# Detecting a Path of Correlations in a Network

Ery Arias-Castro[*] Gábor Lugosi[†] Nicolas Verzelen[‡]

**Abstract**

We consider the problem of detecting an anomaly in the form of a path of correlations hidden in white noise. We provide a minimax lower bound and a test that, under mild assumptions, is able to achieve the lower bound up to a multiplicative constant.

## 1 Introduction

Anomaly detection arises in many applications, including surveillance, the detection of suspicious objects from satellite images or sensor networks, as well as in medical imaging (e.g., tumor detection). While in some applications the object can be assumed to present a larger signal amplitude (e.g., pixel level in images), in other settings it manifests instead as correlations. For example, the object to be detected in an image has different texture but same average pixel amplitude; or in the case of the evolution of the price of a stock, an event could trigger more volatility instead of a monotone change in the value of the stock. The problem of detecting the presence of a subset of observations with different mean from the rest is the so-called *detection-of-means* problem. On the other hand, when one's aim is to detect the presence of a subset of unusually correlated observations, we speak about the *detection-of-correlations* problem.

The detection-of-means problem has been extensively studied in the literature, both applied and theoretical. Papers that develop theory include (Addario-Berry et al., 2010; Arias-Castro et al., 2011, 2008, 2005; Desolneux et al., 2003; Walther, 2010). The literature on detection of correlations is more modest, and while the applied literature is sizable, few papers develop theory beyond the one-dimensional case of change-point detection in times series. In a few recent papers, we developed elements of the first minimax theory, see (Arias-Castro et al., 2012, 2015) and (Arias-Castro et al., 2015). In the present paper we focus on detecting a path of correlations in a general graph. This setting could model an attack in a computer network (Mukherjee et al., 1994; Zhang and Lee, 2000). We develop some minimax theory for this problem.

### 1.1 Context and previous work

The setup is similar to the one studied in (Arias-Castro et al., 2015). We model the data as a Gaussian random field $X = (X_i : i \in \mathcal{V})$, where $\mathcal{V}$ is a set of size $|\mathcal{V}| = n$. The $X_i$'s are assumed to have zero mean and unit variance. Under the null hypothesis $\mathcal{H}_0$, the $X_i$'s are independent. Under the alternative hypothesis $\mathcal{H}_1$, the $X_i$'s are correlated in one of the following ways. Let $\mathcal{C}$ be a class of subsets of $\mathcal{V}$. Each set $S \in \mathcal{C}$ represents a possible anomalous subset of the components of $X$. Specifically, when $S \in \mathcal{C}$ is the anomalous subset of nodes, each $X_i$ with $i \notin S$ is still

[*]Department of Mathematics, University of California, San Diego
[†]ICREA and Department of Economics, Universitat Pompeu Fabra
[‡](corresponding author) INRA, UMR 729 MISTEA, F-34060 Montpellier, FRANCE

independent of all the other variables, while $(X_i)_{i\in S}$ coincides with $(Y_i)_{i\in S}$, where $Y = (Y_i)_{i\in\mathcal{V}}$ is a Gaussian Markov random field. That last assumption on $Y$ is what distinguishes the model of (Arias-Castro et al., 2015) from the ones studied in (Arias-Castro et al., 2012, 2015). In (Arias-Castro et al., 2015), we focus on the problem of detecting a "blob-like" region of correlations in the $d$-dimensional integer lattice, which corresponds to this setting with

$$\mathcal{V} = \{1, \dots, m\}^d , \tag{1}$$

and $\mathcal{C}$ a given class of "blob-like" subsets. This special case covers examples such as the detection of a textured object in an image of dimension $d$.

### 1.2 Detecting paths of correlation in general graphs

Although the case when the components of the observed vector correspond to points in the integer lattice comprises some of the most important cases such as that of detecting an interval of correlations in a signal (for $d = 1$) or that of detecting a textured object in an image ($d = 2$), it does not directly apply to other common situations, such as those arising in sensor networks.

Such, more general, examples have been studied in the context of the detection-of-means problem, with theory available in (Addario-Berry et al., 2010; Arias-Castro et al., 2011, 2008; Sharpnack et al., 2013). However, extending the results of (Arias-Castro et al., 2015) to more general graphs is nontrivial and even difficult to formulate. While general results for general graphs can be rather complicated to establish, we address the concrete example of detecting a *path* of correlations in a general graph.

The setting is similar to that of (Arias-Castro et al., 2015), but also different. Specifically, we consider a general (possibly directed) finite graph $\mathcal{G}$ with vertex set $\mathcal{V}$ and we let $\mathcal{C}$ be the class of subgraphs of $\mathcal{G}$ such that each $S \in \mathcal{C}$ forms a self-avoiding path. When $S \in \mathcal{C}$ is the anomalous path, there is an autoregressive process of order 1 along the path $S$.

### 1.3 Contribution and content

In the spirit of the detection literature cited above, we study the problem of detecting a path of correlations in a graph from a minimax perspective. In Section 2, after formalizing the problem, we derive a lower bound when the correlation parameter is known and then design a testing procedure which achieves this lower bound (up to a multiplicative constant depending on some graph characteristics) without knowledge of the correlation parameter. In Section 3 we specialize our general results to the case of detecting a path of correlations in an integer lattice. The proofs are gathered in Section 4.

## 2 Setting and general results

### 2.1 Formulation of the problem

We are given a finite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ denoting the set of nodes (or vertices) and $\mathcal{E} \subset \mathcal{V}\times\mathcal{V}$ denoting the set of edges. We let $n = |\mathcal{V}|$ denote the number of nodes in the graph. We are also given a class of open, self-avoiding paths of $\mathcal{G}$, denoted $\mathcal{C}$. Recall that an open self-avoiding path in $\mathcal{G}$ is a sequence of nodes $S = (s_1, \dots, s_k) \in \mathcal{V}^k$ such that $(s_j, s_{j+1}) \in \mathcal{E}$ for all $j$ and $s_j \neq s_{j'}$ for all $j \neq j'$. We observe a vector of random variables indexed by $\mathcal{V}$, denoted $X = (X_i)_{i\in\mathcal{V}}$, assumed to be standard normal. Under the null hypothesis all components of $X$ are independent. Under the alternative hypothesis, one of the paths $S \in \mathcal{C}$ is "anomalous", in which case $(X_i)_{i\in S}$

is an autoregressive model of order 1 with correlation coefficient $\psi \in (-1, 1)$, while the other components of $X$ are still independent and $(X_i)_{i\in S}$ and $(X_i)_{i\notin S}$ are also independent. This means that, if $S = (s_1, \ldots, s_k)$, then $X_{s_{j+1}} - \psi X_{s_j}, j = 1, \ldots, k-1$, are independent centered normal random variables with variance $1 - \psi^2$. We denote the distribution of $X$ under $\mathcal{H}_0$ by $\mathbb{P}_0$. We denote the distribution of $X$ under $\mathcal{H}_1$ by $\mathbb{P}_{S,\psi}$ when $S \in \mathcal{C}$ is the anomalous set and $\psi$ is the autocorrelation coefficient.

