Network Discovery via Constrained Tensor Analysis of fMRI Data

Ian Davidson, Sean Gilpin
Computer Science
U.C. Davis
{indavidson|sagilpin}
@ucdavis.edu

Owen Carmichael
Neurology
U.C. Davis
ocarmichael@ucdavis.edu

Peter Walker
Cognitive Psychology
United States Navy
peter.b.walker@navy.mil

ABSTRACT

We pose the problem of network discovery which involves simplifying spatio-temporal data into cohesive regions (nodes) and relationships between those regions (edges). Such problems naturally exist in fMRI scans of human subjects. These scans consist of activations of thousands of voxels over time with the aim to simplify them into the underlying cognitive network being used. We propose supervised and semi-supervised variations of this problem and postulate a constrained tensor decomposition formulation and a corresponding alternating least squares solver that is easy to implement. We show this formulation works well in controlled experiments where supervision is incomplete, superfluous and noisy and is able to recover the underlying ground truth network. We then show that for real fMRI data our approach can reproduce well known results in neurology regarding the default mode network in resting-state healthy and Alzheimer affected individuals. Finally, we show that the reconstruction error of the decomposition provides a useful measure of the network strength and is useful at predicting key cognitive scores both by itself and with clinical information.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data mining

Keywords

fMRI, Tensors, Applications

1. INTRODUCTION

Neuroscience is at a moment in history where mapping the connectivity of the human brain non invasively and **in vivo** has just begun with many unanswered questions. Just as sequencing the human genome offered tremendous opportunities to advance biology, mapping the human brain offers as much if not more opportunity to help humans. Whilst the anatomical structures in the brain have been well known

for decades, how they are used in combination to form task specific networks has still not been completely explored. Understanding what these networks are, how they develop, deteriorate, and vary across individuals will provide a range of benefits from disease diagnosis, to understanding the neural basis of creativity, and in the future brain augmentation.

Data mining has made significant inroads into real world practical applications in industry and the sciences. However most existing work focuses on simple lower-level tasks such as predicting binary labels, clustering and dimension reduction. Take for instance predicting binary labels, even though recent advances in structured, semi-supervised, multi-task and transfer learning further widens the scope of applications, the focus is still labels. This often requires the practitioner to "shoe-horn" their more complex tasks into the algorithm's setting which is the case for neuroscientists.

The focus of this paper is a first attempt to transition to more complex higher-level discovery tasks and in particular eliciting networks from spatio-temporal data represented as a tensor. The aim of our work is algorithms that can take event data in the form of activations over a three dimensional spatial region x, y, z over time t and simplify that data into a network. This involves both aggregation so that the active cohesive regions (nodes) are identified and the formation of relationships (edges) between these regions. The edges and their weights can indicate properties such as information flow, excitation/inhibition or probabilistic relationships. Though we focus on cognitive networks in this paper, we anticipate the work can be applied to other domains such as ocean temperature monitoring and climate modeling. In this paper we shall focus predominantly on node discovery and explore how to create one type of edge. Figure 1 shows a simplified example involving just a slice of brain activity. Here the nodes of the graph represent regions of the slice whose behavior is cohesive over time and the edges indicating that the activations of the regions are synchronized or related in some way. We see this as an initial but important move away from simple pattern recognition and data mining towards true *knowledge discovery*.

We begin this paper by introducing and motivating the three core problems of network discovery: node discovery, edge discovery and network verification along with standard techniques. In the next section we present our constrained tensor decomposition formulation and in the subsequent section our solver. We present extensive experimental results to test the limitations of our approach first on artificial data where the ground truth is known and then fMRI data. Fi-
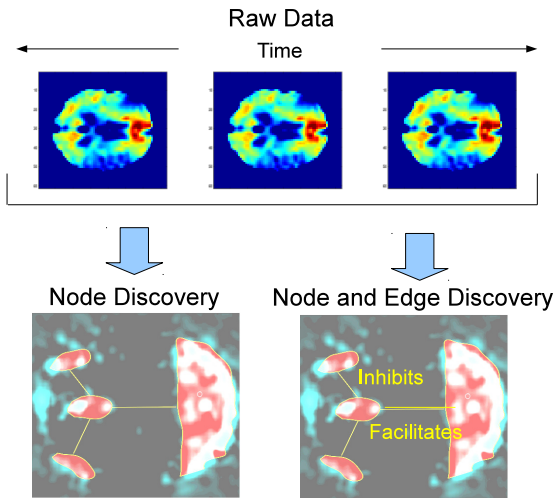
Figure 1: Our aim is to take spatio-temporal data and simultaneously discover nodes and edges to uncover the underlying network the person is using.

nally we discuss related work and then conclude our work. We use the typical notation [12] where ∘ is the tensor product and ⊙ the Khatri-Rao product.

2. PROBLEMS OVERVIEW

We now discuss our proposed contributions at the functional level and go into more detail with respect to formulations and solvers in the next two sections. fMRI is a predominant method for capturing brain activity as it processes information. We view fMRI data as containing a complex interaction of signals and noise [9, 23] with a natural question being to simplify the activity into the underlying cognitive network being used. However this problem is difficult for a number of reasons. Consider the Default Mode Network (DMN)[16] shown in Figure 2 (middle) of a person in rest state and the Blood Oxygenation Level Dependent (BOLD) measurement of the DMN structure centers shown in Figure 2 (right) plus one measurement outside the network in cyan. Not only does the out-of-network background-noise have **greater** intensity than the network signal, each fMRI scan contains over one quarter million such sequences and finding correlated pairs of voxels (pixels in the fMRI images) does not simplify the data to an interpretable level. Rather we wish to state a region of voxels (what we will call a node) form a structure and elicit the information pathways and other relationships between them (what we call edges). However, there could exist many explanations of the data and hence much chance for finding interactions that are superfluous, not interesting or even not anatomically possible.

