Extracting Proximity for Brain Graph Voxel Classification

N. Sismanis [1] D. L. Sussman [2] J. T. Vogelstein [3] W. Gray [4] R. J. Vogelstein [4]
E. Perlman [5] D. Mhembere [5] S. Ryman [6] R. Jung [6] R. Burns [2] C. E. Priebe [2]
N. Pitsianis [1,7] X. Sun [7]

[1]Electrical and Computer Engineering Department, Aristotle University, Thessaloniki, Greece

[2]Applied Mathematics and Statistics Department, Johns Hopkins University, Baltimore MD, USA

[3]Statistical Science and Mathematics Department, Duke University, Durham NC, USA

[4]Johns Hopkins University, Applied Physics Laboratory, Laurel MD, USA

[5]HHMI Janelia Farm Research Park, Ashburn VA, USA

[6]Neurosurgery Department, University of New Mexico, Albuquerque NM, USA

[7]Computer Science Department, Duke University, Durham NC, USA

5 April 2013

5th Panhellenic Conference of Biomedical Technology, Athens, Greece

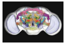
Brain Maps & Connectomes

- Connectome: The totality of neuron connections in a nervous system
- Connectomics: The science concerned with assembling, analyzing connectomes

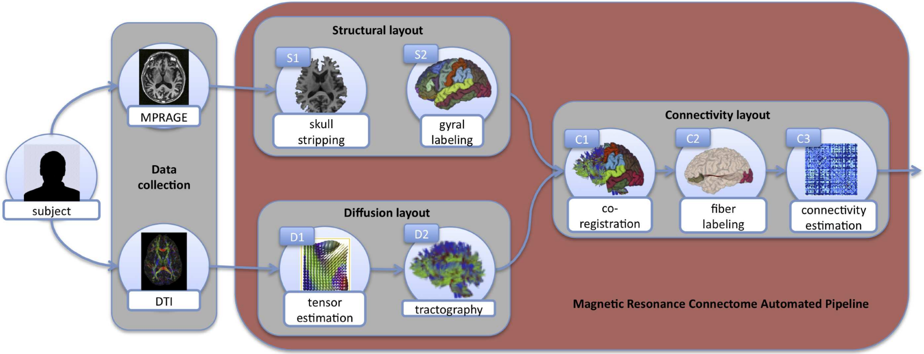
Neural Activity, Association, or Difference between

- anatomical regions
- individual neurons
- physical properties and mental behaviors
  [J. Vogelstein *et al.*, Scient. Rep., 2011]
- spatial cortical regions and functionality
  [R. Desikan *et al.*, NeuroImage, 2006]
- gender
  [J. Vogelstein *et al.*, IEEE Trans. Pattern Anal. and Mach. Intell, 2012]

Scales of Neuron Systems

C.elegans		10^2 neurons
fruit fly		$10^2 \times 10^3$
mouse		$10^2 \times 10^3 \times 10^3$
human		$10^2 \times 10^3 \times 10^3 \times 10^3$

Brain Graph Generation from Imaging

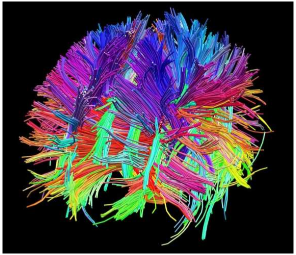


Estimations of brain graphs or connectomes obtained from 3D MRI scans [1]

[1] G. R. Gray *et al.*, IEEE PULSE, 2012

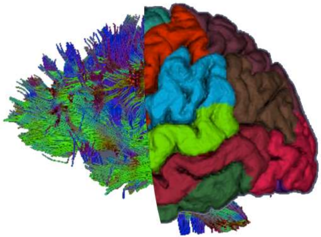
Brain Graph Analysis : Classification

Voxel-Vertex Graph



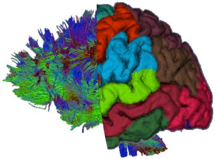
- **voxels** smallest distinguishable partition in a 3D image
- **connections** in PDD chromatic code:
  Interior ↔ Posterior: Green
  Superior ↔ Inferior: Blue
  Left ↔ Right: Red
- `http://www.humanconnectomeproject.org`

Voxel-to-Region Classification



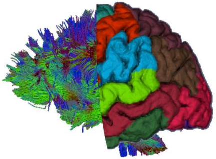
- Voxels in gray matter: classified via adaptive image registration
- Voxels in white matter: highly uncertain, to be labeled by connection, association and inference
- `http://www.humanconnectomeproject.org`