A *test* is a measurable function $f : \mathbb{R}^{\mathcal{V}} \to \{0, 1\}$. When $f(X) = 0$, the test accepts the null hypothesis and it rejects it otherwise. The probability of *type I* error of a test $f$ is $\mathbb{P}_0\{f(X) = 1\}$. Under the alternative $(S, \psi) \in \mathcal{C} \times (-1, 1)$, the probability of *type II* error is $\mathbb{P}_{S,\psi}\{f(X) = 0\}$. In this paper we evaluate tests based on their *worst-case risks.* When the correlation coefficient $\psi$ is known, the risk of a test $f$ corresponding to the class $\mathcal{C}$ is defined as

$$R_{\mathcal{C},\psi}(f) = \mathbb{P}_0\{f(X) = 1\} + \max_{S\in\mathcal{C}} \mathbb{P}_{S,\psi}\{f(X) = 0\} \ .$$

In this case, the minimax risk is defined as

$$R^*_{\mathcal{C},\psi} = \inf_f R_{\mathcal{C},\psi}(f) \ ,$$

where the infimum is over all tests $f$. When $\psi$ is only known to belong to an interval $\mathfrak{I} \subset (-1, 1)$, it is more meaningful to define the risk of a test $f$ as

$$R_{\mathcal{C},\mathfrak{I}}(f) = \mathbb{P}_0\{f(X) = 1\} + \max_{\psi\in\mathfrak{I}} \max_{S\in\mathcal{C}} \mathbb{P}_{S,\psi}\{f(X) = 0\} \ .$$

The corresponding minimax risk is defined as

$$R^*_{\mathcal{C},\mathfrak{I}} = \inf_f R_{\mathcal{C},\mathfrak{I}}(f) \ .$$

When $\psi$ is known (resp. unknown), we say that a test $f$ *asymptotically separates the two hypotheses* if $R_{\mathcal{C},\psi}(f) \to 0$ (resp. $R_{\mathcal{C},\mathfrak{I}}(f) \to 0$), and we say that the hypotheses *merge asymptotically* if $R^*_{\mathcal{C},\psi} \to 1$ (resp. $R^*_{\mathcal{C},\mathfrak{I}} \to 1$), as $n = |\mathcal{V}| \to \infty$. We note that, as long as $\psi \in \mathfrak{I}$, $R^*_{\mathcal{C},\psi} \leq R^*_{\mathcal{C},\mathfrak{I}}$ and that $R^*_{\mathcal{C},\mathfrak{I}} \leq 1$, since the test $f \equiv 1$ that always rejects $\mathcal{H}_0$ has risk equal to 1.

In this paper, we characterize the minimax testing risk in the setting where $\psi$ is known ($R^*_{\mathcal{C},\psi}$) and in the setting when it is unknown ($R^*_{\mathcal{C},\mathfrak{I}}$). That is, we give conditions on $\mathcal{C}$ and $\mathfrak{I}$ under which the hypotheses merge asymptotically so that the detection problem is nearly impossible. We then exhibit a (nonstandard) test that asymptotically separates the hypotheses under essentially the same conditions.

## 2.2 A general lower bound

The main difference between the case of anomalous paths treated here and the case of anomalous blobs studied in (Arias-Castro et al., 2015) is that a lower bound for the latter can be developed based on a subclass of disjoint subsets. Here, however, a reduction to a subclass of disjoint paths is typically too severe. We thus develop a new lower bound tailored to the present situation, which, in particular, allows for possible anomalous subsets to intersect. For any prior distribution $\nu$ on $\mathcal{C}$, the minimax risk is at least as large as the $\nu$-average risk, $R^*_{\mathcal{C},\psi} \geq \bar{R}^*_{\nu,\psi}$, where

$$\bar{R}_{\nu,\psi}(f) = \mathbb{P}_0\{f(X) = 1\} + \sum_{S\in\mathcal{C}} \nu(S)\mathbb{P}_{S,\psi}\{f(X) = 0\} \quad \text{and} \quad \bar{R}^*_{\nu,\psi} = \inf_f \bar{R}_{\nu,\pi}(f) \ . \tag{2}$$

The following result provides a lower bound on the latter.

**Theorem 1.** *Let $\mathcal{C}$ be a class of open self-avoiding paths of $\mathcal{G}$ and let $\nu$ denote some prior over $\mathcal{C}$. Assume that $|\psi| \le 1/9$. Then*

$$\bar{R}^*_{\nu,\psi} \ge 1 - \frac{1}{2}\sqrt{\mathbb{E}_{\nu\otimes\nu}\left[\exp\left(\lambda(\psi)|S \cap T|\right)\right] - 1}\ ,$$

*where*

$$\lambda(\psi) := \frac{1}{4}\left[\left(\frac{1-|\psi|}{1-9|\psi|}\right)^{1/2} - \frac{1+3|\psi|}{1-|\psi|}\right]$$

*and the expectation is with respect to $S, T$ drawn i.i.d. from $\nu$.*

The function $\lambda$ is even and increasing on $(0, 1/9)$, and $\lambda(1/10) = 7/18$. Hence, the bound implies that

$$\bar{R}^*_{\nu,\psi} \ge 1 - \frac{1}{2}\sqrt{\mathbb{E}_{\nu\otimes\nu}\left[\exp\left(\tfrac{7}{18}|S \cap T|\right)\right] - 1}\ , \quad \text{when } |\psi| \le 1/10.$$

*Remark* 1. The condition on $\psi$ is likely to be an artefact of our proof technique. As is standard in the literature, the proof relies on bounding the variance of the likelihood ratio resulting from averaging the alternatives according to $\nu$. A more refined approach such as that of bounding the second moment of the likelihood ratio after truncation—a well-known technique in the detection-of-means setting first proposed by Yuri Ingster—might lead to a sharpening of this result, but the computations for the present case seem to be a highly nontrivial challenge.

