Adaptive Cuts for Extracting Specific White Matter Tracts

Nagesh Adluru, Vikas Singh*, Andrew L. Alexander

Waisman Center & Department of Medical Physics
Department of Biostatistics & Med. Informatics
Department of Computer Sciences
THE UNIVERSITY of WISCONSIN MADISON

IEEE International Symposium on Biomedical Imaging

From DTI to fiber pathways

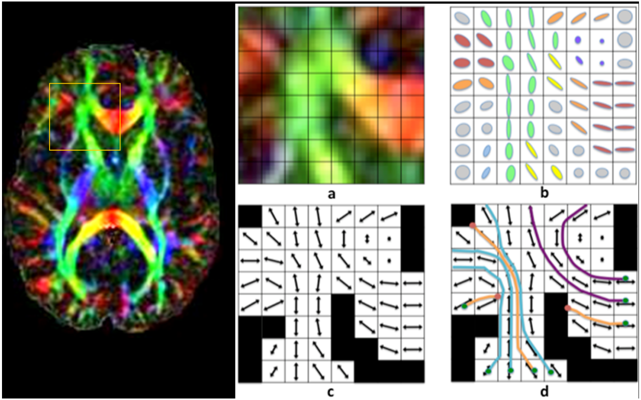


Figure: Stream line tractography: Integrate along the primary eigen vectors of the DTI (Alexander et al.).

From DTI to fiber pathways

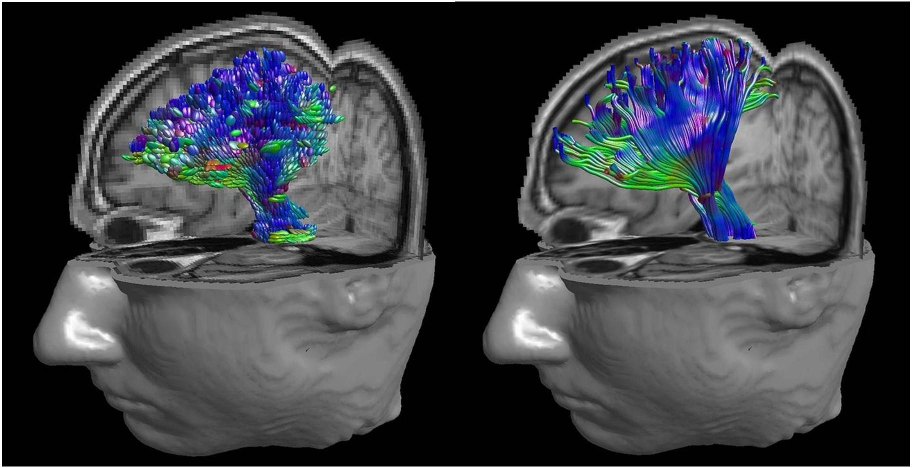


Figure: DTI and stream line tractography results showing the cortico-spinal tracts (Leemans et al. [Explore DTI]).

Goal: Extraction of Specific Tract Bundles

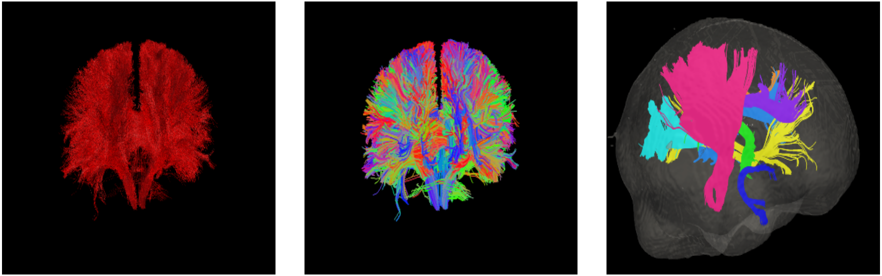


Figure: From whole brain tractography to specific white matter bundles

Project Motivation

1. Evaluate *a priori* hypotheses in white matter studies (Yushkevich et al. 2008, Kunimatsu et al. 2012, Samanez-Larkin 2012, Bendlin 2010)

Project Motivation

1. Evaluate *a priori* hypotheses in white matter studies (Yushkevich et al. 2008, Kunimatsu et al. 2012, Samanez-Larkin 2012, Bendlin 2010)
2. But specific tracts are generally identified manually in the population-specific atlas (Catani et al. 2008.) or modification of bundles from tract clustering (Odonnell et al. 2005.) — bundles correspond to specific bundles of interest?

Project Motivation

1. Evaluate *a priori* hypotheses in white matter studies (Yushkevich et al. 2008, Kunimatsu et al. 2012, Samanez-Larkin 2012, Bendlin 2010)
2. But specific tracts are generally identified manually in the population-specific atlas (Catani et al. 2008.) or modification of bundles from tract clustering (Odonnell et al. 2005.) — bundles correspond to specific bundles of interest?

