From RAGs to Riches: Utilizing Richly Attributed Graphs to Reason from Heterogeneous Data

PI: Joshua Vogelstein
Johns Hopkins University

SIMPLEX Monthly Progress Report
[Aug 1, 2015 – Aug 31, 2015]

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Center for IMAGING SCIENCE
JOHNS HOPKINS UNIVERSITY

INSTITUTE for COMPUTATIONAL MEDICINE

Technical Executive Overview

Technical Accomplishments

- *RAG Embedding (Tensor Factorization):* benchmarked FlashMatrix eigensolver against competitors, we are better :)
- *Data Management (Dense Arrays):* benchmarked NeuroBlaze against OCP, several factor speedup
- *Data Ingest (Diffusion MRI):* processing 20 additional datasets to obtain over 2000 new DTI derived brain graphs
- *RAG Construct (Random Walks):* Developed new RAG generative model and inference technique that significantly outperform previous joint embedding strategies

Other Organizational Updates

- Created umbrella organization called NeuroData, to emphasis more generality over merely connectomes at a single scale
- Launched new website http://neurodata.io, is live, will be fully operational shortly

RAG Embedding – Tensor Factorization

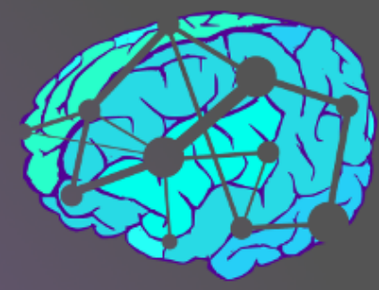
- Benchmarking against other in-memory and some distributed memory eigensolvers
- began exploring possible NMF implementations to backend

singular values	residual
10563	4.067819E-13
10442	1.044935E-13
9531	2.527054E-13
7905	9.656768E-14

Data Management - Dense Arrays

- continuous progress on speeding up and benchmarking
- code: https://github.com/openconnectome/ocpblaze

Premlinary Timings	OCP	Blaze
512x512x16	0.478	0.171
1024x1024x16	1.699	0.709
2048x2048x16	6.696	2.612
4096x4096x16	24.919	8.87

