t)$ in terms of the function ζ_p implicitly defined as

$$\zeta_p := -\lim_{k\to\infty} \frac{1}{k} \log \mathbb{P}\left(S \geq k\right) = -\lim_{k\to\infty} \frac{1}{k} \log \mathbb{P}\left(S = k\right). \tag{3}$$

See (Grimmett, 1999, Sec. 6.3). (Again, the dependency in d is left implicit.) As a function of $p \in (0, p_c)$, ζ_p is continuous, and strictly decreasing, with limits ∞ at $p = 0$ and zero at $p = p_c$ (see Lemma A.1), while $\zeta_p = 0$ for $p \geq p_c$. In the Appendix, we include a proof that

$$\frac{S_m}{\log m} \to \frac{d}{\zeta_p}, \quad \text{in probability,} \tag{4}$$

for a subcritical threshold $p < p_c$.

The convergence result in (4) may be used to bound $S_m(t)$ under the null by taking $p = p_0(t)$. Under the alternative, if we consider a class of hypercubes, then (4) may also be used to bound $S_K(t)$, since K is a scaled version of $\mathbb{V}_m$.

Theorem 1. *In hypercube detection, the test based on $S_m(t)$, with t fixed such that $0 < p_0(t) < p_c$, is asymptotically powerful if $\liminf_{m\to\infty} \theta_m > \theta_*(t)$, and asymptotically powerless if $\limsup_{m\to\infty} \theta_m < \theta_*(t)$, where $\theta_*(t)$ is the unique solution to $\zeta_{p_\theta(t)} = \alpha\zeta_{p_0(t)}$.*

Notice that, when t is fixed, $\zeta_{p_\theta(t)}$ as a function of θ is continuous and strictly strictly decreasing, by the fact that $p_\theta(t)$ is continuous and strictly increasing in θ (Brown, 1986, Cor. 2.6, 2.22) and ζ_p is continuous and strictly decreasing in p (Lemma A.1). Therefore $\theta_*(t)$ in the theorem is well-defined.

If, instead, we consider a class of paths, then (2) may be used to bound $S_K(t)$, since K is a scaled version of the lattice in dimension one. In congruence with (2), we define $\zeta_p^1 = \log(1/p)$.

Theorem 2. *In path detection, the test based on $S_m(t)$, with t fixed such that $0 < p_0(t) < p_c$, is asymptotically powerful if $\liminf_{m\to\infty} \theta_m > \theta_*^+(t)$, and asymptotically powerless if $\limsup_{m\to\infty} \theta_m < \theta_*^-(t)$, where $\theta_*^+(t)$ (resp. $\theta_*^-(t)$) is the unique solution to $d\,\zeta^1_{p_\theta(t)} = \alpha\,\zeta_{p_0(t)}$ (resp. $d\,\zeta_{p_\theta(t)} = \alpha\,\zeta_{p_0(t)}$).*

Note that, in dimension $d \geq 2$, the result is not sharp, as we always have $\theta_*^+(t) > \theta_*^-(t)$. We believe that sharper forms of this result may involve percolation constants that may not be calculated explicitly, and for this reason we have not pursued this.

What if we let $t = t_m \to \infty$, so that $p_0(t_m) \to 0$? Then the test based on $S_m(t_m)$ is powerless under some additional conditions on F_0. For $b, C \geq 0$, consider the following class of approximately exponential power (AEP) distributions[2]

$$\mathrm{AEP}(b, C) = \left\{ F : x^{-b} \log \bar{F}(x) \to -C,\ x \to \infty \right\}.$$

($\bar{F}(x) := 1 - F(x)$ is the survival distribution function of $X \sim F$.) For example, $\mathrm{Exp}(\lambda) \in \mathrm{AEP}(1, \lambda)$ and $\mathcal{N}(\mu, \sigma^2) \in \mathrm{AEP}(2, 1/(2\sigma^2))$, while $\mathrm{Poi}(\lambda)$ behaves roughly as a distribution in $\mathrm{AEP}(1, C)$.

Proposition 1. *Assume that $F_0 \in \mathrm{AEP}(b, C)$ for some $b > 1$ and $C > 0$. In hypercube detection, the test based on $S_m(t)$ is asymptotically powerless when $t = t_m \to \infty$, unless $\theta_m \to \infty$.*

4.2 Supercritical percolation

We consider the supercritical regime, where $p_0(t) > p_c$. (Note that necessarily $d \geq 2$, for $p_c = 1$ in dimension one.) In this setting too, the size of the largest cluster is well-understood. Let Θ_p be the probability that the open cluster at the origin is infinite, and note that $\Theta_p > 0$ for $p > p_c$, by the definition of p_c. We have with probability one that

$$\frac{S_m}{|\mathbb{V}_m|} \to \Theta_p;$$

see (Falconer and Grimmett, 1992, Lemma 2 and proof), (Penrose and Pisztora, 1996, Th. 4), Pisztora (2006). In fact, (with probability $1 - o(1)$) the largest open cluster within $\mathbb{V}_m$ is unique, and the statement above says that it occupies a non-negligible fraction of $\mathbb{V}_m$. With a supercritical choice of threshold, the LOC test is powerless for any θ if the anomalous cluster is too small, specifically if $\alpha < 1/2$ in the setting of hypercube detection. Indeed, we have the following result.

Theorem 3. *In hypercube detection, the test based on $S_m(t)$, with t fixed such that $p_c < p_0(t) < 1$, is asymptotically powerful if $\alpha \geq 1/2$ and $\lim_{m\to\infty} \theta_m m^{(\alpha-1/2)d} = \infty$, and asymptotically powerless if $\alpha < 1/2$ or if $\lim_{m\to\infty} \theta_m m^{(\alpha-1/2)d} = 0$.*

Thus, for the detection of small clusters, a supercritical LOC test is potentially worthless while for larger clusters it improves substantially on the performance of a subcritical LOC test. Note that in the context of path detection, the same arguments show that the LOC test for any choice of supercritical threshold is asymptotically powerless.

4.3 Critical percolation

If our goal is to choose a threshold t so as to maximize the difference in size for the largest open cluster under the null and under an alternative, we are necessarily in the neighborhood of the percolation phase transition, which is to say that $|p - p_c|$ is small. (Again, we assume $d \geq 2$.) The percolation model is not fully understood in the critical regime, and this poses a serious obstacle to a rigorous statistical analysis. See (Grimmett, 1999, Chap. 9) for a general discussion of this percolation regime. We base our discussion on a paper of Borgs et al. (2001). Let $\pi_m(p)$ denote the probability that the open cluster at the origin reaches outside the box $[-m, m]^d$ and let $\xi(p)$ denote the correlation length, defined as

$$\frac{1}{\xi(p)} := -\lim_{m\to\infty} \frac{1}{m} \log \pi_m(p).$$

[2] Sometimes called Subbotin distributions.

Note that, with ξ defined thus, $\xi(p) < \infty$ if and only if $p < p_c$. The critical exponent for (subcritical) correlation length is postulated as

$$\nu := - \lim_{p \nearrow p_c} \frac{\log \xi(p)}{\log |p - p_c|}.$$

It is not known for all dimensions that the limit exists, but it is known that $0 < \nu < \infty$ whenever it exists. It is shown in (Borgs et al., 2001) that, subject to the existence of this limit together with other scaling assumptions, when $p = p_m$ varies with m,

$$S_m \asymp_{\mathrm{P}} \begin{cases} \log m, & \text{if, for some } \nu' > \nu,\ m^{1/\nu'}(p_m - p_c) \to -\infty, \\ m^d, & \text{if, for some } \nu' > \nu,\ m^{1/\nu'}(p_m - p_c) \to \infty, \end{cases} \tag{5}$$

where $X_m \asymp_{\mathrm{P}} Y_m$ means there exists a constant $C \in (0, \infty)$ such that $C^{-1} \leq X_m/Y_m \leq C$ in probability. The scaling assumptions of (Borgs et al., 2001) are believed to hold if and only if the number d of dimensions satisfies $2 \leq d \leq 6$, and they are proved for $d = 2$. (Borgs et al., 2001) is directed at bond percolation only, but similar results are expected for site percolation.

It is known that $\nu = 4/3$ for site percolation on the triangular lattice, see (Smirnov and Werner, 2001), and it is believed that this holds for percolation on any two-dimensional lattice. As described in (Grimmett, 1999, Sect. 10.4), it is believed that $\nu = 1/2$ for $d \geq 6$.

Subject to the assumption that (5) holds, we establish the power of the test based on $S_m(t)$ when choosing $t = t_m$ near criticality. We assume that there exists t_c such that $p_0(t_c) = p_c$, and that $p_0(t)$ is a continuous function of t in a neighborhood of t_c.

Theorem 4. *Let $t_m \geq t_c$ be such that $p_c - p_0(t_m) \asymp m^{-1/\nu'}$ for some $\nu' > \nu$. In hypercube detection and assuming that (5) holds, the test based on $S_m(t_m)$ is asymptotically powerful if $\liminf_{m \to \infty} \theta_m m^{\alpha/\nu'}$ is sufficiently large.*

Compared with a subcritical choice of threshold, which requires that θ_m be bounded away from zero for the test to have any power as seen in Theorem 1, with a near-critical choice of threshold the test is able to detect at polynomially small θ_m. In particular, with a proper choice of threshold, the test is powerful for θ_m of order $m^{-\alpha/\nu'}$ with $\nu' > \nu$. Note that, by Lemma 1, all methods are asymptotically powerless if θ_m is of order $m^{-d\alpha/2}$, implying that $\alpha/\nu \leq d\alpha/2$. We thus obtain the inequality $\nu \geq 2/d$. This may be compared with the scaling relation (Grimmett, 1999, Eq. (9.23)) stating that $d\nu = 2 - a$ where a (< 0) is the percolation critical exponent for the number of clusters per vertex. It is believed that $a = -\frac{2}{3}$ when $d = 2$, and $a = -1$ when $d \geq 6$. Compared with the performance at supercriticality, the test near-criticality (with a proper choice of threshold) is superior if $(\alpha - \frac{1}{2})d < \alpha/\nu$, which is equivalent to $\alpha < (1 - a/2)/(1 - a)$. For example, with $a = -\frac{2}{3}$, the near-critical LOC test is superior when $\alpha < \frac{3}{4}$.

5 The upper level set scan statistic

For a threshold t, let $\mathcal{Q}_m^{(t)}$ denote the (random) class of clusters within $\mathbb{V}_m$ open at t. Also, let $\mathcal{Q}_m^* = \bigcup_t \mathcal{Q}_m^{(t)}$, which is also random. Patil and Taillie (2003) suggest scanning the clusters in $\mathcal{Q}_m^*$. To ease a rigorous mathematical analysis of its performance, we consider the upper level set (ULS) scan at a given threshold t, and use the simple scan described in Section 3. Specifically, in correspondence with (1), we define the (simple) ULS scan statistic at threshold t as

$$U_m(t, k_m) = \max\left\{\sqrt{|K|}(\bar{X}_K - \mu_{0|t}) : K \in \mathcal{Q}_m^{(t)},\ |K| \geq k_m\right\}, \tag{6}$$

where $\mu_{0|t}$ (resp. $\sigma^2_{0|t}$) is the the mean (resp. variance) of $X_v|X_v > t$ when $X_v \sim F_0$, and (k_m) is a non-decreasing sequence of positive integers. The ULS scan statistic of Patil and Taillie (2003) corresponds (in its simple form) to

$$\mathrm{ULS}_m = \max_{t \in \mathbb{R}} \frac{U_m(t,1)}{\sigma_{0|t}}. \tag{7}$$

If $\mu_{0|t}$ and/or $\sigma^2_{0|t}$ are not available, we may use their empirical versions based on the X_v that survive the threshold t. We restrict the scan to clusters of size at least k_m in order to increase power, for the behavior of $U_m(t)$ is, as we shall see, completely driven by the smallest open clusters that are scanned, at least when t is subcritical. We divide the rest of our discussion in terms of subcritical, supercritical and near-critical choices of threshold. We then conclude with a result on the performance of the ULS scan test across all thresholds.

5.1 Subcritical threshold

We start by describing the behavior of $U_m(t, k_m)$ under the null. Let $F_{\theta|t}$ denote the distribution of $X_v|X_v > t$ under F_θ, and let $\mu_{\theta|t}$ and $\Lambda^*_{\theta|t}$ denote its mean and rate function, respectively. Also, when $0 < \beta < 1/\zeta_{p_\theta(t)}$, or $\beta = 0$ and $F_0 \in \mathrm{AEP}(b, C)$ for some $b \geq 2$ and $C > 0$, let $\gamma_{\theta|t}(\beta) := \gamma(F_{\theta|t}, \mu_{0|t}, \zeta_{p_\theta(t)}, \beta)$, where γ is the function defined in Lemma A.9. Note that $\gamma_{\theta|t}(\beta)$ can be computed explicitly in some cases, like the normal location model, and $\gamma_{\theta|t}(\beta) \sim (\mu_{\theta|t} - \mu_{0|t})^2/\zeta_{p_\theta(t)}$ when $\theta \nearrow \theta_c(t)$, defined (when it exists) as the solution to $p_\theta(t) = p_c$.

Lemma 5. *Assume that $\theta \geq 0$ and t is fixed such that $0 < p_\theta(t) < p_c$ and that $k_m/\log m \to d\beta$ for some $\beta \geq 0$. Then under F_θ on $\mathbb{V}_m$, the following holds in probability:*

1. *If $\beta > 1/\zeta_{p_\theta(t)}$, then $U_m(t, k_m) = 0$ for m large enough;*
2. *If $0 < \beta < 1/\zeta_{p_\theta(t)}$, then*
$$(\log m)^{-1/2} U_m(t, k_m) \to (d\gamma_{\theta|t}(\beta))^{1/2}.$$
3. *If $\beta = 0$ and $F_0 \in \mathrm{AEP}(b, C)$ for some $b \geq 1$ and $C > 0$, then*
   - *(a) If $b \geq 2$, the convergence in Part 2 applies;*
   - *(b) If $b < 2$,*
$$k_m^{1/b-1/2} (\log m)^{-1/b} U_m(t, k_m) \to (d/C)^{1/b}.$$

In the last case, where $\beta = 0$, the behavior of $U_m(t)$ is influenced by the very large deviations of $F^{*k}_{\theta|t}$ for $k \geq k_m$. (The symbol $*$ denotes convolution.) We choose to state a result for AEP distributions, for which the very large deviations resemble the large deviations.

Based on Lemma 5, we establish the performance of the ULS scan statistic. We start with arguing that choosing k_m such that $k_m/\log m \to 0$ leads to a test that may potentially have less power than the test based on the largest cluster after thresholding. Indeed, the behavior of the ULS scan statistic does not depend on θ as long as $\theta < \theta_c(t)$.

Proposition 2. *Assume that $F_0 \in \mathrm{AEP}(b, C)$ for some $b \in (1, 2)$ and $C > 0$. In hypercube detection, the test based on $U_m(t, k_m)$, with t fixed such that $0 < p_0(t) < p_c$ and $k_m/\log m \to 0$, is asymptotically powerless if $\limsup_{m \to \infty} \theta_m < \theta_c(t)$.*

For example, in the setting just described with $d = 1$, the ULS scan test has (asymptotically) no power unless $\theta_m \to \infty$, while the test based on the size of the largest cluster after thresholding is, by Theorem 1, asymptotically powerful if $\liminf_{m\to\infty} \theta_m$ is large enough. We therefore choose a sequence k_m comparable in magnitude to $\log m$ and state the performance of the ULS scan test in this case.

Theorem 5. *In hypercube detection, the test based on $U_m(t, k_m)$, with t fixed such that $0 < p_0(t) < p_c$ and $k_m/\log m \to d\beta$ with $0 < \beta < 1/\zeta_{p_0(t)}$, is asymptotically powerful if $\liminf_{m\to\infty} \theta_m > \theta_*(t)$, and asymptotically powerless if $\limsup_{m\to\infty} \theta_m < \theta_*(t)$, where $\theta_*(t)$ is the unique solution to $\alpha\, \gamma_{\theta|t}(\beta) = \gamma_{0|t}(\beta)$.*

Note that $\theta_*(t)$ is well-defined by Lemma A.10 and that $\theta_*(t) < \theta_c$ as long as $\alpha > 0$. In any case, the test based on $U_m(t, k_m)$ with a subcritical threshold t is, in the setting of hypercube detection, asymptotically powerless when $\theta_m \to 0$, just like the LOC test. In essence, the two tests are qualitatively comparable in this setting. This is also true in the context of path detection. Let $\gamma^1_{\theta|t}(\beta)$ denote $\gamma_{\theta|t}(\beta)$ in dimension one.

