# Recovering Graph-Structured Activations using Adaptive Compressive Measurements

Akshay Krishnamuthy
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
akshaykr@cs.cmu.edu

James Sharpnack
Machine Learning Department
Statistics Department
Carnegie Mellon University
Pittsburgh, PA 15213
jsharpna@cs.cmu.edu

Aarti Singh
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213
aartisingh@cs.cmu.edu

***Abstract*—We study the localization of a cluster of activated vertices in a graph, from adaptively designed compressive measurements. We propose a hierarchical partitioning of the graph that groups the activated vertices into few partitions, so that a top-down sensing procedure can identify these partitions, and hence the activations, using few measurements. By exploiting the cluster structure, we are able to provide localization guarantees at weaker signal to noise ratios than in the unstructured setting. We complement this performance guarantee with an information theoretic lower bound, providing a necessary signal-to-noise ratio for any algorithm to successfully localize the cluster. We verify our analysis with some simulations, demonstrating the practicality of our algorithm.**

## I. Introduction

We are interested in recovering the support of a sparse vector $\mathbf{x} \in \mathbb{R}^n$ observed through the noisy linear model:

$$y_i = a_i^T \mathbf{x} + \epsilon_i$$

Where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and $\sum_i ||a_i||^2 \leq m$. This support recovery problem is well known and fundamental to the theory of compressive sensing, which involves estimating a high-dimensional signal vector from few linear measurements [1]. Indeed if $\mathbf{x}$ is a $k$-sparse vector whose non-zero components are $\geq \mu$, it is now well known that one cannot identify these components if $\frac{\mu}{\sigma} = o(\sqrt{\frac{n}{m}\log(n/k)})$ and one can if $\frac{\mu}{\sigma} = \omega(\sqrt{\frac{n}{m}\log n})$, provided that $m \geq k \log n$ [2].

We build upon the classical results of compressive sensing by developing procedures that are *adaptive* and that exploit additional *structure* in the underlying signal. Adaptivity allows the procedure to focus measurements on activated components of the signal while structure can dramatically reduce the combinatorial search space of the problem. Combined, both ideas can lead to significant performance improvements over classical compressed sensing. This paper explores the role of adaptivity and structure in a very general support recovery problem.

Active learning and adaptivity are not new ideas to the signal processing community and a number of papers in recent years have characterized, the advantages and limits of adaptive sensing over passive approaches. One of the first ideas in this direction was *distilled sensing* [3], which uses direct rather than compressive measurements. Inspired by that work, a number of authors have studied adaptivity in compressive sensing and shown similar performance gains [4], [5], [6]. These approaches do not incorporate any notion of structure.

The introduction of structure to the compressed sensing framework has also been explored by a number of authors [7], [8], [9]. Broadly speaking, these structural assumptions restrict the signal to few of the $\binom{n}{k}$ linear subspaces that contain $k$-sparse signals. With this restrictions, one can often design sensing procedures that focus on these allowed subspaces and enjoy significant performance improvements over unstructured problems. We remark that both [7] and [9] develop adaptive sensing procedures for structured problems, but under a more restrictive setting than this study.

This paper continues in both of these directions, exploring the role of adaptivity *and* structure in recovering activated clusters in graphs. We consider localizing activated clusters of nodes whose boundary in the graph is smaller than some parameter $\rho$. This notion of structure is more general than previous studies, yet we are still able to demonstrate performance improvements over unstructured problems.

Our study of cluster identification is motivated by a number of applications in sensor networks measurement and monitoring, including identification of viruses in human or computer networks or contamination in a body of water. In these settings, we expect the signal of interest to be localized, or clustered, in the underlying network, and we want to develop efficient procedures that exploit this cluster structure.

In this paper, we propose two related adaptive sensing procedures for identifying a cluster of activations in a network. We give a sufficient condition on the SNR under which the first procedure exactly identifies the cluster. While this SNR is only slightly weaker than the SNR that is sufficient for unstructured problems, we show, via information theoretic arguments, that one cannot hope for significantly better performance.

