AFRL-OSR-VA-TR-2013-0552

RESOURCE-CONSTRAINED DATA COLLECTION AND FUSION FOR IDENTIFYING WEAK DISTRIBUTED PATTERNS IN NETWORKS

AARTI SINGH

CARNEGIE-MELLON UNIVERSITY

10/15/2013
Final Report

DISTRIBUTION A: Distribution approved for public release.

AIR FORCE RESEARCH LABORATORY
AF OFFICE OF SCIENTIFIC RESEARCH (AFOSR)/RSL
ARLINGTON, VIRGINIA 22203
AIR FORCE MATERIEL COMMAND

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
14-10-2013

2. REPORT TYPE
FINAL

3. DATES COVERED *(From - To)*
15 Jul 2010 - 14 Jul 2013

4. TITLE AND SUBTITLE
Resource-constrained Data Collection and Fusion for Identifying Weak Distributed Patterns in Networks

5a. CONTRACT NUMBER

5b. GRANT NUMBER
AFOSR FA9550-10-1-0382

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
SINGH, AARTI

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CARNEGIE MELLON UNIVERSITY
5000 FORBES AVENUE
PITTSBURGH, PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
801 North Randolph Street
Room 732
Arlington, VA 22203-1977

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution A

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project addressed the problems of detection, localization and estimation of weak and distributed patterns of activation in a large-scale network given access to direct, compressive and adaptive noisy node measurements. Precise information-theoretic limits were identified for these problems that provide necessary conditions on how the signal-to-noise ratio required scales as a function of the number of measurements, the graph size, connectivity and properties such as cut-size of the activated vertices, under a graph-structured normal means model. By leveraging highly inter-disciplinary tools from machine learning, statistics, signal processing and optimization, fast methods were developed that nearly achieve the information-theoretic limits, for general graph structures and classes of activation patterns. Development of such state-of-the-art methods that are both computationally and statistically efficient is crucial to advance AFOSR's ability to monitor, understand and secure modern large-scale networks that are vulnerable to covert attacks.

15. SUBJECT TERMS
graph patterns, structured sparsity, detection, localization, estimation, compressive measurements, adaptive measurements

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		15	AARTI SINGH
					19b. TELEPHONE NUMBER *(include area code)* 412-268-4266

FINAL REPORT: AFOSR GRANT FA9550-10-1-0382

Resource-constrained Data Collection and Fusion for Identifying Weak Distributed Patterns in Networks

Duration: July 15, 2010 to July 14, 2013

Principal Investigator:

Dr. Aarti Singh
8207 Gates Hillman Center, Machine Learning Department,
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
Ph: (412) 268-4266, FAX: (412) 268-3431
Email: aarti@andrew.cmu.edu

Program Manager:

Dr. Tristan Nguyen, AFOSR/RTC (703) 696-7796
DSN 426-7796; FAX (703) 696-7360
E-mail: tristan.nguyen@afosr.af.mil

1 Summary of project goals

The specific challenge tackled in this project is detection, localization and recovery of *weak* and *distributed* patterns of activation in a network. Weak patterns of activation in a network arise in myriad problems including identification of incipient contamination or seismic activity monitored by a sensor network, onset of a virus in the Internet, covert signals in communication networks, or anomalous social activity. Moreover, the distributed nature of these patterns implies that they are undetectable in local signatures of individual nodes, as well as in network-wide aggregates. As a result, the solution to this problem hinges on the development of novel data fusion methods that leverage the structure of the underlying network. Since the number of possible activation patterns can grow exponentially with network size, conventional estimators and detectors such as scan statistic or generalized likelihood ratio that scan over all patterns are computationally intractable. On the other hand, attempts to develop feasible detectors such as fast subset scanning or averaging/thresholding require high Signal-to-Noise Ratios (SNRs). Furthermore, there are constraints on resources such as limits on storage, sensing, communication energy or bandwidth.

The goals of this project were to address the following problems:

1. Determine theoretical limits of detection, localization and recovery of weak distributed activations in large-scale networked systems.

2. Develop practical computationally efficient algorithms that require minimal SNR and measurement resources to identify weak and distributed patterns of network activity.