Theorem 6. *In path detection, the test based on $U_m(t, k_m)$, with t fixed such that $0 < p_0(t) < p_c$ and $k_m/\log m \to d\beta$ with $0 < \beta < 1/\zeta_{p_0(t)}$, is asymptotically powerful if $\liminf_{m\to\infty} \theta_m > \theta^+_*(t)$, and asymptotically powerless if $\limsup_{m\to\infty} \theta_m < \theta^-_*(t)$, where $\theta^+_*(t)$ (resp. $\theta^-_*(t)$) is the unique solution to $\alpha\, \gamma^1_{\theta|t}(\beta) = \gamma_{0|t}(\beta)$ (resp. $\alpha\, \gamma_{\theta|t}(\beta) = \gamma_{0|t}(\beta)$).*

As in Theorem 2, the result is not as sharp.

5.2 Supercritical threshold

We consider the choice of a supercritical threshold, where t is fixed such that $p_0(t) > p_c$. We already saw in Section 4.2 that the largest open cluster is unique and occupies a non-negligible fraction of the entire network. This is actually true both under the null and under an alternative. The ULS scan test based solely on the largest open cluster is comparable to the test based on the grand mean after thresholding. In turn, assuming t is fixed, this test is asymptotically powerful when $m^{(\alpha-1/2)d}\theta_m \to \infty$, and asymptotically powerless if $\alpha \le 1/2$ and θ_m is bounded. (This is easily seen using Chebyshev's inequality.) This is comparable to the LOC test at super-criticality.

In general, the ULS scan statistic includes other (smaller) open clusters. The story of the second largest cluster of supercritical percolation in a box is not yet complete, and for this reason the behavior of the ULS scan statistic is not fully understood. The difficulty arises through the possibility that the second largest cluster in $\mathbb{V}_m$ might lie at its boundary. Whether or not this occurs depends on the outcome of a calculation, yet to be done, of energy/entropy type involving so-called droplets near the boundary of $\mathbb{V}_m$ (see, for example, (Bodineau et al., 2001)). In order to simplify the discussion, we finesse this problem by working where necessary on $\mathbb{V}_m$ with *toroidal boundary conditions*. That is, whenever we make statements concerning supercritical percolation on the graph $\mathbb{V}_m$, we may add edges connecting sites on its boundary as follows: when $d = 2$, for $k = 1, 2, \dots, m$, an additional edge is placed between site $(1, k)$ and site (m, k), and similarly between $(k, 1)$ and (k, m).

In proving exact asymptotics for test statistics under the null, we shall assume toroidal boundary conditions. Our results on asymptotic power do not require such exact results but only orders of magnitude, and these do not need the toroidal assumption. We emphasize that similar exact results are expected to hold with 'free' (that is, without the extra edges) rather than toroidal boundary conditions. Once the percolation picture is better understood, such results will follow in the same manner as those presented in this paper. Our results for the torus are valid also if instead we discount open clusters that touch the boundary of $\mathbb{V}_m$ (details of this are omitted, and the proofs are essentially the same).

When working on the torus, the second largest cluster is controlled via the following calculation. It is proved in (Cerf, 2006) that the limit

$$\delta_p := -\lim_{k\to\infty} k^{-(d-1)/d} \log \mathbb{P}(\infty > S \geq k) = -\lim_{k\to\infty} k^{-(d-1)/d} \log \mathbb{P}(S = k), \tag{8}$$

exists, with $0 < \delta_p < \infty$ for all fixed $p \in (p_c, 1)$. The dependency on d is left implicit.

A result similar to Lemma 5 holds with δ_p playing the role of ζ_p and the exponent of $\log m$ changed in places. It turns out that we only need this result when $\theta = 0$. For $\beta > 0$ and a supercritical t, let $\gamma_{0|t}(\beta) := \gamma(F_{0|t}, \mu_{0|t}, 0, \beta)$ defined in Lemma A.9.

Lemma 6. *Assume that t is fixed such that $p_c < p_0(t) < 1$ and that $k_m/\log m \to d\beta$ and $k_m^{(d-1)/d}/\log m \to d\beta'$ for some $0 \leq \beta, \beta' \leq \infty$. Then under the null, the following holds in probability on the torus $\mathbb{V}_m$:*

1. *If $\beta' > 1/\delta_{p_0(t)}$, then $U_m(t, k_m) = \mathrm{O}(1)$;*
2. *If $0 \leq \beta' < 1/\delta_{p_0(t)}$ and $\beta = \infty$, then*

$$(\log m)^{-1/2} U_m(t, k_m) \to \sigma_{0|t}[2d(1 - \beta'\delta_{p_0(t)})]^{1/2},$$

where $\sigma_{0|t}^2 := \mathrm{Var}\left(F_{0|t}\right)$;

3. *If $\beta < \infty$, the conclusions of Lemma 5 apply. (Note that $\zeta_{p_0(t)} = 0$.)*

Based on Lemma 6, we obtain the following result on the performance of the ULS scan test at supercriticality. As before, we restrict ourselves to the case where $U_m(t, k_m)$ is of order $(\log m)^{1/2}$. We also chose to state a simple result instead of a more precise result with multiple sub-cases. This result holds irrespective of the type of boundary condition assumed on $\mathbb{V}_m$.

Theorem 7. *In hypercube detection, the test based on $U_m(t, k_m)$, with t fixed such that $p_c < p_0(t) < 1$ and $\liminf k_m/\log m > 0$ and $\limsup k_m^{(d-1)/d}/\log m < \alpha d/\delta_{p_0(t)}$, is asymptotically powerful (resp. powerless) if*

$$\theta_m \left[m^{(\alpha-1/2)d} + (\log m)^{d/(2d-2)} \right] (\log m)^{-1/2} \to \infty \quad \textit{(resp. } \to 0\textit{)}.$$

We also mention that the equivalent of Theorem 6 holds here as well.

5.3 Critical threshold

If we choose a threshold as described in Section 4.3, and if (5) is true, then the power of the ULS scan statistic is greatly improved, indeed, as in the case of the LOC test. In fact, one can prove that Theorem 4 remains valid with $S(t_m)$ replaced with $U_m(t_m, k_m)$, as long as $k_m = o(m)^{\alpha d}$ so that the largest open cluster under the alternative is scanned. This boils down to showing that, under the null, the ULS scan statistic is at most a power of $\log m$, which we do in Lemma 7 below. The ULS scan test, however, does not seem to offer any substantial gain in power over the LOC test, as θ_m is still required to be large enough to change the regime of the percolation process within an alternative K from subcritical to supercritical. That said, actually proving this would require information on the smaller open clusters near criticality, which is scarce (and very difficult to obtain)—see Borgs et al. (2001) for some partial results and postulates.

5.4 Across all thresholds

Finally, we discuss the (simple) ULS scan test across all thresholds, as suggested in Patil and Taillie (2003). So as to take advantage of a phase transition near-criticality, we assume as in Section 4.3 that there exists t_c such that $p_0(t_c) = p_c$, and that $p_0(t)$ is a continuous function of t in a neighborhood of t_c. Also, we assume that (5) holds. In Proposition 2, we saw that scanning small clusters may lead to a decrease in power. For this reason, and also to facilitate the analysis, we limit ourselves to clusters of size at least k_m, i.e., we consider the test based on test based on

$$\text{ULS}_m(k_m) = \max_{t \in \mathbb{R}} \frac{U_m(t, k_m)}{\sigma_{0|t}}, \tag{9}$$

where, for definiteness, $U_m(t, k_m)$ is calculated on the torus $\mathbb{V}_m$ when $t < t_c$.

Let $\Gamma_\theta(\beta) = \inf_t \gamma_{\theta|t}(\beta)/\sigma^2_{0|t}$ where, in congruence with Sections 5.1 and 5.2,

$$\gamma_{\theta|t}(\beta) = \begin{cases} \gamma(F_{\theta|t}, \mu_{0|t}, \zeta_{p_\theta(t)}, \beta), & t > t_c \\ \gamma(F_{\theta|t}, \mu_{0|t}, 0, \beta), & t < t_c, \end{cases}$$

with γ being the function defined in Lemma A.9. We first establish the behavior of $\text{ULS}_m(k_m)$ under the null.

Lemma 7. *Let $k_m = \beta \log m$ where $\beta > 0$, and let t_β be such that $d/\beta \le \zeta_{p_0(t_\beta)} < \infty$. Define $\eta(\beta) := \sup\{\sigma_{0|t}/\sigma_{0|s} : s \le t \le t_\beta\}$. Then, under F_0,*

$$(d\Gamma_0(\beta))^{1/2} \le (1 + o_P(1))(\log m)^{-1/2}\text{ULS}_m(k_m) \le \eta(\beta)(d\Gamma_0(\beta))^{1/2}.$$

If in addition, either $\sigma_{0|t}$ is non-decreasing in t or F_0 has no atoms on $(-\infty, t_\beta]$, then in probability under F_0,

$$(\log m)^{-1/2}\text{ULS}_m(k_m) \to (d\Gamma_0(\beta))^{1/2}.$$

The term $o_P(1)$ denotes a random variable that converges to 0 in probability. See the formal definition at the start of the appendix. In fact, a result as precise as Lemma 7 is superfluous, given the behavior of the ULS scan statistic under the alternative at supercriticality and near-criticality, which is polynomial in m. The next theorem does not require the use of toroidal boundary conditions.

Theorem 8. *In hypercube detection and assuming that (5) holds, the test based on $\text{ULS}_m(k_m)$, with $k_m = [\beta \log m]$ for some $\beta > 0$, is asymptotically powerful if $\theta_m m^\lambda \to \infty$, for some $0 < \lambda < \alpha/\nu$ satisfying $\lambda < (\alpha - 1/2)d$ if $\alpha > 1/2$.*

Hence, scanning all thresholds brings the best performance out. Of course, we only need to select as thresholds the distinct node values, and in fact, the lower bound on the cluster size imposes an (implicit) upper bound on the threshold (at least under the null). We mention in passing that the same result holds for the simpler test which only scans the largest open cluster at each threshold.

6 Implementation and numerical experiments

The scan test has been shown to be near-optimal in a wide variety of settings, differing both in terms of network structure and cluster class (Arias-Castro et al., 2011, 2005). It is, however, computationally demanding. For the simple situation of detecting a hypercube, the scan statistic can be computed in $O(N \log N)$ flops, where $N := m^d$ is the network size, if the size of the hypercube is known. If one scans over all possible hypercubes, then computing the scan statistic requires $O(N^2 \log N)$ flops. For non-parametric shapes, the computational cost is even higher. In fact, for the problem of detecting a loopless path, computing the scan statistic corresponds to the Reward-Budget Problem of DasGupta et al. (2006), shown there to be NP-hard. Because the scan statistic is so computationally burdensome, the cluster class is most often taken to be parametric in practice, even though the underlying clusters may take a much wider range of shapes. For instance, discs are the prevalent shape used in disease outbreak detection (Kulldorff and Nagarwalla, 1995), with variants such as ellipses (Hobolth et al., 2002; Kulldorff et al., 2006). For a wide range of parametric shapes, Arias-Castro et al. (2005) recommend a multiscale approximation to the scan statistic. Efforts to move beyond parametric models include tree-based approaches (Kulldorff et al., 2003), simulated annealing (Duczmal and Assunção, 2004) and an exhaustive search among arbitrarily shaped clusters of small size (Tango and Takahashi, 2005).

The LOC test does not assume any parametric form for the anomalous cluster, and in that sense, it is nonparametric. Its computational complexity at a given threshold is of order the number of nodes plus the number of edges in the network (Cormen et al., 2009), so of order $O(N)$ flops for the square lattice.

The ULS scan statistic is also nonparametric. Computing $U_m(t, k_m)$ requires determining $\mathcal{Q}_m^{(t)}$, which we saw takes $O(N)$ flops, and then scanning over $\mathcal{Q}_m^{(t)}$. Since the clusters in $\mathcal{Q}_m^{(t)}$ do not intersect, scanning over them takes order $O(N)$ flops. Therefore, computing ULS_m can be done in $O(M \cdot N)$ flops, where M is the number of distinct values at the nodes. Patil and Taillie (2004) argue that this can be done faster, using the tree structure of $\mathcal{Q}_m^*$, where the root is the entire network $\mathbb{V}_m$ and a cluster $K \in \mathcal{K}_m(t_j)$ is the parent of any cluster $L \in \mathcal{K}_m(t_{j+1})$ such that $L \subset K$, where $t_1 < \cdots < t_M$ denote the distinct values at the nodes.

We complement our theoretical analysis with some small-scale numerical experiments. Specifically, we explore the power properties of the LOC test of Section 4 and the ULS scan test of Section 5 in the context of detecting a hypercube in the two-dimensional square lattice. Patil et al. (2005) are developing sophisticated software implementing the ULS scan statistic for use in real-life situations, with more recent variations Patil et al. (2010). However, this software is not yet available, so we implemented our own (basic) routines.

We used the statistical software `R`[3] with the package `igraph`[4]. Our (basic) implementation of the ULS scan statistic for a given threshold is much slower than both the scan statistic with a given mask and the LOC statistic, especially when there is no constraint on the size of the open clusters to be scanned, i.e., when $k_m = 1$. In all our experiments, we chose the square lattice in dimension $d = 2$ with side-length $m = 500$ for a total of 250,000 nodes, and we considered three alternatives, specifically, squares of side length $\ell \in \{10, 50, 100\}$, corresponding roughly to $\alpha \in \{0.4, 0.7, 0.8\}$. The squares were fixed, away from the boundary of the lattice, as the methods are essentially location independent. (This is rigorously true of the scan statistic.) We assessed the performance of a method in a given situation by estimating its risk, which we define as the sum of the probabilities of type I and type II errors optimized over all rejection regions.

We first ran some experiments to quickly assess the power of the scan test and it agrees very well with the theory, i.e., Lemma 3, though we knew that from previous experience. Specifically, we assumed a normal location model and simulated 100 realizations of the null and each of the three alternatives with $\theta \in \{j/\ell : j = 1, 3, 5, 7, 9\}$. See Figure 2.

Next, we performed some larger experiments to assess the power of the LOC test. We simply

[3]The R Project for Statistical Computing {`http://www.r-project.org`}
[4]The igraph library, by G. Csardi {`http://igraph.sourceforge.net`}

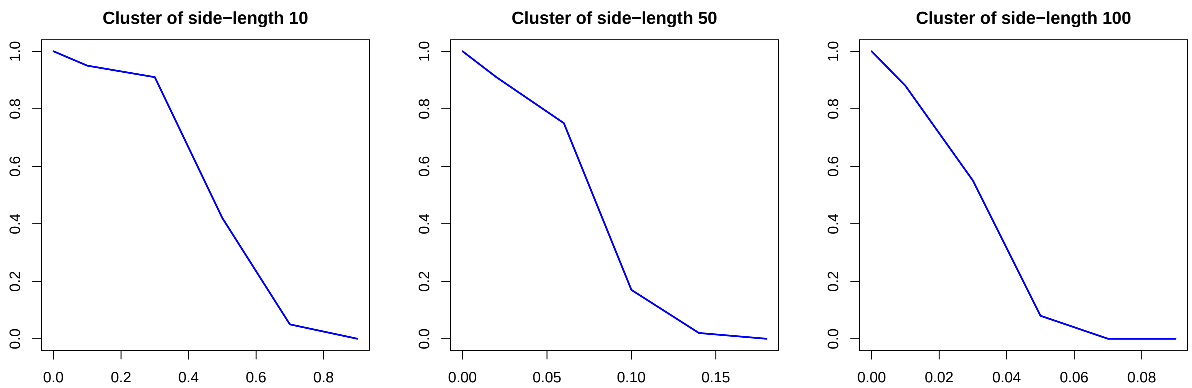

Figure 2: The risk of the scan test against each of the three alternatives. On the x-axis is θ, and the y-axis is the estimated risk based on 100 replicates.

assumed a site percolation model with probability $p \in \{0.05, 0.10, \ldots, 0.90, 0.95\}$. Note that p_c is not known for site percolation in the square lattice, though $p_c \approx 0.5927460$ from extensive numerical experiments (Feng et al., 2008). We simulated the null and each one of the three alternatives with a probability $q \in \{0.05, 0.10, \ldots, 0.90, 0.95\}$, $q > p$, within the anomalous cluster. We replicated each situation 1,000 times. The risk curves are pictured in Figure 3. The test seems to behave similarly above and below criticality. Near criticality, the test is rather erratic. However, when the size of the anomalous cluster is large enough, $\ell = 100$, the risk curve is steepest just under p_c, at $p = 0.55$ in our experiments, with full power against $q \geq 0.65$. In Figure 4 are boxplots of the test statistic for the case of $\ell = 100$ and $p = 0.40$ (subcritical), $p = 0.55$ (near-critical) and $p = 0.70$ (supercritical).

