# Near-optimal and computationally efficient detectors for weak and sparse graph-structured patterns

James Sharpnack[†] and Aarti Singh[*]
Machine Learning Department[†*], Statistics Department[†]
Carnegie Mellon University
Pittsburgh, PA, USA
{jsharpna,aarti}@cs.cmu.edu

***Abstract*—In this paper, we review our recent work on detecting weak patterns that are sparse and localized on a graph. This problem is relevant to many applications including detecting anomalies in sensor and computer networks, brain activity, co-expressions in gene networks, disease outbreaks etc. We characterize such a class of weak and sparse graph-structured patterns by small subsets of weakly activated nodes with a low cut in an underlying known graph. On one hand, the combinatorial nature of this class renders traditional detectors such as GLRT (aka scan statistic) computationally intractable for general graphs. On the other hand, attempts to develop feasible detectors such as fast subset scanning or averaging/thresholding sacrifice statistical efficiency. We describe and compare three detectors for weak graph-structured patterns that are developed using tools from graph theory, optimization and machine learning. These detectors are computationally efficient, applicable to graphs and patterns with general structures and come with precise theoretical guarantees, often achieving near-optimal statistical performance.**

***Index Terms*—graph patterns, structured sparsity, detection**

## I. Introduction

The problem of detecting anomalous patterns of activity in graphs is ubiquitous in modern applications ranging from detecting contamination or seismic activity by sensor networks, virus attacks in computer networks, stimulated activity in brain networks, gene expression in biological networks and disease outbreaks, to name a few. Moreover, the data avaialble to us today is extremely noisy and the scale of graphs of interest continues to explode. As a result, we are faced with the task of detecting weak and sparse graph-structured patterns of activity in a statistically and computationally efficient manner. In this paper, we review our recent work in this direction.

Formally, we consider the following hypothesis testing problem. Given a single noisy observation $y_i = x_i + \epsilon_i$ where $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ at each vertex $i \in \{1, \dots, V\}$ of a known graph $G = (V, E)$, distinguish between the two hypothesis:

$$H_0 : \mathbf{x} = \mathbf{0} \quad vs. \quad H_1 : \mathbf{x} = \mu \mathbf{1}_C$$

Here $\mathbf{x} = \{x_i\}_{i \in V}$ and $C \in \mathcal{C}_{c,\rho} := \{C \subseteq V : |C| = c, |\partial C| \leq \rho\}$ denotes the set of (possibly disconnected) activated vertices with size $|C| = c$ [1] and cut-size $|\partial C| := |(i,j) \in E : i \in C, j \notin C|$ less than or equal to a constant $\rho > 0$. Notice that for a given size of cluster (sparsity level) $c$, smaller values of $\rho$ imply that the set of activated nodes are localized on the graph. The goal is to define computationally efficient detectors that can distinguish between $H_0$ and $H_1$ at very low signal-to-noise ratios (SNRs) $\mu/\sigma$.

This formalization is a variant of the classical normal means model. In our setting, the means are restricted to conform to a graphs structure. While detection under normal means has been well-studied in literature (e.g. see [5]), there has been a renewed interest in the high-dimensional setup where a sparse set of coordinates of the mean are non-zero under the alternative. In the extreme case of a one sparse vector, it is easy to see that the signal-to-noise ratio needs to scale as $\sqrt{2\log(|V|)}$ for asymptotic distinguishability of the null and alternative hypotheses as $|V| \to \infty$. For the general sparse vector case when $c = |V|^\beta$, this problem was studied in [6] and a thresholding procedure known as Higher Criticism was shown to adaptively (without knowledge of the sparsity level $\beta$) achieve optimal SNR scaling of $\sqrt{2r(\beta)\log(|V|)}$, where $r(\beta)$ is a function that only depends on $\beta$. However, in the unstructured setting, the weakest SNR that can be detected needs to grow with the number of vertices (size of the high-dimensional vector) as long as $0 \leq \beta < 1$.