Existing mining and learning algorithms (including our own previous work) do not directly discover networks and have several other key limitations: 1) They simplify the spatial temporal data by measuring an arbitrarily chosen correlation between the voxels' time sequences, 2) They serialize complex problems in an ad hoc fashion and 3) They cannot easily encode existing domain expertise. Our proposed constrained tensor formulation offers the ability to overcome these limitations.

2.1 Node Discovery Problem

This problem can take on a supervised, semi-supervised and unsupervised setting though not in the traditional meaning in our field. In the supervised setting, the network to discover involves coordinated activity among **some combination** of given anatomical structures (their complete geometric boundaries are given) such as a subset of those shown in Figure 2 (left). Since all possible nodes are given along with their boundaries, this is a termed a supervised problem. In the data we will perform our experiments on a canonical anatomical coordinate frame in which the boundaries of 116 known anatomical regions have been defined. In the semi-supervised setting some existing nodes (including their anatomical boundaries) are given, and other nodes whose boundaries are unknown must be discovered to form the completed network. This is desirable since in some demented and injured individuals the brain forms variations of networks. Finally, in the most difficult unsupervised setting, all nodes that are part of the network (the anatomical structures with their precise boundaries) must be discovered. None of these problems have been "solved" for fMRI data and there is much room for approaches that are both computationally efficient and theoretically well understood.

Table 1 shows these types of problems, the current *standard* approaches and applications. As can be seen these approaches are not designed to directly discover networks, but instead the data modified and the output post-processed. For example in our own earlier work [21, 23] we used Pearson correlation to convert the tensor into a 2-Graph and then must visually search amongst the cuts to find the one that closest to the DMN.

2.2 Edge Discovery Problem

Here our aim is to create edges between the nodes. The edges can represent a number of properties between nodes such as synchronized activity, inhibition (decreases activity), facilitation (increases activity), or preceding/succeeding (activity occurs before/after), with the weight encoding the strength of the relationship. Some existing work makes use of hierarchical clustering approaches [10] to first group voxels into nodes, and then cluster the nodes into networks representing synchronized activity. However, this does not allow encoding anatomical knowledge on which regions do not activate in the same network or a graph with heterogeneous edges. More recent innovative work looks at Dynamic Belief Networks [4] and covariance matrix estimation [19] to discover just one type of edge. Neither work seems to be easily extendable to also simultaneously discover nodes.

2.3 Network Verification

Finally, we will explore the important problem of attaching a strength associated with the network. This problem is particularly important in neuroscience problems since networks typically exists in all individuals but to varying degrees/parts. Consider the DMN which will be the focus of our experimental fMRI work. It is known to exist in even the most demented individual, but only in a partial form. Furthermore, it is known that the strength of the network (the activations) varies over individuals being least strong in elderly individuals.
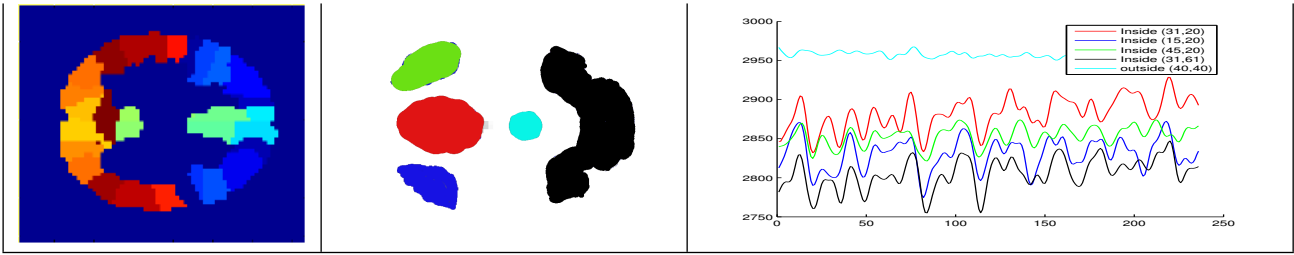
Figure 2: Left: Anatomical structures/nodes in the brain at slice 36. Middle: The default mode network (DMN). Right: Color-coded activation levels and a fifth part (in cyan) outside the DMN.

Problem Setting	Standard Techniques	Uses
Supervised: Given a set of all anatomical regions and their spatial delineations (a node collection), identify active nodes	ICA, Dynamic belief networks	Discover network nodes and assess their strength and variability.
Semi-supervised: Given a node collection, discover additional nodes to form a network	ICA with visual inspection	Assess adaptation of functional networks after focal injury (i.e. traumatic brain injury).
Unsupervised: Discover arbitrary shaped nodes	ICA, tensor and clustering approaches	Discovery of new brain functional networks, brain diseases, or treatments / interventions.

Table 1: The three node discovery problems, standard techniques and applications. We shall focus on the first two in this paper that require guidance. The related work section describes additional emerging techniques.