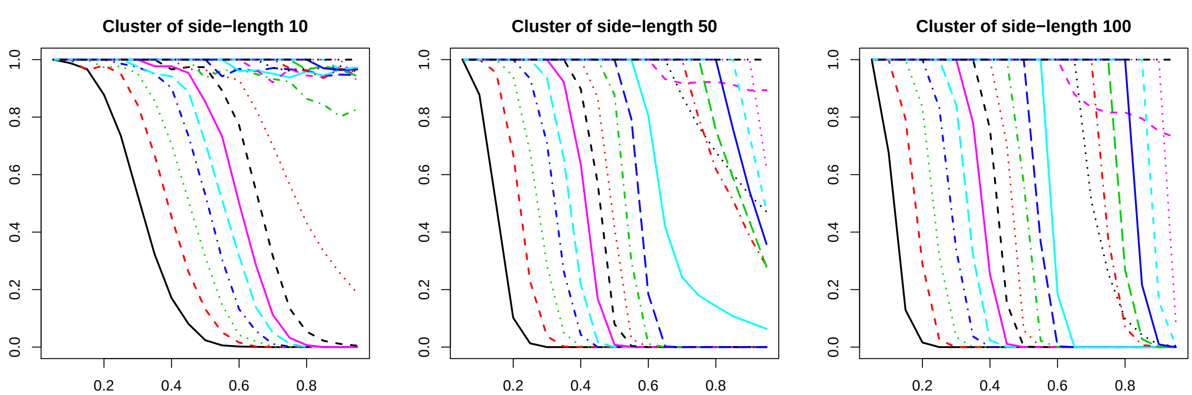

Figure 3: The risk of the LOC test against each of the three alternatives. On the x-axis is the percolation probability q on the anomalous cluster, and the y-axis is the estimated risk based on 1,000 replicates. Each curve corresponds to a different percolation probability p.

If we were to use this test in the context of a normal location model, the correspondence would be $t = \bar{\Phi}^{-1}(p)$ (the threshold) and $\theta = t - \bar{\Phi}^{-1}(q)$, where $\bar{\Phi}$ denotes the normal survival distribution function. In Figure 5 we plot the risk curves in this context for $p \in \{0.40, 0.50, 0.55, 0.60, 0.70\}$. In particular, the test near criticality with $t = \bar{\Phi}^{-1}(0.55) = -0.126$ has full power against the alternative with $\ell = 100$ and $\theta = 0.26$.

Last, we experimented with the ULS scan test. To limit the size of our simulations, we considered

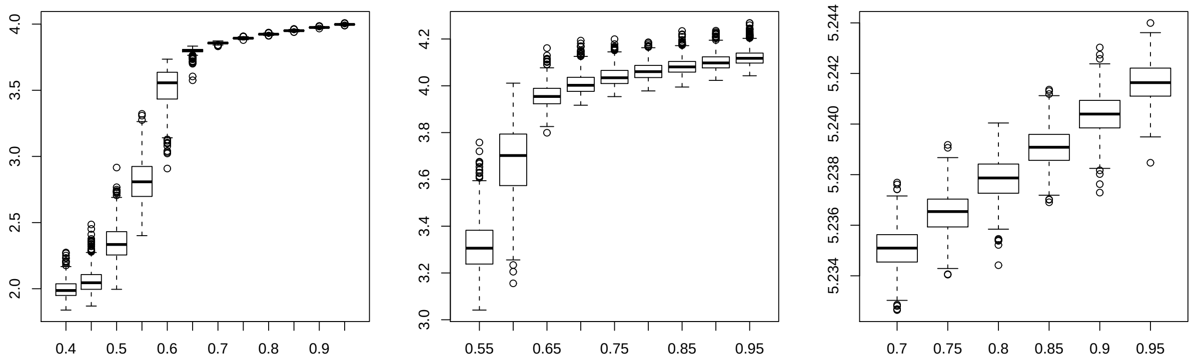

Figure 4: The size of the largest open cluster in $\log_{10}$ scale (y-axis) versus the percolation probability q, for the alternative $\ell = 100$ and $p \in \{0.40, 0.55, 0.70\}$ (from left to right). Each boxplot represent 1,000 replicates.

alternatives with $\theta = \Phi^{-1}(q)$ with $q \in \{0.55, 0.6, 0.65, 0.70, 0.80, 0.90\}$ and chose as thresholds $t = \Phi^{-1}(p)$ with $p \in \{0.40, 0.50, 0.55, 0.60, 0.70\}$. We restricted scanning to open clusters of size not smaller than a tenth (1/10) of the size of largest open cluster, essentially falling in the regime of Part 2 of Lemma 5 and also making the computations much faster. We used 200 replicates. We observe again that the risk curve is sharpest near criticality when the size of the anomalous cluster is sufficiently large, here for $\ell \geq 50$. Compared with the LOC test, the ULS scan test has more power at large θ when the cluster is small $\ell = 10$ (as predicted) and, more interestingly, a little more power when the cluster is larger. Compared to the scan statistic, which knows the size and shape of the anomalous cluster, the ULS scan test with the best choice of threshold (corresponding to $p = 0.55$) requires about three times more signal amplitude.

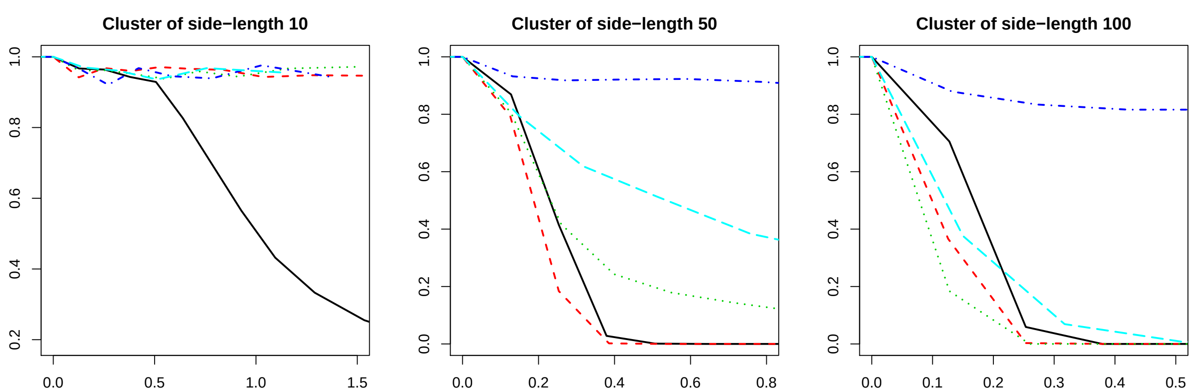

Figure 5: The risk of the LOC test in the context of a normal location model. On the x-axis is θ, and on the y-axis is the estimated risk based on 1,000 replicates. Each curve corresponds to a different threshold t. The black, red, green, blue and cyan curves correspond to $p = 0.40, 0.50, 0.55, 0.60, 0.70$, respectively.

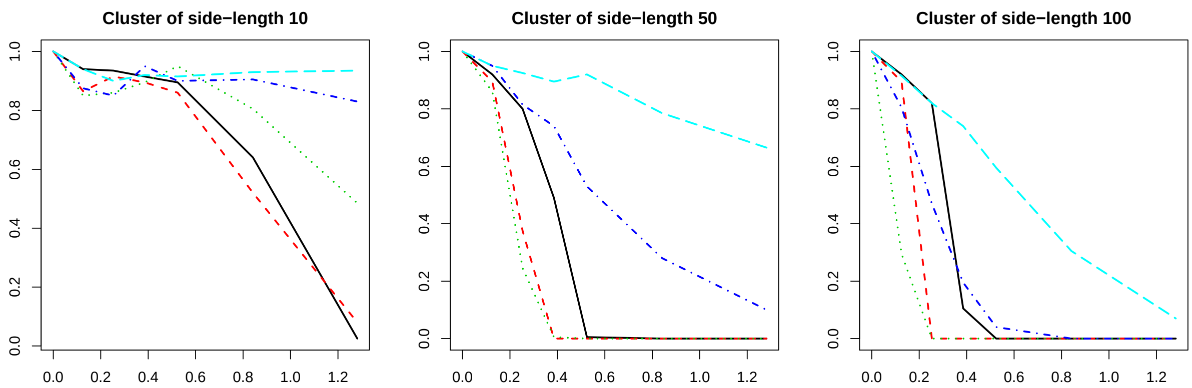

Figure 6: The risk of the ULS scan test against each of the three alternatives. On the x-axis is θ, and on the y-axis is the estimated risk based on 200 replicates. Each curve corresponds to a different threshold t. The black, red, green, blue and cyan curves correspond to $p = 0.40, 0.50, 0.55, 0.60, 0.70$, respectively.

7 Discussion

Our contribution in this paper is a rigorous mathematical analysis of the performance of the LOC test—independently of, and more extensively than (Davies et al., 2010; Langovoy and Wittich, 2011), and of the ULS scan test, both nonparametric and computationally tractable methods. We made abundant use of percolation theory to establish these results. We compared their power with that of the scan statistic, known to be near-optimal in a wide array of settings. While they are comparable in power with the scan statistic for the detection of a path, these tests may be substantially less powerful for the detection of a hypercube. Note however that the scan statistic is provided with knowledge about the shape and size of the anomalous cluster (though the latter is not as important). In theory, we argued that this was the case based on some heuristics and conjectures from percolation theory. Numerically, this appears to be the case when the anomalous cluster is large enough. In our experiments, the ULS scan test was slightly more powerful than the LOC test, and required a θ three to four times larger compared to the scan statistic, the latter having the advantage of knowing the shape and size of the cluster. This is promising and further numerical experiments are needed to evaluate the power of these tests in truly nonparametric settings.

Our theoretical results generalize to other networks that resemble the lattice, with a different critical percolation probability p_c and different functions ζ_p and δ_p. In particular, we used the self-similarity property of the square lattice and the fact that it has polynomial growth. Our results also generalize to other cluster classes. In the setting of the square lattice, our results extend immediately to any class of clusters that include a hypercube of comparable size, e.g., the class $\mathcal{K}_m$ of clusters K of size $|K| = [m^\alpha]^d$ such that there is a hypercube $K_0 \subset K$ with $|K_0|/|K| \geq \omega_m$, where $\omega_m \to 0$ slower than any negative power of m. Also, the class may contain clusters of different sizes, though in that case the worst-case risk is driven by the smallest clusters. (The implementation of the scan statistic may be much more demanding in this case.) The main results of Section 4 only require that $F_\theta(t)$ is twice differentiable in (t, θ), with $\partial_\theta F_\theta(t) < 0$ for all (t, θ), which is for example the case for location models and scale models if F_0 is twice differentiable with a strictly positive first derivate. With some additional work, we may also obtain results for classes of 'thin' clusters as defined in (Arias-Castro et al., 2011). The key is to understand the percolation behavior within and near such clusters. Some results are available for slabs (Grimmett, 1999, Thm 7.2) and more general subgraphs of lattices including 'wedges', and these appear transferable to other 'curved' slabs.

Appendix

Notation

We write $f_m \sim g_m$ as $n \to \infty$ if $f_m/g_m \to 1$. Similarly, we use $\mathrm{O}()$ and $\mathrm{o}()$ notation, and write $f_m \asymp g_m$ as $n \to \infty$ if $f_m = \mathrm{O}(g_m)$ and vice-versa. We will also use their random counterparts, $\sim_{\mathrm{P}}$, $\asymp_{\mathrm{P}}$, $\mathrm{O}_{\mathrm{P}}()$ and $\mathrm{o}_{\mathrm{P}}()$. For example, $Z_m = \mathrm{o}_{\mathrm{P}}(k_m)$ means that $Z_m/k_m \to 0$ in probability and $Z_m = \mathrm{O}_{\mathrm{P}}(k_m)$ means that Z_m/k_m is bounded in probability, which is to say that $\mathbb{P}(|Z_m| \geq k_m l_m) \to 1$ as $m \to \infty$, for any l_m satisfying $l_m \to \infty$. We use $1\{A\}$ to denote the indicator function of the set A. The maximum of k and ℓ is denoted $k \vee \ell$.

A.1 On the size of percolation clusters

We state and prove a couple of results on the sizes of percolation clusters in $\mathbb{Z}^d$. We start by proving some properties of ζ_p. Recall that S denotes the size of the open cluster at the origin. Besides the limit in (3), the following bound holds for $p < p_c$ and all $k \geq 1$:

$$\mathbb{P}_p(S \geq k) \leq (1-p)^2 \frac{k e^{-k\zeta_p}}{(1-e^{-\zeta_p})^2}, \tag{A.1}$$

by (Grimmett, 1999, Eq. (6.80)) adapted to site percolation.

Lemma A.1. *The function ζ_p defined in (3) is continuous and strictly decreasing over $(0, p_c]$, with $\lim_{p\to 0} \zeta_p = \infty$ and $\lim_{p \to p_c} \zeta_p = 0$.*

Proof. Let $0 \leq p < p' \leq 1$. By coupling $\mathbb{P}_p$ and $\mathbb{P}_{p'}$ in the usual way,

$$\mathbb{P}_p(S = k) \geq (p/p')^k \mathbb{P}_{p'}(S = k),$$

so that $\zeta_p \leq \zeta_{p'} + \log(p'/p)$. Applying (Grimmett, 1999, Thm 2.38), to the event $\{S \geq k\}$, we find as in the proof of (Grimmett, 1999, Eq. (6.16)) that $\zeta_p / \log p \leq \zeta_{p'} / \log p'$. In summary,

$$\zeta_p \left(1 - \frac{\log(1/p')}{\log(1/p)}\right) \leq \zeta_p - \zeta_{p'} \leq \log(p'/p). \tag{A.2}$$

Therefore, ζ_p is continuous, and is strictly decreasing on $(0, p_c)$. Moreover, by fixing $p' \in (0, p_c)$ and letting $p \to 0$, we have

$$\zeta_p \geq \zeta_{p'} \frac{\log(1/p)}{\log(1/p')} \to \infty.$$

Finally, by (Grimmett, 1999, Eq. (6.83), (6.56)), $\zeta_p \to 0 = \zeta_{p_c}$ as $p \uparrow p_c$. □

We include next the proof of (4). This is done by standard means and the claim may be strengthened, see also (Grimmett, 1985; Hofstad and Redig, 2006).

Lemma A.2. *Consider site percolation on $\mathbb{Z}^d$ with parameter $p < p_c$, and let S_m denote the size of the largest open cluster within $\mathbb{V}_m$. Then (4) holds, namely,*

$$\frac{S_m}{\log m} \to \frac{d}{\zeta_p}, \quad \textit{in probability.}$$

Proof. Fix $0 < \varepsilon < 1/2$. Let S^v be the size of the open cluster at a node $v \in \mathbb{Z}^d$, which has the same distribution as S. We start with the upper bound. By (3) and the union bound,

$$\mathbb{P}\left(S_m \geq k\right) \leq \sum_{v \in \mathbb{V}_m} \mathbb{P}\left(S^v \geq k\right) = |\mathbb{V}_m| \cdot \mathbb{P}\left(S \geq k\right). \tag{A.3}$$

Hence, for $k_m(\varepsilon) := (1+\varepsilon)(d/\zeta_p)\log m$ and m large enough,

$$\mathbb{P}\left(S_m \geq k_m(\varepsilon)\right) \leq m^d \exp(-(1-\varepsilon/2)\zeta_p k_m(\varepsilon)) \leq m^{-\varepsilon d/4},$$

and the term on the right-hand side converges to zero.