2 Significant work accomplished

This section summarizes the theory and methods developed in this project for the problems of detecting, localizing and estimating weak and distributed graph-structured patterns under 1) a **direct measurement model** and 2) a **compressive and adaptive measurement model**.

2.1 Direct measurement model

Under this model, the observations correspond to a single measurement at each node of a *known* network graph $G = (V, E)$, i.e.,

$$y_i = x_i + \epsilon_i \qquad i = 1, \dots, |V|$$

where x_i is the true underlying activation at node i that is corrupted by additive white Gaussian noise $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

I. **Detection:** The goal of detection is to distinguish between the two hypothesis:

$$\begin{aligned} H_0 : \mathbf{x} &= \mathbf{0} \\ H_1 : \mathbf{x} &= \mu \mathbf{1}_C \end{aligned}$$

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 $|C| = c$ and cut-size $|\partial C| := |(i,j) \in E : i \in$

$C, j \notin C|$ less than or equal to a constant $\rho > 0$. [1] For a given sparsity level c, smaller values of ρ imply that the set of activated nodes are localized on the graph. The goal is to develop computationally efficient detectors that can distinguish between H_0 and H_1 at very low signal-to-noise ratios (SNRs) μ/σ.

The Generalized Likelihood Ratio Test (GLRT) statistic, also known as combinatorial scan statistic, for this hypothesis testing problem is given as:

$$\max_{C \in \mathcal{C}_{c,\rho}} \mathbf{1}_C^\top \mathbf{y}$$

While the GLRT or a scan over an ϵ-net of the class $\mathcal{C}_{c,\rho}$ is optimal in many cases [1, 2, 3], it is computationally intractable. While there has been some work on developing fast graph subset scanning methods [4], these greedy methods sacrifice statistical power. This project developed detectors for weak graph-structured patterns by borrowing 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.

- The **spectral scan statistic** (SSS) developed in [5] is obtained by a convex spectral relaxation of the combinatorial scan statistic, inspired by the relaxation used in spectral clustering algorithm in machine learning. This involves relaxing the cut size constraint using 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. $D_{ii} = \sum_j A_{ij}$. The cut size can be written as $|\partial C| = \mathbf{1}_C^\top \Delta \mathbf{1}_C$, suggesting that the domain of the GLRT can be relaxed to $\mathbf{z}^\top \Delta \mathbf{z}$ where $\mathbf{z} \in \mathbb{R}^{|V|}$ relaxes the vector $\mathbf{1}_C$. The resulting spectral scan statistic is defined as follows where $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{1}^T \mathbf{y}/|V|$

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

  As shown in [5], the convex spectral scan statistic can be solved efficiently in the dual domain by first-order interior point methods. The SNR required by the SSS is characterized as follows. Here $a = \omega(b)$ denotes that $a/b \to \infty$.

  Theorem 1. *[5] 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 λ_i are the eigenvalues of the graph Laplacian matrix Δ sorted in ascending order.

  This result suggests that the SNR required by SSS scales with the complexity of the pattern class (cut-size to size ratio ρ/c, or equivalently the surface to volume ratio, of the activated vertices), as well as the complexity of the graph (decay of Laplacian eigenvalues). The graph spectrum and this bound is evaluated for specific low-cut and

[1] Some of the methods developed apply to more general composite null hypotheses that allow for piece-wise constant activations, but for simplicity we focus on this setup in the report.

sparse patterns on specific graphs (e.g. subtrees of activation in a tree graph, squares of activation in a 2-dimensional torus or multi-resolution groups in Kronecker graphs) in [5]. An extension of this work, the Graph Ellipsoid Scan Statistic (GESS), was recently developed which upper and lower bounds the SSS, and enables tighter performance bounds. Work on GESS is currently in preparation for submission [6].

Some information-theoretic lower bounds for this problem are also derived in [5, 7] which reveal that while the SSS and GESS are nearly-optimal for non-sparse activations (large c), their performance is suboptimal for sparse patterns, except for very specific graphs. The remaining two detectors described below overcome this drawback and perform better with a small set of activated vertices.

- The **Lovász extended scan statistic** (LESS) [8] is another relaxation of the GLRT obtained as follows. The GLRT can be written in terms of the binary vector $\mathbf{z} = \mathbf{1}_C \in \{0,1\}^{|V|}$ as

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

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 GLRT as follows

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

which is called the LESS. In [8], convex analysis has been used to derive the dual program to the LESS, and it is shown that LESS can be solved efficiently using methods for finding graph cuts. The SNR required by LESS 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.