For the second procedure, we perform a more refined analysis and show that the required SNR depends on how our algorithmic tool captures the cluster structure. In some cases this can lead to consistent recovery at much weaker SNR. The second procedure can also be adapted to recover a large fraction of the cluster. We also explore the performance of our procedures via an empirical study. Our results demonstrate the gains from exploiting both structure and adaptivity in support recovery problems. Due to space restrictions, all proofs are available in an extended version of the paper [10].

| Setting | Necessary | Sufficient |
|---|---|---|
| Passive, unstructured | $\sqrt{\frac{n}{m}\log(n/k)}$ [2] | $\sqrt{\frac{n}{m}\log n}$ [2] |
| Adaptive, unstructured | $\sqrt{\frac{n}{m}}$ [11] | $\sqrt{\frac{n}{m}\log k}$ [5] |
| Adaptive, structured | $\sqrt{\frac{n}{m}}$ (Thm. 4) | $\sqrt{\frac{n}{m}\log(\rho\log n)}$ (Prop. 3) |

TABLE I: Compressed Sensing landscape.

| Graph | Structure | Necessary | Sufficient |
|---|---|---|---|
| 2-d Lattice | Rectangle | $\frac{1}{k}\sqrt{\frac{n}{m}}$ [9] | $\frac{1}{k}\sqrt{\frac{n}{m}}$ [9] |
| Rooted Tree | Rooted subtree | $\sqrt{\frac{k}{m}}$ [12] | $\sqrt{\frac{k}{m}\log k}$ [7] |
| Arbitrary | Best case | | $\frac{1}{k}\sqrt{\frac{n}{m}\log((\rho+k)\log n)}$ |

TABLE II: Adaptive, Structured, Compressed Sensing.

We put our results in context of the compressed sensing landscape in Tables I and II. Here $k$ is the cluster size and, for the structured setting, $\rho$ denotes the number of edges leaving the cluster. Near-optimal procedures for passive and adaptive unstructured support recovery were analyzed in [2] and [5] respectively. Our work provides both upper and lower bounds for the adaptive structured setting. Focusing on different notions of structure, Balakrishnan et. al. give necessary and sufficient conditions for recovering a small square of activiations in a grid [9] while Soni and Haupt analyze the recovery of tree-sparse signals [7], [12]. Our work provides guarantees that depends on how well the signal is captured by our algorithmic construction. In the worst case, we guarantee exact recover with an SNR of $\sqrt{\frac{n}{m}\log(\rho\log n)}$ (Proposition 3) and in the best case, we can tolerate an SNR of $\frac{1}{k}\sqrt{\frac{n}{m}\log((\rho+k)\log n)}$ (Theorem 5). It is worth mentioning that [7] obtains better results than ours, but study a very specific setting where the graph is a rooted tree and the signal is rooted subtree.

## II. Main Results

Let $C^\star$ denote a set of activated vertices in a known graph $G=(V,E)$ on $n$ nodes with maximal degree $d$. We observe $C^\star$ through noisy compressed measurements of the vector $\mathbf{x}=\mu\mathbf{1}_{C^\star}$, that is we may select sensing vectors $a_i\in\mathbb{R}^n$ and observe $y_i=a_i^T\mathbf{x}+\epsilon_i$ where $\epsilon_i\sim\mathcal{N}(0,\sigma^2)$ independently. We require $\sum_i||a_i||^2\le m$ so that the total sensing energy, or budget, is at most $m$. We allow for *adaptivity*, meaning that the procedure may use the measurements $y_1,\dots,y_{i-1}$ to inform the choice of the subsequent vector $a_i$. Our goal is to develop procedures that successfully recover $C^\star$ in a low signal-to-noise ratio regime.