3. TENSOR DECOMPOSITION WITH TEMPLATE/PATTERN CONSTRAINTS

All visualization of results in this paper are presented as analyzing a slice of the brain (dimensions a and b) over time t so they can be easily visualized but all results of course are easily generalized to high order tensors. *If needed $\mathbf{a_i}$ and $\mathbf{b_i}$ can be vectorized to represent any arbitrary shaped region.*

For illustrative purposes and without loss of generality let $\mathcal{X}$ (2D space × time) be a three-mode tensor representing the fMRI data for a mid-level slice of the brain. We shall decompose (simplify) this tensor into f factors ($\hat{\mathcal{X}} = \mathcal{X}_1 + \mathcal{X}_2 \ldots + \mathcal{X}_f$) using a PARAFAC model (though more complex decompositions could be used). Let factor i be defined by the outer product of three *factor vectors* ($\mathcal{X}_i = \mathbf{a_i} \circ \mathbf{b_i} \circ \mathbf{t_i}$) and for brevity we write $\hat{\mathcal{X}} = A \circ B \circ T$ with the factor vectors being stacked column-wise so that each factor matrix has f columns. Then $\mathcal{X}_i$ can potentially represent a region of the brain with the outer product of $\mathbf{a_i}$ and $\mathbf{b_i}$ being the active region (i.e. a single area shown in Figure 2 left) and $\mathbf{t_i}$ their activations over time (right hand side of Figure 2). However, with unconstrained tensor decomposition $\mathcal{X}_i$ is typically not a spatially contiguous region nor does it necessarily match an anatomical region. To achieve this guidance is introduced.

We propose a constrained tensor decomposition where the objective function is complemented by adding constraints as shown in equation 1. The addition of guidance help rule out solutions that are non-actionable by restricting them to be consistent with known domain knowledge and expectations. In section 4 we discuss our solvers to address this new variation of tensor decomposition.

3.1 Node Discovery

It is worth noting that unconstrained tensor decomposition results for node discovery are poor for fMRI data since, for example, many spatially adjacent voxels in the same structure are not active in the same factor which is anatomically not possible. Pre-processing the tensor by applying wavelets [2] can alleviate this and could complement our work, though in practice this pre-processing was time intensive and yielded only marginally better results.

Supervised. Here the potential groups/nodes/structures of the brain have been identified. A group is collection of voxels/activations that are known to behave cohesively and are defined by a matrix/mask with all given matrices being $Q_1 \ldots Q_m$. Such matrices can be used to represent predefined regions that will comprise the possible nodes in the network and in our work will consist of the 116 known anatomical regions in the brain. Examples of the Q matrices used in this work can be seen in Figure 2 (left) with one matrix for each anatomical region. However, most networks only consist of a small number of nodes so we wish to use only a subset of Q matrices in the decomposition. We can encode which structures/nodes/matrices are to be used with a vector $\mathbf{w}$ with one component/entry per factor. We can represent how closely the discovered factors match these matrices with some deviation allowed which is upper-bounded by ϵ which is proportional to the number of voxels outside the given group. The formulation for supervised node discovery is in equation 1.

$$\underset{\mathbf{w},A,B,T}{argmin} \quad ||\mathcal{X} - \sum_i w_i \mathbf{a_i} \circ \mathbf{b_i} \circ \mathbf{t_i}||_F + ||\mathbf{w}||$$

subject to

$$w_1(\mathbf{a_1}\mathbf{b_1}^T - Q_1)|| \le \epsilon_1$$
$$\vdots$$
$$w_m||(\mathbf{a_m}\mathbf{b_m}^T - Q_m)|| \le \epsilon_m$$

$$(1)$$

The output of this computation will be the smallest set of groups/nodes (defined by $\mathbf{a_i} \circ \mathbf{b_i}$) that best summarize the
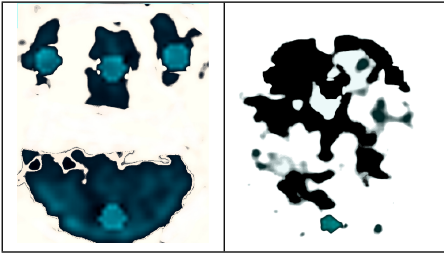
Figure 3: Supervised Network Discovery. For a healthy person (left) and a demented person (right) the top four nodes of the network found amongst the possible 116 nodes/structures. Compare with Figure 2 (left).
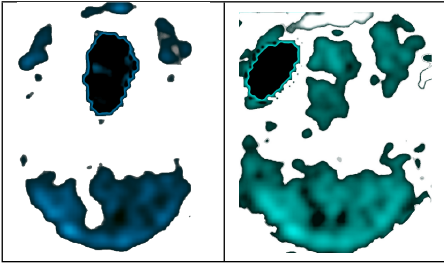

Figure 4: Semi-supervised Network Discovery Problem: Given just one node (highlighted and colored black) in the network what are the other nodes of the networks? In both cases the remaining three parts of the DMN are found. Compare with Figure 2 (left).

fMRI scan with $\mathbf{t_i}$ stating the activations over time. The penalty term $||\mathbf{w}||_1$ introduces a sparsity constraint to enforce that the simplest structure is discovered. Furthermore, by rank ordering the factors by their entry in $\mathbf{w}$ we can determine the most important nodes in the network. Figure 3 (left) shows an example of our work where the Q matrices are just of **the cores** of the anatomical structures. The top four most important nodes as per the $\mathbf{w}$ matrix and the corresponding Q matrices are shown and are the DMN. An important negative result is that the DMN is **not typically discovered** for demented individual (Figure 3 right).

Semi-Supervised Network Discovery. This is a specialization of the above problem with the same objective function except the problem is "relaxed" in that there are more factors than there are **used** Q matrices. This is particularly useful in the setting when we know one node in the network, but not the others. In particular, we wish to find a f node network, but require at most r of these to be from the given nodes (Q matrices). We can achieve this by placing an *additional* constraint such as $||\mathbf{w^{1\cdots r}}||_F \leq r$ (where $\mathbf{w^{1\cdots r}}$ masks only the first r factors which are the only ones with constraints). In this way the discovered network is the most simplest combination of given nodes and discovered nodes that match the brain activity. Figure 4 shows some results where we specify only one part of the network (shown in blacked and highlighted) and the tensor decomposition fills in the remaining parts of the network successfully.