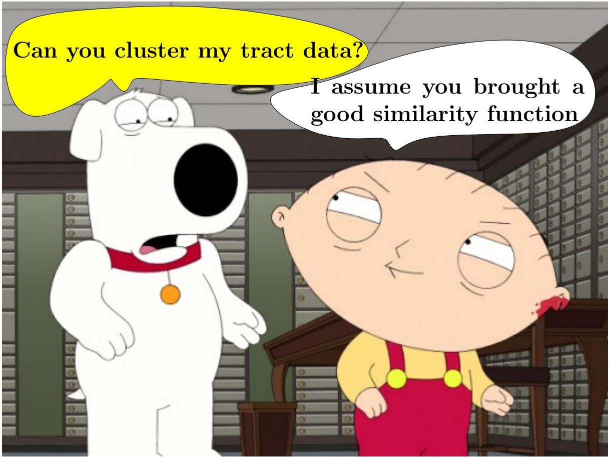
1. Tract clustering algorithms depend on some variations of the Hausdorff distance (Odonnell et al. 2006 and Odonnell et al. 2007) or *B*-splines (Maddah et al. 2005)

Project Motivation

1. Evaluate *a priori* hypotheses in white matter studies (Yushkevich et al. 2008, Kunimatsu et al. 2012, Samanez-Larkin 2012, Bendlin 2010)
2. But specific tracts are generally identified manually in the population-specific atlas (Catani et al. 2008.) or modification of bundles from tract clustering (Odonnell et al. 2005.) — bundles correspond to specific bundles of interest?

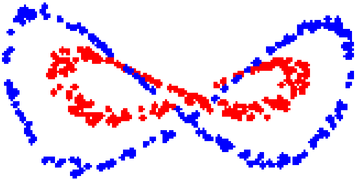
1. Tract clustering algorithms depend on some variations of the Hausdorff distance (Odonnell et al. 2006 and Odonnell et al. 2007) or *B*-splines (Maddah et al. 2005)
2. Meta similarities using path similarities in MST, shortest geodesic paths, divergences (Tsai et al. 2007, Wasserman et al. 2009) – what is a good similarity function?

The general practical issue with clustering



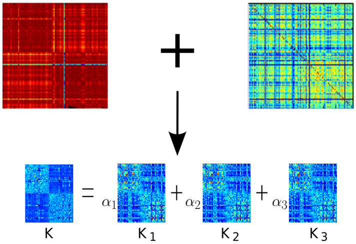
Outline

- Normalized Cuts



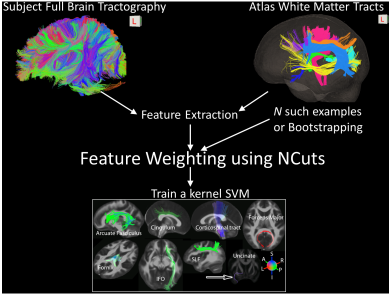
Outline

- Normalized Cuts
- Kernel Learning
  - Models
  - Optimization



Outline

- Normalized Cuts
- Kernel Learning
  - Models
  - Optimization
- Tract Clustering in Diffusion Tensor Images



Related Work – which clustering objective?

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. ?, NO. ?, ? 2007

Automatic Tractography Segmentation Using a High-Dimensional White Matter Atlas

Lauren J. O'Donnell and Carl-Fredrik Westin, *Member, IEEE*

***Abstract*— We propose a new white matter atlas creation method that learns a model of the common white matter structures present in a group of subjects. We demonstrate that our atlas creation method, which is based on group spectral clustering of tractography, discovers structures corresponding to expected white matter anatomy such as the corpus callosum, uncinate fasciculus, cingulum bundles, arcuate fasciculus, and corona radiata. The white matter clusters are augmented with expert anatomical labels and stored in a new type of atlas that we call a high-dimensional white matter atlas. We then show how to perform automatic segmentation of tractography from novel subjects by extending the spectral clustering solution, stored in the atlas, using the Nystrom method. We present results regarding the stability of our method and parameter choices. Finally we give results from an atlas creation and automatic segmentation experiment. We demonstrate that our automatic tractography**

Diffusion MRI can be used to create a representation of white matter tracts in the brain via a process called tractography [1], [2], [3], [4], [5] that estimates white matter tract trajectories by following likely tract directions. The two main tractography strategies are probabilistic (an attempt to describe all possible fiber trajectories [6], [7], [8]) and streamline (which locally chooses the most probable fiber trajectory [2], [3], [4], [5]). Tractography methods have been adapted to various representations of diffusion and fiber tract orientation data, including diffusion tensor MRI (DTI), q-ball [9], and multiple tensor models.

The most common combination of diffusion data and fiber tractography method today is DTI and streamline tractography.

Related Work – which clustering objective?