Computational Challenges in Classification



- A huge number of vertices (potentially 100 billion)
- A large of classes (70) Desikan *et al.*, NeuroImage, 2006

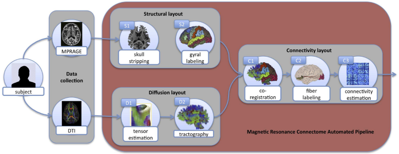
Computational Challenges in Classification



- ▷ A huge number of vertices (potentially 100 billion)
- ▷ A large of classes (70) Desikan *et al.*, NeuroImage, 2006

- ▷ Noisy data
- ▷ Partially available connectivity and labels
- ▷ Complex geometry
- ▷ Individual variation

Recent Advances in Voxel-to-Region Classification



A Magnetic Resonance Connectome Automated Pipeline [2]

B Spectral Embedding of Graphs [3], using also efficient SVD package

C Universally Consistent Latent Position Estimation and Vertex Classification for Random Dot Product Graphs [4]

---

[2] Grey *et al.* IEEE PULSE, 2012

[3] Rohe *et al.* Annals of Statist. 2011

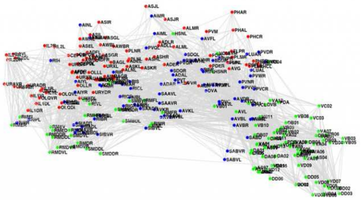
[4] Sussman, *et al.* in preprint, 2012

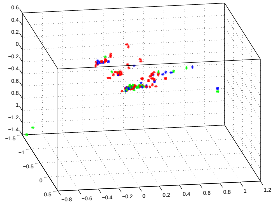
Spectral Embedding of a Brain Graph

Spectral embedding : a graph placed in an Euclidean space with d chosen singular-vectors of adjacency or Laplacian matrix as the axes

$$A^2 = U\,\Sigma^2\,U^\top \quad \text{truncated to} \quad A_d^2 = U_d\,\Sigma_d^2\,U_d^\top$$

- ◇ Σ, U : singular-value, singular-vector matrices
- ◇ $\mathbb{R}^d$ by U_d as a feature space : encoding connections, revealing latent info.
- ◇ each vertex coded with a d-vector
- ◇ each edge associated with a pair of d-vectors
- ◇ a metric established for similarity/dissimilarity, a critical connection to standard classification methods





Data and figure source: http://www.openconnectomeproject.org/

Proximity Analysis in an Embedding Vector Space : Status and Gaps

- Proximity analysis so far *limited* to the *sequential* use of k-NN search in a *low-dimension* embedding space
- Highly efficient, robust k-NN search *all at once* is needed, especially for large data sets in relatively high-dimensional space

All-k-NN : Among an ensemble of N points in a d-dimensional Euclidean space, locate for each and every point its k nearest neighbors, according to a distance metric

- Exact search methods prohibitively expensive
- Resort to approximate methods (statistical, numerical or both)

All-k-NN Search : Exact Methods Prohibitively Expensive

Quadratic Scaling in N by naíve use of one-to-all k-NN search for each point

$$O(d \cdot N^2)$$

C. Elegans :	$d \cdot 100^2$
Fruit Fly :	$d \cdot 100^2 \cdot 10^6$
Mouse :	$d \cdot 100^2 \cdot 10^6 \cdot 10^6$
Human :	$d \cdot 100^2 \cdot 10^6 \cdot 10^6 \cdot 10^6$

Exponential Growth with d (*dimension curse*) by spatial partition/binning [5], limited to low-dimension spectral embedding

$$O(2^d\ N) \begin{cases} \leq O(d\,N^2) & \text{when} \quad d < \log N \\ > O(d\,N^2) & \text{otherwise} \end{cases}$$

[5] P. B. Callahan *et al.*, Jurn. ACM, 1995; J.Sankaranarayanan *et al.*, Comput. Graph. 2007

All-k-NN Search : Status of Approximate Methods

- By randomized projections for locality sensitive hashing , $O(\lambda N d^2) + O(k\, d\, \lambda N)$ [6]
  λ: number of hash tables

- randomized kd-trees, $O(N \log N) + O(\alpha\, d\, N)$ [7]
  α: number of tree nodes traversed

- Hierarchical k-means, $O(N \log N) + O(\alpha\, d\, N)$ [8]

Shortcomings

low dimension assumption

limited in parallel execution

poor data locality

---

[6] Indyk and Motwani, STOC, 1998; M. Trad *et al.*, ICMR, 2012
[7] Silpa-Anan and Hartley, CVPR, 2008
[8] Fukunaga and Narendra, IEEE Trans. Comput. 1975

