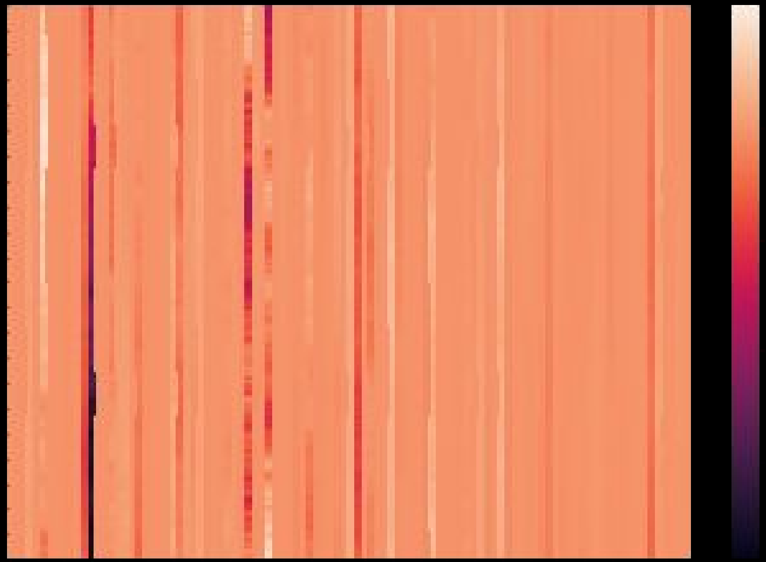
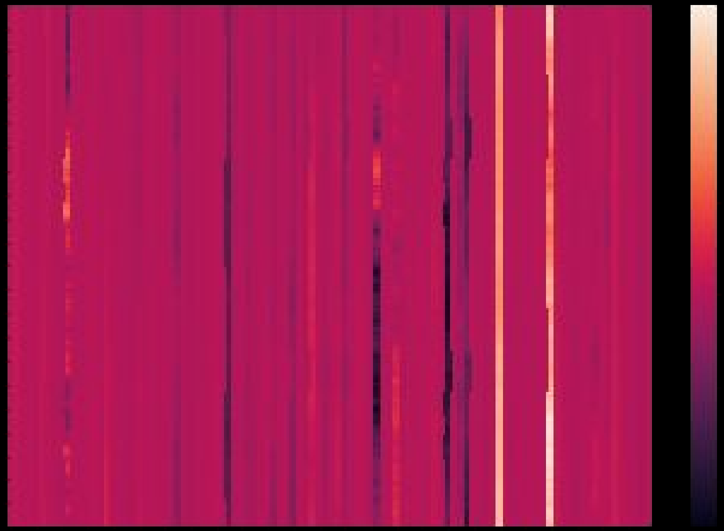
Week of 10/16 Deliverables





Red Lemurs

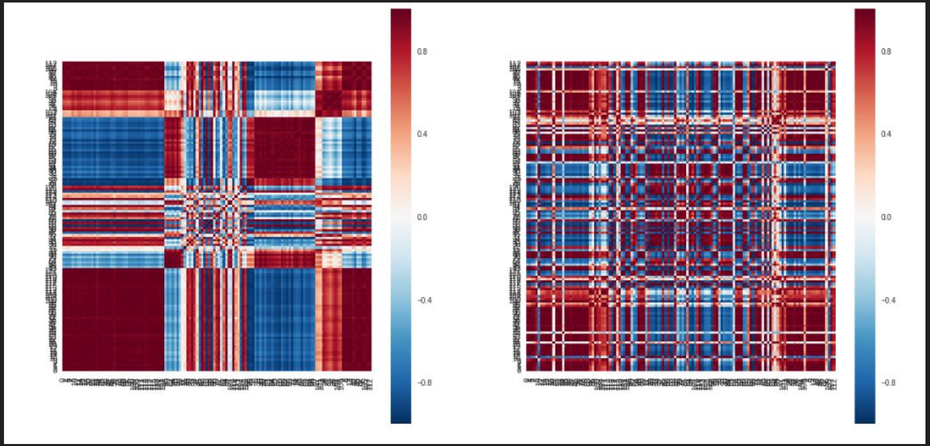
Deliverables

- Clustering algorithms paper reading + basic implementation
  - notebook, .md notes
- P-factor paper reading + exploration of possible implementation
  - .md notes
- Edit and write scripts to move and load HBNB data easily (½)
  - script
- Multidimensional Time Series (MTS) Distance function implementation
  - jupyter notebook
- Generalized Distance Function + Visualization
  - https://nbviewer.jupyter.org/github/NeuroDataDesign/lemur-f17s18/blob/master/docs/notebooks/vidurkailash/Generalized%20Distance%20Function.ipynb
- Maximize Core Availability
  - https://github.com/NeuroDataDesign/lemur-f17s18/blob/master/docs/notebooks/vidurkailash/Multiprocessing%20Optimization%20.ipynb