We will require the set $C^\star$, which we will henceforth call a *cluster*, to have small cut-size in the graph $G$. Formally:

$$C^\star\in\mathcal{C}_\rho=\{C:|\{(u,v):u\in C,v\notin C\}|\le\rho\}$$

Our algorithmic tool for identification of $C^\star$ is a **dendrogram** $\mathcal{D}$, a hierarchical partitioning of $G$. A dendrogram is a tree of blocks $\{D\}$ where each block is a connected set of vertices in $G$. For now, we state the critical properties that we require of $\mathcal{D}$. We will see one way to construct such dendrograms in Section II-C. Note that results in [13] imply that one can construct a suitable dendrogram for any graph.

**Assumption 1.** *Let $\mathcal{D}$ be a dendrogram for $G$. We assume that*

1) *The root of $\mathcal{D}$ is $V$, the set of all vertices, and the leaves of the dendrogram are all of the singletons $\{v\}$, $v\in G$. The sets corresponding to the children of a block $D$ form a partition of the elements in $D$ while preserving graph connectivity in each cluster.*
2) *$\mathcal{D}$ has degree at most $d$, the maximum degree in $G$.*
3) *$\mathcal{D}$ is approximately balanced. Specifically the child of any block $D$ has size at most $|D|/2$.*
4) *The height $L$ of $\mathcal{D}$ is at most $\log_2(n)$* [1].

**Algorithm 1** Exact Recovery

**Require:** Dendrogram $\mathcal{D}$ and sensing budget $m$, failure probability $\delta$.
set $\alpha=\frac{m}{4n\log_2\rho}$, $\tau=\sigma\sqrt{2\log((d\rho L+1)/\delta)}$.
(1) Let $D$ be the root of $\mathcal{D}$.
(2) Obtain $y_D=\sqrt{\alpha}\mathbf{1}_D^T\mathbf{x}+\epsilon_D$
(3) If $y_D\ge\mu\sqrt{\alpha}|D|-\tau$ add $D$ to the estimate $\hat{C}$.
(4) If $\tau\le y_D\le\mu\sqrt{\alpha}|D|-\tau$ recurse on (2)-(4) with $D$'s children.
Output $\hat{C}$.

By the fact that each block of $\mathcal{D}$ is a connected set of vertices, we immediately have the following proposition:

**Proposition 2.** *For any $C^\star$ in $\mathcal{C}_\rho$ at most $\rho$ blocks are **impure** at any level in $\mathcal{D}$. A block $D$ is impure if $0<|D\cap C^\star|<|D|$.*

### A. Universal Guarantees

With a dendrogram $\mathcal{D}$, we can sense with measurements of the form $\mathbf{1}_D$ for a parent block $D$ and recursively sense on the children blocks to identify the activated vertices. This procedure has the same flavor as the compressive binary search procedure [4]. Specifically, fix a threshold $\tau$ and energy parameter $\alpha$ and when sensing on block $D$ obtain the measurement

$$y_D=\sqrt{\alpha}\mathbf{1}_D^T\mathbf{x}+\epsilon_D \tag{1}$$

If $\tau<y_D<\mu\sqrt{\alpha}|D|-\tau$ continue sensing on $D$'s children, otherwise terminate the recursion. See Algorithm 1. At a fairly weak SNR and with appropriate setting for $\tau$ and $\alpha$, we can show that this procedure will exactly identify $C^\star$:

**Proposition 3.** *Set $\tau=\sigma\sqrt{2\log((d\rho L+1)/\delta)}$. If the SNR satisfies:*

$$\frac{\mu}{\sigma}\ge\sqrt{\frac{8}{\alpha}\log\left(\frac{d\rho L+1}{\delta}\right)} \tag{2}$$

*then with probability $\ge1-\delta$, Algorithm 1 recovers $C^\star$ and using a sensing budget of at most $3n\alpha\log_2(d\rho)$.*

We must set $\alpha$ appropriately so we do not exceed our budget of $m$. With the correct choice, the SNR requirement is:

$$\frac{\mu}{\sigma}\ge\sqrt{\frac{24n}{m}\log_2(d\rho)\log\left(\frac{d\rho L+1}{\delta}\right)}$$

Algorithm 1 performs similarly to the adaptive procedures for unstructured support recovery. For constant $\rho$, the SNR requirement is $\omega(\sqrt{\frac{n}{m}\log\log_2 n})$ which is on the same order

[1] Unless explicitly stated, log will denote the natural logarithm.

as the compressive binary search procedure [4] for recovering 1-sparse signals. For $k$-sparse signals, the best results [6], [5] require SNR of $\sqrt{\frac{n \log k}{m}}$ which can be much worse than our guarantee when $k \geq \log n$ and $\rho$ is small.