For the lower bound, consider $N = \lceil m^d/(\log m)^{2d} \rceil$ nodes $v_1, \ldots, v_N \in \mathbb{V}_m$ separated from each other and the boundary of $\mathbb{V}_m$ by at least $\frac{1}{2}(\log m)^2$. Let $k_m(\varepsilon) := (1-\varepsilon)(d/\zeta_p)\log m$. For sufficiently large m, the events $E_i := \{|S^{v_i}| \leq k_m(\varepsilon)\}$ are independent. Therefore, for large m,

$$\begin{aligned} \mathbb{P}\left(S_m \leq k_m(\varepsilon)\right) &\leq \left(1 - \mathbb{P}\left(S \geq k_m(\varepsilon)\right)\right)^N \\ &\leq \left(1 - \exp(-(1+\varepsilon/2)\zeta_p k_m(\varepsilon))\right)^N \\ &\leq \exp(-m^{\varepsilon d/2}/(\log m)^{2d}), \end{aligned} \tag{A.4}$$

and the last term on the right-hand side tends to zero as $m \to \infty$. ☐

The following result describes the behavior of size of the open cluster at the origin when p is small. It may be made more precise, but this is not pursued here.

Lemma A.3. *There exists $c > 0$ depending only on d such that, for $p \in (0, (2c)^{-1})$,*

$$p^k \leq \mathbb{P}_p(S \geq k) \leq \tfrac{1}{2}(cp)^k, \quad \forall k \geq 1.$$

Proof. An *animal* is a connected subgraph of $\mathbb{Z}^d$ containing the origin. The lower bound comes from considering the probability that any given animal of size k is open. For the upper bound, by the union bound we have $\mathbb{P}_p(S = k) \leq |\mathcal{A}_k| p^k$, where $\mathcal{A}_k$ is the set of animals with k vertices. There is a constant $c > 0$ such that $|\mathcal{A}_k| \leq c^k$, so that

$$\mathbb{P}_p(S \geq k) \leq \sum_{\ell \geq k} c^\ell p^\ell = \frac{(cp)^k}{1 - cp} \leq \tfrac{1}{2}(cp)^k,$$

when $cp < \frac{1}{2}$. ☐

Next is a result on the number of open clusters of a given size and is valid for all $p \in (0,1)$.

Lemma A.4. *Consider site percolation on $\mathbb{Z}^d$ with parameter p, and let $N_m(k)$ denote the number of open clusters of size k within $\mathbb{V}_m$. Then, for $k \geq 1$,*

$$\frac{(m-2k)^d}{k} \mathbb{P}\left(S = k\right) \leq \mathbb{E}\left(N_m(k)\right) \leq \frac{m^d}{k} \mathbb{P}\left(\infty > S \geq k\right),$$

In addition, for $k, \ell \geq 1$,

$$\left|\mathrm{Cov}\left(N_m(k), N_m(\ell)\right)\right| \leq 3^{d+1}(k+\ell)^d \mathbb{E}\left(N_m(k \vee \ell)\right).$$

Hence, for $k \geq 1$,

$$\mathrm{Var}\left(N_m(k)\right) \leq 6^{d+1} k^d \mathbb{E}\left(N_m(k)\right).$$

Proof. Let S_m^v be the size of the open cluster at v within the box $\mathbb{V}_m$. Then

$$N_m(k) = \sum_{v \in \mathbb{V}_m} X^v(k), \tag{A.5}$$

where $X^v(k) = k^{-1} 1\{S_m^v = k\}$. We immediately have

$$\mathbb{E}\left(N_m(k)\right) \leq \sum_{v \in \mathbb{V}_m} \frac{1}{k} \mathbb{P}\left(\infty > S^v \geq k\right) = \frac{|\mathbb{V}_m|}{k} \mathbb{P}\left(\infty > S \geq k\right).$$

For the lower bound, we count only nodes away from the boundary, obtaining

$$\mathbb{E}\left(N_m(k)\right) \geq |\mathbb{V}_m(k)| \frac{1}{k} \mathbb{P}\left(S = k\right),$$

where $\mathbb{V}_m(k) := \{k, \ldots, m-k\}^d$.

We turn now to the covariances. By (A.5),

$$\begin{aligned}\mathrm{Cov}\left(N_m(k), N_m(\ell)\right) &= \sum_{v,w \in \mathbb{V}_m} \mathrm{Cov}\left(X^v(k), X^w(l)\right) \\ &= \sum_{\substack{v,w \in \mathbb{V}_m \\ \|v-w\| \leq k+\ell}} \mathrm{Cov}\left(X^v(k), X^w(l)\right),\end{aligned}$$

since $X^v(k)$ and $X^w(\ell)$ are independent if $\|v-w\| > k+\ell$, where $\|\cdot\|$ denotes ℓ^∞-norm. Now,

$$\begin{aligned}\left|\mathrm{Cov}\left(X^v(k), X^w(\ell)\right)\right| &= \left|\mathbb{E}\left(X^w(\ell) \mid X^v(k) = k^{-1}\right) - \mathbb{E}\left(X^w(\ell)\right)\right| \mathbb{E}\left(X^v(k)\right) \\ &\leq \frac{1}{\ell} \mathbb{E}\left(X^v(k)\right),\end{aligned}$$

so that

$$\left|\mathrm{Cov}\left(N_m(k), N_m(\ell)\right)\right| \leq \frac{1}{\ell}(2k+2\ell+1)^d \mathbb{E}\left(N_m(k)\right),$$

and the second claim of the lemma follows. □

We now describe some properties of the open clusters within $\mathbb{V}_m$ in the supercritical regime. In this regime, it is known that, with probability one, there is a unique infinite open cluster in $\mathbb{Z}^d$, denoted Q_∞, see for example (Grimmett, 1999, Sec. 8.2). With high probability, the largest open cluster within $\mathbb{V}_m$ is a subgraph of this infinite open cluster. And we display below some additional information on its size S_m.

Lemma A.5. *Suppose that $p > p_c$. There is a constant $C > 0$ such that, with probability at least $1 - \exp(-Cm^{d-1})$, there is a unique largest open cluster within $\mathbb{V}_m$, and it is a subgraph of Q_∞. Moreover, as $m \to \infty$, its size S_m satisfies*

$$\frac{S_m - \mathbb{E}\left(S_m\right)}{\sqrt{\mathrm{Var}\left(S_m\right)}} \to \mathcal{N}(0,1), \quad \text{in distribution,}$$

with $\mathbb{E}\left(S_m\right) \sim \Theta_p |\mathbb{V}_m|$ and $\mathrm{Var}\left(S_m\right) \sim \sigma^2 |\mathbb{V}_m|$ for some $\sigma^2 > 0$ depending on (d,p).

Proof. For the first part and the limiting behavior of $\mathbb{E}(S_m)$ as $m \to \infty$, see the discussion of (Penrose and Pisztora, 1996, Th. 4 and Th. 6) and the beginning of this section. For the weak limit and the limit size of the variance of S_m, see for example (Penrose, 2001, Th. 3.2). □

Next, we describe some properties of the smaller open clusters. Let $S_m^{(2)}$ be the size of the largest open cluster of $\mathbb{Z}^d$ that is contained entirely within $\mathbb{V}_m$.

Lemma A.6. *Suppose that $p > p_c$. There exists a positive constant δ_p such that*

$$\frac{S_m^{(2)}}{(\log m)^{d/(d-1)}} \to \left(\frac{d}{\delta_p}\right)^{d/(d-1)}, \quad \textit{in probability.}$$

For any $c > 0$ there exists $\sigma_i = \sigma_i(p, c) > 0$ such that, the following holds. With probability tending to one, there exist at least $\sigma_1 m^d \exp[-\sigma_2 (\log m)^{(d-1)/d}]$ open clusters of size $[c \log m]$ of $\mathbb{Z}^d$ lying within $\mathbb{V}_m$.

Our results on exact asymptotics in the supercritical phase concern $\mathbb{V}_m$ with toroidal boundary conditions. One effect of the removal from $\mathbb{V}_m$ of its boundary is that the asymptotics of the largest cluster coincide with those of S_m, and similarly for the second largest cluster $S_m^{(2)}$. In the proof of Theorem 7, we shall need an upper bound on the size of the second largest cluster inside a box with 'free' boundary conditions. We do not explore this in detail here, since it relies on extensions of arguments of (Kesten and Zhang, 1990)—see also (Grimmett, 1999, Proof of Thm 8.65)—that have not been not fully explored in the literature. Instead, we note that the the second largest open cluster in a supercritical percolation model on $\mathbb{V}_m$ with free boundary conditions has size of order $\mathrm{O_P}((\log m)^{d/(d-1)})$.

Proof. It was proved in (Cerf, 2006) that the limit

$$\delta_p := -\lim_{k\to\infty} k^{-(d-1)/d} \log \mathbb{P}(S = k) \tag{A.6}$$

exists and is strictly positive and finite when $p_c < p < 1$. It is an elementary exercise that δ_p defined thus is equal to that of (8). See also (Grimmett, 1999, Sect. 8.6). The first part of the lemma follows by the same proof as used in Lemma A.2.

As in the proof of Lemma A.4, the mean number μ_m of clusters of size $k := [c \log m]$ satisfies

$$\frac{m^d}{c \log m} \exp\left(-\delta^1 (c \log m)^{(d-1)/d}\right) \le \mu_m \le \frac{m^d}{[c \log m]} \exp\left(-\delta^2 (c \log m)^{(d-1)/d}\right)$$

for positive constants δ^i. The number of such clusters has variance no larger than $Ck^d \mu_m$ for some $C < \infty$. The claim follows by Chebyshev's inequality. □

A.2 Some distributional properties

We gather here some results for AEP and exponential families of distributions. Our first result is on the size of the maximum of an i.i.d. sample from an AEP distribution.

Lemma A.7. *Let $F \in \mathrm{AEP}(b, C)$ for some $b > 0$ and $C > 0$. Then, for $X_1, \dots, X_n \overset{\text{iid}}{\sim} F$,*

$$\frac{\max(X_1, \dots, X_n)}{(\log n)^{1/b}} \to C^{-1/b}, \quad \textit{in probability.}$$

Proof. Fix $\varepsilon \in (0,1)$ and define $x_n(\varepsilon) = ((1-\varepsilon)(\log n)/C)^{1/b}$. For n large enough, we have, by independence,

$$\begin{aligned}\mathbb{P}\left(\max(X_1,\ldots,X_n) \le x_n(\varepsilon)\right) &\le (1-\bar{F}(x_n(\varepsilon)))^n \\ &\le (1-\exp(-(1+\varepsilon)Cx_n(\varepsilon)^b))^n \\ &\le \exp(-n^{\varepsilon^2}) \to 0.\end{aligned}$$

Now, redefine $x_n(\varepsilon) = ((1+\varepsilon)(\log n)/C)^{1/b}$. For n large enough, we have, by the union bound,

$$\begin{aligned}\mathbb{P}\left(\max(X_1,\ldots,X_n) \ge x_n(\varepsilon)\right) &\le n\bar{F}(x_n(\varepsilon)) \\ &\le n\exp(-(1-\varepsilon/3)Cx_n(\varepsilon)^b) \\ &\le n^{-\varepsilon/3} \to 0.\end{aligned}$$

□

Next, we describe the behavior at infinity of the logarithmic moment generating function and rate function of an AEP distribution.

Lemma A.8. *Let $F \in \mathrm{AEP}(b,C)$ for some $b \ge 1$ and $C > 0$, with logarithmic moment generating function Λ and rate function Λ^*. Then, as $\theta \to \infty$,*

$$\theta^{-\frac{b}{b-1}}\Lambda(\theta) \to C(b-1)(Cb)^{-\frac{b}{b-1}}, \quad b > 1; \tag{A.7}$$

$$(\log(1/(C-\theta)))^{-1}\Lambda(\theta) \to 1, \qquad b = 1; \tag{A.8}$$

and, as $x \to \infty$,

$$x^{-b}\Lambda^*(x) \to C. \tag{A.9}$$

Proof. Assume F is continuous for concreteness and let φ denote its moment generating function. We focus on the upper bound in (A.7)—obtaining the bound in (A.8) is analogous—and deduce the lower bound in (A.9). Let $b > 1$, $C/2 < A < C$, and let $x_1 > 0$ be such that $\bar{F}(x) \le \exp(-Ax^b)$ for all $x > x_1$. We start from the following bound

$$\varphi(\theta) = \int_{-\infty}^{\infty} \theta\exp(\theta x)\bar{F}(x)dx \le \exp(\theta x_1) + \int_{x_1}^{\infty} \theta\exp(\theta x - Ax^b)dx.$$

We again divide the integral into $x \le x_2$ and $x > x_2$, where $x_2 := (2\theta/A)^{\frac{1}{b-1}}$. For $x \le x_2$, we bound $\exp(\theta x - Ax^b)$ by its maximum over $(0,\infty)$. For $x > x_2$, $\exp(\theta x - Ax^b) \le \exp(-(C/4)x^b)$. Letting $B = A(b-1)(Ab)^{-\frac{b}{b-1}}$ and assuming θ is large enough that $x_2 > x_1$, we get

$$\int_{x_1}^{\infty} \theta\exp(\theta x - Ax^b)dx \le (x_2 - x_1)\theta\exp\left(B\theta^{\frac{b}{b-1}}\right) + \theta\int_{x_2}^{\infty}\exp(-(C/4)x^b)dx.$$

Hence, when $\theta \to \infty$,

$$\varphi(\theta) = \mathrm{O}(\theta^{\frac{b}{b-1}})\exp\left(B\theta^{\frac{b}{b-1}}\right). \tag{A.10}$$

Taking logs and letting $\theta \to \infty$, we get

$$\limsup_{\theta\to\infty} \theta^{-\frac{b}{b-1}}\Lambda(\theta) \le A(b-1)(Ab)^{-\frac{b}{b-1}}.$$

Then letting A tend to C, we obtain the upper bound in (A.7).

Now, for x exceeding the mean of F, $\Lambda^*(x) = \sup_{\theta \geq 0}(\theta x - \Lambda(\theta))$, and starting from (A.10), we obtain

$$\Lambda^*(x) \geq \sup_{\theta \geq 0}(\theta x - B\theta^{\frac{b}{b-1}}) - \log 2 = Ax^b - \log 2.$$

Therefore,

$$\varliminf_{x \to \infty} x^{-b}\Lambda^*(x) \geq A.$$

Then letting A tend to C, we obtain the lower bound in (A.9). □

We now define γ, first appearing in Section 5.1. Our function γ depends on certain quantities listed in the following lemma. It depends also on the quantity ζ, which we will take as that defined in (3). It is only through its dependence on ζ that γ is affected by the geometry of $\mathbb{V}_m$.

Lemma A.9. *Consider a distribution F on the real line, possibly discrete but not a point mass, with finite mean μ and finite moment generating function at some positive $\theta > 0$, and let Λ^* denote its rate function. Let $\nu \leq \mu$, and fix $\beta, \zeta \in [0, \infty)$.*

1. *Assume $\zeta \neq 0$. If $0 < \beta < 1/\zeta$, or $\beta = 0$ and $F \in \mathrm{AEP}(b, C)$ for some $b \geq 2$ and $C > 0$, there is a unique solution $\gamma = \gamma(F, \nu, \zeta, \beta)$ to the following equation*

$$\inf_{\beta < s < 1/\zeta} \left[s\Lambda^* \left(\nu + \sqrt{\gamma/s} \right) + s\zeta \right] = 1.$$

2. *Assume $\zeta = 0$. The above holds so long as $\nu = \mu$ (and with $1/\zeta$ interpreted as ∞).*

Proof. Let $M = \sup\{x : \Lambda^*(x) < \infty\}$. Since F is not a point mass, $\mu < M \leq \infty$. Define

$$G(s, \gamma) = s\Lambda^* \left(\nu + \sqrt{\gamma/s} \right) + s\zeta.$$

Note that $G(s, \gamma)$ is finite (resp. infinite) if $\gamma/s < (M - \nu)^2$ (resp. $\gamma/s > (M - \nu)^2$). Also, $G(s, \gamma)$ and its derivatives are continuous wherever G is finite, and hence are uniformly continuous on any compact subset of $[0, \infty)^2$ on which G is finite. Furthermore, $G(s, \gamma)$ is strictly increasing in γ on the interval $(0, s(M - \nu)^2)$. Let

$$L_\beta(\gamma) = \inf_{\beta < s < 1/\zeta} G(s, \gamma). \tag{A.11}$$

Thus $L_\beta(\gamma)$ is finite if $\gamma\zeta < (M - \nu)^2$, and infinite when $<$ is replaced by $>$. Furthermore, for $\gamma < (M - \nu)^2/\zeta$, the infimum is achieved at some value s_γ of s in a neighborhood of which $G(s, \gamma) < \infty$.