Clustering Algos

- To a large extent, need to find good distance function/mapping for a clustering algorithm to work
    - Eg, DBSCAN did not work with high dimensional set
- Algos:
    - Fast Greedy Algorithm (like Kruskal's)
    - Walktrap
    - Label propagation
- Currently implemented spectral clustering

Clustered Correlation Matrix vs Normal Correlation Matrix for Resting State



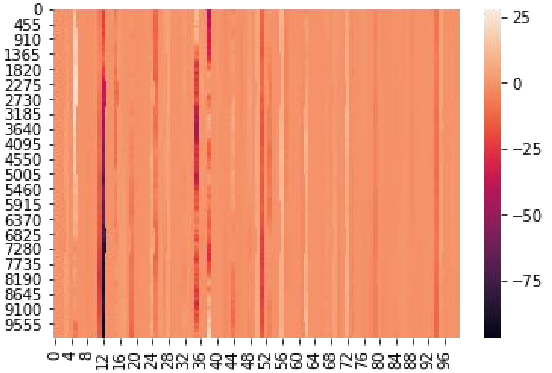
Clustered

Normal

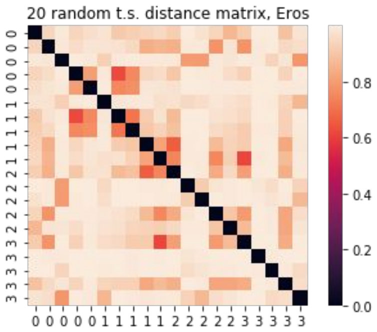
MTS Algorithm Implementation

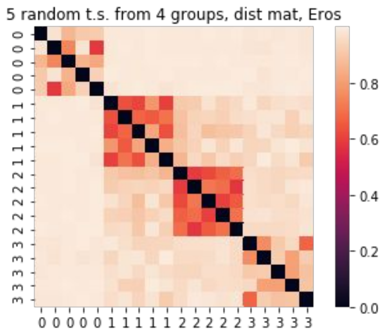
Toy Data: A simple stochastic process

1. Sample $\Theta \sim \mathcal{N}^d(\vec{0}, s * I_d)$
2. Transform $\tilde{\Theta} = e^{\Theta}/||\Theta||$
3. Sample $x_0 \sim \mathcal{N}^d(\vec{0}, I_d\tilde{\Theta})\sqrt{T}$
4. For i from 1 to T
   - Sample $c \sim \mathcal{N}^d(\vec{0}, I_d\tilde{\Theta})$
   - $x_i = x_{i-1} + c$

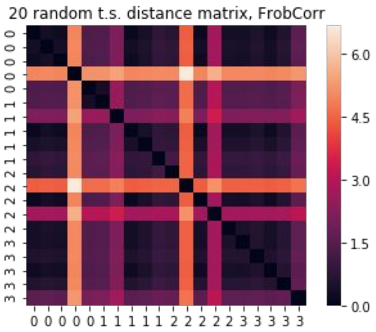


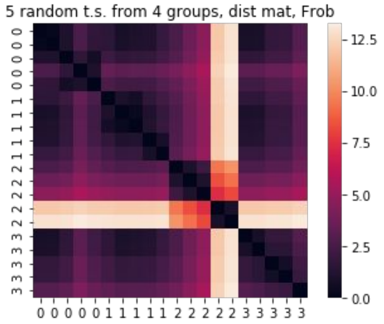
Extended fRObeniuS norm (EROS)