Theorem 2. *[8] The Lovász extended scan statistic asymptotically distinguishes H_0 from H_1 if*

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

By Foster's theorem, 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 ϵ-nearest neighbor graphs). For these cases, a comparison with information-theoretic lower bounds suggests that LESS is nearly optimal [2]. If $r_{\max} \approx \rho/d \ll c$, the active nodes are localized on the graph and the detector takes advantage of structured sparsity. On the other hand, if $r_{\max} \approx \rho/d \approx c$ the pattern is not localized and the SNR requirement degrades gracefully to $\sqrt{\log|V|}$ (up to log factors), which is characteristic of unstructured tests (that do not leverage knowledge of the graph) such as the max statistic or Higher Criticism [9].

[2] The necessary SNR for sparse patterns essentially scales like $\sqrt{(\rho/d_{\max}c)\log|V|}$ [7]

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.

- The **graph wavelet statistic** [10, 7] can be obtained by constructing an orthonormal wavelet basis $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_{|V|}]$ for the graph with the property that every pattern in $\mathcal{C}_{c,\rho}$ has a sparse representation in terms of the basis coefficients. Projecting the node observations onto such a basis would concentrate the signal energy in a few coefficients while the noise distribution remains the same, thus boosting the SNR. This leads to natural detectors based on thresholding the maximum wavelet coefficient

$$\max_{\mathbf{b} \in \mathbf{B}} \mathbf{b}^T \mathbf{y}$$

which is equivalent to scanning over an epsilon-net of $\mathcal{C}_{c,\rho}$.

For hierarchically-structured network patterns characterized by a latent tree graph, such an orthonormal unbalanced Haar wavelet basis was developed [10]. This construction was then extended to low-cut activation patterns on general graph structures by leveraging the spanning tree of a graph to correspond to the latent tree [7]. Specifically, for general graphs, the graph wavelet construction relies on the uniform spanning tree (UST) which can be constructed in time nearly linear in the number of vertices for most graphs using the Aldous-Broder algorithm [11]. Given a UST, the wavelet construction iterates the following steps: finding a balancing vertex, removing it from the uniform spanning tree, 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. The wavelet construction is summarized 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. The SNR required by the UST wavelet detector is given as follows.

Theorem 3. *[7] 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.

This performance bound is similar to that of LESS, and similar to LESS the UST wavelet