As mentioned earlier, in many applications, the observation vector is structured to conform to an underlying graph. If the cluster of activated vertices under the alternate hypothesis is known a priori, we have a simple alternate hypothesis and Neyman-Pearson lemma states that the Likelihood Ratio Test (LRT) or matched filtering (comparing $\mathbf{1}^T\mathbf{y}$ to an appropriate threshold) is optimal. This test is equivalent to averaging the observations of the activated vertices which reduces the noise variance by a factor of $c$, yielding an optimal SNR of $\sqrt{2\log(|V|)/c} = \sqrt{2\log(|V|)/|V|^\beta}$ for asymptotic distinguishability of the hypotheses. Thus, given oracle knowledge of the activated cluster, it is possible to detect sparse graph-structured patterns at much weaker SNR. In the absence of this knowledge, a popular approach to handle composite hypothesis tests is to scan over all permissible clusters under the alternative hypothesis. This approach is known as the scan statistic or Generalized Likelihood Ratio Test (GLRT) and involves comparing the statistic $\max_{C \in \mathcal{C}_{c,\rho}} \mathbf{1}_C^T\mathbf{y}/\sqrt{c}$ to an appropriate threshold. However, the graph scan statistic [7] or GLRT are computationally tractable only in specific graph

[1]For ease of presentation, in this paper we assume $c \leq |V|/2$. For modifications without this restriction, see [1], [2], [3], [4].

topologies and specific pattern classes. Efforts in this direction include detecting an interval in a line graph or geometric shapes such as rectangles, disks or ellipses in a lattice graph [8], path of activation in a tree or lattice [9], or nonparametric shapes in a lattice graph [10]. In these settings, scanning over the entire pattern class or over an epsilon-net for the pattern class is often feasible and has been shown to have near-optimal statistical performance. However, for the general setting of arbitrary graphs, the combinatorial nature of the pattern class $\mathcal{C}_{c,\rho}$ renders these detectors infeasible, either because the scan involves too many patterns or due to lack of constructive ways to obtain an epsilon-net. While there has been some work on developing fast graph subset scanning methods [11], these greedy methods sacrifice statistical power. The detectors we will describe are computationally efficient and statistically optimal for general graphs. In [12], the authors consider the complete graph and study detection under some combinatorial classes of patterns such as cliques, bi-cliques, spanning trees and perfect matchings. They establish lower bounds on the performance of any detector, and provide upper bounds for some simple but sub-optimal detectors such as averaging all node observations and thresholding. We demonstrate that our methods outperform global averaging and thresholding that do not take advantage of the graph structure.

## II. Computationally efficient detectors for graph-structured patterns

The combinatorial scan statistic or GLR for the hypothesis testing problem introduced in previous section is given as $\max_{C\in\mathcal{C}_{c,\rho}} \frac{\mathbf{1}_C^T\mathbf{y}}{\sqrt{c}}$. An epsilon-scan statistic is defined similarly where the max is taken over an epsilon net $\mathcal{N}_\epsilon$ of the class $\mathcal{C}_{c,\rho}$, where $\mathcal{N}_\epsilon$ satisfies the property that for every $C \in \mathcal{C}_{c,\rho}$ there exists a $C' \in \mathcal{N}_\epsilon$ such that $C'$ provides an $\epsilon$-approximation to $C$ under an appropriate notion of distance between them [10]. As mentioned earlier, both these statistics while optimal are computationally intractable.

In this section, we describe three computationally efficient detectors for the graph-structured normal means test. The first two detectors are obtained by considering convex relaxations of the combinatorial scan statistic. The third detector can be thought of as a feasible construction of an appropriate epsilon-scan statistic via the notion of graph wavelets.

### A. Spectral scan statistic

The spectral scan statistic was introduced in [1] with a slight modification proposed in [2]. The spectral scan statistic is a relaxation of the generalized likelihood ratio (GLR). This serves as the simplest relaxation of the GLR as it relaxes the domain $\mathbf{1}_C \in \{0,1\}^{|V|}$ such that $|C| = c$ to $\mathbf{z} \in \mathbb{R}^{|V|}$. In order to relax the cut size constraint we will introduce the graph Laplacian. The graph Laplacian matrix $\Delta = D - A$ where $A$ denotes the adjacency matrix of the graph and $D$ is a diagonal matrix with vertex degrees on the diagonal i.e. $D_{ii} = \sum_j A_{ij}$. Furthermore, we can rewrite the graph cut constraint $|\partial C| = \mathbf{1}_C^\top \Delta \mathbf{1}_C$, indicating that we can relax the domain of the GLR to $\mathbf{z}^\top\Delta\mathbf{z}$ where $\mathbf{z} \in \mathbb{R}^{|V|}$ relaxes the vector $\mathbf{1}_C$. For details, refer to [1]. The resulting spectral scan statistic is defined as follows where $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{1}^T\mathbf{y}/|V|$

$$\hat{s} = \sup_{\mathbf{z}\in\mathbb{R}^{|V|}} (\mathbf{z}^\top\tilde{\mathbf{y}})^2 \text{ s.t. } \mathbf{z}^\top\mathbf{\Delta}\mathbf{z} \le \rho, \|\mathbf{z}\| \le 1, \mathbf{z}^\top\mathbf{1} = 0.$$