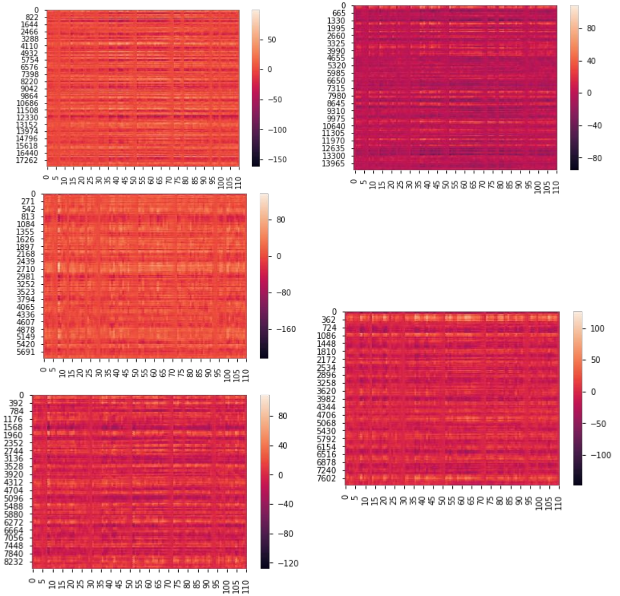
Frobenius Norm of Correlation Matrices



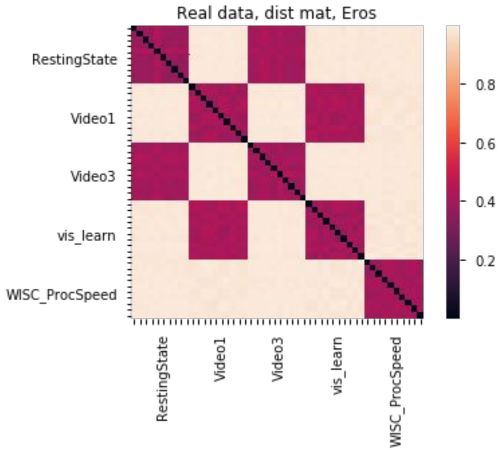


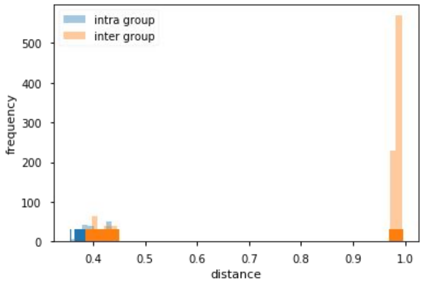
MTS Algorithm Implementation

Real Data: EEG time series from 5 paradigms

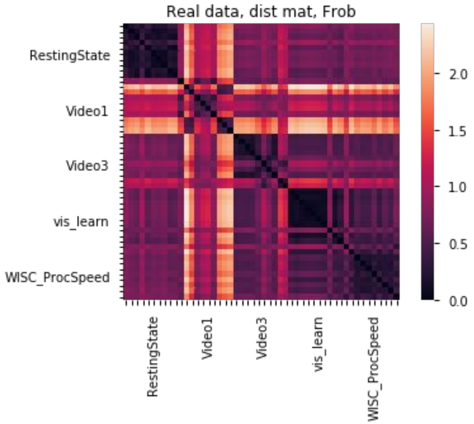


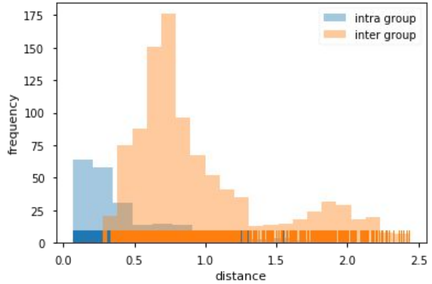
Extended fRObeniuS norm (EROS)