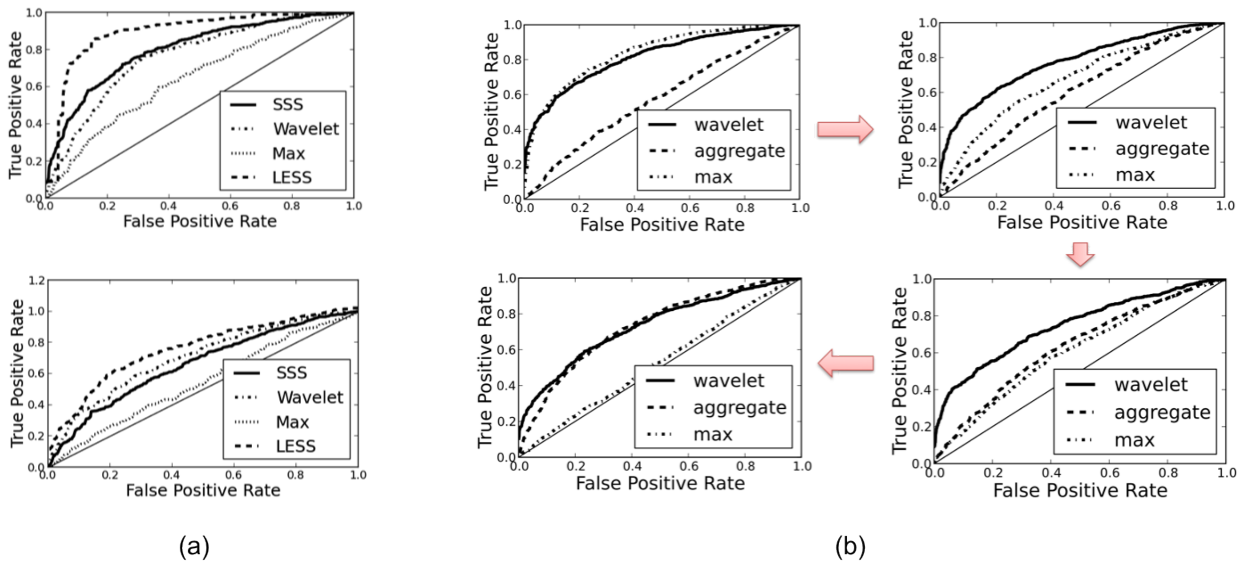

Figure 1: (a) ROC curves for spectral scan statistic (SSS), uniform spanning tree wavelet statistic (Wavelet), the maximum statistic, $\max_i |y_i|$, (Max), and Lovász extended scan statistic (LESS). The graphs used are square 2D Torus (top), and ϵ-NN graph (bottom) with $\epsilon \approx |V|^{-1/3}$; 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.

detector is nearly optimal for many graphs and pattern classes. It also takes advantage of structured sparsity and degrades gracefully for unstructured settings.

A comparison of the three detectors will appear in [12] for graph-structured patterns simulated over a 2-dimensional torus and ϵ-NN random graph. Fig. 1(a) reports the true positive rate versus the false positive rate as the threshold varies (also known as the receiver operating curve or ROC.) The LESS provides a tight relaxation and hence performs better than SSS. The wavelet detector, though theoretically optimal, suffers from additional log factors which make its performance slightly inferior to LESS. For each graph, all of the developed detectors dominate the max statistic, indicating that one cannot ignore graph structure and hope to detect at optimal SNRs.

To demonstrate that the proposed detectors degrade gracefully when the cut size to cluster size becomes large, the wavelet detector is compared to two unstructured detectors based on the maximum and global average of all observations. The global 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 graph structure cannot be leveraged).

II. **Estimation:** The goal of estimation is to de-noise the node observations and recover the underlying activation pattern $\mathbf{x}$ accurately in mean-square-error (MSE). In this problem, $\mathbf{x}$ does not necessarily correspond to a binary activation. Instead, we focus on the class of activations that are smooth with respect to the graph G, i.e. if two nodes are connected by an edge, their activations are similar. This can be formalized by considering the class of patterns

$$\mathcal{X}_\rho = \{\mathbf{x} : \mathbf{x}^\top \Delta \mathbf{x} = \sum_{(i,j)\in E} |x_i - x_j| \leq \rho\}$$

where Δ denotes the graph Laplacian, as before. Such activation patterns (with a specific ρ) also arise with high probability when sampled from a Gaussian Graphical model or an Ising model [13].

Patterns that are smooth over a known graph can be denoised by projection onto the Graph Laplacian Eigenbasis. Consider the spectral decomposition of the Graph Laplacian $\Delta = \mathbf{U}\Lambda\mathbf{U}^\top$, and denote the first k eigenvectors (corresponding to the smallest eigenvalues) of Δ by $\mathbf{U}_{[k]}$. Define the estimator

$$\hat{\mathbf{x}}_k = \mathbf{U}_{[k]}\mathbf{U}_{[k]}^\top \mathbf{y}$$

which is a hard thresholding of the projection of node measurements onto the graph Laplacian eigenbasis. This estimator reduces to some well-known estimators for specific graphs, e.g. for regular grids aka lattice graphs, the Laplacian eigenbasis correspond to Fourier basis and for hierarchical graphs, the Laplacian eigenbasis correspond to the Wavelet basis. The following theorem bounds the MSE of this estimator.

Theorem 4. *[13] The maximum MSE of the Projected Graph Laplacian estimator can be bounded as*

$$\sup_{\mathbf{x}\in\mathcal{X}_\rho} \mathbb{E}[\|\hat{\mathbf{x}} - \mathbf{x}\|^2] \leq \min(|V|, \rho/\lambda_{k+1}) + k\sigma^2$$

where $\lambda_1 \leq \lambda_2 \leq \ldots$ are the ordered eigenvalues of Δ.