As shown in [1], the convex spectral scan statistic can be solved efficiently in the dual domain by first-order interior point methods.

### B. Lovász extended scan statistic

The Lovász extended scan statistic (LESS) is another relaxation of the computationally infeasible generalized likelihood ratio. Notice that the GLR can be written in terms of the binary vector $\mathbf{z} = \mathbf{1}_C \in \{0,1\}^{|V|}$,

$$\max_{\mathbf{z}\in\{0,1\}^{|V|}} \frac{\mathbf{z}^\top\mathbf{y}}{\sqrt{c}} \text{ s.t. } \sum_{(i,j)\in E} I\{z_i \neq z_j\} \le \rho, \mathbf{1}^\top\mathbf{z} = c$$

Submodularity is the combinatorial analogue of convexity, and it turns out that the cut size ($|\partial C|$) is submodular. For every submodular function there exists a convex relaxation, called the Lovász extension. The Lovász extension of $|\partial C| = \sum_{(i,j)\in E} I\{z_i \neq z_j\}$ is the total variation $\sum_{(i,j)\in E} |z_i - z_j|$. Thus, it is natural to relax the GLR program as follows

$$\hat{l} = \max_{\mathbf{z}\in[0,1]^{|V|}} \frac{\mathbf{z}^\top\mathbf{y}}{\sqrt{c}} \text{ s.t. } \sum_{(i,j)\in E} |z_i - z_j| \le \rho, \mathbf{1}^\top\mathbf{z} = c \quad (1)$$

which we call LESS. In [3], convex analysis was used to derive the dual program to the LESS, and it was shown that LESS can be solved efficiently using graph cuts.

### C. Uniform spanning tree wavelet statistic

Yet another computationally efficient detector can be obtained by constructing a wavelet basis $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_{|V|}]$ for the graph. The basis elements we construct are orthonormal and have the property that every pattern in $\mathcal{C}_{c,\rho}$ has a sparse representation in terms of the basis coefficients. The graph wavelet construction we propose relies on a uniform spanning tree of the graph. A uniform spanning tree is a spanning tree drawn at random from the set of all distinct spanning trees of the graph, and can be constructed in time nearly linear in the number of vertices for most graphs using the Aldous-Broder algorithm. A detailed explanation of the wavelet construction can be found in [4].

Given a uniform spanning tree, the wavelet construction iterates the following steps: finding a balancing vertex, removing it from the uniform spanning tree and forming a basis that spans the resulting connected components, and recursing on the remaining subtrees. A balancing vertex is one such that the remaining connected components, after its removal from the tree, are at most half the size of the graph. A simple algorithm that travels in the direction of the largest subtree at a vertex can be used to find this in nearly $O(|V|)$ time. We summarize the wavelet construction algorithm in Algorithm 1, which takes as input the connected subtrees $\mathcal{S} = \{\mathcal{T}_i\}_{i=1}^{d_v}$ after the removal of the balancing vertex $v$, where $d_v$ is the degree of vertex $v$.

**Algorithm 1** FormWavelets

**Require:** $\mathcal{S} = \{\mathcal{T}_i\}_{i=1}^{d_v}$

(1) Let $\mathcal{T}_1 = \cup_{i \leq |\mathcal{S}|/2} \mathcal{T}_i$ and $\mathcal{T}_2 = \cup_{i > |\mathcal{S}|/2} \mathcal{T}_i$.