CONCEPTS, INNOVATIONS, AND TECHNIQUES

Interactive Diffusion Tensor Tractography Visualization for Neurosurgical Planning

Alexandra J. Golby, MD*†
Gordon Kindlmann, PhD†
Isaiah Norton, BS*
Alexander Yarmarkovich, PhD†
Steven Pieper, PhD†
Ron Kikinis, MD†

Brigham and Women's Hospital, Departments of *Neurosurgery and †Radiology, Harvard Medical School, Boston, Massachusetts

Correspondence:
Alexandra J. Golby, MD,
Department of Neurosurgery,
Brigham and Women's Hospital,
75 Francis St,
Boston, MA 02115.
E-mail: agolby@bwh.harvard.edu

Received, January 20, 2010.
Accepted, March 5, 2010.

BACKGROUND: Diffusion tensor imaging (DTI) infers the trajectory and location of large white matter tracts by measuring the anisotropic diffusion of water. DTI data may then be analyzed and presented as tractography for visualization of the tracts in 3 dimensions. Despite the important information contained in tractography images, usefulness for neurosurgical planning has been limited by the inability to define which are critical structures within the mass of demonstrated fibers and to clarify their relationship to the tumor.

OBJECTIVE: To develop a method to allow the interactive querying of tractography data sets for surgical planning and to provide a working software package for the research community.

METHODS: The tool was implemented within an open source software project. Echo-planar DTI at 3 T was performed on 5 patients, followed by tensor calculation. Software was developed that allowed the placement of a dynamic seed point for local selection of fibers and for fiber display around a segmented structure, both with tunable parameters. A neurosurgeon was trained in the use of software in < 1 hour and used it to review cases.

RESULTS: Tracts near tumor and critical structures were interactively visualized in 3 dimensions to determine spatial relationships to lesion. Tracts were selected using 3 methods: anatomical and functional magnetic resonance imaging-defined regions of interest, distance from the segmented tumor volume, and dynamic seed-point spheres.

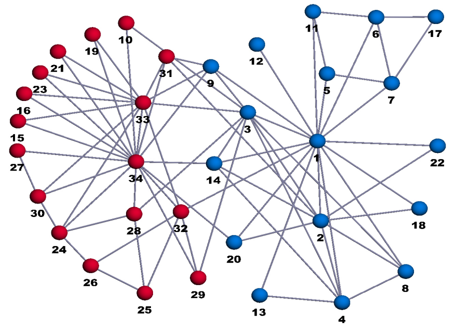
CONCLUSION: Interactive tractography successfully enabled inspection of white matter structures that were in proximity to lesions, critical structures, and functional cortical areas, allowing the surgeon to explore the relationships between them.

KEY WORDS: Diffusion tensor imaging, Magnetic resonance imaging, Neurosurgery, Surgical planning, Tractography

Neurosurgery 68:496–505, 2011 DOI: 10.1227/NEU.0b013e3182061ebb www.neurosurgery-online.com

Partitining Graphs

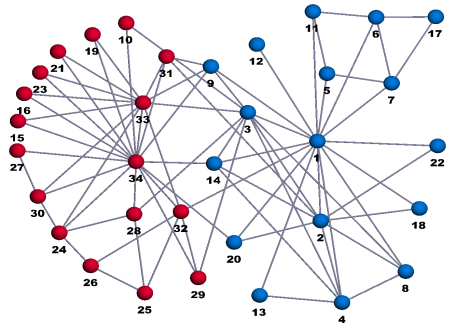
Zachary's Karate Club (Temple University)



34 members of a karate club in the 1970s.

Partitining Graphs

Zachary's Karate Club (Temple University)

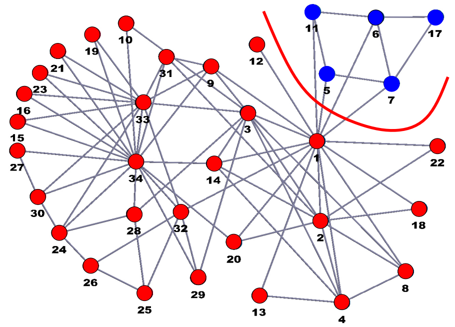


34 members of a karate club in the 1970s.

Partition a graph into **two** disjoint partitions such that the number of discarded edges is small.

Partitining Graphs

Zachary's Karate Club (Temple University)

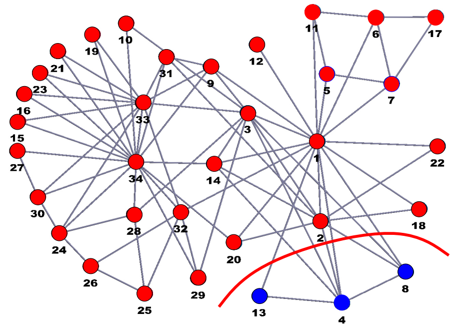


34 members of a karate club in the 1970s.