3.2 Adding Simultaneous Edge Discovery

The limitation with the previous mentioned formulations is that it discovers only *active* groups/nodes but they need not have anything in common regarding activation (i.e. in the temporal aspect). A simple way to infer edges between nodes is to post-process and examine the respective $\mathbf{t}$ vectors and determine if they are similar using some sort of distance metric. However, a more elegant way is to simultaneously discover nodes and edges. We can change the above formulation to **discover interactions** between groups with similar activation levels by adding the constraint in equation 2 to discover 'co-occur' edges and the constraint in equation 3 to discover 'inhibit' edges. It is easy to encode other types of constraints to represent other types of edges, but the computational challenge we will defer to future work is to discover multiple types of edges simultaneously. In this paper we focus on only finding one edge type at a time.

$$subject\,to \quad \sum_{i,j} w_i.w_j ||\mathbf{t_i} - \mathbf{t_j}|| \leq \epsilon \qquad (2)$$

$$subject\,to \quad \sum_{i,j,i\neq j} w_i.w_j ||\mathbf{t_i}^T.\mathbf{t_j}|| \leq \epsilon \qquad (3)$$

4. SOLVERS

Our constrained decomposition formulations are not solvable by the two popular existing tensor toolkits by Kolda (MATLAB tensor toolbox) and Bro (nway). A variety of approaches to solve tensor decomposition exists [8] and we chose to explore alternating least squares (ALS) because it is known to be resilient method for tensor decompositions that is easy to interpret [12] and when compared to six other competing techniques was the best performing in terms of decomposition error [8]. Our experimental results support this conclusion and we find that ALS does not take appreciatively more time to converge with constraints than when there are no constraints. The addition of constraints to the tensor decomposition requires a more complex algorithm shown in Figure 5. This is an ALS algorithm but at each step to update each of $\mathbf{A}, \mathbf{B}, \mathbf{T}$ and $\mathbf{w}$ is now a constrained optimization problem. We describe two methods of addressing this and choose the latter since it scales better.

4.1 Solving Each Sub-Problem in Figure 5

Without loss of generality consider the sub-problem presented in Equation 4:

$$\underset{A}{\arg\min} \quad ||X_1 - AK^T||_F$$

subject to

$$w_1||(\mathbf{a_1}\mathbf{b_1}^T - Q_1)|| \leq \epsilon_1 \qquad (4)$$

$$\vdots$$

$$w_m||(\mathbf{a_m}\mathbf{b_m}^T - Q_m)|| \leq \epsilon_m$$

$$K = (T \odot B \odot w)$$

We can solve the sub-problem directly by simply plugging the problem directly into `cvx`. However, in practice `cvx` does not scale to handle problems with large number of constraints as we require. Instead of solving for all columns of

Constrained PARAFAC Alternating Least Squares
Input:
$\mathcal{X}$: The tensor to decompose.
α: The minimum change in error.
$Q_i, i = 1, \ldots, m$: Spatial patterns.
Output: w, A, B, T

1. Calculate matricizations X_A, X_B, X_T, X_w

2. Solve and set $A =$

$$\underset{A}{\arg\min} \quad ||X_A - A(T \odot B \odot w)^T||_F$$

$$\text{subject to} \quad w_i||(a_i b_i^T - Q_i)|| \leq \epsilon_i; i = 1, \ldots, m$$

3. Solve and set $B =$

$$\underset{B}{\arg\min} \quad ||X_B - B(T \odot A \odot w)^T||_F$$

$$\text{subject to} \quad w_i||(a_i b_i^T - Q_i)|| \leq \epsilon_i; i = 1, \ldots, m$$

4. Solve and set $T =$

$$\underset{T}{\arg\min} \quad ||X_T - T(B \odot A \odot w)^T||_F$$

5. Solve and set $w =$

$$\underset{w}{\arg\min} \quad ||X_w - w(T \odot B \odot A)^T||_F + ||w||_1$$

$$\text{subject to} \quad w \geq 0$$

6. $\Delta||\mathcal{X} - A \odot B \odot T|| < \alpha$

    (a) Not Satisfied: Goto step 2.
    (b) Satisfied: **return** w, A, B, T.

Figure 5: The constrained ALS algorithm for solving the canonical decomposition. The algorithm is shown for an order three tensor but is easily changed to an arbitrary order tensor.

A, **B**, **T** or **w** simultaneously, we can solve for each column of say **A** separately. To solve for column i of **A** we first calculate the residuals left over from subtracting the effects of other columns in **A** then solve least squares problems using the residuals (constrained only by those constraints on column i) rather than the original tensor as shown below. This change will not affect the quality of solution, rather it just creates not k sub-problems (if the tensor to decompose is k dimensional) but rather kf where f is the number of factors in our decomposition. It is faster because though there are more sub-problems there are far **fewer constraints per sub-problem.** Formally:

$$\mathbf{a_i} = \underset{\mathbf{a}}{\arg\min} \quad ||R - \mathbf{a}\mathbf{k_i}^T||_F$$

$$\text{subject to} \quad w_i||(\mathbf{a}\mathbf{b_i}^T - Q_i)|| \leq \epsilon_i \tag{5}$$

$$k_i = (t_i \odot b_i \odot w_i)$$

5. EXPERIMENTS

We present two sets of experimental results in this section. The first on artificial data allows us to perform a series of controlled experiments to better understand the strengths and limitations of our work. The second is on real fMRI data of healthy and demented (Alzheimer's) individuals at rest and allows us to test if the assumptions our work makes are realistic and applicable to this problem.