Frobenius Norm of Correlation Matrices





P-factor exploration

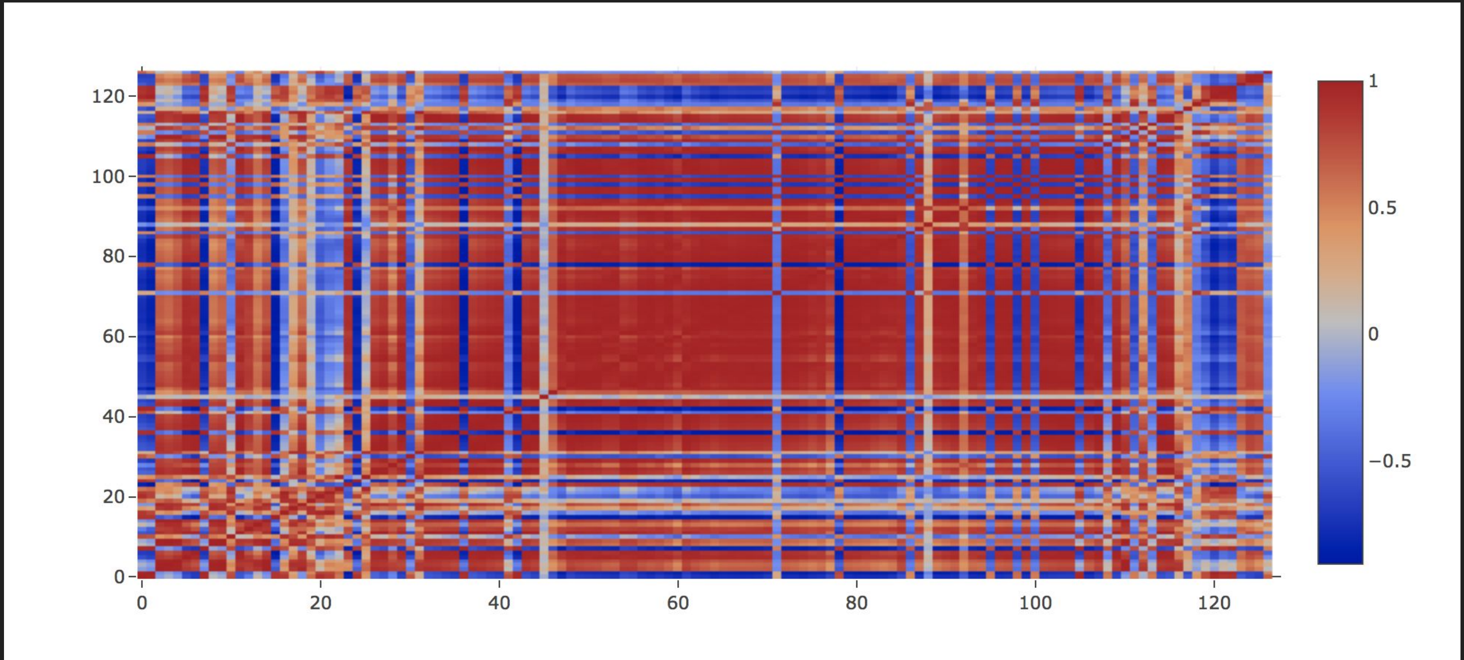
- feature scores were extracted from self-report and diagnostic interview to calculate p-factor scores
- Structural correlates to p-factor:
    - Negative correlation with white matter in bilateral pons
    - Negative correlation with gray matter in occipital lobe
    - Whole brain intrinsic connectivity in visual cortex
- CWAS (using resting state fMRI) -- for whole brain connectome
    - Pseudo:
        - Calculate correlation and normalized to contain Z score. Compute distance matrix using distance matrix = 1 - correlation. Compute Gower’s centered matrix from distance matrix and perform MDMR for the voxel. Then determine significance of MDMR statistics with permutation tests.

Pulling and Reformatting Data

- Script to pull data to local directory (1)
- Reformat into BIDS format (½)
    - Understand draft specification (1)
    - Code for reformatting and compressing
- Next week: Incorporate into Python package, code for loading data into RAM.

Generalized Distance Function + Visualization

DoD: https://nbviewer.jupyter.org/github/NeuroDataDesign/lemur-f17s18/blob/master/docs/notebooks/vidurkailash/Generalized%20Distance%20Function.ipynb



Maximize Core Availability

DoD: https://github.com/NeuroDataDesign/lemur-f17s18/blob/master/docs/notebooks/vidurkailash/Multiprocessing%20Optimization%20.ipynb

```
        b = data[j, :]
        d = metric(a, b)
        dist.append(d)
    return k1, k2, dist


ts_start = ts()
with Pool(n_processes) as pool:
    for k1, k2, res in pool.imap_unordered(proc, range(0, k_max, k_step)):
        dist[k1:k2] = res
        print("{:.0f} minutes, {:,}..{:,} out of {:,}".format(
            (ts() - ts_start)/60, k1, k2, k_max))


print("Elapsed %.0f minutes" % ((ts() - ts_start) / 60))
print("Saving...")
np.savez("dist.npz", dist=dist)
print("DONE")
```

```
0 minutes, 20..40 out of 4,950
0 minutes, 0..20 out of 4,950
0 minutes, 60..80 out of 4,950
0 minutes, 100..120 out of 4,950
0 minutes, 40..60 out of 4,950
0 minutes, 80..100 out of 4,950
0 minutes, 120..140 out of 4,950
0 minutes, 140..160 out of 4,950
0 minutes, 160..180 out of 4,950
0 minutes, 180..200 out of 4,950
0 minutes, 200..220 out of 4,950
0 minutes, 220..240 out of 4,950
0 minutes, 240..260 out of 4,950
0 minutes, 260..280 out of 4,950
0 minutes, 280..300 out of 4,950
0 minutes, 300..320 out of 4,950
0 minutes, 320..340 out of 4,950
```

Next Week

- Run algorithm in p-factor paper on whole-brain intrinsic connectivity versus gender/age
    - Take k fMRI scans and compute connectivity matrix as done in the p-factor paper. (thoughts on k?) instead of relating connectivity to p-factor, relate to age (float) or gender (binary 0, 1)
- Further explore MTS based on feedback, look at data further to see why distances of Eros were strange. Compare also on gender / age. Implement another algorithm in same fashion.
- Test various clustering algorithms on distance matrices produced from metrics, using available ones from SKLearn