Partition a graph into **two** disjoint partitions such that the number of discarded edges is small.

Partitining Graphs

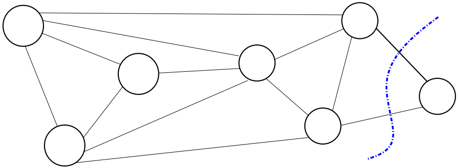
Zachary's Karate Club (Temple University)



34 members of a karate club in the 1970s.

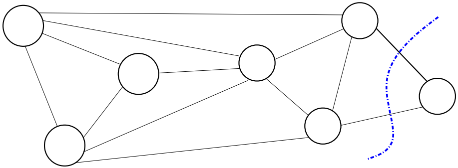
Partition a graph into **two** disjoint partitions such that the number of discarded edges is small.

The Normalized Cuts argument (Shi and Malik, 2000)



"shrinking": simply minimizing the sum of edge weights could put a single node in one cluster

The Normalized Cuts argument (Shi and Malik, 2000)



"shrinking": simply minimizing the sum of edge weights could put a single node in one cluster

The Fix: Ask for Balanced Partitions

Minimize the cut of the graph normalized by the weight of the partitions,

$$\frac{\text{Cut}(\mathcal{V}_1, \mathcal{V}_2)}{\text{Weight}(\mathcal{V}_1)} + \frac{\text{Cut}(\mathcal{V}_2, \mathcal{V}_1)}{\text{Weight}(\mathcal{V}_2)} \tag{1}$$

Generalized Version of Normalized Cuts

- Let V be the set of all vertices
- Let $\mathcal{V}_i$ be the i-th cluster

$$\text{Cut}\left(\mathcal{V}_1, \cdots, \mathcal{V}_k\right) = \sum_i \frac{\text{Weight}(\mathcal{V}_i, \bar{\mathcal{V}}_i)}{\text{Weight}(\mathcal{V}_i, V)} \tag{2}$$

$$= \sum_i \frac{\mathbf{x}_i^T (D - W)\mathbf{x}_i}{\mathbf{x}_i^T (D)\mathbf{x}_i} \tag{3}$$

$$= \sum_i \left(\mathbf{X}^T (D - W)\mathbf{X}\right)^{-1} \left(\mathbf{X}^T (D)\mathbf{X}\right) \tag{4}$$

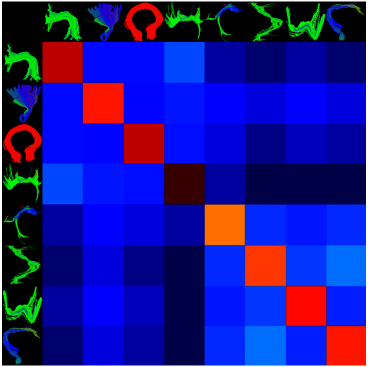
W is a weight matrix, and D is degree matrix (sum of all edges of a node).

Solve the **eigen-value** problem:

$$(D - W)\mathbf{y} = \lambda D\mathbf{y}$$

The story so far . . .

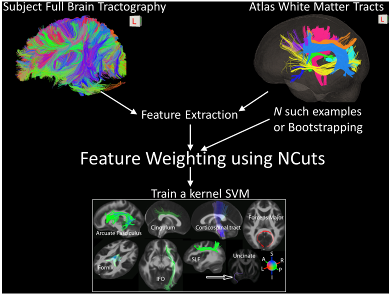
"Give me a good affinity matrix, W and the model will provide a good clustering solution."



White Matter Bundles from the Pittsburgh Brain Competition

Outline

- ~~Normalized Cuts~~
- Kernel Learning
  - Motivation
  - Models
  - Optimization
- Tract Clustering in Diffusion Tensor Images



The Learning Version of the Problem (Shortreed 2005, Bach 2003, Mukherjee 2010)

Let $\mathcal{D}(\mathcal{V}_1) = \mathcal{C}(\mathcal{V}_1, \mathcal{V}_1)$,

$$\max_{\mathcal{V}_1, \mathcal{V}_2} \sum_{k=1}^{2} \frac{\mathcal{D}(\mathcal{V}_k)}{\mathcal{C}(\mathcal{V}_k, \mathcal{V} \setminus \mathcal{V}_k)} = \max_{\mathcal{V}_1, \mathcal{V}_2} \sum_{k=1}^{2} \frac{\sum_{p,q \in \mathcal{V}_k} \mathcal{K}_{pq}}{\sum_{p \in \mathcal{V}_k, q \notin \mathcal{V}_k} \mathcal{K}_{pq}}, \tag{5}$$

Similarity across segments is small, similarity within the segment is large.