5.1 Artificial Data With Known Ground Truth

In this section we attempt to answer three core questions to better understand our work:

1. How does the approach perform with incomplete guidance (i.e. the core of the underlying nodes are given)?

2. Can the approach handle incorrect guidance, that is, if wrong nodes are given, can the approach ignore them?

3. How robust is the approach to noise. Can networks corrupted by noise still be recovered given the ground truth nodes?

Another important question to address is what types of networks can be discovered in terms of geometric shapes and number of such shapes. Since we can vectorize any matrix any arbitrary shape can be represented, but to demonstrate our claim the ground truth networks we will use will be: all squares, all ellipses and then a combination of squares and ellipses. In all experiments there is a Q matrix for each node in the network given unless otherwise specified.

Complete versus Partial Guidance. To test this issue we generate twenty ground truth networks with an example ground truth network we use shown in Figure 6. To facilitate partial guidance we give our algorithm only a small central fraction/proportion of the node itself (encoded in the Q matrices) and compare the difference between the ground truth network and the network we discover. Figure 7 shows the results as the fraction of the nodes given to the algorithm increases. The error is reported as a percentage of the total possible error where an error occurs if a voxel/pixel in the ground truth network is not in the discovered network or vice versa. See section 5.2 for details. The graph can be read as follows, the x-axis value tells us how much of the underlying ground truth network is revealed to the algorithm whilst the y-axis tells us how much of the network is never retrieved or incorrectly retrieved. We see an important trend that as the proportion of the node provided increases not only does the error decrease, but also the variance of the error.

Ignoring Incorrect Guidance. To test this issue we generate twenty ground truth networks with an example ground truth network we use shown in Figure 8. To simulate giving incorrect guidance we start with one superfluous and irrelevant node (i.e. the node is not present in the ground truth network) and anticipate our work will ignore it and steadily increase the number of false nodes that the algorithm has available to choose from. We then compare the difference between the ground truth network and the network we discover (see section 5.2 for how). Figure 9 shows that as the number of false/incorrect nodes given to the algorithm increases the error does increase, but only minimally.

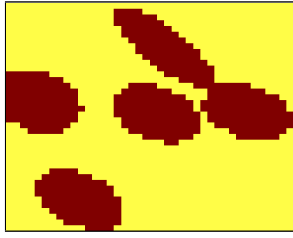All code will be freely available at `www.cs.ucdavis.edu/~davidson`

Figure 6: An example ground truth network for our experiments on network discovery with varying amounts of guidance.
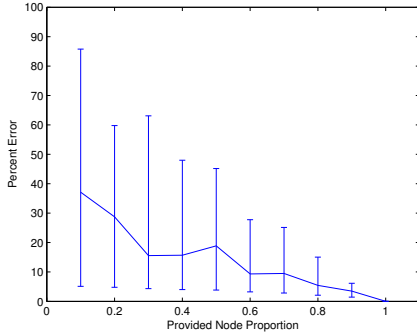

Figure 7: Plot of the results for our experiments on network discovery with varying amounts of guidance.
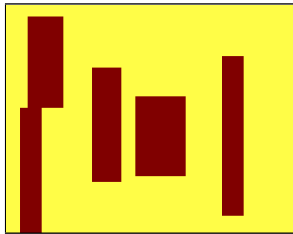

Figure 8: An example ground truth network for our experiments on network discovery with varying amounts of false nodes given.
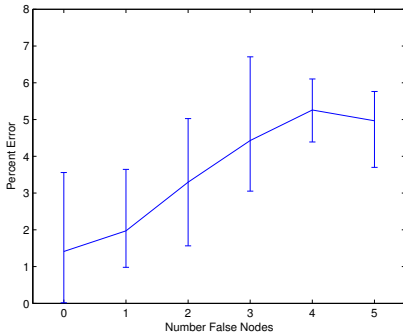

Figure 9: Plot of our experimental results on network discovery with varying amounts of false nodes given.
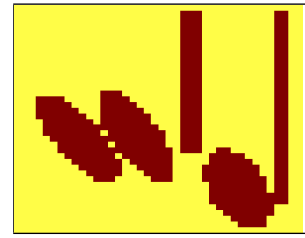

Figure 10: An example ground truth network for our experiments on network discovery with varying amounts of noise.
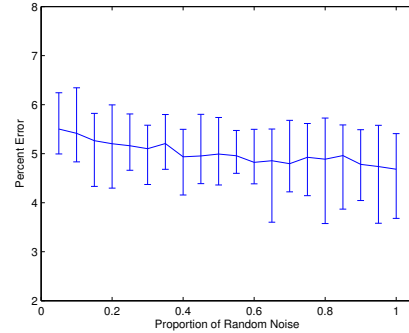

Figure 11: Plot of the results for our experiments on network discovery with varying amounts of noise.

Robustness to Noise. This is perhaps the most important result since most guidance will be in some idealized form whilst the observed data will inherently be flawed. An example ground truth network we use is shown in Figure 10 and as before we generate twenty such networks to test the performance of our algorithm. As mentioned, the idealized nodes are typical, but the observed activity will contain noise. To facilitate noise we take the idealized nodes and perturb them by adding noise uniformly throughout the tensor. Figure 11 shows the results as the amount of the noise given to the algorithm increases. Error is measured as before and discussed in section 5.2. The total magnitude of uniformly sampled random noise added is reported with respect to the total network size. The results show that our method is relatively unaffected by noise, but that somewhat surprisingly the method does slightly worse with a small amount of noise as compared to a large amount of noise. We believe this is due to the fact that a small amount of uniformly sampled noise appears less like uniform noise than a larger sample and that our method is **more** likely to try to explain the noise in the smaller samples.