The two terms in the bound indicate a tradeoff between the amount of signal discarded (first term) and the amount of noise retained (second term) by projecting onto the first k Laplacian eigenbasis. By evaluating the eigenspectrum of various graphs, it is possible to establish an appropriate scaling of k with graph size $|V|$ and the amount of noise that can be tolerated while ensuring MSE consistent recovery i.e. $MSE \to 0$ as the graph size $|V| \to \infty$. For many example graphs, it is observed that the tolerable noise level scales as $\sigma^2 = o(p^\gamma)$, where $\gamma \in (0,1)$ characterizes the strength of network interactions [13]. For example, for lattice graphs the noise tolerance results if the node degrees scale as $\gamma \log |V|$ (higher γ implying more neighbors per node), for hierarchical graphs this requires that the non-zero interactions exist until level $\gamma \log |V|$ going bottom-up (higher γ implying interactions between nodes at coarse scales in the hierarchy), and for Erdos-Renyi graphs the noise tolerance results if probability of an edge scales as $|V|^{(\gamma-1)}$ (higher γ implying more connectivity).

III. **Localization:** The goal of localization is to identify the set of edges across which the true underlying activation differs, i.e.

$$\partial C = \{(i,j) \in E : x_i \neq x_j\}$$

based on noisy observations $\{y_i\}_{i=1}^{|V|}$.

This problem can be solved via the "edge lasso" which arises as a special case of the generalized fused lasso optimization as described in literature [14]

$$\min_{\hat{\mathbf{x}}} \ \frac{1}{2}\|\mathbf{y} - \hat{\mathbf{x}}\|^2 + \lambda\|\mathbf{D}\hat{\mathbf{x}}\|_1$$

where the matrix $\mathbf{D} \in \mathbb{R}^{|E|\times|V|}$ specifies the constraints imposed by the graph structure. Specifically, each row of the matrix $\mathbf{D}$ corresponds to an edge $(i, j) \in E$ and the entries are zero except for a $+1$ for node i and -1 for node j. Thus, the optimization seeks to find a least square fit to the noisy observations while penalizing the ℓ_1 norm of the differences of measurements across edges in G. This project investigated conditions under which the edge lasso is sparsistent i.e. the edges over which $\hat{\mathbf{x}}$ differs agree exactly with the edge set ∂C, asymptotically for large graph sizes $|V| \to \infty$.

Theorem 5. *[15] Let A denote the maximally connected components of C. For each A, consider the following notion of degree of connectivity:*

$$\rho(A) := \max_{W \subset A} \frac{|\partial \bar{W} \cap \partial A|}{|\partial W \cap \partial \bar{W}|} \frac{|W|}{|A|}$$

Also let $\overline{\partial C}$ denote the set of edges that are not in ∂C and $\Delta^\dagger$ denotes the pseudo-inverse of the graph Laplacian. If for each A, $\rho(A) = o(1)$,

$$\frac{\mu}{\sigma} = \omega\left(\frac{|\partial A|}{|A|}|||D_{\overline{\partial C}}\ \Delta^\dagger_{\overline{\partial C}}|||_{2,\infty}\sqrt{\log(|\overline{\partial C}|)}\right) \quad \textit{and} \quad \frac{\mu}{\sigma} = \omega\left(\frac{1}{\sqrt{|A|}}\right)$$

then the edge lasso is sparsistent.

The theorem provides general conditions for the success of edge lasso. While these conditions are hard to comprehend directly, evaluating them for specific graphs provides useful insights. As shown in [15], for 1-d and 2-d lattice graphs, the conditions imply that edge lasso succeeds at the same SNR (up to log factors) as thresholding the difference of observations at nodes connected by an edge. On the other hand, for more structured graphs such as the nested complete graph (c.f. [15]) if the activated vertices have low connectivity as per $\rho(A)$ (e.g. see Fig. 2), then edge lasso can localize the activated vertices at much lower SNR.

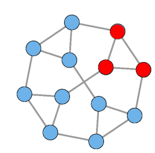

Figure 2: A nested complete graph with a low connectivity activated subgraph.