### 2.3 A general upper bound

A natural approach in related testing problems is the generalized likelihood ratio test. When $\psi$ is known, this test is based on rejecting the null hypothesis for large values of

$$\max_{S\in\mathcal{C}} X_S^\top (\mathbf{I}_S - \mathbf{\Gamma}_S^{-1}(\psi))X_S\ ,$$

where $\mathbf{\Gamma}_S(\psi)$ denotes the covariance matrix of an autoregressive model of order 1 indexed by $S$ and with parameter $\psi$ and $X_S = \sum_{i\in S} X_i$. Establishing a useful performance bound for this test appears surprisingly challenging due to our lack of understanding of concentration properties of the test statistic under the null hypothesis. In particular, our effort to combine the union bound with a standard concentration bound for Gaussian quadratic forms (i.e., Gaussian chaoses of order 2) were inconclusive. The situation is even more complicated when $\psi$ is unknown.

However, we were able to craft and analyze an ad-hoc test based on pairwise comparisons of consecutive values along a path. For simplicity assume that all paths in $\mathcal{C}$ are of same length $k$. (When this is not the case, typically the test needs to be repeated for all possible lengths and the resulting multiple testing situation is resolved by applying Bonferroni's method.) Fix a threshold $t > 0$. For $S \in \mathcal{C}$ seen as a sequence of indices $S = (s_j : j = 1, \dots, k) \subset \mathcal{V}$, define

$$V_{t,S} = \sum_{j=1}^{k-1} V_{t,S}(j)\ , \quad V_{t,S}(j) = \mathbb{I}\{|X_{s_{j+1}} - X_{s_j}| \le \sqrt{2}t\}\ ,$$

and consider the statistic and the test

$$V_t^* = \max_{S\in\mathcal{C}} V_{t,S}\ , \quad f = \mathbb{I}\{V_t^* > k/2\}\ . \tag{3}$$

**Proposition 1.** *Let $\mathcal{C}$ be a class of paths of length $k$. For $t \in \mathbb{R}$, define $p_t = 2F(t) - 1$ where $F$ denotes the standard normal distribution function. Let $t > 0$ be largest such that $h(2p_t) \geq \frac{8}{k}\log(|\mathcal{C}|) \vee 1$ where $h(x) := x - \log(x) - 1$. If $\psi$ exceeds $1 - (t/F^{-1}(4/5))^2$, then the test $f$ defined in (3) satisfies $R_{\mathcal{C},\psi}(f) \to 0$ as $k \to \infty$.*

*Remark* 2. The proposition only applies to positive $\psi$ and in fact the test (3) is only useful in that case. To handle the case where $\psi$ is negative, we use instead the same procedure except that $V_{t,S}(j)$ is replaced with

$$\tilde{V}_{t,S}(j) = \mathbb{I}\{|X_{s_{j+1}} + X_{s_j}| \leq \sqrt{2}t\} \ ,$$

The resulting test achieves a similar performance. In practice, if the sign of $\psi$ is a priori unknown, one can simply combine these two tests using a Bonferroni correction. In the rest of the paper we only consider $\psi > 0$.

*Remark* 3. Computing the test statistic $V_t^*$ is difficult, even when the starting point is known. Indeed, this problem is known as the *prize collecting salesman problem* or *bank robber problem* or *reward-budget problem*. Even in the case when the underlying graph is the integer lattice there are no known polynomial-time algorithms that solve it, although polynomial approximations do exist, see (DasGupta et al., 2006). An alternative is the test based on the length of the longest path of significant adjacent correlations. This is inspired from some proposals in the detection-of-means setting (Arias-Castro et al., 2006; Arias-Castro and Grimmett, 2013). Quick calculations suggest that the test achieves a comparable theoretical performance, although we expect this test to be less powerful in practice.

# 3 Special case: the lattice

Consider the integer lattice (1) in dimension $d \geq 3$. The story is a little different when $d = 2$, and we refer the reader to the treatment in (Arias-Castro et al., 2008) carried out in the detection-of-means setting, as we expect a similar phenomenon to hold in the present context. For simplicity, to guarantee that all nodes play a symmetric role, we take the lattice to be a torus.

## 3.1 Known starting point

Suppose first that the departing vertex of the path is known, meaning that $\mathcal{C}$ is the class of all self-avoiding paths with $k$ nodes in $\mathcal{V}$ starting at some given $v_0 \in \mathcal{V}$. In that case, when $d \geq 3$, there is a constant $C > 0$ such that, when $|\psi| \leq C$, the risk is at least 1/2. To see this, let $\nu$ be a prior on $\mathcal{C}$ that has exponential intersection tails, which means that there exist some constants $\eta \in (0, 1)$ and $C_0 > 0$ such that

$$\mathbb{P}_{\nu\otimes\nu}(|S \cap T| \geq \ell) \leq C_0\eta^\ell, \quad \forall \ell \geq 1 \ , \tag{4}$$

where $S, T$ are i.i.d. from $\nu$. This concept was introduced by (Benjamini et al., 1998, Theorem 1.3) where it is shown that such a prior exists on infinite paths ($k = \infty$) in the infinite $d$-dimensional integer lattice ($m = \infty$) when $d = 3$. In fact, it is constructed with support on oriented paths of taking steps in $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. Obviously, in the finite case, it suffices to restrict such a prior on the first $k-1$ steps and property (4) still holds when $k \leq m$, which we assume henceforth.[1] Note that we can use the same prior when $d \geq 3$, by embedding the 3-dimensional integer lattice into the $d$-dimensional integer lattice. Because of this, we can take $\eta$ to be a numeric constant, not even depending on the dimension.

[1] Clearly, an upper bound on $k$ is needed for there are no self-avoiding paths when $k > n$.