Assume first that $\beta > 0$. It may be seen as follows that $L_\beta(\gamma)$ is continuous and strictly increasing in γ on the interval $[0, (M - \nu)^2/\zeta)$. Let $0 \leq \gamma < \gamma' < (M - \nu)^2/\zeta$. Then

$$0 \leq L_\beta(\gamma') - L_\beta(\gamma) \leq G(s_\gamma, \gamma') - G(s_\gamma, \gamma), \tag{A.12}$$

and continuity follows from the properties of G noted above. Similarly,

$$L_\beta(\gamma') - L_\beta(\gamma) \geq G(s_{\gamma'}, \gamma') - G(s_{\gamma'}, \gamma). \tag{A.13}$$

and strict monotonicity follows similarly.

It suffices to prove that $L_\beta(\gamma)$ takes values smaller than 1 and *finite* values larger than 1. The first claim follows from the fact that, with $\gamma = \beta(\mu-\nu)^2$,

$$L_\beta(\gamma) \le G(\beta,\gamma) = \beta\zeta < 1.$$

We turn to the second claim, and shall make use of two general properties of rate functions which follow from (Dembo and Zeitouni, 2010, Eq. (2.2.10), Lem. 2.2.20). It is standard that $\Lambda^*(\mu+x) \sim \frac{1}{2}(x/\sigma)^2$ as $x \downarrow 0$, where $\sigma^2 > 0$ is the variance of F. Therefore,

$$\exists T \in (0, M) \text{ such that } \Lambda^*(\mu+x) \ge \tfrac{1}{4}(x/\sigma)^2 \text{ when } 0 \le x \le T. \tag{A.14}$$

With T chosen thus, by convexity,

$$\exists A > 0 \text{ such that } \Lambda^*(\mu+x) \ge Ax \text{ when } x \ge T. \tag{A.15}$$

Assume first that $\zeta > 0$ and $M = \infty$. By (A.15), for sufficiently large γ,

$$\infty > L_\beta(\gamma) \ge \inf_{\beta<s<1/\zeta} \Big[sA(\nu-\mu+\sqrt{\gamma/s}) + s\zeta \Big] \ge A\big(\beta(\nu-\mu)+\sqrt{\gamma\beta}\big) > 1.$$

Suppose next that $\zeta > 0$ and $M < \infty$. Let $0 < \gamma < (M-\nu)^2/\zeta$. Since $\Lambda^*(\nu+\sqrt{\gamma/s}) = \infty$ if $s < \gamma/(M-\nu)^2 =: \beta_0(\gamma)$,

$$\begin{aligned} \infty > L_\beta(\gamma) &\ge \beta_0 \inf_{\beta_0<s<1/\zeta} \Lambda^*(\nu+\sqrt{\gamma/s}) + \beta_0\zeta \\ &= \beta_0\Lambda^*(\nu+\sqrt{\gamma\zeta}) + \beta_0\zeta. \end{aligned} \tag{A.16}$$

The limit of this, as $\gamma \uparrow (M-\nu)^2/\zeta$, is strictly greater than 1.

Now let $\zeta = 0$ and $\nu = \mu$, and note that $L_\beta(\gamma) < \infty$ for all $\gamma \ge 0$. Suppose $M \le \infty$ and $\gamma > 0$. By dividing the infimum in (A.11) according to whether or not $\sqrt{\gamma/s} < T$, we find that

$$\begin{aligned} \infty > L_\beta(\gamma) &\ge \min\left\{ \inf_{\beta<s<\gamma/T^2} s\Lambda^*(\mu+\sqrt{\gamma/s}), \inf_{s>\gamma/T^2} s\Lambda^*(\mu+\sqrt{\gamma/s}) \right\} \\ &\ge \min\left\{ A\sqrt{\gamma\beta}, \tfrac{1}{4}\gamma/\sigma^2 \right\}, \end{aligned}$$

by (A.14)–(A.15). This diverges as $\gamma \to \infty$.

When $\beta = 0$, some of the arguments fail as $G(s,\gamma)$ may not be continuous at $(0,0)$. Assume that $F \in \mathrm{AEP}(b,C)$ for some $b \ge 2$ and $C > 0$. Note that $M = \infty$ by Lemma A.8. If $b = 2$, $G(s,\gamma) \to C\gamma$ when $\gamma > 0$ is fixed and $s \to 0$, by Lemma A.8, and taking this limit as an extension at $s = 0$, the same arguments used in the case $\beta > 0$ apply. If $b > 2$, we need slightly different arguments. As before, let s_γ be a minimizer of $G(s,\gamma)$. We have that s_γ is well-defined for all γ and strictly positive, since G is uniformly continuous on any compact of $(0,1/\zeta] \times [0,\infty)$ and $G(s,\gamma) \sim C\gamma^{b/2} s^{1-b/2} \to \infty$ when $s \to 0$. Hence, we may proceed as before in (A.13)-(A.12), obtaining that $L_0(\gamma)$ is strictly increasing and continuous. As before, we turn to proving that L_0 takes values below 1 and finite values above 1. First, with $\gamma = (\mu-\nu)^2/(2\zeta)$ and $s = 1/(2\zeta)$,

$$L_0(\gamma) \le G(s,\gamma) = \gamma\zeta/(\mu-\nu)^2 = 1/2 < 1.$$

Next, showing that L_0 takes finite values above 1 is done exactly as before, except that (A.14) is replaced by

$$G(s,\gamma) \sim Cs^{1-b/2}\gamma^{b/2} \ge C\zeta^{b/2-1}\gamma^{b/2}, \quad \gamma \to \infty.$$

by Lemma A.8. □

The following result describes the variations of γ (defined in Lemma A.9) with the parameter of an exponential family.

Lemma A.10. *Consider a natural exponential family of distributions $(F_\theta, \theta \geq 0)$ and let μ_θ and Λ_θ^* denote the mean and the rate function of F_θ, respectively. Let ζ_θ be a continuous and decreasing function of θ. Then, for any fixed $0 < \beta < 1/\zeta_0$, $\gamma_\theta := \gamma(F_\theta, \mu_0, \zeta_\theta, \beta)$ is continuous and strictly increasing in θ. Moreover, if $\zeta_\theta \to 0$ when $\theta \to \theta_c$, then $\gamma_\theta \to \infty$ when $\theta \to \theta_c$.*

Proof. First, note that $\mu_\theta \geq \mu_0$ (Brown, 1986, Cor. 2.22) so that γ_θ is well-defined. That γ_θ is strictly increasing comes from the fact that both ζ_θ and $\Lambda_\theta^*(a)$ ($a > \mu_\theta$ fixed) are decreasing. The latter can be seen from

$$\Lambda_\theta^*(a) = -\lim_{k\to\infty} \frac{1}{k} \log \mathbb{P}_\theta(\bar{X}_k \geq a),$$

where $\bar{X}_k$ is the average of the sample of size k from F_θ, (Brown, 1986, Cor. 2.22) and the fact that the distribution of $\bar{X}_k$ as θ varies forms a natural exponential family with parameter $k\theta$. That γ_θ is continuous comes from the continuity of ζ_θ and $\Lambda_\theta^*(a)$ (in (θ, a)).

For the behavior near θ_c, notice that $\Lambda_\theta^*(a) = 0$ for $a \leq \mu_\theta$, so that $G(1/(2\zeta_\theta), \gamma) = 1/2$ for any $\gamma \leq (\mu_\theta - \mu_0)^2/(2\zeta_\theta)$. Then combine this with the fact that μ_θ is strictly increasing in θ to see that γ_θ is of order at least $1/\zeta_\theta$. In fact, it is easy to see that $\gamma_\theta \sim (\mu_\theta - \mu_0)^2/\zeta_\theta$ when $\theta \nearrow \theta_c$. □

A.3 Main proofs

A.3.1 Proof of Theorem 1

By monotonicity, it is enough to assume that $\theta_m = \theta$ for all m. Fix t and, for short, let $p = p_0(t)$ and $p' = p_\theta(t)$. First, assume that $\theta > \theta_*$, so that $\zeta_{p'} < \alpha\zeta_p$. Fix B such that $1/\zeta_p < B < \alpha/\zeta_{p'}$ and consider the test with rejection region $\{S_m(t) \geq dB \log m\}$. Under $\mathbb{H}_0^m$, we have $S_m(t) = (1 + o_P(1))(d/\zeta_p)\log m$ by (4), so that $\mathbb{P}(S_m(t) \geq dB \log m) \to 0$. Under $\mathbb{H}_{1,K}^m$, $S_m(t) \geq S_K(t) = (1 + o_P(1))(\alpha d/\zeta_{p'})\log m$, so that $\mathbb{P}(S_m(t) \geq dB \log m) \to 1$. Hence, this test is asymptotically powerful.

Now assume that $\theta < \theta_*$, so that $\zeta_{p'} > \alpha\zeta_p$ and there is B such that $\alpha/\zeta_{p'} < B < 1/\zeta_p$. Let $K^c = \mathbb{V}_m \setminus K$. It is enough to show that, under both $\mathbb{H}_0^m$ and $\mathbb{H}_{1,K}^m$, $S_m(t) = S_{K^c}(t)$ with probability tending to one, so that the values at the nodes in K have no influence on $S_m(t)$. Indeed, let J be a hypercube within $\mathbb{V}_m$ of sidelength $[m/3]$ which does not intersect K. Then $S_{K^c}(t) \geq S_J(t)$ and the distribution of $S_J(t)$ is the same, both under $\mathbb{H}_0^m$ and $\mathbb{H}_{1,K}^m$. In addition, $\mathbb{P}(S_J(t) \geq dB \log m) \to 1$ by (4). Now, let L be the set of nodes within (supnorm) distance $(\log m)^2$ from K, so that L is a hypercube of sidelength $[m^\alpha] + [2(\log m)^2]$ containing K in its interior. Under the event $\{S_m(t) \leq (\log m)^2\}$, $S_m(t) \neq S_{K^c}(t)$ only when $S_L(t) > S_{K^c}(t)$. The distribution of $S_L(t)$ under the null is stochastically bounded by its distribution under $\mathbb{H}_{1,K}^m$, which is itself bounded by its distribution under $\mathbb{H}_{1,L}^m$. Even under the latter, $\mathbb{P}(S_L(t) \geq dB \log m) \to 0$ by (4). We then conclude using the fact that $\mathbb{P}\left(S_m(t) \leq (\log m)^2\right) \to 1$, again by (4).

A.3.2 Proof of Theorem 2

We use the notation and follow the arguments of Section A.3.1. In addition, let $\zeta_{p'}^1 = \log(1/p')$, i.e., the function ζ in dimension one. When $\theta > \theta_*^+$, we consider $1/\zeta_p < B < \alpha/d\zeta_{p'}^1$. Under $\mathbb{H}_0^m$, we still have $S_m(t) = (1 + o_P(1))(d/\zeta_p)\log m$. Under $\mathbb{H}_{1,K}^m$, $S_m(t) \geq S_K(t) = (1 + o_P(1))(\alpha/\zeta_{p'})\log m$, since K is isomorphic to a subinterval of the one-dimensional lattice. We conclude as before that the test with rejection region $\{S_m(t) \geq dB \log m\}$ is asymptotically powerful.

When $\theta < \theta_*^-$, we consider $\alpha/d\zeta_{p'} < B < 1/\zeta_p$. As before, let L be the set of nodes within (supnorm) distance $(\log m)^2$ from K, so that L is now a band. As before, it suffices to prove that $\mathbb{P}(S_L(t) \geq dB \log m) \to 0$ under $\mathbb{H}_{1,L}^m$. Though (4) cannot be applied as L is not isomorphic to a square lattice, its proof via the union bound and (3) applies. Indeed, fix $\eta > 0$ small enough that $(1-\eta)\zeta_{p'} dB > \alpha$. Then, for m large enough, we have

$$\begin{aligned}\mathbb{P}(S_L(t) \geq dB \log m) &\leq |L| \cdot \mathbb{P}(S \geq dB \log m) \\ &\leq \mathrm{O}(m^\alpha (\log m)^{2(d-1)}) \ \exp(-(1-\eta)\zeta_{p'} dB \log m) \\ &= \mathrm{O}(\log m)^{2(d-1)} \ \exp((\alpha - (1-\eta)\zeta_{p'} dB) \log m) \to 0.\end{aligned}$$

A.3.3 Proof of Proposition 1

Let $k_m(\varepsilon) = (1-\varepsilon) d \log(m)/\log(1/p_0(t_m))$ with $\varepsilon > 0$ fixed. We first show that $S_m(t_m) \geq k_m(\varepsilon)$ with probability tending to one under $\mathbb{H}_0^m$. We use the notation and arguments provided in the proof of Lemma A.2. As in (A.4),

$$\begin{aligned}\mathbb{P}(S_m(t_m) < k_m(\varepsilon)) &\leq (1 - \mathbb{P}(S \geq k_m(\varepsilon)))^N \\ &\leq \left(1 - p_0(t_m)^{k_m(\varepsilon)}\right)^N \\ &\leq \exp\left(-m^{\varepsilon d}/(\log m)^{2d}\right) \to 0,\end{aligned}$$

where the second inequality holds for m large enough by Lemma A.3.

Assume that $\theta_m \leq \theta < \infty$ for all m. Proceeding as in Section A.3.1 and using the slightly larger region L, it is enough to show that, for ε small enough, $S_L(t_m) \leq k_m(\varepsilon)$ when $X_v \sim F_\theta$ for all $v \in L$. Using the union bound and the fact that $|L| = \mathrm{O}(m)^{\alpha d}$, we have

$$\mathbb{P}(S_L(t_m) \geq k_m(\varepsilon)) \leq |L| \cdot \mathbb{P}(S \geq k_m(\varepsilon)) \leq \mathrm{O}(m)^{\alpha d} (c p_\theta(t_m))^{k_m(\varepsilon)}, \tag{A.17}$$

where the last inequality is due to Lemma A.3 (and c is the constant that appears there). By integration by parts, for $\theta > 0$ and $\varepsilon \in (0,1)$ fixed, we have $p_\theta(t) \leq p_0((1-\varepsilon)t)$ for sufficiently large t. Indeed, for t large enough,

$$\begin{aligned}p_\theta(t) &= \exp(\theta t - \Lambda(\theta)) p_0(t) + \int_t^\infty \theta \exp(\theta x - \Lambda(\theta)) p_0(x) dx \\ &\leq \exp(\theta t - \Lambda(\theta) - C(1-\varepsilon/3)^b t^b) + \int_t^\infty \theta \exp(\theta x - \Lambda(\theta) - C(1-\varepsilon/3)^b x^b) dx \\ &\leq \exp(-C(1-\varepsilon/2)^b t^b) \\ &\leq p_0((1-\varepsilon)t),\end{aligned}$$

where we used the fact that $b > 1$ in line 3 and the fact that $\log p_0(t) \sim -Ct^b$ as $t \to \infty$ (since $F_0 \in \mathrm{AEP}(b, C)$) in lines 2 and 4. The last property implies also that $p_0((1-\varepsilon)t) \leq p_0(t)^{(1-\varepsilon)^{b+1}}$ for large t. Hence, for m large enough, $p_\theta(t_m) \leq p_0(t_m)^{(1-\varepsilon)^{b+1}}$, so that taking logs in (A.17), we get

$$\log \mathbb{P}(S_L(t_m) \geq k_m(\varepsilon)) \leq \mathrm{O}(1) + (d \log m)\left(\alpha + \mathrm{O}(\log p_0(t_m))^{-1} - (1-\varepsilon)^{b+2}\right) \to -\infty,$$

when $\varepsilon < 1 - \alpha^{1/(b+2)}$. (Remember that $\alpha < 1$ and that $p_0(t_m) \to 0$ so that the middle term is small.)