5.2 Evaluation Measure

To evaluate the quality of nodes discovered we compare learned nodes (N) to the ground truth nodes (Q^*) using reconstruction error. We pair the learned nodes with the true nodes by picking the assignment (M) that minimizes the total reconstruction error across all nodes. This is summarized in equation 6 and is an assignment problem easily solvable

in MATLAB in polynomial time.

$$error = \min_{M_1...M_k} \sum_{i=1}^{k} ||N_i - Q_{M_i}^*||_F \quad (6)$$

5.3 Experiments with fMRI

We present experimental results on fMRI data of eight healthy and eight demented (Alzheimer's) individuals at rest. Each individual has been interviewed and measured using a series of cognitive tests to measure Episodic, Executive, Semantic and Spatial scores (whose range is from -2.5 to +2.5) and are categorized as `Normal` or having full-set Alzheimer's (`Demented`). When at rest state, an individual's brain activity exhibits the default mode network (DMN) shown in Figure 2 (middle) in some form and in various degrees in all people. In normal individuals it is expected to be fully intact and well exhibited whilst for demented individuals it may be only partially formed and weak. These insights are well known and extensively published in the literature [10] and we expect our method to be able to verify these results. The ability to determine the strength of the DMN from the scan is then akin to being able to predict the progression of Alzheimer's which we show is possible by predicting the cognitive scores (see Table 5).

Experiments To Discover The Nodes in a Network. Here we take all sixteen scans and attempt to discover the DMN under two settings. The DMN consists of four nodes and hence in the completely supervised setting we expect the top four most important nodes (according to **w**) to be the DMN masks for the normal people and less so with the demented individuals. We provided our algorithm with all 116 anatomical regions/masks of the brain encoded each in its own Q matrix and performed a tensor decomposition and selected the top four factors (according to the **w** vector) to determine which nodes/structures were active in the brain during the scan. Table 2 (columns 2 and 3) shows the fraction of the various nodes/parts of the DMN found for Normal and Demented people in the top four factors. We clearly see that the DMN is completely discovered in seven out of the eight healthy individuals with the eighth individual's prefrontal region being ranked fifth according to **w**. For demented individuals the DMN in its entirety is only found twice in the eight patients.

Figure 12 shows the actual network reconstructed from the top four factors for some healthy individuals and Figure 13 for the demented individuals. These plots are created using a reconstructed tensor from the first four factors $\hat{\mathcal{X}} = \mathcal{X}_1 + \mathcal{X}_2 + \mathcal{X}_3 + \mathcal{X}_4$. Table 2 (columns four and five) shows our results where we repeat the above experiments except in a semi-supervised setting. For each scan, we decomposed the tensor using four factors but only provided three nodes from the DMN as guidance and required the algorithm to discover the boundaries of the fourth missing node. This produced four experiments per scan since each of the nodes was left out in turn.

Experiments For Edge Detection. Here we present results on edge detection by adding in the additional constraint shown in equation 2 to supervised network discovery. These experiments reuse the experimental settings used to produce the supervised results in the previous sub-section. Our results (see Table 3) show that for healthy individuals, the connectivity between the four parts of the DMN are close to one. On only one occasion was the Prefrontal region not

Table 2: Supervised and Semi-Supervised Setting: Fraction of times parts of DMN are in top four factors (according to w) for various sub-populations. 116 possible nodes were given to the algorithm (see left column of Table 3 for some examples). There are eight scans per sub-population.

Struc-ture	Super-vised Normal	Super-vised Demented	Semi Sup-ervised Normal	Semi Sup-ervised Demented
Prefr.	88%	50%	71%	31%
Post. C.	100%	63%	94%	50%
Inf. P.	100%	38%	91%	28%
Med. T.	100%	25%	97%	25%

Table 3: Mean Pearson correlation for healthy individuals for top four nodes of a network discovered in each individual. The nodes/structures denoted by "*" are part of the DMN. There are no entries for nodes 6,7 and 8 since they never appear in the top four factors.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Prefront*	1.0	0.87	0.91	0.94	0.63			
(2) Post. Cin.*	0.87	1.0	0.93	0.89	0.45			
(3) Inf. Par.*	0.91	0.93	1.0	0.84	0.3			
(4) Med.Tem.*	0.94	0.89	0.84	1.0	0.1			
(5) Hypothal.	0.63	0.45	0.3	0.1	1.0			
(6) Pallium								
(7) Hippo.								
(8) Basal Gan.								

part of the network induced, instead the Hypothalamus was found to be active in one individual. For demented individuals (see Table 4), the correlation between the top four nodes (according to **w**) spanned eight different nodes and we see that the DMN was the most strongest interconnected nodes but was rarely found in its entirety.