(2) Form the following basis element and add it to $\mathbf{B}$:

$$\mathbf{b} = \frac{\sqrt{|\mathcal{T}_1||\mathcal{T}_2|}}{\sqrt{|\mathcal{T}_1| + |\mathcal{T}_2|}} \left[ \frac{1}{|\mathcal{T}_1|} \mathbf{1}_{\mathcal{T}_1} - \frac{1}{|\mathcal{T}_2|} \mathbf{1}_{\mathcal{T}_2} \right]$$

(3) Recurse at (1) with $\mathcal{S} \leftarrow \{\mathcal{T}_i\}_{i \leq |\mathcal{S}|/2}$ and $\mathcal{S} \leftarrow \{\mathcal{T}_i\}_{i > |\mathcal{S}|/2}$ separately.

This algorithm returns an unbalanced Haar wavelet basis for a general graph, and leads to the natural detector based on thresholding the maximum wavelet coefficient. Thus the uniform spanning tree wavelet statistic is given as $\max_{\mathbf{b} \in \mathbf{B}} \mathbf{b}^T \mathbf{y}$ and is equivalent to scanning over an epsilon-net of $\mathcal{C}_{c,\rho}$.

## III. Statistical efficiency

In this section, we summarize the statistical performance of the three computationally efficient detectors for general graph structures. We will show that our detectors have near-optimal statistical performance in many settings.

In [1], [2], [3], [4], we provide precise bounds on the power and size of these tests for finite graph size. For brevity, in this paper we will summarize statistical performance by the critical SNR $\mu/\sigma$ scaling needed to ensure asymptotic distinguishability, i.e. to ensure

$$\lim_{|V| \to \infty} P_{\mathbf{0}}(T > th) + \sup_{C \in \mathcal{C}_{c,\rho}} P_C(T < th) = 0$$

where $P_{\mathbf{0}}$ denotes probability with respect to $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$ and $P_C$ is with respect to $\mathbf{y} \sim \mathcal{N}(\mu \mathbf{1}_C, \sigma^2 I)$.

The first result characterizes the critical SNR required by the spectral scan statistic in terms of the graph spectrum, as characterized by eigenvalues of the graph Laplacian matrix $\Delta$.

**Theorem 1.** *[1] The spectral scan statistic asymptotically distinguishes $H_0$ from $H_1$ if*

$$\frac{\mu}{\sigma} = \omega \left( \sqrt{\frac{1}{c} \sum_{i=2}^{|V|} \min \left( 1, \frac{\rho}{c \lambda_i} \right)} \right)$$

*where $\lambda_i$ denote the eigenvalues of the graph Laplacian matrix $\Delta$ sorted in ascending order.*

Here $a = \omega(b)$ denotes that $a/b \to \infty$. This result suggests that the critical SNR required by the spectral scan statistic scales with the complexity of the pattern class (cut-size to size ratio $\rho/c$ of patterns, or equivalently the surface to volume ratio of the activated cluster), as well as the complexity of the graph (decay of Laplacian eigenvalues). We evaluate the graph spectrum and this bound for specific low-cut and sparse patterns on specific graphs (e.g. subtrees of activation in a tree graph or squares of activation in a 2-dimensional torus) in [1]. Furthermore, the choice of threshold is given that achieves this rate. In [1], we also show that while the bound captures the dependence on the cut-size to size ratio and is near-optimal for non-sparse activations, the performance of the spectral scan statistic is suboptimal for sparse patterns. The remaining two detectors we discuss overcome this drawback and perform better with a small cluster size, $c$.

The next two results characterize the performance of the Lovász extended scan statistic and uniform spanning tree wavelet statistic. In both cases, the critical SNR depends on $r_{\max}$ the maximum effective resistance of the graph cut induced by a pattern in $\mathcal{C}_{c,\rho}$. Formally, $r_{\max} = \max_{C \in \mathcal{C}_{c,\rho}} \sum_{e \in \partial C} r_e$ where $r_e$ is the effective resistance of the edge $e$. The effective resistance of an edge $e = (i, j)$ is the potential difference required to create a unit flow between vertices $i$ and $j$. Formally, $r_e = (\delta_i - \delta_j)^T \Delta^\dagger (\delta_i - \delta_j)$, where $\Delta^\dagger$ is pseudo-inverse of the graph Laplacian and $\delta_i$ the Dirac delta function.