Thus consider a prior $\nu$ satisfying (4). Then, for any $a > 0$ small enough that $e^a \eta < 1$,

$$\begin{aligned}
\mathbb{E}_{\nu\otimes\nu}[\exp(a|S\cap T|)] &= \sum_{\ell\geq 1} e^{a\ell}\, \mathbb{P}_{\nu\otimes\nu}(|S\cap T| = \ell) \\
&= e^a + \sum_{\ell\geq 2} (e^{a\ell} - e^{a(\ell-1)})\, \mathbb{P}_{\nu\otimes\nu}(|S\cap T| \geq \ell) \\
&\leq \Xi(a) := e^a + C_0 \frac{(e^a-1)e^a\eta^2}{1-e^a\eta} \ .
\end{aligned}$$

We use this upper bound in Theorem 1 with $a = \lambda(|\psi|)$ and get that $\bar{R}^*_{\nu,\psi} \geq 1/2$ when $|\psi|$ is sufficiently small that $\eta \exp(\lambda(|\psi|)) < 1$ and $\Xi(\lambda(|\psi|)) \leq 2$. This is possible since $\lambda(|\psi|) \to 0$ as $|\psi| \to 0$ and $\Xi(a) \to 1$ as $a \to 0$.

Conversely, there exists a positive constant depending only on the dimension $d$ such that, if $\psi$ is larger than that constant, then the test defined in (3) asymptotically separates the hypotheses. This comes from a simple application of Proposition 1 together with the fact that, in the $d$-dimensional integer lattice, there are at most $(2d)^{k-1}$ paths of length $k$ starting at a given node. Following Remark 2, we can handle the case when $|\psi|$ is large enough in a similar way. From this discussion, we arrive at the following.

**Corollary 1.** *Consider the $d$-dimensional integer lattice (1) with $d \geq 3$ seen as a torus. Let $\mathcal{C}$ denote the class of self-avoiding paths of length $k \leq m$ starting at a known location. Assume the autocorrelation coefficient $\psi$ is fixed. There exist constants $0 < C_1 \leq C_2 < 1$ depending only on $d$ such that, when $|\psi| < C_1$, $\liminf_{k\to\infty} R^*_{\mathcal{C},\psi} > 0$, while when $|\psi| > C_2$, $\lim_{k\to\infty} R^*_{\mathcal{C},\psi} = 0$.*

*Remark* 4. As long as $k \leq m$, the size of the lattice does not matter since the starting location is known. Therefore, an asymptotic analysis is necessarily in terms of large $k$.

*Remark* 5. We conjecture that the constants in Corollary 1 are identical, meaning, that $C_1 = C_2$. Our arguments show that this is so if we replace lim inf with lim sup in the statement, for in that case one can take

$$C_1 = C_2 = C_\ddagger := \inf\left\{C > 0 : \lim_{k\to\infty} R^*_{\mathcal{C},\psi} = 0 \text{ when } |\psi| > C\right\}\ , \tag{5}$$

and the corollary implies that $C_\ddagger \in (0, 1)$.

### 3.2 Unknown starting point

In the general case, the location of the possible anomalous paths is unknown. In other words, $\mathcal{C}$ is the class of all self-avoiding paths with $k$ nodes in $\mathcal{V}$. Assume that $m = n^{1/d}$ is a multiple of $2k$ for simplicity. To define the prior, we partition the lattice into hypercubes of side length $2k - 1$, indexed by $J$, and let $v_j$ denote the center of the hypercube $j \in J$. The number of such hypercubes satisfies $|J| \sim (m/2k)^d = n/(2k)^d$. Still in dimension $d \geq 3$, let $\nu_j$ be a prior on self-avoiding paths starting at $v_j$ satisfying (4) and let $\nu$ be the even mixture of all these priors. Noting that paths starting from different origin nodes cannot intersect, for any $a > 0$ small enough that $e^a\eta < 1$,

$$\mathbb{E}_{\nu\otimes\nu}[\exp(a|S\cap T|)] = 1 - \frac{1}{|J|} + \frac{1}{|J|}\mathbb{E}_{\nu_j\otimes\nu_j}[\exp(a|S\cap T|)] \leq 1 + \frac{\Xi(a)}{|J|}\ ,$$

where $j \in J$ is arbitrary. We use this upper bound in Theorem 1 with $a = \lambda(|\psi|)$ and get that $\bar{R}^*_{\nu,\psi} \geq 1 - \frac{1}{2}\sqrt{2/|J|}$ when $|\psi|$ is so small that $\eta\exp(\lambda(|\psi|)) < 1$ and $\Xi(\lambda(|\psi|)) \leq 2$. Noting that $|J| \to \infty$ when $m \gg k$ (i.e., the size of the grid dominates the path length), we see that $\bar{R}^*_{\nu,\psi} \to 1$.

Conversely, assume that $k/\log n \geq C_0$ for some positive constant $C_0$. Then there exists another positive constant depending only on the dimension $d$ and $C_0$ such that, if $\psi$ is larger than that constant, then the test defined in (3) asymptotically separates the hypotheses. This comes from Proposition 1 and the fact that, in the $d$-dimensional integer lattice (1) with a total of $n = m^d$ nodes, there are at most $n(2d)^{k-1}$ paths of length $k$. Following Remark 2, we can handle the case where $|\psi|$ is large enough in a similar way. We thus arrive at the following.

**Corollary 2.** *Consider the context of Corollary 1 but now assuming that the starting location is unknown and that $k/\log m \geq C_0$ for some constant $C_0 > 0$. There exist constants $0 < C_1 \leq C_2 < 1$ depending only on $d$ and $C_0$ such that, when $|\psi| < C_1$, $\lim_{m\to\infty} R^*_{\mathcal{C},\psi} = 1$, while when $|\psi| > C_2$, $\lim_{m\to\infty} R^*_{\mathcal{C},\psi} = 0$.*

*Remark* 6. As in Remark 5, we conjecture that we may take $C_1 = C_2$, defined as in (5). (This definition would lead to a different constant in the present setting.)

*Remark* 7. The constants in Corollary 1 and Corollary 2 are implicit. Our analysis provides some nontrivial bounds on these constants. However, it is not precise enough to lead to the exact values. We note that the same is true in the (simpler) detection-of-means setting (Arias-Castro et al., 2008). Also, we reveal a regime where the minimal correlation coefficient $\psi$ allowing asymptotic hypotheses separation (i.e., $R^*_{\mathcal{C},\psi} \to 0$) is bounded away from 0 and 1 when both $k$ and $n$ go to infinity. (The situation is qualitatively different in dimension $d = 2$ and we refer the reader to (Arias-Castro et al., 2008) for a detailed treatment of that case in the detection-of-means setting.)