2.2 Compressive and adaptive measurement model

So far the focus has been on the direct measurement model. This project also explored the used of compressive and adaptive measurements to minimize the resource budget needed for detection and localization of graph-structured patterns. Under the compressive and adaptive measurement model, each observation corresponds to a (random/passive or sequentially designed/active) linear combination of the node measurements, i.e.

$$y_i = \mathbf{a}_i^\top \mathbf{x} + \epsilon_i \qquad i = 1, \ldots, m$$

where the total sensing budget $\sum_i \|\mathbf{a}_i\|^2 \leq m$.

First, the specific case of a $k_1 \times k_2$ block of activation in a $n_1 \times n_2$ lattice graph structure was considered, i.e. $\mathbf{x} = \mu \mathbf{1}_C$ where C corresponds to a $k_1 \times k_2$ contiguous block [16]. The precise tradeos between the various problem parameters, SNR and the number of measurements required to reliably detect and localize the block of activation were characterized. The sufficient conditions are complemented with information theoretic lower bounds. A summary of known results for the vector case and results of this project for the block-structured case are provided in Tables 1 and 2, respectively. Contrary to results in compressed sensing of sparse vectors, where it has been shown that neither adaptivity nor structure help reduce the SNR or number of measurements needed [17, 18, 19, 20], results of this project shows that for reliable **localization** the minimum SNR needed (or equivalently the number of compressive measurements needed) is strongly influenced by both structure and the ability to choose measurements adaptively. However, for **detection** neither adaptivity nor structure reduce the requirement on the SNR.

Table 1: Known results for a k-sparse length n vector

	Detection	Localization
Passive	$\frac{\mu}{\sigma} \asymp \sqrt{\frac{n}{mk^2}}$	$\frac{\mu}{\sigma} \asymp \sqrt{\frac{n \log n}{m}}$, [21] $m \succ k \log n$
Active	[17]	$\frac{\mu}{\sigma} \asymp \sqrt{\frac{n}{m}}$ [18, 19, 20]

Table 2: Findings for a $k_1 \times k_2$ block of activation in a $|V| = n_1 \times n_2$ lattice [16]

	Detection	Localization
Passive	$\frac{\mu}{\sigma} \asymp \sqrt{\frac{n_1 n_2}{m k_1^2 k_2^2}}$	$\frac{\mu}{\sigma} \asymp \sqrt{\frac{n_1 n_2}{m \min(k_1, k_2)}}$
Active		$\frac{\mu}{\sigma} \asymp \max\Big(\sqrt{\frac{n_1 n_2}{m k_1^2 k_2^2}},$

These scalings are verified in Figure **??** where plotting the probability of successful localization vs. SNRs rescaled with predicted scaling, aligns all the curves.

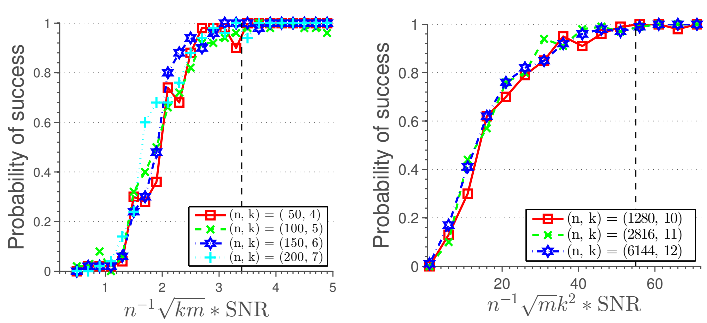

Figure 3: Probability of successful localization of a sparse square block of activation in a square lattice vs. SNRs rescaled with predicted scaling, for 100 passive compressive measurements (left) and 500 adaptive compressive measurements (right), averaged over 100 simulation runs.

The upper bound for the detection problem is achieved by a simple detector based on thresholding the average of measurements obtained using passively designed, constant-valued linear measurements with $\mathbf{a}_i(j) = 1/\sqrt{n_1 n_2}$ for all i and j. The upper bound for passive localization is obtained using a procedure that searches over all contiguous blocks of size $k1 \times k2$ and outputs the one minimizing the squared error. Finally, the upper bound for active localization is attained by a compressive binary search procedure on a collection of cyclically shifted non-overlapping blocks that partition the lattice graph. Details of the procedure are available in [16].