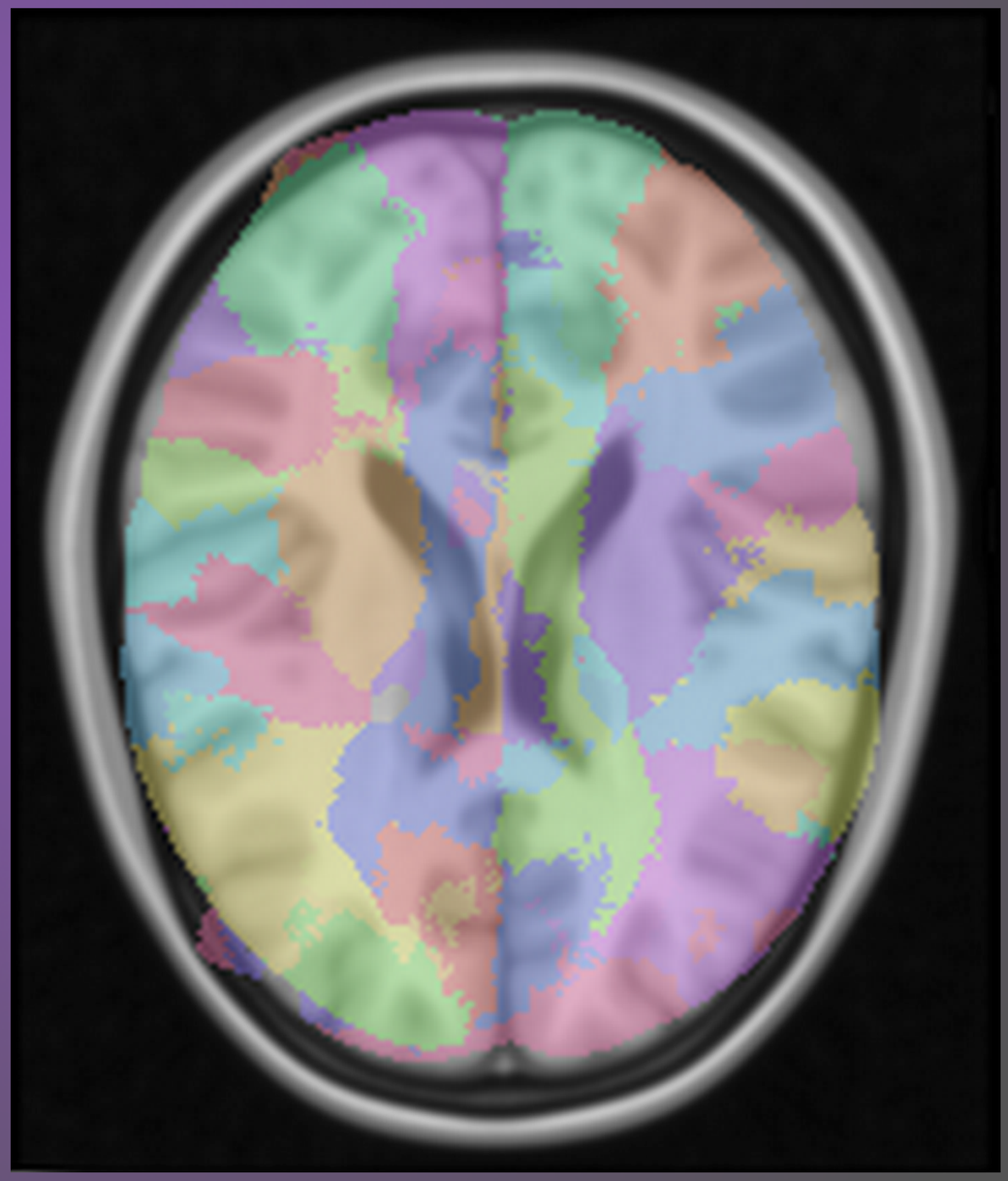


Data Management - Sparse Arrays

- Began exploring surface compression
- Began exploring surface visualization
- Began exploring a surface database

Data Ingest - Diffusion MRI

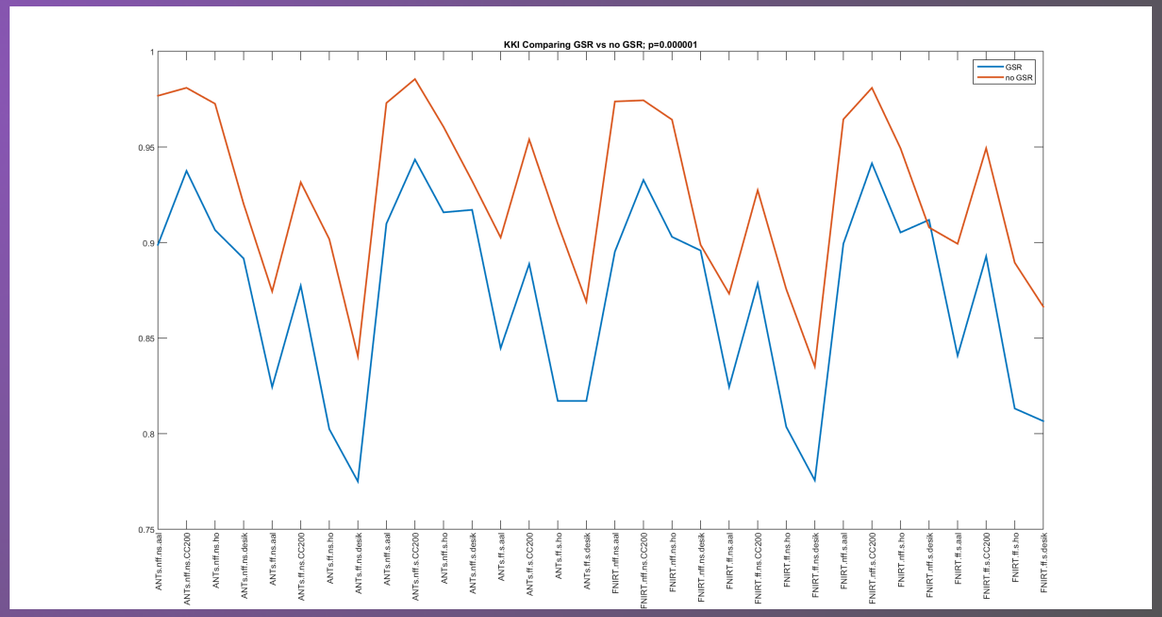
- Ran our first pediatric brain via Web-service
- Feedback: “Again, I think your program will revolutionize the way we do research"
- Running all data from the CoRR dataset
- Deploying AMI to enable others to modify and run ad will



http://dsp061.pha.jhu.edu/ocp/overlay/0.4/dsp061.pha.jhu.edu/mniatlas/desikan/dsp061.pha.jhu.edu/mniatlas/image/xy/0/0,182/0,218/91/