**Theorem 2.** *[3] The Lovász extended scan statistic asymptotically distinguishes $H_0$ from $H_1$ if*

$$\frac{\mu}{\sigma} = \omega \left( \sqrt{\frac{\max(r_{\max}, \log(|V|)) \log(|V|)}{c}} \right)$$

**Theorem 3.** *[4] The uniform spanning tree wavelet statistic asymptotically distinguishes $H_0$ from $H_1$ if*

$$\frac{\mu}{\sigma} = \omega \left( \sqrt{\frac{r_{\max} \log(d_{\max}) \log^2(|V|)}{c}} \right)$$

*where $d_{\max}$ is the maximum degree of the graph $G$.*

By Foster's theorem [13], [14], the effective resistance of a cut is $\approx \rho/d$ where $d$ is the average degree of a vertex. This intuition can be formalized for specific graphs such as edge transitive graphs (including the lattice and complete graphs) and random geometric graphs (such as $k$-nearest neighbor and $\epsilon$-nearest neighbor graphs). For details see [3], [4]. Hence, if the maximum effective resistance $r_{\max} \approx \rho/d \ll c$, the active nodes are localized on the graph leading to structured sparsity. In this case, akin to the performance of the oracle estimator, we see that these detectors take advantage of structured sparsity, and allow critical SNR to decrease with increasing graph size $|V|$ if the cluster size $c = |V|^\beta$. On the other hand, if $r_{\max} \approx \rho/d \approx c$ the pattern is not localized and performance of these methods degrades gracefully to that of unstructured tests (up to log factors), requiring critical SNR to grow with graph size.

While our detectors outperform the unstructured tests in structured sparse settings, comparing our bounds for the Lovász extended scan statistic and uniform spanning tree wavelet statistic with that of the oracle, we see that (modulo log factors) the critical SNR for the oracle is better by a factor of $r_{\max}$. Prior work e.g. [10] demonstrates that for some simple graphs such as the lattice, it is possible to achieve oracle performance (modulo log factors). This raises the question whether our bounds can be further improved. We answer this in the negative by demonstrating in [3], [4] a matching lower bound on the performance of any detector for general graphs which is indeed worse by a factor of $r_{\max}$ than the oracle

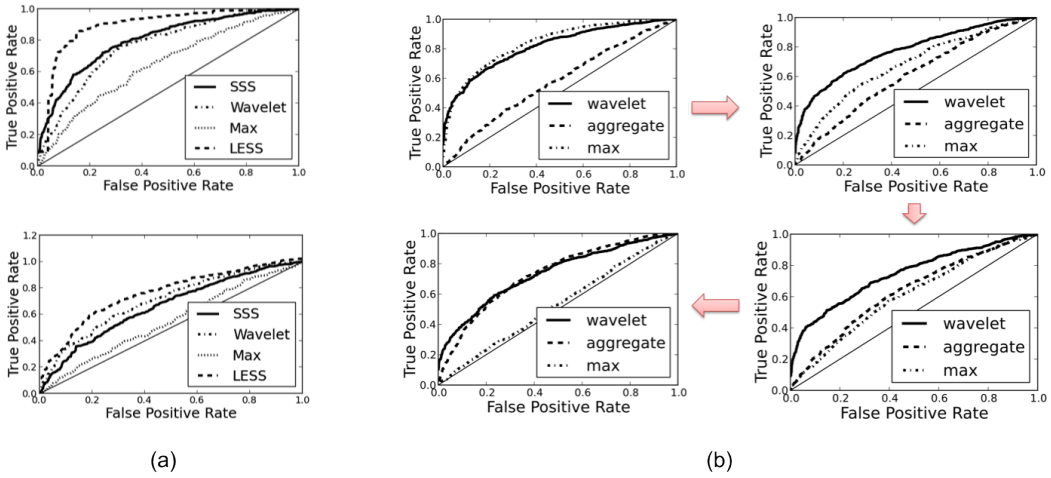


Fig. 1. (a) ROC curves for spectral scan statistic (SSS), uniform spanning tree wavelet statistic (Wavelet), the maximum statistic, $\max_i |y_i|$, (Max), and Lovász extended scan statistic (LESS). The graphs used are square 2D Torus (top), and $\epsilon$-NN graph (bottom) with $\epsilon \approx |V|^{-1/3}$; with $\mu = 4, 3$ respectively, $|V| = 225$, and $c \approx |V|^{1/2}$. (b) Comparison of wavelet detector with maximum and aggregate statistic on a torus with increasing size of activated cluster, for a fixed cut size.

performance. This suggests that for general graph structures and complex pattern classes such as $\mathcal{C}_{c,\rho}$, it is not possible to exactly achieve oracle performance and *there is a price for not knowing the location of activated vertices.*

## IV. Comparison and simulations

We compare the performance of our methods by simulating graph structured patterns over a 2-dimensional torus and $\epsilon$-NN random graph in Fig. 1(a). We report the true positive rate versus the false positive rate as the threshold varies (also known as the receiver operating curve or ROC.) For each graph, our methods dominate the max statistic, indicating that one cannot ignore graph structure and hope to detect optimally.