Thus, the procedure does enjoy small benefit from exploiting structure, but the generality of our set up precludes more substantial performance gains. Indeed, we are able to show that one cannot do much better than Algorithm 1. This information theoretic lower bound is a simple consequence of the results from [11].

**Theorem 4.** *Fix any graph $G$ and suppose $\rho \geq d$. If:*

$$\frac{\mu}{\sigma} = o\left(\sqrt{\frac{n}{m}}\right)$$

*then* $\inf_{\hat{C}} \sup_{C^\star \in \mathcal{C}_\rho} \mathbb{P}[\hat{C} \neq C^\star] \to \frac{1}{2}$ *so that no procedure can reliably estimate $C^\star \in \mathcal{C}_\rho$.*

The lower bound demonstrates one of the fundamental challenges in exploiting structure in the cluster recovery problem: since $\mathcal{C}_\rho$ is not parameterized by cluster size, in the worst case, one should not hope for performance improvements that depend on cluster size or sparsity. More concretely, if $\rho \geq d$, the set $\mathcal{C}_\rho$ contains all singleton vertices, reducing to a completely unstructured setting. Here, the results of [4] imply that to exactly recover a cluster of size one, it is necessary to have SNR of $\sqrt{\frac{n}{m}}$. This is one argument for the lower bound.

While our lower bound relies on singletons, they are not the only challenging facet of the problem. Another is the generality of the graph $G$. Indeed, nothing in our setup prevents $G$ from being a complete graph on $n$ vertices, in which case there is no structure, so one should not expect stronger results.

The inherent difficulty of this problem is not only information theoretic, but also computational. The typical way to exploit structure is to scan across the possible signal patterns, using the fact that the search space is highly restricted as in the Generalized Likelihood Ratio Test or the Scan Statistic. In the cluster setting, Karger proved that the number of cuts of size $\rho$ is on the order of $\Theta(n^\rho)$ [14], meaning that restricting signals to $\mathcal{C}_\rho$ does not significantly reduce the search space.

Even if we could sweep across all cuts of size $\rho$, without further assumptions on $G$ or $\mathcal{C}_\rho$ there could be a number of clusters that disagree with $C^\star$ on only a few vertices, and distinguishing between these would require high SNR. As a concrete example, if we are interested in localizing a contiguous chain of activations in the line graph, an adaptation of the lower bound in [9] shows that if $\frac{\mu}{\sigma} = o(\max\{\frac{1}{k}\sqrt{\frac{n-k}{m}}, \sqrt{\frac{1}{m}}\})$ then localization is impossible. The second term arises from the overlap between the contiguous blocks. It is independent of $n$, but also independent of $k$, showing that exploiting structure does not significantly help when distinguishing clusters that differ only by a few vertices.

### *B. Cluster-Specific Guarantees*

The main performance bottleneck for Algorithm 1 comes from testing whether a block of size 1 is active or not. If there are no such singleton blocks, meaning that the cluster $C^\star$ is grouped into large blocks in $\mathcal{D}$, we might expect that Algorithm 1 or a variant can succeed at lower SNR. We formalize this idea here, analyzing an algorithm whose performance depends on how $C^\star$ is partitioned across the dendrogram $\mathcal{D}$.