The Learning Version of the Problem (Shortreed 2005, Bach 2003, Mukherjee 2010)

Let $\mathcal{D}(\mathcal{V}_1) = \mathcal{C}(\mathcal{V}_1, \mathcal{V}_1)$,

$$\max_{\mathcal{V}_1,\mathcal{V}_2} \sum_{k=1}^{2} \frac{\mathcal{D}(\mathcal{V}_k)}{\mathcal{C}(\mathcal{V}_k, \mathcal{V} \setminus \mathcal{V}_k)} = \max_{\mathcal{V}_1,\mathcal{V}_2} \sum_{k=1}^{2} \frac{\sum_{p,q\in\mathcal{V}_k} \mathcal{K}_{pq}}{\sum_{p\in\mathcal{V}_k, q\notin\mathcal{V}_k} \mathcal{K}_{pq}}, \tag{5}$$

Similarity across segments is small, similarity within the segment is large.

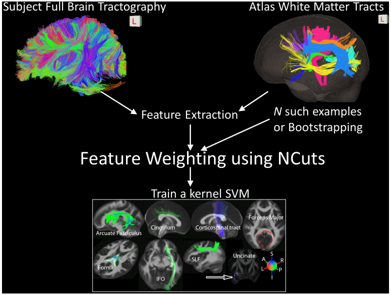
Basis kernels are $\{\mathcal{K}^{(1)}, \cdots, \mathcal{K}^{(d)}\}$, and $\hat{\mathcal{K}} = \mathcal{K}^{\alpha} = \sum_{l=1}^{d} \alpha_l \mathcal{K}^{(l)}$
Linearly combine kernels to make distribution NCuts friendly:

$$f(\alpha) = \sum_{t=1}^{2} \frac{\sum_{p,q\in\mathcal{V}_t} \mathcal{K}^{\alpha}_{pq}}{\sum_{p\in\mathcal{V}_t, q\notin\mathcal{V}_t} \mathcal{K}^{\alpha}_{pq}} \tag{6}$$

$$= \sum_{t=1}^{2} \frac{\sum_{l=1}^{d} \alpha_l \sum_{p,q\in\mathcal{V}_t} \mathcal{K}^{l}_{pq}}{\sum_{l=1}^{d} \alpha_l \sum_{p\in\mathcal{V}_t, q\notin\mathcal{V}_t} \mathcal{K}^{l}_{pq}} \tag{7}$$

Outline

- ~~Normalized Cuts~~
- Kernel Learning
  - ~~Motivation~~
  - Models
  - Optimization
- Tract Clustering in Diffusion Tensor Images



Reformulation as a Ratio Optimization

$$f(\alpha) = \sum_{t=1}^{2} \frac{\sum_{l=1}^{d} \alpha_l \sum_{p,q \in \mathcal{V}_t} \mathcal{K}^l_{pq}}{\sum_{l=1}^{d} \alpha_l \sum_{p \in \mathcal{V}_t, q \notin \mathcal{V}_t} \mathcal{K}^l_{pq}} \qquad (8)$$

(inter-class similarity) $\mathbf{U} = [u(t,l)] \in \mathbb{R}^{2\times d}$ where $u(t,l) = \sum_{p \in V_t, q \notin V_t} \mathcal{K}^l_{pq}$

(intra-class similarity) $\mathbf{V} = [v(t,l)] \in \mathbb{R}^{2\times d}$ where $v(t,l) = \sum_{p,q \in V_t} \mathcal{K}^l_{pq}$

Expressing our objective in terms of **U** and **V**:

$$\max_{\alpha} \quad f(\alpha) = \sum_{t=1}^{2} \frac{\sum_{l=1}^{d} v(t,l)\alpha_l}{\sum_{l=1}^{d} u(t,l)\alpha_l} \quad \text{subject to} \quad \sum_{l=1}^{d} \alpha_l = 1, \alpha_l \geq 0$$

For two classes denote:

$$\hat{v}(l) = \sum_{t=1}^{2} v(t,l); \quad \hat{u}(l) = u(1,l) = u(2,l),$$

Reformulation as a Ratio Optimization

$$\max_{\alpha} \quad f(\alpha) = \sum_{t=1}^{2} \frac{\sum_{l=1}^{d} v(t,l)\alpha_l}{\sum_{l=1}^{d} u(t,l)\alpha_l} \quad \text{subject to} \quad \sum_{l=1}^{d} \alpha_l = 1, \alpha_l \geq 0$$

Substituting,

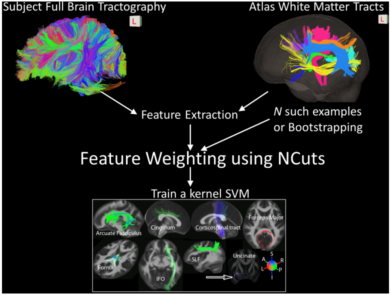
$$\max_{\alpha} f(\alpha) = \qquad \max \frac{\hat{v}^T\alpha}{\hat{u}^T\alpha} = \min \frac{\hat{u}^T\alpha}{\hat{v}^T\alpha} \quad \text{s.t.} \ \sum_{l=1}^{d} \alpha_l = 1, \alpha_l \geq 0.$$