Data Ingest - functional MRI

- Ran 64 different pipelines on 4 different datasets
- Collecting the results now
- There are steps that, so far, clearly are superior although historically have been thought inferior



RAG Construction

- Novel RAG generative model
- Novel inference algorithm motivated by generative model
- Under a simple simulation, new method achieves optimal performance, old methods achieve 86% or 28%

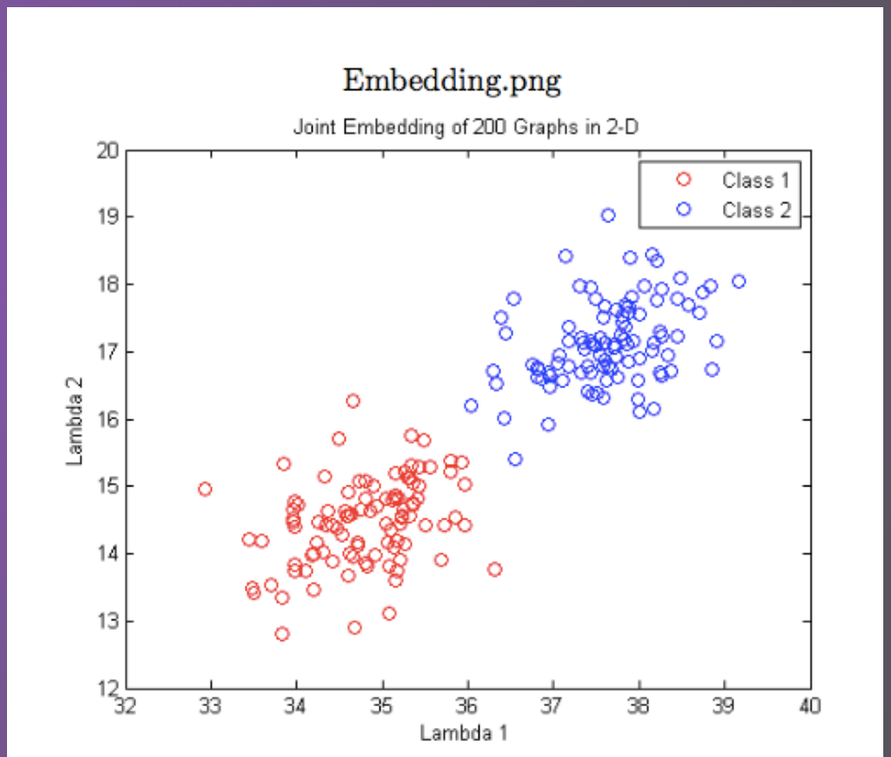
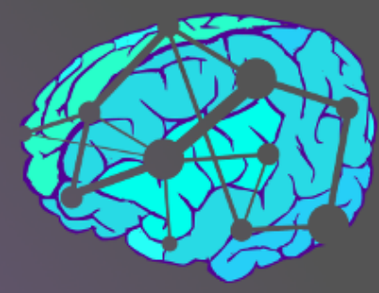
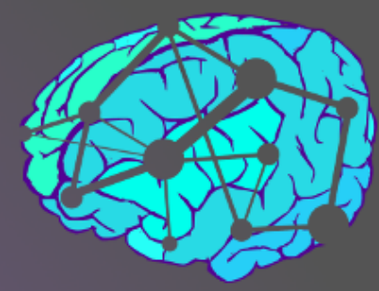


Figure 1: Awesome Image

Potential Next Steps



- **RAG Embedding**
  - Tensor Factorization: backend NMF
- **Data Management**
  - Dense Arrays: continue developing NeuroBlaze to v0.1
  - Sparse Arrays: add functionality to query & visualize surfaces
- **Data Ingest**
  - Diffusion MRI - ingest several benchmark datasets
  - functional MRI - ingest fMRI data as a separate channel for each subject
- **RAG Construct**
  - Continue testing/ exploring new model and inference method



Software Development

Software

- FlashGraph: https://github.com/icoming/FlashGraph
- NeuroBlaze: https://github.com/openconnectome/ocpblaze
- NeuroSurf: https://github.com/openconnectome/surface-extractor

Programmatic challenges and requests for PM action

Key challenges and/or associated risk:

- For RAG construction,we have devised a novel model and inference method, which has compelling preliminary results, and requires significant additional theoretical and numerical analysis.

Request for action:

No action requested.