While sequentially designed adaptive compressive measurements yield improvements for the simple case of a block-structured activation in a lattice graph, it wasn't clear whether similar improvements hold for general activation patterns and graph structures. This question was explored in [22] for patterns with low cut-sizes on general graph structures. The results indicate that in general no significant gains over unstructured settings are possible for localizing the activated vertices, however if the activation pattern coincides with a dendrogram over the graph, then the graph structure can be exploited to design adaptive compressive measurements that yield SNR improvements, or equivalently savings in the number of measurements needed. Two methods are proposed in [22] that perform modifications of a compressive binary search over the dendrogram. Comparing these methods to sequentially designed compressed sensing algorithm (SDC) from [23] (which does not exploit structure, but has near-optimal performance for localization of non-zero entries in unstructured sparse vectors) indicates the importance of exploiting structure (see Figure 4).

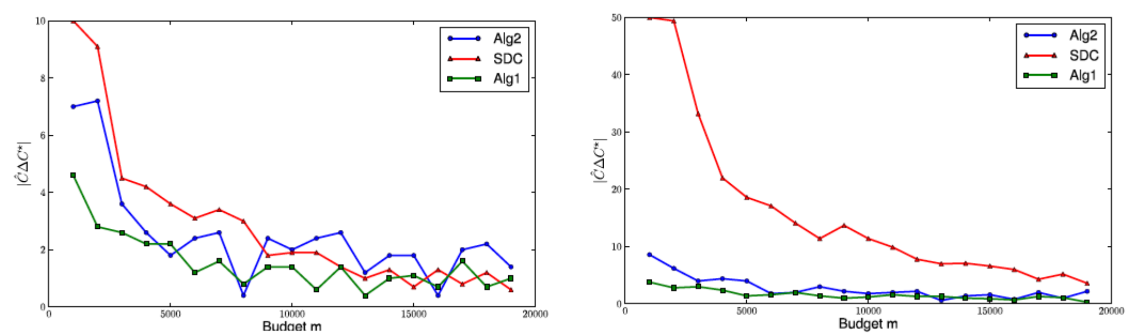

Figure 4: Localization error for proposed Algorithms 1, 2, and SDC from [23] which does not exploit structure. When the activation is very small $k = 10$ (left), all algorithms perform the same, but when activation size $k = 50$, exploiting structure leads to significant improvement. Here G is a 512 node line graph and $\rho = 2$, resulting in one connected activated subgraph.

3 Conclusion

This project addressed the problems of detection, localization and estimation of weak and distributed patterns of activation in a large-scale network given access to direct, compressive and adaptive noisy node measurements. Information-theoretic limits were identified for these problems, along with computationally efficient methods that nearly achieve the limits, for general graph structures and classes of activation patterns. Development of such state-of-the-art methods that are both computationally and statistically efficient is crucial to advance AFOSR's ability to monitor, understand and secure modern large-scale networks. The methods developed leveraged highly inter-disciplinary tools, and resulted in publications including invited papers and oral presentations at NIPS, AISTATS, Asilomar and GlobalSIP, some of the most prominent conferences in machine learning, statistics and signal processing.

4 People involved in various aspects of project

Graduate Students:

- James Sharpnack (PhD student, Machine Learning Department; now postdoc, University of California - San Diego)
- Akshay Krishnamurthy (PhD student, Computer Science Department)

Faculty Collaborator:

- Alessandro Rinaldo (Assistant Professor (now Associate), Statistics Department)

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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

[6] ——, "Detecting activity in graphs via the Graph Ellipsoid Scan Statistic," 2013, in preparation.

[7] J. Sharpnack, A. Krishnamurthy, and A. Singh, "Detecting activations over graphs using spanning tree wavelet bases," in *Artifical Intelligence and Statistics (AISTATS)*, 2013.

[8] ——, "Near-optimal anomaly detection in graphs using Lovász Extended Scan Statistic," in *Neural Information Processing Systems (NIPS)*, 2013.

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

[10] A. Singh, R. Nowak, and R. Calderbank, "Detecting weak but hierarchically-structured patterns in networks," in *Artificial Intelligence and Statistics (AISTATS)*, 2010.

[11] D. Aldous, "The random walk construction of uniform spanning trees and uniform labelled trees," *SIAM Journal on Discrete Mathematics*, vol. 3, no. 4, pp. 450–465, 1990.