$\mathcal{X} = \{X^{(1)}, \cdots, \mathcal{X}^{(N)}\}$ comes with "correct" partition(s)
Create $\hat{u}_{(j)}$ and $\hat{v}_{(j)}$ for each training example $x_j \in \mathcal{X}$.
Then we obtain a function of multiple ratios,

$$\max_{\alpha} f(\alpha) = \quad \min \sum_{j \in X} \frac{\hat{u}_j^T\alpha}{\hat{v}_j^T\alpha} \text{subject to} \qquad \sum_{l=1}^{d} \alpha_l = 1, \quad \alpha_l \geq 0.$$

Outline

- ~~Normalized Cuts~~
- Kernel Learning
  - ~~Motivation~~
  - ~~Models~~
  - Optimization
- Tract Clustering in Diffusion Tensor Images



Convex Relaxations

$$\max_{\alpha} f(\alpha) = \quad \min \sum_{j \in X} \frac{\hat{u}_j^T \alpha}{\hat{v}_j^T \alpha} \text{subject to} \qquad \sum_{l=1}^{d} \alpha_l = 1, \quad \alpha_l \geq 0.$$

Convex Relaxations

$$\min \sum_{j \in X} \frac{\hat{u}_j^T \alpha}{\hat{v}_j^T \alpha}(|X|-1) = \min \sum_{(g,h)\in\Phi} \frac{\hat{u}_g^T \alpha}{\hat{v}_g^T \alpha} + \frac{\hat{u}_h^T \alpha}{\hat{v}_h^T \alpha}$$

$$= \min \sum_{(g,h)\in\Phi} \frac{\alpha^T(\hat{u}_g \hat{v}_h^T + \hat{u}_h \hat{v}_g^T)\alpha}{\alpha^T \hat{v}_g \hat{v}_h^T \alpha}$$

$$= \min \sum_{(g,h)\in\Phi} \frac{\alpha^T A_{gh} \alpha}{\alpha^T B_{gh} \alpha}$$

(Minimize Gap) $\min \sum_{g \neq h} \delta_{gh}$

subject to $\alpha^T(A_{gh} - B_{gh})\alpha \leq \delta_{gh}, \quad \sum_{l=1}^{d} \alpha_l = 1.$

Convex Relaxations

$$\min \sum_{j \in X} \frac{\hat{u}_j^T \alpha}{\hat{v}_j^T \alpha}(|X|-1) = \min \sum_{(g,h)\in\Phi} \frac{\hat{u}_g^T \alpha}{\hat{v}_g^T \alpha} + \frac{\hat{u}_h^T \alpha}{\hat{v}_h^T \alpha}$$

$$= \min \sum_{(g,h)\in\Phi} \frac{\alpha^T(\hat{u}_g \hat{v}_h^T + \hat{u}_h \hat{v}_g^T)\alpha}{\alpha^T \hat{v}_g \hat{v}_h^T \alpha}$$

$$= \min \sum_{(g,h)\in\Phi} \frac{\alpha^T A_{gh} \alpha}{\alpha^T B_{gh} \alpha}$$

(Minimize Gap) $\min \sum_{g \neq h} \delta_{gh}$

subject to $\alpha^T(\mathcal{J}_{gh})\alpha \leq \delta_{gh}, \quad \sum_{l=1}^{d} \alpha_l = 1.$

Convex Relaxations

Standard QP

$$
\begin{aligned}
\text{(StQP)} \quad \min \quad & \sum_{g \neq h} \alpha^T \mathcal{J}_{gh} \alpha \\
\text{subject to} \quad & \sum_{l=1}^{d} \alpha_l = 1, \alpha \geq 0
\end{aligned}
$$

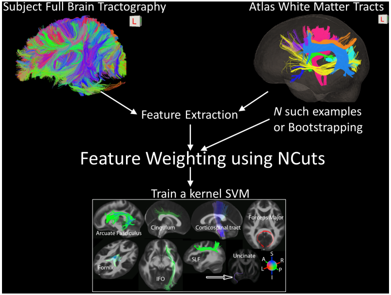
Let $\mathcal{J} = \sum_{g \neq h} \mathcal{J}_{gh}$ and $Q = (\mathcal{J} + \mathcal{J}^T)/2$,

SDP Relaxations

$$
\begin{aligned}
\text{(SDP)} \min \quad & \operatorname{tr}(QZ) \\
\text{subject to} \quad & \sum_{l=1}^{d} \sum_{l'=1}^{d} Z_{ll'} = 1, \\
& Z \succeq 0, \quad Z \geq 0.
\end{aligned}
$$