AkNN-RARE: All-k-NN Search with RAndom REflections

We developed a fast and robust algorithm for All-k-NN search

$$O(d\, h\, N \log(N)) + O(k\, h\, d^2\, N)$$

- use h random distance-preserving coordinate transforms with *Householder reflections*
- sort along each axis, in parallel
- merge kNN among all axes

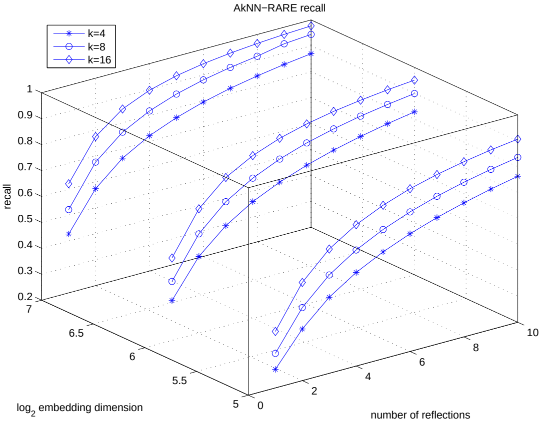
Advantages

- defying the dimension curse :
  superlinear in N, quadratic in d, a small number h sufficient for desired accuracy
- simple data structure, regardless geometric, relational structures
- high parallel potential, at multiple levels
- high degree of data locality, hashing free
- simple program structures, hassle free

Experimental Results with AkNN-RARE

- Data collected at the Mind Research Network (MRN), New Mexico
- Data labeled via adaptive image registration and inference techniques
- Data size : about a half million voxel-vertices
- Performance evaluation of AkNN-RARE
  - Accuracy metric : RECALL
  - Comparison with FLANN, a popular package for kNN search

Accuracy and Efficiency of AkNN-RARE



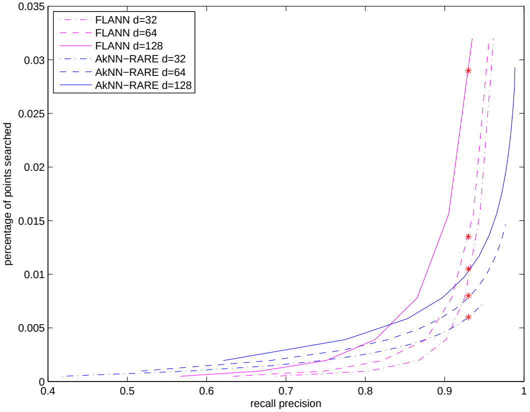
- Total arithmetic complexity

$$O(d\,h\,N \log N) + O(d^2\,k\,h\,N)$$

- RECALL precision: percent of correct kNN found
- High recall precision with only $h = 10$ transformations
- Reflection transformations can be executed concurrently
- Problem size shown $N = 500,000$
- Enabling high-dim. space embedding
  $32 \leq d \leq 128$

Comparison with FLANN

FLANN : based on randomized *kd*-trees, widely used for *k*NN search [9]



- Cost : actual number of pairwise distances calculated
- At a higher level of recall precision, A*k*NN-RARE incurs much lower cost

[9] Muja and Lowe, VISSAPP, 2009

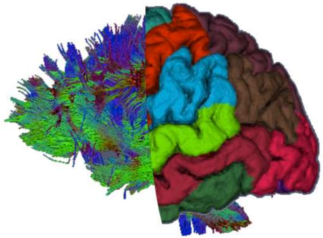
Recap: Extracting Proximity for Brain Graph Voxel Classification

Draw upon recent advance

- A Magnetic Resonance Connectome Automated Pipeline
- B Spectral Embedding of Graphs, using also efficient SVD package
- C Universally Consistent Latent Position Estimation and Vertex Classification for Random Dot Product Graphs

We developed

- D a fast, robust algorithm, enabling proximity extraction
  - at increasingly larger scale toward 100 billion
  - in sufficiently high-dim. info.-encoding space
  - on high accuracy demand
  - utilizing highly parallel computing architectures



Acknowledgments

- The authors at AUTh acknowledge the support of Marie Curie International Reintegration Program, EU
- J. Vogelstein acknowledges the support of Research Program in Applied Neuroscience and the London Institute for Mathematical Science Subcontract on HDTRA1 $-$ 11 $-$ 1 $-$ 0048 and NIH $RO1ES017436$
- R. Jung and S. Ryman acknowledge the John Templeton Foundation-Grant #22156: The Neuroscience of Scientific Creativity
- J. T. Vogelstein, R. J. Vogelstein and W. Gray acknowledge the Research Program on Applied Neuroscience NIH/NINDS $5R01NS056307$