Experiments for Network Verification. Discovering networks is an important problem, but it is also important to attach a measure of strength to the network. We explore the idea of using the reconstruction error $||\mathcal{X} - \sum_f \mathcal{X}_f||$ as such a measure both by itself and with other information. We expect that the reconstruction error of the data given the ground truth is inversely related to the strength of the network. We can verify this claim by noting that cognitive scores can be viewed as a surrogate of the strength of the DMN [16] and see if we can predict the cognitive scores from the reconstruction error. Furthermore, we can compare how well the reconstruction error is predictive of the network strength by noting that cognitive score are easily predictive by three clinical pieces of data (age, education level and gender) [16] using linear regression. We also tried state of the art methods to predict cognitive scores from fMRI data [22, 18]. Table 5 shows the predicted cognitive scores using the clinical information, just the reconstruction error, the reconstruction error and clinical information and the two state of the art methods that make use of machine learning from features of the scan (rather than the entire scan as we
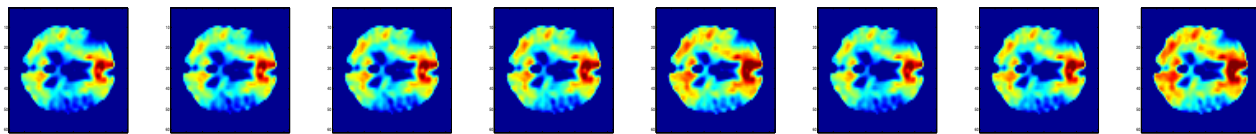
Figure 12: The typical networks (red and yellow) discovered using the top four factors for the eight healthy individuals. Best viewed in color.
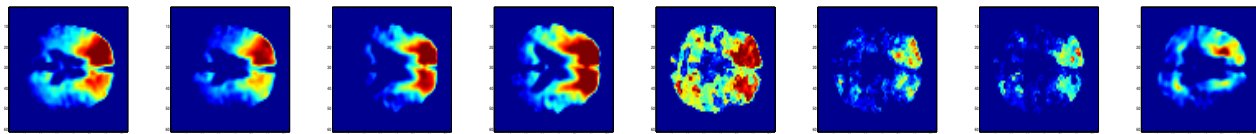

Figure 13: The typical networks (red and yellow) discovered using the top four factors for the eight demented individuals. Best viewed in color.

Table 4: Mean Pearson correlation for demented individuals for top four nodes of a network discovered in each individual. The nodes/structures denoted by "*" are part of the DMN.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Prefront*	1.0	0.45	0.52	0.31	0.19	0.32	0.45	0.17
(2) Post. Cin.*	0.45	1.0	0.14	0.21	0.56	0.11	0.19	0.1
(3) Inf. Par.*	0.52	0.14	1.0	0.44	0.13	0.17	0.05	0.1
(4) Med.Tem.*	0.31	0.21	0.44	1.0	0.1	0.05	0.01	0.0
(5) Hypothal.	0.19	0.56	0.13	0.1	1.0	0.12	0.11	0.21
(6) Pallium.	0.32	0.11	0.17	0.05	0.12	1.0	0.15	0.31
(7) Hippo.	0.45	0.19	0.05	0.01	0.11	0.15	1.0	0.15
(8) Basal Gan.	0.17	0.1	0.1	0.0	0.21	0.31	0.15	1.0

Table 5: Mean residual error for each set of predictors and several state of the art methods of predicting cognitive scores from image data.

Test	Clin-ical	Decomp. Error	Clinical+ Decomp.	[22]	[18]
Episodic	1.00	0.27	0.19	0.36	0.45
Executive	1.03	0.26	0.22	0.41	0.39
Semantic	1.01	0.20	0.20	0.38	0.41
Spatial	0.99	0.24	0.26	0.35	0.43

do). We can immediately see that the reconstruction error is far better at predicting cognitive scores than the clinical information and together produces even better results.

6. RELATED WORK AND COMPARISON

Network Discovery. Most work on network discovery focuses on either node identification or edge identification but rarely both and to our knowledge not simultaneously discovering both as we propose. Furthermore, finding a network with heterogeneous edges to our knowledge has not been addressed. A variety of current data mining tools have been used to determine, in an unsupervised setting, clusterings of voxels (what we refer to as nodes/regions) that appear to exhibit relatively high connectivity or covary with each other in systematic ways [20][5][15]Another stream of research focuses on just identifying connectivity between given brain regions using dynamic belief networks [4], and using sparse inverse covariance estimation [19][11]. Some supervised learning work could be used to predict the existence of a network. Such work performs dimension reduction, followed by classification or regression, on the 4D matrix provided by BOLD fMRI [14] [7]. However, probing functional connectivity goes far beyond a yes or no question of whether two regions are functionally connected; activity in one region can lag, excite another region, or coordinate only in certain circumstances [13].

Tensors, Clustering and fMRI Data. There appears to be a considerable body of work on tensor analysis and fMRI data but much of this is due to naming conventions. The area of diffusion tensor analysis is misleadingly named and is in fact an imaging method rather than a method of analysis. A similar situation exists for tensor-based morphometry (TBM) and other developing imaging methods.

There has been work on fMRI analysis using tensor decomposition [17] but to our knowledge this work is not focused on inferring networks. Beckmann and collaborators [1] have extended ICA to factor in time series data and refer to this as a "Tensor" form of ICA but its contribution is really to make ICA parameter free. The novel work of combining wavelets and Tensors (TWave) [2] explores discovering clusters and classification (but not network discovery as described in this work) in fMRI data. The author's correctly point out that regular tensor decomposition methods such as PARAFAC are limited and will fail due to scalability issues and that they ignore the spatial locality requirements to analyzing brain imaging data. The authors pre-process the data using a wavelet analysis which helps scale and overcome spatial locality issues and then apply regular PARAFAC. In contrast our work overcomes the same limitations by introducing constraints. The constraints used in the supervised and semi-supervised setting enforce properties such as spatial continuity and the multi-mode level constraints ensure applicability to anatomical regions.