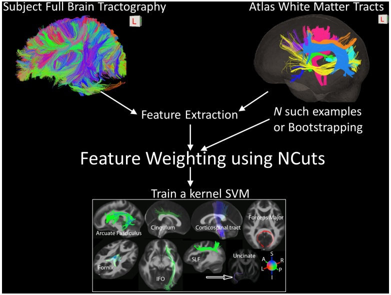
Outline

- ~~Normalized Cuts~~
- ~~Kernel Learning~~
  - ~~Motivation~~
  - ~~Models~~
  - ~~Optimization~~
- Tract Clustering in Diffusion Tensor Images



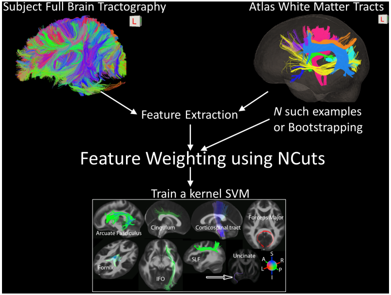
Outline

- ~~Normalized Cuts~~
- ~~Kernel Learning~~
  - ~~Motivation~~
  - ~~Models~~
  - ~~Optimization~~
- Tract Clustering in Diffusion Tensor Images



Outline

- ~~Normalized Cuts~~
- ~~Kernel Learning~~
  - ~~Motivation~~
  - ~~Models~~
  - ~~Optimization~~
- Tract Clustering in Diffusion Tensor Images



Tract Clustering from Diffusion Tensor Images

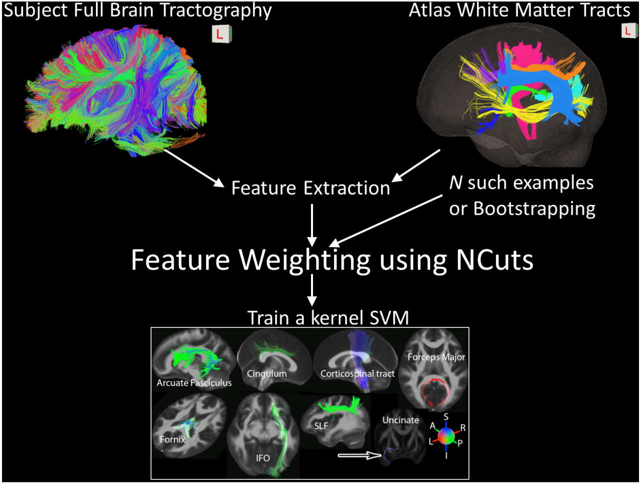
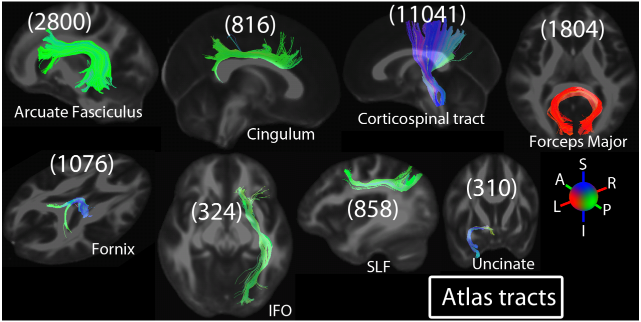


Figure: The overview of our adaptive cuts framework

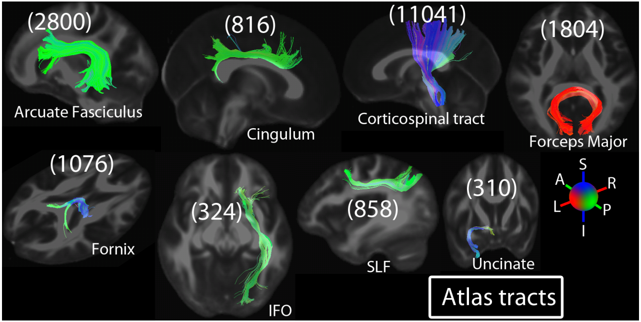
Tract Clustering from Diffusion Tensor Images



Eight different white matter pathways from Pittsburgh Brain Connectivity (PBC).

- Chamfer/Haussdorf distance, followed with MST based heuristic
- B-spline similarity
- Similarity based on Cosine series representation

Tract Clustering from Diffusion Tensor Images



Eight different white matter pathways from Pittsburgh Brain Connectivity (PBC).

Q: Which similarity measure to use?

A: Let us use them all.

Experimental Design: Generating Multiple Similarities

- Six different features to define fiber similarities
- Cosine bases (Chung et al. 2010) for every individual fiber:
  - 3D coordinates
  - Curvature
  - Torsion functions
- Include midpoints to generate six different similarity measures
- Using Cosine bases allows us to compute tract-tract similarity very efficiently avoiding pointwise comparisons.