**Algorithm 2** Approximate Recovery

**Require:** Dendrogram $\mathcal{D}$, sensing budget parameters $\alpha, \beta$.
Set $\alpha, z$ as in Theorem 5. Initialize $\mathcal{K} = \emptyset$.
(1) Let $D$ be the root of $\mathcal{D}$.
(2) Obtain $y_D = \sqrt{\alpha}\mathbf{1}_D^T\mathbf{x} + \epsilon_D$.
(3) If $y_D \geq z$ add $D$ to $\mathcal{K}$ and recurse on (1)-(3) with $D$s children.
Construct $U$ an orthonormal basis for span$\{\mathbf{1}_D\}_{D\in\mathcal{K}}$.
Sense $\mathbf{y} = \sqrt{\beta}U^T\mathbf{x} + \epsilon$ and form $\hat{\mathbf{x}} = U\mathbf{y}/\sqrt{\beta}$.
Output $\hat{C} = \text{argmax}_{C\subseteq[n]} \frac{\mathbf{1}_C^T\hat{\mathbf{x}}}{||\hat{\mathbf{x}}||\sqrt{|C|}}$.

We quantify this dependence with the notion of *maximal* blocks $D \in \mathcal{D}$ which are the largest blocks that are completely active. Formally $D$ is *maximal* if $D \cap C^\star = D$ and $D$'s parent is impure, and we denote this set of maximal blocks $\mathcal{M}$. If the maximal blocks are all large, then we can hope to obtain significant performance improvements.

The algorithm consists of two phases. The first phase (the adaptive phase) is similar to Algorithm 1. With a threshold $z$, and energy parameter $\alpha$, we sense on a block $D$ with

$$y_D = \sqrt{\alpha}\mathbf{1}_D^T\mathbf{x} + \epsilon_D$$

If $y_D > z$ we sense on $D$'s children and we construct a **pruned dendrogram** $\mathcal{K}$ of all blocks $D$, for which $y_D > z$. The pruned dendrogram is much smaller than $\mathcal{D}$ but it retains a large fraction of $C^\star$.

Since we have significantly reduced the dimensionality of the problem we can now use a passive localization procedure to identify $C^\star$ at a low SNR. In the passive phase, we construct an orthonormal basis $U$ for the subspace:

$$\{\mathbf{1}_D : D \in \mathcal{K}\}$$

With another energy parameter $\beta$, we observe $y_i = \sqrt{\beta}u_i^T\mathbf{x} + \epsilon_i$ for each basis vector $u_i$ and form the vector $\mathbf{y} = \sqrt{\beta}U^T\mathbf{x} + \epsilon$ by stacking these observations. We then construct the vector $\hat{\mathbf{x}} = U\mathbf{y}/\sqrt{\beta}$. With the vector $\hat{\mathbf{x}}$ we solve the following optimization problem to identify the cluster ($[n] = \{1, \ldots, n\}$):

$$\hat{C} = \text{argmax}_{C\subseteq[n]} \frac{\mathbf{1}_C^T\hat{\mathbf{x}}}{||\hat{\mathbf{x}}||\sqrt{|C|}}$$

which can be solved by a simple greedy algorithm. A detailed description is in Algorithm 2. For a more concise presentation, in the following results, we omit the dependence on the maximum degree of the graph, $d$. This localization guarantee is stated in terms of the distance $d(\hat{C}, C^\star) \triangleq 1 - \frac{|\hat{C}\cap C^\star|}{\sqrt{|\hat{C}||C^\star|}}$.

**Theorem 5.** *Set $z$ so that $\mathbb{P}[\mathcal{N}(0,1) > \sigma z] \leq \frac{\sqrt{5}-1}{d}$ and*[2]*:*

$$\alpha = \frac{m}{n\log_2((\rho+k)\log n)}, \beta = \frac{m}{(\rho+k)\text{polylog}(n,\rho)}$$

[2]We provide exact definitions of $\alpha$ and $\beta$ in the supplementary material [10].