A.3.4 Proof of Theorem 3

Let $\mathbb{E}_\theta$ denote the expectation of X_v under F_θ. By Lemma A.5, under the null,

$$\frac{S_m(t) - \mathbb{E}_0(S_m(t))}{\sqrt{\mathrm{Var}_0(S_m(t))}} \to \mathcal{N}(0,1), \tag{A.18}$$

with $\mathrm{Var}_0(S_m(t))$ of order m^d. Write $p := p_0(t)$ and $p' := p_{\theta_m}(t)$.

We consider the alternative with anomalous cluster K as a two-stage percolation process, where the first stage is percolation on $\mathbb{V}_m$ with probability p, as under the null, and the second stage is percolation on the closed nodes within K, i.e., $K \setminus \{v : X_v > t\}$, with (conditional) probability $(p'-p)/(1-p)$. An open cluster at the first stage is called *small* if it is not a largest open cluster.

We may assume, except where noted below, that $\theta_m \to 0$. Since

$$\frac{\partial}{\partial\theta} \log p_\theta(t) = \mathbb{E}_\theta(X_v | X_v > t) - \mathbb{E}_\theta(X_v),$$

which is positive at $\theta = 0$ by choice of t, there exists $c \in (0,\infty)$ such that

$$p' - p \sim c\theta_m \quad \text{as} \quad m \to \infty. \tag{A.19}$$

Let $\Delta_m \geq 0$ be the difference between the sizes of the largest clusters under the null and alternative. For $x \in K$, let F_x be the sum of the sizes of all small clusters of the entire lattice that contain some neighbor of x. Note that $\Delta_m \leq \sum_{x \in D}(1 + F_x)$ where D is the set of $x \in K$ that are closed at the first stage and open at the second. Therefore, Δ_m has expectation bounded above by

$$\mathbb{E}(\Delta_m) \leq \left(\frac{p'-p}{1-p}\right) |K|(1 + 2d\mu_p), \tag{A.20}$$

where $\mu_p < \infty$ is the mean size of a finite open cluster in the infinite lattice.

By (A.19) and the above, $\mathbb{E}(\Delta_m) \leq C\theta_m m^{\alpha d}$ for some $C < \infty$. By Markov's inequality, $\Delta_m = \mathrm{O_P}(\theta_m m^{\alpha d})$.

Thus, if $\theta_m m^{(\alpha - 1/2)d} \to 0$, then $\Delta_m / \sqrt{\mathrm{Var}_0(S_m(t))} \to 0$, implying that the same central limit law as (A.18) holds under the alternative, so that the test based on the largest open cluster is asymptotically powerless. We must also consider the case when $\theta_m \not\to 0$, for which a similar argument is valid.

Now assume that $\alpha \geq 1/2$ and $\theta_m m^{(\alpha-1/2)d} \to \infty$. By (Grimmett, 1999, Thm 8.99) and standard properties of the largest cluster in a box (to be found in, for example, (Falconer and Grimmett, 1992)), with probability tending to one the largest open cluster increases in size by at least $C_1(p'-p)|K|$ for some $C_1 = C_1(p) > 0$. By (A.19), this has order $\theta_m m^{\alpha d}$. Since

$$\frac{\theta_m m^{\alpha d}}{\sqrt{\mathrm{Var}_0(S_m(t))}} \sim C_2 \theta_m m^{(\alpha-1/2)d} \to \infty$$

for some $C_2 = C_2(p) > 0$, the test based on the largest open cluster is asymptotically powerful.

A.3.5 Proof of Theorem 4

We may assume without loss of generality that $\theta_m \to 0$ as $m \to \infty$. By (5) and the assumption on t_m, we have that $S_m(t_m) \asymp_{\mathrm{P}} \log m$ under the null. Now $p_\theta(t)$ is infinitely differentiable in θ, with each derivative continuous in t, and with

$$\left.\frac{\partial p_\theta(t)}{\partial\theta}\right|_{\theta=0} = p_0(t)\left[\mathbb{E}_0(X_v|X_v > t) - \mathbb{E}_0(X_v)\right] \geq \frac{p_c}{2}\left[\mathbb{E}_0(X_v|X_v > t_c) - \mathbb{E}_0(X_v)\right] > 0,$$

uniformly for t in a neighborhood of t_c. Therefore, there exists $C > 0$ such that

$$\frac{\partial p_\theta(t)}{\partial \theta} \geq 1/C, \quad \text{and} \quad \left|\frac{\partial^2 p_\theta(t)}{\partial \theta^2}\right| \leq C,$$

for (θ, t) in some neighborhood of $(0, t_c)$. Hence,

$$p_\theta(t) - p_0(t) \geq \theta/C - C^2\theta^2/2 \geq \theta/(2C),$$

on such a neighborhood. Let A and B be such that $p_c - p_0(t_m) \leq Am^{-\alpha/\nu'}$ and $\theta_m \geq Bm^{-\alpha/\nu'}$, and assume that $B > 2AC$, which we may by the statement of the theorem. Since $\theta_m \to 0$ and $t_m \to t_c$,

$$m^{\alpha/\nu''}(p_{\theta_m}(t_m) - p_c) \geq m^{\alpha/\nu''}\left[\frac{\theta_m}{2C} + (p_0(t_m) - p_c)\right] \geq \left[\frac{B}{2C} - A\right] m^{\alpha(1/\nu'' - 1/\nu')} \to \infty,$$

for $\nu'' < \nu'$ and sufficiently large m. By (5) applied to $K \in \mathcal{K}_m$, it follows that $S_K(t_m) \asymp_{\mathrm{P}} m^{\alpha d}$ under the alternative. Consequently, the test with rejection region $\{S_m(t_m) \geq (\log m)^2\}$ is asymptotically powerful.

A.3.6 Proof of Lemma 5

Part 1. This follows immediately from Lemma A.2.

We therefore focus on the remaining two parts. We use the abbreviated notation $F := F_{\theta|t}$, $\Lambda^* := \Lambda^*_{\theta|t}$, $\mu := \mu_{\theta|t}$, $\zeta := \zeta_{p_\theta(t)}$, $\gamma := \gamma_{\theta|t}(\beta)$, $U_m := U_m(t, k_m)$, and we write $\nu := \mu_{0|t}$. Let $Y_k = X_k - \nu$. As in Lemma A.4, let $N_m(k)$ denote the number of open cluster of size k within $\mathbb{V}_m$, and define

$$G_k(x) = \mathbb{P}\left(k^{1/2}\bar{Y}_k \leq x\right),$$

where $\bar{Y}_k = \bar{X}_k - \nu$ and $\bar{X}_k$ is the average of an i.i.d. sample of size k from F. By the independence of $\bar{Y}_K$ and $\bar{Y}_L$ for $K, L \in \mathcal{Q}_m^{(t)}$ distinct, we have

$$\mathbb{P}(U_m \leq x) = \mathbb{E}\left(\prod_{k \geq k_m} G_k(x)^{N_m(k)}\right) = \mathbb{E}\left(\exp[-R_m(x)]\right),$$

where

$$R_m(x) := -\sum_{k \geq k_m} N_m(k) \log(1 - \bar{G}_k(x)).$$

We turn, therefore, to bounding $R_m(x)$.

Part 2. Define $x_m = \sqrt{\gamma d \log m}$ and fix $\varepsilon > 0$. For the lower bound, let ℓ_m be the closest integer to $ad \log m$ between k_m and $(d/\zeta) \log m$, where

$$a = \operatorname*{arg\,min}_{\beta < s < 1/\zeta} \left[s\Lambda^*\left(\nu + \sqrt{\gamma/s}\right) + s\zeta\right]. \tag{A.21}$$

We have

$$R_m((1-\varepsilon)x_m) \geq T_m := N_m(\ell_m)\bar{G}_{\ell_m}((1-\varepsilon)x_m),$$

and we shall show that, for ε fixed, $T_m \to \infty$ in probability. Fix $\eta > 0$. On the one hand, we use Lemma A.4 and (3), to get

$$\mathbb{E}(N_m(\ell_m)) \geq \frac{(m - 2\ell_m)^d}{\ell_m} \mathbb{P}(S = \ell_m) \geq m^d \exp(-(1+\eta)\zeta\ell_m),$$

for m large enough. On the other hand, we use Cramér's Theorem (Dembo and Zeitouni, 2010, Thm 2.2.3) to get

$$\begin{aligned}\bar{G}_{\ell_m}((1-\varepsilon)x_m) &\geq \mathbb{P}\left(\bar{Y}_{\ell_m} \geq (1-\varepsilon/2)\sqrt{\gamma/a}\right)\\ &\geq \exp\left(-(1+\eta)\ell_m\Lambda^*\left[\nu+(1-\varepsilon/2)\sqrt{\gamma/a}\right]\right),\end{aligned}$$

for m large enough. By the definition of γ, $a\Lambda^*\left[\nu+\sqrt{\gamma/a}\right]+a\zeta = 1$, and hence, for ε small enough,

$$a\zeta + a\Lambda^*\left[\nu+(1-\varepsilon/2)\sqrt{\gamma/a}\right] < 1,$$

by strict monotonicity, as in the proof of Lemma A.9. Hence, for η small enough,

$$\ell_m\zeta + \ell_m\Lambda^*\left[\nu+(1-\varepsilon/2)\sqrt{\gamma/a}\right] \leq (1-\eta)d\log m.$$

It follows that

$$\mathbb{E}\left(T_m\right) \geq m^{\eta^2 d}.$$

To bound the corresponding variance, we use Lemma A.4 to obtain

$$\operatorname{Var}\left(T_m\right) \leq \mathrm{O}(\log m)^d\, \mathbb{E}\left(T_m\right),$$

and it follows by Chebyshev's Inequality that, indeed, $T_m \to \infty$ in probability.

Since $T_m \geq 0$, $\exp(-T_m) \to 0$ in L^1, and therefore

$$\mathbb{P}\left(U_m \leq (1-\varepsilon)x_m\right) \to 0.$$

We show next that $\mathbb{E}\left(R_m((1+\varepsilon)x_m)\right) \to 0$, and this will imply the claim of Part 2. Fix $\eta > 0$. We have that

$$R_m((1+\varepsilon)x_m) \leq T_m + 2Z_m, \tag{A.22}$$

where

$$T_m := 2\sum_{k=k_m}^{k_m^{(\eta)}} N_m(k)\bar{G}_k((1+\varepsilon)x_m)$$

and Z_m is the number of clusters of size exceeding $k_m^{(\eta)} := [(1+\eta)(d/\zeta)\log m]$. We note first that, as in the proof of Lemma A.4, for large m,

$$\mathbb{E}\left(Z_m\right) \leq m^d\exp(-\tfrac{1}{2}\zeta k_m^{(\eta)}) \to 0. \tag{A.23}$$

We turn next to T_m, and shall show that, for ε fixed and η small enough, $\mathbb{E}\left(T_m\right) \to 0$. On the one hand, we use Lemma A.4 and (3), to get

$$\mathbb{E}\left(N_m(k)\right) \leq m^d\exp(-(1-\eta)\zeta k),$$

for m large enough. On the other hand, by Chernoff's Bound,

$$\bar{G}_k((1+\varepsilon)x_m) \leq \exp\left(-k\Lambda^*\left[\nu+(1+\varepsilon)x_m/\sqrt{k}\right]\right).$$

Together, we obtain

$$\begin{aligned}
\mathbb{E}\left(T_m\right) &\leq 2 \sum_{k=k_m}^{k_m^{(\eta)}} m^d \exp\left(-(1-\eta)\left[k\zeta + k\Lambda^*\left(\nu + (1+\varepsilon)x_m/\sqrt{k}\right)\right]\right) \\
&\leq \mathrm{O}(\log m)\, \exp\left(d \log m - (1-\eta) \min_{k_m \leq k \leq k_m^{(\eta)}} \left[k\zeta + k\Lambda^*\left(\nu + (1+\varepsilon)x_m/\sqrt{k}\right)\right]\right) \\
&\leq \mathrm{O}(\log m)\, \exp\left((1-(1-\eta)A) d \log m\right),
\end{aligned}$$

where

$$A := \inf_{\beta < a < (1+\eta)/\zeta} \left[a\Lambda^*\left(\nu + (1+\varepsilon)\sqrt{\gamma/a}\right) + a\zeta\right]. \tag{A.24}$$

As in the proof of Lemma A.9, $A = A(\varepsilon,\eta)$ is continuous in (ε,η) and strictly increasing in ε. Since $A(0,0) = 1$ by definition of γ, for ε fixed, $-h := 1-(1-\eta)A(\varepsilon,\eta) < 0$ for η small enough, in which case $\mathbb{E}\left(T_m\right) \leq m^{-hd/2} \to 0$ as m increases.

By (A.22)–(A.23), we have that $\mathbb{E}\left(R_m((1+\varepsilon)x_m)\right) \to 0$. By Jensen's inequality,

$$\mathbb{P}\left(U_m \leq (1+\varepsilon)x_m\right) \geq \exp(-\mathbb{E}\left(R_m((1+\varepsilon)x_m)\right)) \to 1,$$

and the proof of this part is complete.

Part 3. We build on the arguments provided so far, which apply essentially unchanged, except in two places. In the lower bound, instead of Cramér's Theorem, we use

$$\bar{G}_k(x) \geq \bar{F}(x/\sqrt{k})^k,$$

combined with the asymptotic behavior for $\bar{F}$. And in the upper bound, the evaluation of A defined in (A.24) is done differently when $b < 2$.

Part 3(a). When $b > 2$, we have $a > 0$ in (A.21) (with $\beta = 0$), since

$$h(s) := s\Lambda^*\left(\nu + \sqrt{\gamma/s}\right) + s\zeta \asymp s^{1-b/2} \to \infty,$$

for γ fixed and $s \to 0$, by Lemma A.8. When $b = 2$, we take a small enough if the minimum is at $a = 0$. Then the other arguments in Part 2 apply unchanged.

Part 3(b). By the same calculations, $a = 0$ in (A.21), since $h(s) > 0$ for all $s > 0$, and $h(s) \asymp s^{1-b/2} \to 0$ when $s \to 0$, because $b < 2$. This would make $A = 0$ in (A.24) for any $\varepsilon > 0$, making the arguments for the upper bound collapse. Instead, redefine $x_m = (Cd\log m)^{1/b} k_m^{1/2-1/b}$. Since $x_m/\sqrt{k} \to \infty$ uniformly over $k \leq k_m^{(\eta)}$, for $\eta > 0$ fixed, we have

$$k\zeta + k\Lambda^*\left(\nu + (1+\varepsilon)x_m/\sqrt{k}\right) \geq k\zeta + (1-\eta)Ck^{1-b/2}(1+\varepsilon)^b x_m^b,$$

for m large enough, by Lemma A.8. Then the term on the right-hand side takes its minimum over $k_m \leq k \leq k_m^{(\eta)}$ at $k = k_m$, and from here, the remaining arguments apply.

A.3.7 Proof of Proposition 2

Assume, for simplicity, that $\theta_m = \theta < \theta_c$ for all m. The key point is that $F_{\theta|t} \in \mathrm{AEP}(b, C)$. Indeed, we have $\bar{F}_{\theta|t}(x) = \bar{F}_\theta(x)/\bar{F}_\theta(t)$, where the denominator is constant in x, and, integrating by parts,

$$\bar{F}_\theta(x) = \exp(\theta x - \Lambda(\theta))\bar{F}_0(x) + \int_x^\infty \theta \exp(\theta y - \Lambda(\theta))\bar{F}_0(y) dy.$$

From here, we reason as in the proof of Proposition 1, using the fact that $\log \bar{F}_0(y) \sim -Cy^b$ when $y \to \infty$, with $b > 1$. Hence, $F_{\theta|t}$ and $F_{0|t}$ have same (first-order) asymptotics, so there is nothing distinguishing the asymptotic behavior of U_m under the null and under an alternative. In detail, we proceed as in Section A.3.1, with the enlarged hypercube L, and show that, in probability under $\mathbb{H}_{1,L}^m$,

$$\limsup_{m\to\infty} k_m^{1/b-1/2}(\log m)^{-1/b} U_L < (d/C)^{1/b},$$

where U_L is the ULS scan statistic restricted to open clusters within L. Since L is a scaled version of $\mathbb{V}_m$, $F_{\theta|t} \in \mathrm{AEP}(b, C)$ and $p_\theta(t) < p_c$, Lemma 5 applies to yield

$$k_m^{1/b-1/2}(\alpha \log m)^{-1/b} U_L \to (d/C)^{1/b}.$$

We then conclude with the fact that $\alpha < 1$.