# 4 Proofs

## 4.1 Preliminaries

Let $\mathbf{\Gamma}(\psi)$ denote the covariance operator of an (infinite) autoregressive model of order 1 with parameter coefficient $\psi$. The operator $\mathbf{\Gamma}(\psi)$ is positive definite and invertible when $|\psi| < 1$. Note that any $S \in \mathcal{C}$ is homomorphic to $\{1, \dots, |S|\}$, and identifying the two, we have $(\mathbf{\Gamma}_S(\psi))_{i,j} = \psi^{|i-j|}$, where $\mathbf{\Gamma}_S(\psi)$ is the principal submatrix of the covariance operator $\mathbf{\Gamma}(\psi)$ indexed by $S$.

Any autoregressive process $Y = (Y_i)_{i\in\mathbb{Z}}$ of order 1 with parameter $\psi$ can be represented as a Gaussian Markov random field (GMRF) on the line. We have the decomposition

$$Y_i = \phi Y_{i-1} + \phi Y_{i+1} + \epsilon_i \ , \tag{6}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_\phi^2)$ is independent of $(Y_j)_{j\neq i}$ and

$$\phi := \frac{\psi}{1+\psi^2} \ , \qquad \sigma_\phi^2 := \frac{1-\psi^2}{1+\psi^2} \ . \tag{7}$$

We start with some simple remarks relating autoregressive processes of order 1 to GMRFs. See (Guyon, 1995, Sect. 1.3) for more details. This representation of a stationary autoregressive process enables us to adapt some of the arguments developed in (Arias-Castro et al., 2015) for stationary GMRFs.

**Lemma 1.** *Identify $S$ with $(1, \dots, |S|)$ and consider any $\psi \in (-1, 1)$. Then*

$$(\mathbf{\Gamma}_S^{-1}(\psi))_{i,j} = \begin{cases} 1/\sigma_\phi^2 & \text{if } i = j \text{ and } i \in \{2, \dots, |S|-1\}, \\ 1/(1-\psi^2) & \text{if } i \in \{1, |S|\}, \\ -\phi/\sigma_\phi^2 & \text{if } |i-j| = 1, \\ 0 & \text{if } |i-j| > 1. \end{cases}$$

*Proof.* We leave $\psi$ implicit and write $\mathbf{\Gamma}$ for $\mathbf{\Gamma}(\psi)$. Denote by $Y_S$ the restriction of the stationary autoregressive process $Y$ to $S$. First consider any index $i \in \{2, \ldots, |S|-1\}$. By the Markov property (6), conditionally to $(Y_{i-1}, Y_{i+1})$, $Y_i$ is independent to all the remaining variables. Thus, the conditional distribution of $Y_i$ given $(Y_j)_{j \neq i}$ is the same as the conditional distribution of $Y_i$ given $(Y_j)_{j \in S \setminus \{i\}}$. This conditional distribution characterizes the $i$-th row of the inverse covariance matrix $\mathbf{\Gamma}_S^{-1}$. More precisely, the conditional variance $\sigma_\phi^2$ of $Y_i$ given $Y_S$ is $[(\mathbf{\Gamma}_S^{-1})_{i,i}]^{-1}$. Furthermore, $-(\mathbf{\Gamma}^{-1})_{i,j}/(\mathbf{\Gamma}^{-1})_{i,i}$ is the $j$-th parameter of the conditional regression (6) of $Y_i$ given $Y^{(i)}$, and therefore we conclude that $(\mathbf{\Gamma}^{-1})_{i,i} = (\sigma_\phi^2)^{-1} = (\mathbf{\Gamma}_S^{-1})_{i,i}$ and $(\mathbf{\Gamma}^{-1})_{i,j}/(\mathbf{\Gamma}^{-1})_{i,i}$ equals $-\phi$ if $|j-i| = 1$ and is zero otherwise.

Now consider the case $i = |S|$, $i = 1$ being handled similarly. Since $\mathbf{\Gamma}_S^{-1}$ is a symmetric matrix, we only have to compute $(\mathbf{\Gamma}_S^{-1})_{|S|,|S|}$ and $(\mathbf{\Gamma}_S^{-1})_{|S|,1}$. By definition of autoregressive processes, we have

$$Y_{|S|} = \psi Y_{|S|-1} + \omega_{|S|} \ ,$$

where $\omega_{|S|} \sim \mathcal{N}(0, 1-\psi^2)$ is independent of $(Y_1, \ldots Y_{|S|-2})$. The above expression characterizes the conditional regression of $Y_{|S|}$ given $(Y_1, \ldots Y_{|S|-1})$. Arguing as previously, we conclude that $(\mathbf{\Gamma}^{-1})_{|S|,1} = 0$ and $(\mathbf{\Gamma}^{-1})_{|S|,|S|} = 1/(1-\psi^2)$. □

We let $\|\mathbf{A}\|$ denote the operator norm of a matrix $\mathbf{A}$.

**Lemma 2.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be (complex or real) matrices of same dimensions. Let* $\mathrm{col}(\mathbf{A})$ *index the column vectors of* $\mathbf{A}$ *that are nonzero. Then*

$$|\operatorname{Tr}(\mathbf{A}^\top \mathbf{B})| \leq |\mathrm{col}(\mathbf{A}) \cap \mathrm{col}(\mathbf{B})| \|\mathbf{A}\| \|\mathbf{B}\| \ .$$

*Proof.* Define the index set $J = \mathrm{col}(\mathbf{A}) \cap \mathrm{col}(\mathbf{B})$. Let $\mathbf{A}_J$ denote the submatrix of $\mathbf{A}$ with columns indexed by $J$, and define $\mathbf{B}_J$ similarly. We then have

$$|\operatorname{Tr}(\mathbf{A}^\top \mathbf{B})| = |\operatorname{Tr}(\mathbf{A}_J^\top \mathbf{B}_J)| \leq |J| \|\mathbf{A}_J^\top \mathbf{B}_J\| \leq |J| \|\mathbf{A}_J\| \|\mathbf{B}_J\| \leq |J| \|\mathbf{A}\| \|\mathbf{B}\| \ .$$ □