| Setting | $\frac{\mu}{\sigma}$ |
|---|---|
| One maximal block | $\omega\left(\frac{1}{k}\sqrt{\frac{n}{m}\log(k\log n)}\right)$ |
| Uniform sizes | $\omega\left(\frac{\rho}{k}\sqrt{\frac{n}{m}\log(k\log n)}\right)$ |
| Worst Case | $\omega\left(\sqrt{\frac{n}{m}\log(k\log n)}\right)$ |

TABLE III: Instantiations of Theorem 5

*where* $k = |C^\star|$. *If:*

$$\frac{\mu}{\sigma} = \omega\left(\frac{(\rho+k)\text{polylog}(n,\rho)}{\sqrt{mk}} + \sqrt{\frac{n\log_2((\rho+k)\log n)}{m|M_{\min}|^2}}\right)$$

*where* $M_{\min} = argmin_{M\in\mathcal{M}} M$, *then* $d(\hat{C}, C^\star) \to 0$ *and the budget is* $O(m)$.

The SNR requirement in the theorem decomposes into two terms, corresponding to the two phases of the algorithm, and our choice of $\alpha$ and $\beta$ distribute the sensing budget evenly over the terms, allocating $O(m)$ energy to each. Note however, that the first term, corresponding to the passive phase, has a logarithmic dependence on $n$ while the second term, corresponding to the adaptive phase, has a polynomial dependence, so in practice one should allocate more energy to the adaptive phase. With our allocation, the second term will usually dominate, particularly for small $\rho$ and $k$, which is a regime of interest. Then the required SNR is:

$$\frac{\mu}{\sigma} = \omega\left(\frac{1}{|M_{\min}|}\sqrt{\frac{n}{m}\log_2((\rho+k)\log n)}\right)$$

To more concretely interpret the result, we present sufficient SNR scalings for three scenarios in Table III. We think of $\rho \ll |C^\star|$. The most favorable realization is when there is only one maximal block of size $k$. Here, there is a significant gain in SNR over unstructured recovery or even Proposition 3.

Another interesting case is when the maximal blocks are all at the same level in the dendrogram. In this case, there can be at most $\rho d$ maximal blocks since each of the parents must be impure and there can only be $\rho$ impure blocks per level. If the maximal blocks are approximately the same size, then $|M_{\min}| \approx k/\rho$, and we arrive at the requirement in the second row of Table III. Again we see performance gains from structure, although there is some degradation.

Unfortunately, since the bound depends on $M_{\min}$, we do not always realize such gains. When $M_{\min}$ is a singleton block (one node), our bound deteriorates to the third row of Table III. We remark that modulo $\log\log$ factors, this matches the SNR scaling for the unstructured (sparse) setting. It also nearly matches the lower bound in Theorem 4.

Theorem 5 shows that the size of $|M_{\min}|$ is the bottleneck to recovering $C^\star$. If we are willing to tolerate missing the small blocks we can sense at lower SNR.

**Corollary 6.** *Let* $\tilde{C} = \bigcup_{M\in\mathcal{M},|M|\geq t} M$ *and* $k = |C^\star|$. *If:*

$$\frac{\mu}{\sigma} = \omega\left(\frac{(\rho+k)\text{polylog}(n,\rho)}{\sqrt{mk}} + \frac{1}{t}\sqrt{\frac{n}{m}\text{polylog}(n,\rho,j,t)}\right)$$

*then with probability* $1-o(1)$, $d(\hat{C},\tilde{C}) \to 0$ *and* $n\to\infty$.

**Algorithm 3** FindBalance

**Require:** $\mathcal{T}$ a subtree of $G$ and initialize $v \in \mathcal{T}$ arbitrarily .
(1) Let $T'$ be the component of $\mathcal{T}\backslash\{v\}$ of largest size.
(2) Let $w$ be the unique neighbor of $v$ in $T'$.
(3) Let $T''$ be the component of $\mathcal{T}\backslash\{w\}$ of largest size.
(4) Stop and return $v$ if $|T''| \geq |T'|$.
(5) $v \leftarrow w$. Repeat at (1).