A.3.8 Proof of Theorem 5 and Theorem 6

The proof of Theorem 5 is parallel to that of Theorem 1 in Section A.3.1, here using Lemma 5 in place of Lemma A.2. Note that we use the fact that, for t and $\beta > 0$ fixed, $\gamma_{\theta|t}(\beta)$ is continuous and strictly increasing in θ. This comes from Lemma A.10 and the fact that, when t is fixed, $F_{\theta|t}$ is also a natural exponential family with parameter θ. Similarly, the proof of Theorem 6 is parallel to that of Theorem 2 in Section A.3.2. Further details are omitted.

A.3.9 Proof of Lemma 6

The proof is parallel to that of Lemma 5. In particular, we use the notation introduced there and only sketch where the arguments differ (though never substantially).

Part 1. In this case, by Lemma A.5 and Lemma A.6, there is only one open cluster with size k_m or larger, and the result follows from, e.g., Chebyshev's inequality.

Part 2. Define $x_m = \sqrt{2\sigma^2 d(1-\delta\beta')\log m}$ and fix $\varepsilon > 0$. For the lower bound, we have

$$R_m((1-\varepsilon)x_m) \geq T_m := N_m(k_m)\bar{G}_{k_m}((1-\varepsilon)x_m),$$

Fix $\eta > 0$. By Lemma A.4 (still valid) and (8),

$$\mathbb{E}\left(N_m(k_m)\right) \geq m^d \exp(-(1+\eta)\delta k_m^{(d-1)/d}),$$

for m large enough. And by Cramér's Theorem and the fact that $\Lambda^*(x) \sim x^2/(2\sigma^2)$ when x is small,

$$\begin{aligned}\bar{G}_{k_m}((1-\varepsilon)x_m) &\geq \exp\left(-(1+\eta)k_m\Lambda^*\left[(1-\varepsilon)x_m/\sqrt{k_m}\right]\right)\\ &\geq \exp\left(-(1+\eta)(1-\varepsilon/2)x_m^2/(2\sigma^2)\right),\end{aligned}$$

for m large enough. Hence,

$$\mathbb{E}\left(T_m\right) \geq \exp\left(d\log m - (1+\eta)(\delta k_m^{(d-1)/d} + (1-\varepsilon/2)x_m^2/(2\sigma^2))\right) \geq m^{\varepsilon d(1-\delta\beta')/4},$$

for m large enough and η small enough. For the variance, we use Lemma A.4 to get

$$\mathrm{Var}\left(T_m\right) \leq \mathrm{O}(\log m)^{d^2/(d-1)}\,\mathbb{E}\left(T_m\right)$$

We then conclude by Chebyshev's Inequality.

We now show that $R_m((1+\varepsilon)x_m) \to 0$ in probability. Equation (A.22) holds with $k_m^{(\eta)} := [(1+\eta)(d/\delta)\log m]^{d/(d-1)}$. As before,

$$\mathbb{E}\,(Z_m) \le m^d \exp\{-\tfrac{1}{2}\delta(k_m^{(\eta)})^{(d-1)/d}\} \to 0 \quad \text{as } m \to \infty.$$

By Lemma A.4 and (8),

$$\mathbb{E}\,(N_m(k)) \le m^d \exp(-(1-\eta)\delta k^{(d-1)/d}),$$

for m large enough. (The absence of a boundary to $\mathbb{V}_m$ is being used here. The tail behavior of percolation clusters near the boundary of a box is not yet fully understood. See the remark in Section 5.2.) And by Chernoff's Bound and the behavior of Λ^* near the origin,

$$\bar{G}_k((1+\varepsilon)x_m) \le \exp\left(-(1+\varepsilon)x_m^2/(2\sigma^2)\right),$$

for any $k \ge k_m$. Thus,

$$\begin{aligned}\mathbb{E}\,(T_m) &\le 2\sum_{k=k_m}^{k_m^{(\eta)}} m^d \exp\left(-(1-\eta)\delta k^{(d-1)/d} - (1+\varepsilon)x_m^2/(2\sigma^2)\right) \\ &\le \mathrm{O}(\log m)^{d/(d-1)}\, m^{-\varepsilon d(1-\delta\beta')/4},\end{aligned}$$

for m large enough and η small enough.

Part 3. This part is even more similar to what we did in the proof of Lemma 5. The behavior of U_m is driven by the open clusters of size of order $\log m$ and the only difference is that the term in $k^{(d-1)/d}$ from the bounds on $N_m(k)$ is negligible. Details are omitted.

A.3.10 Proof of Theorem 7

Without loss of generality, we assume θ_m is bounded. By Lemma 6 and our assumptions on k_m, under the null, $U_m := U_m(t, k_m) \sim_{\mathrm{P}} A(\log m)^{1/2}$, for a finite constant $A > 0$. We now consider the alternative where the anomalous cluster is K.

The contribution of the largest open cluster Q_m is

$$\begin{aligned}\sqrt{|Q_m|}(\bar{X}_{Q_m} - \mu_{0|t}) = \frac{|Q_m \cap K|}{\sqrt{|Q_m|}}(\bar{X}_{Q_m\cap K} - \mu_{\theta_m|t}) + \frac{|Q_m \cap K^c|}{\sqrt{|Q_m|}}(\bar{X}_{Q_m\cap K^c} - \mu_{0|t}) \\ + \frac{|Q_m \cap K|}{\sqrt{|Q_m|}}(\mu_{\theta_m|t} - \mu_{0|t}).\end{aligned}$$

On the right-hand side, the first term is of order $\mathrm{o}_{\mathrm{P}}(1)$ and the second term is of order $\mathrm{O}_{\mathrm{P}}(1)$, by Chebyshev's inequality and the fact that, with probability tending to one, $|Q_m \cap K| \asymp |K|$ and $|Q_m| \asymp |\mathbb{V}_m|$, by Lemma A.5. The last term is of (exact) order $\mathrm{O}(\theta_m m^{(\alpha-1/2)d})$, by the fact that $\mu_{\theta|t}$ is differentiable at $\theta = 0$ with derivative equal to $\sigma_{0|t}^2 > 0$. Therefore, the ULS scan test is asymptotically powerful when $\liminf \theta_m m^{(\alpha-1/2)d}(\log m)^{-1/2}$ is large enough. (Note that this requires $\alpha > 1/2$.) If instead, $\limsup \theta_m m^{(\alpha-1/2)d}(\log m)^{-1/2} \to 0$, the scan over Q_m may be ignored and we need to consider smaller clusters.

By Lemma A.6 and the upper bound on k_m, the second largest cluster entirely within K is scanned and its contribution is of order $\mathrm{O}(\theta_m(\log m)^{d/(2d-2)})$, by the same arguments that established the contribution of the largest open cluster. Hence, the ULS scan test is asymptotically powerful when $\liminf \theta_m(\log m)^{d/(2d-2)-1/2}$ is large enough. If instead, $\theta_m(\log m)^{d/(2d-2)-1/2} \to 0$, the test is asymptotically powerless. Indeed, let L be the set of nodes within distance $(\log m)^3$ from K and let U_L be the result of scanning the open clusters of size at least k_m and entirely within L. As argued in the proof of Proposition 2, this time using Lemma A.6, it is enough to show that $U_L \leq A(\log m)^{1/2}$ with probability tending to one under $\mathbb{H}_{1,L}^m$. For any open cluster Q entirely within L,

$$\sqrt{|Q|}(\bar{X}_Q - \mu_{0|t}) = \sqrt{|Q|}(\bar{X}_Q - \mu_{\theta_m|t}) + \sqrt{|Q|}(\mu_{\theta_m|t} - \mu_{0|t}),$$

so that

$$U_L \leq \max_Q \sqrt{|Q|}(\bar{X}_Q - \mu_{\theta_m|t}) + o_P(1),$$

where the maximum is over open clusters of size at least k_m and entirely within L, and the second term is $o_P(1)$ by Lemma A.6 and the size of θ_m. Though $\theta_m \to 0$ varies, this maximum may be handled exactly as in Lemma 6, so that it is $\sim_P A(\alpha \log m)^{1/2}$ and we conclude.

A.3.11 Proof of Lemma 7

We only prove the more refined part. We use abbreviated notation as before, in particular, we omit the subscript 0, using $F_t = F_{0|t}$, $\sigma_t = \sigma_{0|t}$, etc. The lower bound is obtained via $\text{ULS}_m \geq U_m(t^*)/\sigma_{t^*}$, where t^* defines $\Gamma(\beta)$, and applying Lemmas 5 or 6 to $U_m(t^*)$ depending on whether $t^* > t_c$ or $t^* < t_c$. For simplicity, we assume that $t^* \neq t_c$. If $t^* = t_c$, then we consider a nearby threshold and argue by continuity. For the upper bound, we prove that $\mathbb{P}(\text{ULS}_m \geq x_m) \to 0$, where $x_m := \sqrt{g \log m}$ and $g > G := (d\Gamma(\beta))^{1/2}$.

As t increases, clusters are created and then destroyed in the coupled percolation processes. Suppose the removal at time t from the percolation process of vertex v creates some cluster $Q_t(w)$ at some neighbor w of v. If $\text{ULS}_m \geq x_m$, there must exist a vertex v and a neighbor w such that the cluster formed at w at time X_v contributes at some future time $t' > X_v$ an amount at least x_m to ULS_m. By conditioning on v, X_v, and w, one obtains that

$$\mathbb{P}(\text{ULS}_m \geq x_m) \leq o(1) + \int_{-\infty}^{t_\beta} \mathbb{P}\left(\bigcup_{v \in \mathbb{V}_m} \bigcup_{w \in \partial v} \Omega_t(w)\right) dF(t), \tag{A.25}$$

where the o(1) term covers the probability that the cluster at time $-\infty$, namely $\mathbb{V}_m$, determines ULS_m, or that a cluster at threshold $t > t_\beta$ is of size at least $k_m := \beta \log m$; ∂v is the neighbor set of v; and $\Omega_t(w)$ is the event that:

1. $k := |Q_t(w)|$ satisfies $k \geq \beta \log m$,
2. there exists a time $t' \geq t$ such that $Q_t(w)$ still exists at time t', and
3. $Y_t(k) - \mathbb{E}(Y_{t'}(k)) \geq x_m \sigma_{t'} \sqrt{k}$, where $Y_t(k)$ is the sum of a k-sample from F_t.

Assume (briefly) that σ_t is non-decreasing, and note that μ_t is automatically non-decreasing. Then as in the proofs of Lemmas 5 and 6, and using similar notation,

$$\begin{aligned}\sum_{v \in \mathbb{V}_m} \sum_{w \in \partial v} \mathbb{P}(\Omega_t(w)) &\leq \sum_{v \in \mathbb{V}_m} \sum_{w \in \partial v} \mathbb{P}\left(k := |Q_t(w)| \geq \beta \log m,\ Y_t(k) - \mathbb{E}(Y_t(k)) \geq x\sigma_t\sqrt{k}\right) \\ &\leq 2d\ \mathbb{E}(R_t(x_m)), \quad R_t(x) := \sum_{k \geq k_m} N_t(k)\bar{G}_t(k, x),\end{aligned}$$

where $N_t(k)$ is the number of t-open clusters of size k and

$$\bar{G}_t(k, x) = \mathbb{P}\left(Y_t(k) - \mathbb{E}(Y_t(k)) \geq x\sigma_t\sqrt{k}\right).$$

Therefore, by (A.25),

$$\mathbb{P}\left(\mathrm{ULS}_m \geq x_m\right) \leq \mathrm{o}(1) + 2d\left(\int_{-\infty}^{t_c-h} + \int_{t_c+h}^{t_\beta} \mathbb{E}(R_t(x_m))\, dF(t)\right) + F(t_c+h) - F(t_c-h), \quad \text{(A.26)}$$

for any $h > 0$. We bound $\mathbb{E}(R_t(x_m))$ as we did in the proofs of Lemmas 5 and 6. Explicitly, when $t_c + h \leq t \leq t_\beta$, we use Lemma A.4 and (A.1), to get

$$\begin{aligned}\mathbb{E}(N_t(k)) &\leq (1-p(t))^2 \frac{k e^{-k\zeta_{p(t)}}}{(1-e^{-\zeta_{p(t)}})^2} \\ &\leq C(h,\beta)\, k \exp(-k\zeta_{p(t_c+h)}), \quad C(h,\beta) := \frac{(1-p(t_\beta))^2}{(1-e^{-\zeta_{p(t_c+h)}})^2}.\end{aligned}$$

We use Chernoff's Bound on $\bar{G}_t(k,x)$, to obtain

$$\mathbb{E}(R_t(x_m)) \leq C(h,\beta)(k_{m,t}^h)^2 \exp\left((1-A_t)d\log m\right) + \exp(-hd\log(m)/2),$$

where $k_{m,t}^h := (1+h)(d/\zeta_{p(t)})\log m$,

$$A_t := \inf_{\beta < s < (1+h)/\zeta_{p(t)}} \left[s\Lambda_t^*\left(\mu + \sqrt{g/s}\right) + s\zeta_{p(t)}\right],$$

as in (A.24), and the last term is the probability that a there is a t-open of size exceeding $k_{m,t}^h$. Note that $A_t > 1$ for all $t_c + h \leq t \leq t_\beta$ because $g > G$. By continuity of A_t, $A_+ := \inf\{A_t : t_c + h \leq t \leq t_\beta\} > 0$. Hence, we have the following bound for all $t_c + h \leq t \leq t_\beta$,

$$\mathbb{E}(R_t(x_m)) \leq C(h,\beta)[(1+h)(d/\zeta_{p(t_c+h)})\log m]^2 m^{-(A_+-1)d} + \exp(-hd\log(m)/2).$$

When $t \leq t_c - h$, we simply use the fact that

$$\sum_k \mathbb{E}(N_t(k)) \leq |\mathbb{V}_m| = m^d,$$

and bound $\bar{G}_t(k,x)$ in the same way. We get

$$\mathbb{E}(R_t(x_m)) \leq \exp\left((1-A_t)d\log m\right),$$

where

$$A_t := \inf_{\beta < s} s\Lambda_t^*\left(\mu + \sqrt{g/s}\right).$$

Again, $A_t > 1$ for $t < t_c - h$ and $A_t \to A_{-\infty} > 1$ as $t \to -\infty$. Hence, by continuity of A_t, $A_- := \inf\{A_t : t < t_c - h\} > 0$, so that

$$\mathbb{E}(R_t(x_m)) \leq m^{-(A_- - 1)d},$$

valid for all $t < t_c - h$. Hence, the two integrals in (A.26) tend to zero with m. We then let $h \to 0$ so that $F(t_c + h) - F(t_c - h) \to 0$, since F is continuous at t_c.

Assume now that F has no atoms on $(-\infty, t_\beta]$. Then σ_t is continuous on $(-\infty, t_\beta]$, and in fact, is uniformly continuous since $\sigma_t \to \sigma$ when $t \to -\infty$. Since it is positive on that interval (because $\sigma_t = 0$ implies that F_t is a point mass), $\underline{\sigma} := \min\{\sigma_t : t \leq t_\beta\} > 0$. Since $g > G$ we can find $c > 0$ such that $g' := g(1-c)^2 > G$, and also $\eta > 0$ such that

$$|\sigma_s - \sigma_t| \leq c\underline{\sigma}, \qquad \text{if } |s-t| \leq \eta,\ s,t \leq t_\beta. \quad \text{(A.27)}$$

Let $x'_m = \sqrt{g' \log m}$. We say that a cluster Q *scores at time* s if it exists at time s and in addition

$$|Q| \geq \beta \log m, \qquad \sum_{v \in Q} X_v \geq |Q|\mu_s + x_m \sigma_s \sqrt{|Q|}.$$

Without loss of generality, assume that t_c is not an integer multiple of η. Fix two neighbors $v, w \in \mathbb{V}_m$, and a time $t \leq t_\beta$. If $\Omega_t(w)$ occurs then either:

(a) $Q_t(w)$ scores at some time $s \in [t, n_t\eta]$, where $n_t \in \mathbb{Z}$ satisfies $(n_t - 1)\eta \leq t < n_t\eta$, or

(b) there exists $n \geq n_t$ and $s \in [n\eta, (n+1)\eta)$ such that $Q_{n\eta}(w)$ scores at time s.