Computation times between our methods and these other tensor decomposition methods are comparable when pre-

processing of the tensors is also included. In practice decomposing a single fMRI scan ($60 \times 50 \times 70$ voxels) over 236 snapshots took less than a minute on an i7 - 4 core machine. This is directly due to the constraints that result in faster convergence due to limiting the search space.

7. CONCLUSION

We pose the problem of network discovery which involves discovering regions of cohesive behavior (the nodes) and the relationship between those regions (the edges). We pose three sub-problems: node discovery, edge discovery and network strength verification. We postulated a constrained tensor decomposition formulation for this problem and show that it readily facilies a supervised setting where the algorithm must chose a small subset of given nodes that best match the activity and a semi-supervised setting where the algorithm must chose a subset of given nodes and also discover new nodes. Such a constrained tensor decomposition formulation is not solvable using the existing toolkits and we pose our own algorithm which is a ALS formulation with each step being a constrained optimization readily solvable in MATLAB. We show that our work can discover networks in the presence of incomplete guidance (Figure 7), false/misleading nodes (Figure 9) and noise (Figure 11). This last result is particularly important. It shows that even if the guidance given is in an idealized form, it can be used to discover networks in noisy data.

We then explored fMRI data for two sub-populations: healthy and demented. We were able to show that our methods were able to discover the nodes in the DMN in healthy people with considerable regularity (see Table 2) as well as the connectivity between the nodes (see Table 3). We also explored the idea of using the reconstruction error as a measure of network strength and found that it was able to predict the well known surrogates for network strength, cognitive scores in far greater precision that well known clinical information (see Table 5).

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] C.F. Beckmann and S.M. Smith, Tensorial extensions of independent component analysis for multisubject FMRI analysis, NeuroImage 25 (2005) 294-311.

[2] M. Barnathan, V. Megalooikonomou, C. Faloutsos, S. Faro, F.B. Mohamed FB, TWave: high-order analysis of functional MRI, Neuroimage. 2011 537-48.

[3] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 10(3):186–198, 2009.

[4] J. Burge, T. Lane, H. Link, s. Qiu, and V. Clark, Discrete Dynamic Bayesian Network Analysis of fMRI Data, Human Brain Mapping 30 (2009).

[5] D. Cordes, V. Haughton, J. D. Carew, K. Arfanakis, and K. Maravilla. Hierarchical clustering to measure connectivity in fmri resting-state data. *Magnetic Resonance Imaging*, 20(4):305 – 317, 2002.

[6] N. A. Dennis, et. al. Temporal lobe functional activity and connectivity in young adult carriers. *Alzheimer's and Dementia*, 6(4):303 – 311, 2010.

[7] N. U. F. Dosenbach, et. al. Prediction of individual brain maturity using fmri. *Science*, 329(5997):1358–1361, 2010.

[8] N. K. M. Faber, R. Bro, and P. K. Hopke, Recent developments in CANDECOMP/PARAFAC algorithms: A critical review, Chemometrics and Intelligent Laboratory Systems, 65 (2003)

[9] C. Genovese, N. Lazar, T. Nichols, Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage, 15, 870-8, (2002).

[10] M. Greicius. Resting-state functional connectivity in neuropsychiatric disorders. *Curr Opin Neurol*, 21(4):424–430, 2008.

[11] S. Huang, J. Li, L. Sun, J. Liu, T. Wu, K. Chen, A. Fleisher, E. Reiman and J. Ye Learning Brain Connectivity, NIPS 2009.

[12] T. Kolda and B. Bader, Tensor Decompositions and Applications, SIAM Review 2008.

[13] P.-J. Lahaye, et. al. Functional connectivity: studying nonlinear, delayed interactions between bold signals. *NeuroImage*, 20(2):962 – 974, 2003.

[14] F. D. Martino, et. al. Classification of fmri independent components using ic-fingerprints and support vector machine classifiers. *NeuroImage*, 34(1):177 – 194, 2007.

[15] S. J. Peltier, T. A. Polk, and D. C. Noll. Detecting low-frequency functional connectivity in fmri using a self-organizing map (som) algorithm. *Human Brain Mapping*, 20(4):220–226, 2003.

[16] M. Raichle, A. Snyder, A default mode of brain function: a brief history of an evolving idea. Neuroimage, 37, 1083-1090. 2007.

[17] A. Stegman, Comparing independent component analysis and the PARAFAC model for artificial multi-subject fMRI data, Technical Report, University of Groningen, The Netherlands, Feb. 2007.

[18] C.M. Stonnington et. al., Predicting clinical scores from magnetic resonance scans in Alzheimer's disease, Neuroimage. 2010 Jul 15;51(4):1405-13.

[19] L. Sun, R. Patel, J. Liu, K. Chen, T. Wu, J. Li, E. Reiman, J. Ye, Mining Brain Region Connectivity for Alzheimer's Disease Study via Sparse Inverse Covariance Estimation, KDD 2009.

[20] V. G. van de Ven, et. al. Functional connectivity as revealed by spatial independent component analysis of fmri measurements during rest. *Human Brain Mapping*, 22(3):165–178, 2004.

[21] X. Wang, B. Qian, I. Davidson, Flexible Constrained Spectral Clustering: Algorithms and Applications, Journal of Knowledge Discovery and Data Mining (DMKD), November 2012.

[22] J. L. Woodard, et. al.,Prediction of Cognitive Decline in Healthy Older Adults using fMRI, Journal of Alzheimers Disease. 2010 January 1; 21(3).

[23] P. Walker, I. Davidson, Exploring new methodologies for the analysis of fMRI following closed-head Injuries. In D. D. Schmorrow, C. M. Fidopiastis (Eds.), Springer. (2011)