**Algorithm 4** BuildDendrogram

**Require:** $\mathcal{T}$ is a spanning tree of $G$.
Initialize $\mathcal{D} = \{\{v : v \in \mathcal{T}\}\}$.
Let $v$ be the output of FindBalance applied to $\mathcal{T}$.
Let $\mathcal{T}_1, \ldots, \mathcal{T}_{d_v}$ be the connected component of $\mathcal{T} \setminus v$ and add $v$ to the smallest component.
Add $\{v : v \in \mathcal{T}_i\}$ for each $i$ as children of $\mathcal{T}$ to $\mathcal{D}$.
Recurse at (2) for each $\mathcal{T}_i$ as long as $|\mathcal{T}_i| \geq 2$.

In particular, we can recover all maximal blocks of size $t$ with SNR on the order of $\tilde{O}(\frac{1}{t}\sqrt{\frac{n}{m}})$, which clearly shows the gain in exploiting structure in this problem.

### C. Constructing Dendrograms

A general algorithm for constructing a dendrogram parallels the construction of spanning tree wavelets in [13]. Given a spanning tree $\mathcal{T}$ for $G$, the root of the dendrogram is $V$, and the children are the subtrees around a balancing vertex $v \in \mathcal{T}$. The dendrogram is built recursively by identifying balancing vertices and using the subtrees as children. See Algorithm 4 for details. It is not hard to verify that a dendrogram constructed in this way satisfies Assumption 1.

## III. Experiments

We conducted two simulation studies to verify our theoretical results and examine the performance of our algorithms empirically. First, we empirically verify the SNR scaling in Proposition 3. In the second experiment, we compare both of our algorithms with the algorithm from [5], which is an unstructured adaptive compressed sensing procedure with state-of-the-art performance.

In Figure 1 we plot the probability of successful recovery of $C^\star$ as a function of a rescaled parameter. This parameter $\theta(n,m,\rho,\frac{\mu}{\sigma}) = \frac{\mu}{\sigma}\sqrt{\frac{m}{n\log_2\rho\log(\rho\log n)}}$ was chosen so that the

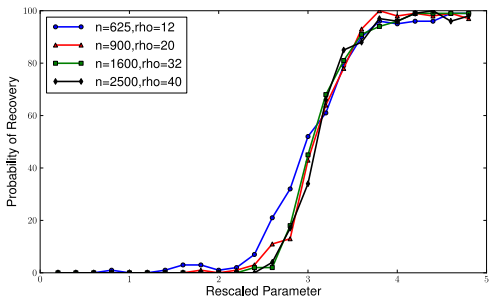


Fig. 1: Probability of success for Algorithm 1 as a function of the $\theta = \frac{\mu}{\sigma}\sqrt{\frac{m}{n\log_2\rho\log(\rho\log(n))}}$ for the torus.

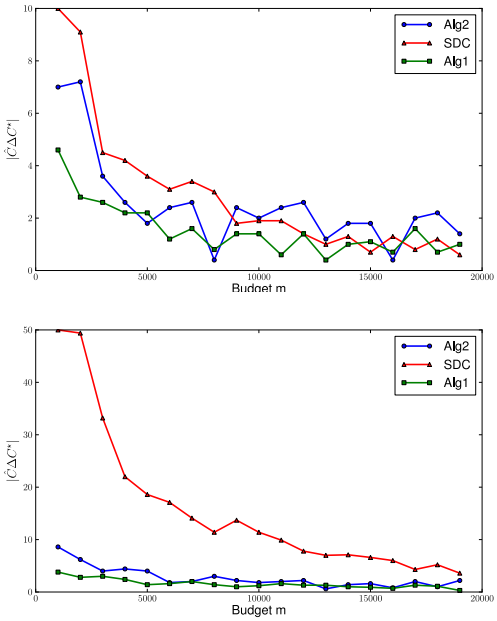


Fig. 2: Error as a function of $m$ for $n = 512$ and $k = 10, 50$ (top, bottom) demonstrating the gains from exploiting structure. Here $G$ is a line graph and $\rho = 2$, resulting in one connected cluster.

condition on the SNR in Proposition 3 is equivalent to $\theta = c$ for some constant $c$. Proposition 3 then implies that with this rescaling, the curves should all line up, which is the phenomenon we observe in Figure 1. Here $G$ is the two dimensional torus and $\mathcal{D}$ was constructed using Algorithm 4.