## 4.2 Proof of Theorem 1

Recall that with $\psi$ fixed, $\mathbf{\Gamma}_S$ denotes the covariance matrix of an autoregressive model of order 1 of length $|S|$ and with parameter $\psi$. As explained in (Arias-Castro et al., 2015), the $\nu$-average risk is

$$\bar{R}_\nu^* = 1 - \frac{1}{2} \mathbb{E}_0 |L_\nu(X) - 1| \geq 1 - \frac{1}{2}\sqrt{\mathbb{E}_0[L_\nu(X)^2] - 1} \ , \tag{8}$$

where

$$L_\nu(x) = \sum_{S \in \mathcal{C}} \nu(S) L_S(x) \ ,$$

with

$$L_S(x) = \exp\left(\tfrac{1}{2} x_S^\top (\mathbf{I}_S - \mathbf{\Gamma}_S^{-1}) x_S - \tfrac{1}{2} \log\det(\mathbf{\Gamma}_S)\right) \ . \tag{9}$$

We need to upper bound

$$\mathbb{E}_0[L_\nu(X)^2] = \sum_{S,T \in \mathcal{C}} \nu(S)\nu(T)\, \mathbb{E}_0[L_S(X) L_T(X)] \ . \tag{10}$$

Unlike in the setting that concerns (Arias-Castro et al., 2015), here two subgraphs $S$ and $T$ in the support of the prior $\nu$ may not be disjoint, and if $S \cap T \neq \emptyset$, $L_S(X)$ and $L_T(X)$ are not independent.

Before we proceed, we leave the dependency in $X$ implicit and we formally index the elements of $S$ in $\{1,\dots,k\}$. By Lemma 1, we have

$$1 \le \left(\mathbf{\Gamma}_S^{-1}\right)_{1,1} = \left(\mathbf{\Gamma}_S^{-1}\right)_{k,k} = \frac{1}{1-\psi^2} \le \left(\mathbf{\Gamma}_S^{-1}\right)_{i,i} = \frac{1}{\sigma_\phi^2}\ , \quad \forall i \in \{2,\dots k-1\}\ ,$$

and

$$\left(\mathbf{\Gamma}_S^{-1}\right)_{i,i+1} = \left(\mathbf{\Gamma}_S^{-1}\right)_{i,i-1} = -\frac{\phi}{\sigma_\phi^2}\ ,$$

while all the other entries of $\mathbf{\Gamma}_S^{-1}$ are zero. Hence, the matrix $\mathbf{A}_S := \mathbf{\Gamma}_S^{-1} - \mathbf{I}_S$ satisfies

$$\|\mathbf{A}_S\| \le \sup_{j\in S}\sum_{i\in S}|(\mathbf{A}_S)_{i,j}| \le \frac{2|\phi|}{\sigma_\phi^2} + \frac{1-\sigma_\phi^2}{\sigma_\phi^2} = \frac{2|\psi|}{1-|\psi|}\ ,$$

where we used the expressions (7) of $\phi$ and $\sigma_\phi$ . Let us fix $S$ and $T$ two anomalous subsets in $\mathcal{C}$. In the following $\widetilde{\mathbf{\Gamma}}_S$ (resp. $\widetilde{\mathbf{\Gamma}}_T$) denotes the covariance of $X_{S\cup T}$ when $X \sim \mathbb{P}_{S,\psi}$ (resp. $X \sim \mathbb{P}_{T,\psi}$). Note that the restriction of $\widetilde{\mathbf{\Gamma}}_S$ to $S\times S$ is exactly $\mathbf{\Gamma}_S$ whereas its restriction to $(T\setminus S)\times(T\setminus S)$ is the identity matrix. We have

$$\begin{aligned}
\mathbb{E}[L_S L_T] &= \mathbb{E}\left[\exp\left(X_{S\cup T}^\top(\mathbf{I}_{S\cup T} - \tfrac12\widetilde{\mathbf{\Gamma}}_S^{-1} - \tfrac12\widetilde{\mathbf{\Gamma}}_T^{-1})X_{S\cup T} - \tfrac12\log\det(\mathbf{\Gamma}_S) - \tfrac12\log\det(\mathbf{\Gamma}_T)\right)\right] \\
&= \exp\left(-\tfrac12\log\det(\widetilde{\mathbf{\Gamma}}_S^{-1} + \widetilde{\mathbf{\Gamma}}_T^{-1} - \mathbf{I}_{S\cup T}) + \tfrac12\log\det(\mathbf{\Gamma}_S^{-1}) + \tfrac12\log\det(\mathbf{\Gamma}_T^{-1})\right)\ ,
\end{aligned}$$

using the fact that

$$\|\mathbf{I}_{S\cup T} - \tfrac12\widetilde{\mathbf{\Gamma}}_S^{-1} - \tfrac12\widetilde{\mathbf{\Gamma}}_T^{-1}\| \le \frac12\|\mathbf{A}_S\| + \frac12\|\mathbf{A}_T\| = \|\mathbf{A}_S\| \le \frac{2|\psi|}{1-|\psi|}\ , \tag{11}$$

we have that the last term is strictly less than 1/2 since $|\psi| \le 1/5$. Define $\widetilde{\mathbf{A}}_S = \mathbf{I}_{S\cup T} - \widetilde{\mathbf{\Gamma}}_S$ and $\widetilde{\mathbf{A}}_T$ similarly. Using these bounds, together with the fact that, for a symmetric matrix $\mathbf{B}$ with operator norm strictly less than 1,

$$\log\det(\mathbf{I}-\mathbf{B}) = \mathrm{Tr}\log(\mathbf{I}-\mathbf{B}) = -\sum_{\ell=1}^{\infty}\frac{1}{\ell}\,\mathrm{Tr}(\mathbf{B}^\ell)\ ,$$

we get

$$\begin{aligned}
\Lambda :&= -\tfrac12\log\det(\widetilde{\mathbf{\Gamma}}_S^{-1} + \widetilde{\mathbf{\Gamma}}_T^{-1} - \mathbf{I}_{S\cup T}) + \tfrac12\log\det(\mathbf{\Gamma}_S^{-1}) + \tfrac12\log\det(\mathbf{\Gamma}_T^{-1}) \\
&= \frac12\sum_{\ell=1}^{\infty}\frac{1}{\ell}\left(\mathrm{Tr}\left[(\widetilde{\mathbf{A}}_S + \widetilde{\mathbf{A}}_T)^\ell\right] - \mathrm{Tr}\left[\widetilde{\mathbf{A}}_S^\ell\right] - \mathrm{Tr}\left[\widetilde{\mathbf{A}}_T^\ell\right]\right) \\
&= \frac12\sum_{\ell=2}^{\infty}\frac{1}{\ell}\sum_{(s,t)\in Q_\ell}\mathrm{Tr}\left[\widetilde{\mathbf{A}}_S^{s_1}\widetilde{\mathbf{A}}_T^{t_1}\cdots\widetilde{\mathbf{A}}_S^{s_\ell}\widetilde{\mathbf{A}}_T^{t_\ell}\right]\ ,
\end{aligned}$$

where

$$Q_\ell = \left\{(s,t) \in (\{0,1\}^\ell \setminus \{0\}^\ell)^2 : s_1 + \cdots + s_\ell + t_1 + \cdots + t_\ell = \ell\right\}\ .$$