Ensemble of pairwise kernel SVMs

Once we obtain the feature weights (α^*), there are several options of using these weights in extracting specific tracts:

1. Apply weights, obtain weighted kernel, then a *k*-way multicut or iteratively perform normalized cuts.
2. Use weighted kernel within a multi-class classifier.
3. We use an ensemble of *pairwise* classifiers (such as Arcuate vs. IFO, Arcuate vs. UNC), after learning the kernel.

Experimental Results (Qualitative)

Training data is kept separate, test on two distinct test brains.

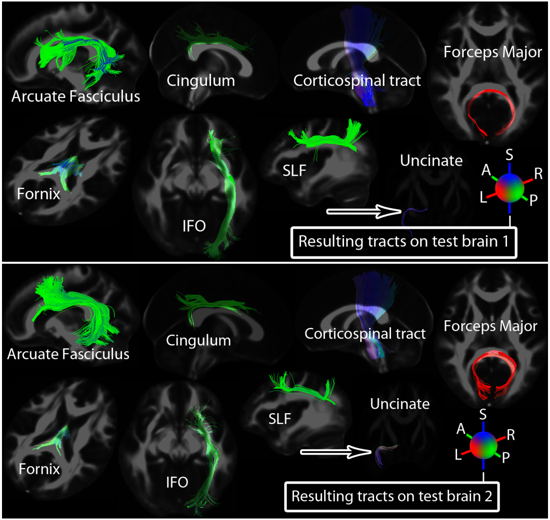
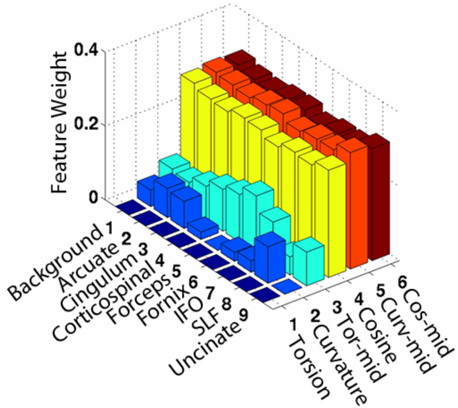


Figure: The eight different tracts extracted by the proposed framework from a whole-brain tractography on two test brains.

Experimental Results (Interpretability)



Curvature plays more important role for association fibers such as asoarcuate fasciculus and cingulum, which have a bend going from posterior to anterior part of the brain, compared to projection fibers such as corticospinal tract.

Experimental Results (Sum of Kernels vs. ACuts)

Comparison	Dice	Kappa
ACuts vs. Sum	$p = 0.0146$	$p = 0.0169$

Table: Pairwise *t*-test results show that there is statistically significant improvement when using ACuts compared to using plain sum of features when we do not pre-select features.

Experimental Results (Single Feature vs. ACuts)

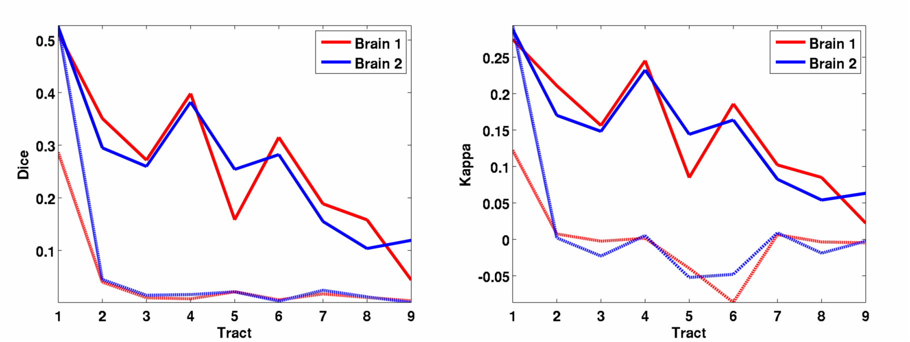


Figure: Paired t-test shows there is statistically significant improvement ($p < 10^{-5}$) when using ACuts compared to using only curvature. We also observe (although not as much) statistically significant improvement ($p < 0.05$) when comparing against only Cosine encoding of the 3D coordinates.

Conclusions

- Adaptive similarity measure can be learnt for extracting different white matter tracts.
- Do not have to solve large eigen decompositions because normalized cuts objective is mainly used in learning and then we can use SVMs.
- **Code available**, and not difficult to adapt for specific applications as long as a black box procedure for generating similarities is available `http://brainimaging.waisman.wisc.edu/~adluru/AC`

Thanks to the following collaborators:

- Moo K. Chung
- Richard J. Davidson
- Chris Hinrichs
- Sterling C. Johnson
- Lopamudra Mukherjee
- Jiming Peng

Funding:

- NIH R21-AG034315
- NIH R01-AG021155
- NSF IIS 1116584