The latter possibility arises when $Q_t(w)$ scores at some time s not belonging to the interval $[t, n_t\eta)$. Writing $[n\eta, (n+1)\eta)$ for the interval containing s, $Q_t(w)$ must exist at the start of this interval, which is to say that $Q_t(w) = Q_{n\eta}(w)$.

The probability of (a) is no larger than

$$\mathbb{P}\left(k := |Q_t(w)| \geq \beta \log m,\ \exists s \in [t, n_t\eta] : Y_t(k)/k \geq \mu_s + x_m\sigma_s/\sqrt{k}\right). \tag{A.28}$$

By (A.27) and the fact that μ_s is non-decreasing,

$$\mu_s + \frac{x_m \sigma_s}{\sqrt{k}} \geq \mu_t + \frac{x'_m \sigma_t}{\sqrt{k}}, \tag{A.29}$$

so that (A.28) is no greater than

$$\mathbb{P}\left(k := |Q_t(w)| \geq \beta \log m,\ Y_t(k)/k \geq \mu_t + x'_m\sigma_t/\sqrt{k}\right). \tag{A.30}$$

Arguing similarly, part (b) has probability no greater than

$$\sum_{t/\eta < n < t_\beta/\eta} \mathbb{P}\left(k := |Q_t(w)| \geq \beta \log m,\ Y_{n\eta}(k)/k \geq \mu_{n\eta} + x'_m\sigma_{n\eta}/\sqrt{k}\right). \tag{A.31}$$

We divide the integral in (A.25) as follows

$$\int_{-\infty}^{t_\beta} = \int_{-\infty}^{-1/h} + \int_{-1/h}^{t_c - h} + \int_{t_c - h}^{t_c + h} + \int_{t_c + h}^{t_\beta}.$$

The first integral is bounded by $F(-1/h)$ and the third integral by $F(t_c + h) - F(t_c - h)$, both terms vanishing as $h \to 0$. For the second and fourth integrals, we do exactly as before, separately for (A.30) and (A.31)—for the latter, the sum has at most $(t_\beta + 1/h)/\eta + 1$ terms in the second integral and at most $(t_\beta - t_c - h)/\eta + 1$ terms in the fourth integral.

A.3.12 Proof of Theorem 8

By Lemma 7, $\text{ULS}_m(k_m)$ is of order at most $\sqrt{\log m}$ under the null. Now consider the alternative with anomalous cluster K. If $0 < (\alpha - 1/2)d < \alpha/\nu$, consider the contribution of the largest open cluster at supercritical threshold t and reason as in the proof of Theorem 7. Otherwise, consider the contribution of the largest open cluster at a threshold t_m such that $p_c - p_0(t_m) \asymp m^{-\lambda/\alpha}$. As in Theorem 4, the largest open cluster will be comparable in size to, and occupy a substantial portion of K. Reasoning again as in the proof of Theorem 7, the contribution is of order $m^{\alpha d/2}\theta_m \geq m^{\alpha/\nu}\theta_m \geq m^{\alpha/\nu - \lambda}$, which grows as a positive power of m.

B The scan statistic as the generalized likelihood ratio

We show that the simple scan statistic defined in (1) approximates the scan statistic of Kulldorff (1997), which is strictly speaking the generalized likelihood ratio (GLR), defined as follows. The log-likelihood under $\mathbb{H}_{1,K}^m$ is given by

$$\text{loglik}(K, \theta, \theta_0) := |K| \left(\theta \bar{X}_K - \log \varphi(\theta)\right) + |K^c| \left(\theta_0 \bar{X}_{K^c} - \log \varphi(\theta_0)\right).$$

Assuming θ and θ_0 are both unknown, the log GLR is defined as

$$\max_{K \in \mathcal{K}_m} \sup_{\theta > \theta_0} \text{loglik}(K, \theta, \theta_0) - \sup_{\theta_0} \text{loglik}(\mathbb{V}_m, \theta_0, \theta_0),$$

which is equal to

$$\max_{K \in \mathcal{K}_m} \left[|K|\, \Lambda^*(\bar{X}_K) + |K^c|\, \Lambda^*(\bar{X}_{K^c}) - |\mathbb{V}_m| \Lambda^*(\bar{X}_{\mathbb{V}_m})\right]_+ . \tag{B.1}$$

(The subscript $_+$ denotes the positive part.)

Under the normal location model, $\Lambda^*(x) = x^2/2$ and (B.1) is equal to

$$\max_{K \in \mathcal{K}_m} \frac{|\mathbb{V}_m|\, |K|}{|\mathbb{V}_m| - |K|} (\bar{X}_K - \bar{X}_{\mathbb{V}_m})_+^2.$$

(We used the fact that $\bar{X}_K \geq \bar{X}_{K^c} \Leftrightarrow \bar{X}_K \geq \bar{X}_{\mathbb{V}_m}$.) If $k_m^+ := \max\{|K| : K \in \mathcal{K}_m\}$ satisfies $k_m^+/|\mathbb{V}_m| \to 0$, which is the case in our examples, the fraction above is equal to $|K|(1+\text{O}(k_m^+/|\mathbb{V}_m|))$. Moreover, knowing that there is always a cluster K such that $\bar{X}_K \geq \bar{X}_{\mathbb{V}_m}$, we get that the square root of (B.1) is approximately equal to

$$\max_{K \in \mathcal{K}_m} \sqrt{|K|}(\bar{X}_K - \bar{X}_{\mathbb{V}_m}), \tag{B.2}$$

which is the version of (1) when μ_0 is unknown. (Note that $\bar{X}_{\mathbb{V}_m} = \mu_0 + \text{O}(|\mathbb{V}_m|)^{-1/2}$, by the Central Limit Theorem, so that (B.2) is within $\text{O}(k_m^+/|\mathbb{V}_m|)^{1/2}$ from (1).) This approximation is actually valid more generally, at least in a way that suffices for the asymptotic analysis we perform in this paper. Indeed, with $\sigma_0^2 = \text{Var}_0(X_v)$, we have $\Lambda^*(x) = (x - \mu_0)^2/(2\sigma_0^2) + \text{O}(x - \mu_0)^3$ in the neighborhood of μ_0. Assuming that $k_m^- := \min\{|K| : K \in \mathcal{K}_m\}$ satisfies $k_m^- \to \infty$, which is the case in our examples, the approximation of the square root of (B.1) by (B.2) is valid under the null, since $\bar{X}_K = \mu_0 + \text{O}(k_m^-)^{-1/2}$ and $\bar{X}_{K^c}, \bar{X}_{\mathbb{V}_m} = \mu_0 + \text{O}(|\mathbb{V}_m|)^{-1/2}$, by the Central Limit Theorem and the fact that $k_m^- \to \infty$ and $k_m^+/|\mathbb{V}_m| \to 0$. The same applies under the alternative if $\theta_m \to 0$, so that $\mu_{\theta_m} := \mathbb{E}_{\theta_m}(X_v) \to \mu_0$ and therefore $\bar{X}_K$ for any $K \in \mathcal{K}_m$. When θ_m is bounded away from zero, the two statistics, square root of (B.1) and (B.2), are both of order $\sqrt{|K|}$, where K denotes the cluster under the alternative (or in the case of the ULS scan, the largest open cluster within the anomalous cluster). All together, this is enough to conclude that the tests respectively based on (B.1) and (1) behave similarly.

Acknowledgements

The authors thank Ganapati Patil for providing additional references on the upper level set scan statistic, and Mikhail Langovoy for alerting the authors of his work on the LOC test. They are grateful to Thierry Bodineau and Raphaël Cerf for their remarks on the second largest cluster in supercritical percolation. EAC was partially supported by grants from the National Science Foundation (DMS-0603890) and the Office of Naval Research (N00014-09-1-0258), as well as a Hellman Fellowship. GRG was partially supported by a grant from the Engineering and Physical Sciences Research Council.

References

Arias-Castro, E., E. J. Candès, and A. Durand (2011). Detection of an anomalous cluster in a network. *Ann. Statist. 39*(1), 278–304.

Arias-Castro, E., E. J. Candès, H. Helgason, and O. Zeitouni (2008). Searching for a trail of evidence in a maze. *Ann. Statist. 36*(4), 1726–1757.

Arias-Castro, E., D. Donoho, and X. Huo (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory 51*(7), 2402–2425.

Balakrishnan, N. and M. V. Koutras (2002). *Runs and scans with applications.* Wiley Series in Probability and Statistics. New York: Wiley-Interscience [John Wiley & Sons].

Bodineau, T., D. Ioffe, and Y. Velenik (2001). Winterbottom construction for finite range ferromagnetic models: an L^1 approach. *J. Statist. Phys. 105*, 93–131.

Borgs, C., J. T. Chayes, H. Kesten, and J. Spencer (2001). The birth of the infinite cluster: finite-size scaling in percolation. *Comm. Math. Phys. 224*, 153–204.

Brown, L. D. (1986). *Fundamentals of statistical exponential families with applications in statistical decision theory.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 9. Hayward, CA: Institute of Mathematical Statistics.

Cerf, R. (2006). The Wulff crystal in Ising and percolation models. In J. Picard (Ed.), *Ecole d'Eté de Probabilités de Saint Flour*, Volume 1878 of *Lecture Notes in Mathematics*. Berlin: Springer.

Chen, J. and X. Huo (2006). Distribution of the length of the longest significance run on a Bernoulli net and its applications. *J. Amer. Statist. Assoc. 101*(473), 321–331.

Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein (2009). *Introduction to algorithms* (Third ed.). Cambridge, MA: MIT Press.

Culler, D., D. Estrin, and M. Srivastava (2004). Overview of sensor networks. *IEEE Computer 37*(8), 41–49.

DasGupta, B., J. P. Hespanha, J. Riehl, and E. Sontag (2006, November). Honey-pot constrained searching with local sensory information. *Nonlinear Analysis: Hybrid Systems and Applications 65*(9), 1773–1793.

Davies, P. L., M. Langovoy, and O. Wittich (2010). Detection of objects in noisy images based on percolation theory. *Submitted.*

Dembo, A. and O. Zeitouni (2010). *Large deviations techniques and applications*, Volume 38 of *Stochastic Modelling and Applied Probability.* Berlin: Springer-Verlag. Corrected reprint of the second (1998) edition.

Duczmal, L. and R. Assunção (2004). A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Comput. Statist. Data Anal. 45*(2), 269–286.

Erdős, P. and A. Rényi (1970). On a new law of large numbers. *J. Analyse Math. 23*, 103–111.

Falconer, K. J. and G. R. Grimmett (1992). On the geometry of random Cantor sets and fractal percolation. *J. Theor. Probab. 5*, 465–485.

Feng, X., Y. Deng, and H. W. J. Blöte (2008, Sep). Percolation transitions in two dimensions. *Phys. Rev. E 78*(3), 031136.

Geman, D. and B. Jedynak (1996). An active testing model for tracking roads in satellite images. *IEEE Trans. Pattern Anal. Mach. Intell. 18*, 1–14.

Grimmett, G. (1999). *Percolation* (Second ed.), Volume 321 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Berlin: Springer-Verlag.

Grimmett, G. R. (1985). The largest components in a random lattice. *Studia Sci. Math. Hungar. 20*(1-4), 325–331.

Heffernan, R., F. Mostashari, D. Das, A. Karpati, M. Kulldorff, and D. Weiss (2004). Syndromic surveillance in public health practice, New York City. *Emerging Infectious Diseases 10*(5), 858–864.

Hills, R. (2001, July/August). Sensing for danger. *Science & Technology Review*, 11–17.

Hobolth, A., J. Pedersen, and E. B. V. Jensen (2002). A deformable template model, with special reference to elliptical templates. *J. Math. Imaging Vision 17*(2), 131–137. Special issue on statistics of shapes and textures.

Hofstad, R. v. d. and F. Redig (2006). Maximal clusters in non-critical percolation and related models. *J. Statist. Phys. 122*, 670–703.

Jain, A., Y. Zhong, and M. Dubuisson-Jolly (1998). Deformable template models: A review. *Signal Processing 71*(2), 109–129.

Kesten, H. and Y. Zhang (1990). The probability of a large finite cluster in supercritical Bernoulli percolation. *Ann. Probab. 18*, 537–555.

Kulldorff, M. (1997). A spatial scan statistic. *Comm. Statist. Theory Methods 26*(6), 1481–1496.

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *J. Roy. Statist. Soc. Ser. A 164*(1), 61–72.

Kulldorff, M., Z. Fang, and S. J. Walsh (2003). A tree-based scan statistic for database disease surveillance. *Biometrics 59*(2), 323–331.

Kulldorff, M., L. Huang, L. Pickle, and L. Duczmal (2006). An elliptic spatial scan statistic. *Stat Med 25*(22), 3929–43.

Kulldorff, M. and N. Nagarwalla (1995). Spatial disease clusters: detection and inference. *Statistics in medicine 14*(8), 799–810.

Langovoy, M. and O. Wittich (2011). Multiple testing, uncertainty and realistic pictures. Technical report, EURANDOM.

Li, D., K. Wong, Y. H. Hu, and A. Sayeed (2002, Mar). Detection, classification, and tracking of targets. *Signal Processing Magazine, IEEE 19*(2), 17–29.

McInerney, T. and D. Terzopoulos (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis 1*(2), 91–108.

Patil, G., J. Balbus, G. Biging, J. JaJa, W. Myers, and C. Taillie (2004). Multiscale advanced raster map analysis system: Definition, design and development. *Environmental and Ecological Statistics 11*(2), 113–138(26).

Patil, G., S. Joshi, and R. Koli (2010). Pulse, progressive upper level set scan statistic for geospatial hotspot detection. *Environmental and Ecological Statistics 17*, 149–182.

Patil, G. P., R. Modarres, W. L. Myers, and P. Patankar (2006). Spatially constrained clustering and upper level set scan hotspot detection in surveillance geoinformatics. *Environ. Ecol. Stat. 13*(4), 365–377.

Patil, G. P., R. Modarres, and P. Patankar (2005). The ULS software, version 1.0. Center for Statistical Ecology and Environmental Statistics. Department of Statistics, Pennsylvania State University.

Patil, G. P. and C. Taillie (2003). Geographic and network surveillance via scan statistics for critical area detection. *Statist. Sci. 18*(4), 457–465.

Patil, G. P. and C. Taillie (2004). Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environ. Ecol. Stat. 11*(2), 183–197.

Penrose, M. D. (2001). A central limit theorem with applications to percolation, epidemics and Boolean models. *Ann. Probab. 29*(4), 1515–1546.

Penrose, M. D. and A. Pisztora (1996). Large deviations for discrete and continuous percolation. *Adv. in Appl. Probab. 28*(1), 29–52.

Perone Pacifico, M., C. Genovese, I. Verdinelli, and L. Wasserman (2004). False discovery control for random fields. *J. Amer. Statist. Assoc. 99*(468), 1002–1014.

Pisztora, A. (2006). Surface order large deviations for Ising, Potts and percolation models. *Prob. Th. Rel. Fields 104*, 427–466.

Pozo, D., F. Olmo, and L. Alados-Arboledas (1997). Fire detection and growth monitoring using a multitemporal technique on avhrr mid-infrared and thermal channels. *Remote Sensing of Environment 60*(2), 111–120.

Rotz, L. and J. Hughes (2004). Advances in detecting and responding to threats from bioterrorism and emerging infectious disease. *Nature Medicine 10*, S130–S136.

Smirnov, S. and W. Werner (2001). Critical exponents for two-dimensional percolation. *Math. Res. Lett. 8*, 729–744.

Tango, T. and K. Takahashi (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics 4*(1), 11.

Wagner, M., F. Tsui, J. Espino, V. Dato, D. Sittig, R. Caruana, L. Mcginnis, D. Deerfield, M. Druzdzel, and D. Fridsma (2001). The emerging science of very early detection of disease outbreaks. *Journal of Public Health Management and Practice 7*(6), 51–59.

Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist. 38*(2), 1010–1033.