For $(s,t) \in Q_\ell$, there is $j$ such that either $s_j = t_j = 1$ or $t_{j-1} = s_j = 1$. For example, assuming the former holds, we apply Lemma 2 to get

$$\begin{aligned}
\mathrm{Tr}\left[\widetilde{\mathbf{A}}_S^{s_1}\widetilde{\mathbf{A}}_T^{t_1}\cdots\widetilde{\mathbf{A}}_S^{s_\ell}\widetilde{\mathbf{A}}_T^{t_\ell}\right] &= \mathrm{Tr}\left[\big(\widetilde{\mathbf{A}}_S^{s_1}\widetilde{\mathbf{A}}_T^{t_1}\cdots\widetilde{\mathbf{A}}_S^{s_j}\big)\big(\widetilde{\mathbf{A}}_T^{t_j}\cdots\widetilde{\mathbf{A}}_S^{s_\ell}\widetilde{\mathbf{A}}_T^{t_\ell}\big)\right] \\
&\le |S\cap T| \|\widetilde{\mathbf{A}}_S^{s_1}\widetilde{\mathbf{A}}_T^{t_1}\cdots\widetilde{\mathbf{A}}_S^{s_j}\| \|\widetilde{\mathbf{A}}_T^{t_j}\cdots\widetilde{\mathbf{A}}_S^{s_\ell}\widetilde{\mathbf{A}}_T^{t_\ell}\| \\
&\le |S\cap T| \|\widetilde{\mathbf{A}}_S\|^{s_1}\|\widetilde{\mathbf{A}}_T\|^{t_1}\cdots\|\widetilde{\mathbf{A}}_S\|^{s_\ell}\|\widetilde{\mathbf{A}}_T\|^{t_\ell} \\
&\le |S\cap T|\zeta^\ell\ , \quad \text{where } \zeta := \frac{2|\psi|}{1-|\psi|}\ .
\end{aligned}$$

Note that the last line comes from (11).

With this, together with the fact that $|Q_\ell| \le \binom{2\ell}{\ell}$ and $(1-x)^{-1/2} = \sum_{n\ge 0}\binom{2n}{n}(x/4)^n$ for $x \in (0,1)$, we obtain

$$\Lambda \le \frac{1}{2}|S\cap T| \sum_{\ell\ge 2}\frac{1}{\ell}\binom{2\ell}{\ell}\zeta^\ell \le \frac{1}{4}|S\cap T|\left[(1-4\zeta)^{-1/2} - 1 - 2\zeta\right]\ ,$$

since $|\psi| < 1/9$. We then conclude with (10) and the expression of $\zeta$ in terms of $\psi$.

### 4.3 Proof of Proposition 1

*Under the null.* We first control $V_t^*$ under the null hypothesis. Decompose the statistics $V_{t,S}$ into

$$V_{t,S} = V_{1,t,S} + V_{2,t,S}\ , \text{ where } V_{1,t,S} := \sum_{j=1}^{\lfloor k/2\rfloor} V_{t,S}(2j) \text{ and } V_{2,t,S} := \sum_{j=2}^{\lceil k/2\rceil} V_{t,S}(2j-1)\ ,$$

so that all the terms in $V_{1,t,S}$ (resp. $V_{2,t,S}$) are independent. It suffices to prove that, with probability going to one, $V_{1,t}^* := \max_{S\in\mathcal{C}} V_{1,t,S}$ is smaller than $\lfloor k/2\rfloor/2$ and $V_{2,t}^* := \max_{S\in\mathcal{C}} V_{2,t,S}$ is smaller than $\lfloor k/2\rfloor/2$. By symmetry, we only consider $V_{1,t}^*$.

Recall that $n = |\mathcal{V}|$ and that $\tau$ denotes the maximum degree of $\mathcal{G}$. A simple counting argument gives $|\mathcal{C}| \le n\tau^{k-1}$. Also, under the null hypothesis, $V_{1,t,S} \sim \mathrm{Bin}(\lfloor k/2\rfloor, p_t)$ for any $S \in \mathcal{C}$. Thus, for any $S \in \mathcal{C}$, with the union bound, we have

$$\mathbb{P}_0(V_{1,t}^* \ge v) \le |\mathcal{C}|\,\mathbb{P}_0(V_{1,t,S} \ge v) \le n\tau^{k-1}\,\mathbb{P}\left(\mathrm{Bin}(\lfloor k/2\rfloor, p_t) \ge v\right)\ .$$

Define $b_t = 1/(2p_t)$ and $\varphi(b) = b(\log b - 1) + 1$. Choosing $v = v_t = \lceil k/2\rceil b_t p_t$, and using Bennett's inequality, the right-hand side is bounded by

$$\begin{aligned}
n\tau^{k-1}\exp\big(-\lfloor k/2\rfloor p_t\varphi(b_t)\big) &= \exp\big(\log n + (k-1)\log\tau - \lfloor k/2\rfloor p_t\varphi(b_t)\big) \\
&= \exp\big(\log n + (k-1)\log\tau - \lfloor k/2\rfloor \tfrac{1}{2}h(2p_t)\big) \\
&\le \exp\big(-kh(2p_t)/5\big) \to 0\ ,
\end{aligned}$$

where the inequality holds eventually as the sample size increases. Thus the test has a type I error tending to zero.