[12] J. Sharpnack and A. Singh, "Near-optimal and computationally efficient detectors for weak and sparse graph-structured patterns," in *GlobalSIP*, 2013.

[13] ——, "Identifying graph-structured activation patterns in networks," in *Neural Information Processing Systems (NIPS)*, 2010.

[14] R. J. Tibshirani and J. Taylor, "The solution path of the generalized lasso," *Annals of Statistics*, vol. 39, no. 3, pp. 1335–1371, 2011.

[15] J. Sharpnack, A. Rinaldo, and A. Singh, "Sparsistency of the edge lasso over graphs," in *Artificial Intelligence and Statistics (AISTATS)*, 2012.

[16] S. Balakrishnan, M. Kolar, A. Rinaldo, and A. Singh, "Recovering block-structured activations using compressive measurements," submitted, Available at http://arxiv.org/abs/1209.3431.

[17] E. Arias-Castro, "Detecting a vector based on linear measurements," *Electronic Journal of Statistics*, vol. 6, pp. 547–558, 2012.

[18] E. Arias-Castro, E. Candès, and M. Davenport, "On the fundamental limits of adaptive sensing," *arXiv:1111.4646*, 2011.

[19] M. Davenport and E. Arias-Castro, "Compressive binary search," *arXiv:1202.0937*, 2012.

[20] M. Malloy and R. Nowak, "Near-optimal compressive binary search," *arXiv:1203.1804*, 2012.

[21] M. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5728–5741, 2009.

[22] A. Krishnamurthy, J. Sharpnack, and A. Singh, "Recovering graph-structured activations using adaptive compressive measurements," in *Asilomar Conference on Signals, Systems, and Computers*, 2013.

[23] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Sequentially designed compressed sensing," in *IEEE Statistical Signal Processing Workshop (SSP)*, 2012.

5 Publications resulting from the project

Published/Accepted Papers

(Available at `http://www.cs.cmu.edu/~aarti/pubs.html`)

1. J. Sharpnack, A. Krishnamurthy, and A. Singh, "Near-optimal anomaly detection in graphs using Lovasz Extended Scan Statistic," in Neural Information Processing Systems (NIPS), 2013, *to appear.*

2. J. Sharpnack and A. Singh, "Near-optimal and computationally efficient detectors for weak and sparse graph-structured patterns," in GlobalSIP, 2013, *to appear, invited paper.*

3. A. Krishnamurthy, J. Sharpnack, and A. Singh, "Recovering graph-structured activations using adaptive compressive measurements," in Asilomar Conference on Signals, Systems, and Computers, 2013, *to appear, invited paper, finalist for best student paper.*

4. J. Sharpnack, A. Krishnamurthy, and A. Singh, "Detecting activations over graphs using spanning tree wavelet bases," in Artificial Intelligence and Statistics (AISTATS), 2013, *oral presentation.*

5. J. Sharpnack, A. Rinaldo, and A. Singh, "Changepoint detection over graphs with the spectral scan statistic," in Artificial Intelligence and Statistics (AISTATS), 2013.

6. J. Sharpnack, A. Rinaldo, and A. Singh, "Sparsistency of the edge lasso over graphs," in Artificial Intelligence and Statistics (AISTATS), 2012.

7. J. Sharpnack and A. Singh, "Identifying graph-structured activation patterns in networks," in Neural Information Processing Systems (NIPS), 2010, *oral presentation.*

8. A. Singh, R. Nowak, and R. Calderbank, "Detecting weak but hierarchically-structured patterns in networks," in Artificial Intelligence and Statistics (AISTATS), 2010, *oral presentation.*

Papers submitted/in preparation

1. J. Sharpnack, A. Rinaldo, and A. Singh, "Detecting activity in graphs via the Graph Ellipsoid Scan Statistic," 2013, in preparation.

2. S. Balakrishnan, M. Kolar, A. Rinaldo, and A. Singh, "Recovering block-structured activations using compressive measurements," submitted, Available at `http://arxiv.org/abs/1209.3431`.

PhD Thesis

1. J. Sharpnack. Graph structured normal means inference. Available at `http://www.cs.cmu.edu/~aarti/pubs/JamesThesis.pdf`