In Figure 2 we plot the error, measured by $|\hat{C}\Delta C^{\star}|$, as a function of $m$ for three algorithms. We use both Algorithms 1 and 2 as well as the sequentially designed compressed sensing algorithm (SDC) [5], which does not exploit structure, but has near-optimal performance for unstructured sparse recovery. Here $G$ is the line graph, $\mathcal{D}$ is the balanced binary dendrogram, and $\rho = 2$ so each signal is a contiguous block.

In the first figure, $k = 10$ and since the maximal clusters are necessarily small, there should be little benefit from structure. Indeed, we see that all three algorithms perform similarly. This demonstrates that in the absence of structure, our procedures perform comparably to existing approaches for unstructured recovery. When $k = 50$ (the second figure), we see that both Algorithms 1 and 2 outperform SDC, particularly at low SNRs. Here, as predicted by our theory, Algorithm 2 can identify a large part of the cluster at very low SNR by exploiting the cluster structure. In fact Algorithm 1 empirically performs well in this regime although we do not have theory to justify this.

## IV. Conclusion

We explore the role of structure and adaptivity in the support recovery problem, specifically in localizing a cluster of activations in a network. We show that when the cluster has small cut size, exploiting this structure can result in performance improvements in terms of signal-to-noise ratios sufficient for cluster recovery. If the true cluster $C^{\star}$ coincides with a dendrogram over the graph, then weaker SNRs can be tolerated. These results do not contradict the necessary conditions for this problem, which shows that one cannot do much better than the unstructured setting for exact recovery.

While our work contributes to understanding the role of structure in compressive sensing, our knowledge is still fairly limited. We now know of some specific instances where structured signals can be localized at very weak SNR, but we do not have a full characterization of this effect. Our goal was to give such a precise characterization, but the generality of our set-up resulted in an information-theoretic barrier to demonstrating significant performance gains. An interesting direction for future research is to precisely quantify settings that are not too general nor very specific when structure can lead to improved sensing performance and to develop algorithms that enjoy these gains.

## Acknowledgements

This research is supported in part by AFOSR under grant FA9550-10-1-0382 and NSF under grant IIS-1116458. AK is supported in part by a NSF Graduate Research Fellowship.

## References

[1] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

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

[3] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *Information Theory, IEEE Transactions on*, vol. 57, no. 9, pp. 6222–6235, 2011.

[4] M. A. Davenport and E. Arias-Castro, "Compressive binary search," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 1827–1831.

[5] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Sequentially designed compressed sensing," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*. IEEE, 2012, pp. 401–404.

[6] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," in *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*. IEEE, 2012, pp. 1935–1939.

[7] A. Soni and J. Haupt, "Efficient adaptive compressive sensing using sparse hierarchical learned dictionaries," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. IEEE, 2011, pp. 1250–1254.

[8] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982–2001, 2010.

[9] S. Balakrishnan, M. Kolar, A. Rinaldo, and A. Singh, "Recovering block-structured activations using compressive measurements," *arXiv preprint arXiv:1209.3431*, 2012.

[10] A. Krishnamurthy, J. Sharpnack, and A. Singh, "Recovering graph-structured activations using adaptive compressive measurements," *arXiv preprint arXiv:1305.0213*, 2013.

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

[12] A. Soni and J. Haupt, "On the fundamental limits of recovering tree sparse vectors from noisy linear measurements," *arXiv preprint arXiv:1306.4391*, 2013.

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

[14] D. R. Karger, "Minimum cuts in near-linear time," *Journal of the ACM (JACM)*, vol. 47, no. 1, pp. 46–76, 2000.