*Under the alternative.* We now consider the alternative hypothesis. Let $(S,\psi) \in \mathcal{C}\times(-1,1)$ denote the anomalous path and the autoregressive parameter. Denote $S = (s_1,\ldots,s_k)$. By definition, $V_t^* \ge V_{t,S}$. Define $Z_j := (X_{s_{j+1}} - X_{s_j})/\sqrt{2(1-\psi)}$ for any $j \in \{1,\ldots,k-1\}$. We have $Z_j \sim \mathcal{N}(0,1)$

and $\mathbb{E}[Z_j Z_{j'}] = \psi^{|j-j'|-1}(\psi-1)/2$ for $j \neq j'$. Define $q_t(\psi) := 2F(t/\sqrt{1-\psi}) - 1$. Computing the first moment of $V_{t,S}$, we obtain

$$\mathbb{E}_S[V_{t,S}] = (k-1)q_t(\psi) \geq \frac{3}{5}(k-1) \to \infty \ .$$

In order to conclude, it suffices to prove that $\mathrm{Var}_S(V_{t,S}) \ll \mathbb{E}_S^2[V_{t,S}]$. Fix any $j \neq j'$. Denoting $a = \mathbb{E}[Z_j Z_{j'}]$, we define $U = (Z_j - aZ_{j'})/\sqrt{1-a^2}$. The next argument shows that for any $x$ smaller than $t/\sqrt{1-\psi}$ in absolute value, the probability that $|Z_j| \leq t(1-\psi)^{-1/2}$ conditionally to $Z_i = x$ where $|x| \leq t(1-\psi)^{-1/2}$ is close to $q_t(\psi)$. Indeed,

$$\begin{aligned}
\mathbb{P}\left[|Z_j| \leq \frac{t}{\sqrt{1-\psi}}\Big| Z_i = x\right] &\leq \mathbb{P}\left[\frac{-t-ax}{\sqrt{1-\psi}\sqrt{1-a^2}} \leq U \leq \frac{t-ax}{\sqrt{1-\psi}\sqrt{1-a^2}}\right] \\
&\leq \mathbb{P}\left[|U| \leq \frac{t}{\sqrt{1-\psi}\sqrt{1-a^2}}\right] \\
&= 2\,\mathbb{P}\left[U \leq \frac{t}{\sqrt{1-\psi}}\right] + 2\,\mathbb{P}\left[\frac{t}{\sqrt{1-\psi}} \leq U \leq \frac{t}{\sqrt{1-\psi}\sqrt{1-a^2}}\right] \\
&= q_t(\psi) + 2\left[F\left(\frac{t}{\sqrt{1-\psi}\sqrt{1-a^2}}\right) - F\Big(\frac{t}{\sqrt{1-\psi}}\Big)\right] \\
&\leq q_t(\psi) + 2\frac{t}{\sqrt{1-\psi}}\left[\frac{1}{\sqrt{1-a^2}} - 1\right]\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{t^2}{2(1-\psi)}\right] \\
&\leq q_t(\psi) + a^2 \ ,
\end{aligned}$$

where we used the fact that $U$ is standard normal in the second and fourth line, a Taylor development (of order 1) in the fifth line, and the fact that $|a| \leq 1/2$ and $xe^{-x^2} \leq (2e)^{-1/2}$ in the last line. As a consequence,

$$\mathrm{Cov}\left(\mathbb{I}\{|Z_i| \leq \tfrac{t}{\sqrt{1-\psi}}\}, \mathbb{I}\{|Z_j| \leq \tfrac{t}{\sqrt{1-\psi}}\}\right) \leq q_t(\psi)\psi^{2|i-j|-2}(1-\psi)^2 \ ,$$

for all $i \neq j$. We conclude that

$$\mathrm{Var}_S(V_{t,S}) \leq (k-1)\left[q_t(\psi)(1-q_t(\psi)) + 2q_t(\psi)\frac{1-\psi}{1+\psi}\right] \leq 3\,\mathbb{E}_S[V_{t,S}] \ll \mathbb{E}_S^2[V_{t,S}] \ .$$

Thus the test also has a type II error tending to zero.

### Acknowledgements

This work was partially supported by the US National Science Foundation (DMS 1223137, DMS 1120888) and the French Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration). The third author was supported by the Spanish Ministry of Science and Technology grant MTM2012-37195.

## References

Addario-Berry, L., N. Broutin, L. Devroye, and G. Lugosi (2010). On combinatorial testing problems. *Ann. Statist. 38*(5), 3063–3092.

Arias-Castro, E., S. Bubeck, and G. Lugosi (2012). Detection of correlations. *Ann. Statist. 40*(1), 412–435.

Arias-Castro, E., S. Bubeck, and G. Lugosi (2015). Detecting positive correlations in a multivariate sample. *Bernoulli 21*, 209–241.

Arias-Castro, E., S. Bubeck, G. Lugosi, and N. Verzelen (2015). Detecting markov random fields hidden in white noise. *arXiv preprint arXiv:1504.06984*.

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.

Arias-Castro, E., D. L. Donoho, and X. Huo (2006). Adaptive multiscale detection of filamentary structures in a background of uniform random points. *The Annals of Statistics 34*(1), 326–349.

Arias-Castro, E. and G. R. Grimmett (2013). Cluster detection in networks using percolation. *Bernoulli 19*(2), 676–719.

Benjamini, I., R. Pemantle, and Y. Peres (1998). Unpredictable paths and percolation. *Ann. Probab. 26*(3), 1198–1211.

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.

Desolneux, A., L. Moisan, and J.-M. Morel (2003). Maximal meaningful events and applications to image analysis. *Ann. Statist. 31*(6), 1822–1851.

Guyon, X. (1995). *Random fields on a network*. Probability and its Applications (New York). New York: Springer-Verlag. Modeling, statistics, and applications, Translated from the 1992 French original by Carenne Ludeña.

Mukherjee, B., L. Heberlein, and K. Levitt (1994). Network intrusion detection. *Network, IEEE 8*(3), 26–41.

Sharpnack, J., A. Singh, and A. Krishnamurthy (2013). Detecting activations over graphs using spanning tree wavelet bases. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pp. 536–544.

Sharpnack, J. L., A. Krishnamurthy, and A. Singh (2013). Near-optimal anomaly detection in graphs using lovász extended scan statistic. In *Advances in Neural Information Processing Systems*, pp. 1959–1967.

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

Zhang, Y. and W. Lee (2000). Intrusion detection in wireless ad-hoc networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pp. 275–283. ACM.