To demonstrate that our methods degrade gracefully when the cut size to cluster size becomes large, we compare the wavelet detector to two unstructured detectors based on the maximum and average of all observations. The aggregate statistic is expected to work well when the cluster size is very large. Fig. 1(b) shows that, for a fixed cut size, the wavelet detector degrades to the aggregate and maximum tests for very large and very small cluster sizes respectively, but outperforms them when the pattern is localized on the graph (not globally spread or too sparse such that structure cannot be leveraged).

## V. Conclusion

In this paper, we present three computationally efficient detectors for weak graph-structured patterns. Two of the detectors (spectral scan statistic and Lovász extended scan statistic) are obtained by convex relaxations of the combinatorial scan statistic or GLRT. The third detector uses a graph wavelet construction and is akin to a constructive epsilon-net scan statistic. All detectors leverage the graph-structured sparsity to operate at very weak signal-to-noise ratios. Our experiments support these results. Furthermore, our results indicate that the spectral scan statistic is nearly optimal for non-sparse patterns under certain graph models, while the Lovász extended scan statistic and uniform spanning tree wavelet statistic are nearly optimal for any pattern sparsity given only a cut-size constraint.

## Acknowledgement

The authors thank Alessandro Rinaldo and Akshay Krishnamurthy, our co-authors on some of the methods. This work is supported by AFOSR grant FA9550-10-1-0382.

## References

[1] J. Sharpnack, A. Rinaldo, and A. Singh, "Changepoint detection over graphs with the spectral scan statistic," in *Artifical Intelligence and Statistics (AISTATS)*, 2013.

[2] ——, "Detecting activity in graphs via the graph ellipsoid scan statistic," 2013, submitted.

[3] J. Sharpnack, A. Krishnamurthy, and A. Singh, "Near-optimal anomaly detection in graphs using lovász extended scan statistic," 2013, submitted.

[4] ——, "Detecting activations over graphs using spanning tree wavelet bases," in *Artifical Intelligence and Statistics (AISTATS)*, 2013.

[5] Y. Ingster and I. Suslina, *Nonparametric goodness-of-fit testing under Gaussian models*. Springer Verlag, 2003, vol. 169.

[6] J. Jin and D. L. Donoho, "Higher criticism for detecting sparse heterogeneous mixtures," *Annals of Statistics*, vol. 32, no. 3, pp. 962–994, 2004.

[7] C. E. Priebe, "Scan statistics on graphs," Johns Hopkins University, Baltimore, MD 21218-2682, Tech. Rep. 650, 2004.

[8] E. Arias-Castro, D. Donoho, and X. Huo, "Near-optimal detection of geometric objects by fast multiscale methods," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2402–2425, 2005.

[9] E. Arias-Castro, E. Candes, H. Helgason, and O. Zeitouni, "Searching for a trail of evidence in a maze," *The Annals of Statistics*, vol. 36, no. 4, pp. 1726–1757, 2008.

[10] E. Arias-Castro, E. Candes, and A. Durand, "Detection of an anomalous cluster in a network," *The Annals of Statistics*, vol. 39, no. 1, pp. 278–304, 2011.

[11] S. Speakman and D. B. Neill, "Fast graph scan for scalable detection of arbitrary connected clusters," in *Proceedings of the 2009 International Society for Disease Surveillance Annual Conference*, 2010.

[12] L. Addario-Berry, N. Broutin, L. Devroye, and G. Lugosi, "On combinatorial testing problems," *The Annals of Statistics*, vol. 38, no. 5, pp. 3063–3092, 2010.

[13] R. Foster, "The average impedance of an electrical network," *Contributions to Applied Mechanics (Reissner Anniversary Volume)*, pp. 333–340, 1949.

[14] P. Tetali, "Random walks and the effective resistance of networks," *Journal of Theoretical Probability*, vol. 4, no. 1, pp. 101–109, 1991.