

Connectome Coding

Laura Balzano & Joshua T. Vogelstein

Challenge, Impact and Innovation

A human brain is perhaps the most complex grapefruit sized object in the known universe, with $\mathcal{O}(10^8)$ neurons and $\mathcal{O}(10^{13})$ synapses [?]. Moreover, magnetic resonance imaging (MRI) provides access to the living brain at high spatiotemporal resolution, $\sim 1 \text{ mm}^3$ and $\sim 1 \text{ Hz}$. Yet, after 30 years of well-funded research, there remains to be a single clinically useful brain imaging based biomarker for any psychiatric diagnosis [?]. Evidence is building that most psychiatric disorders are fundamentally “connectopathies”, that is, disorders of connectivity [?]. Based on this, **we conjecture that the lack of clinically useful brain imaging biomarkers stems from an absence of adequate statistical models of brain connectivity.** More specifically, both functional and diffusion MRI data can yield estimates of human brain connectivity, or “connectomes”. The challenge is to build statistical models linking brain connectivity to phenotype. Learning such mappings requires overcoming both computational and statistical bottlenecks.

We propose to develop statistical network models specifically for populations of networks. This is in stark contrast to nearly all extant models of networks, which typically model a single network (for example, the Facebook network), rather than populations thereof. While models of populations of other data types are prevalent in neuroscience (such as population activity), we are unaware of any such population models of networks.

Rationale

Neural coding is a rich field devoted to understanding the relationship between neural *activity* and perception. By analogy, connectome coding will become a burgeoning field devoted to understanding the relationship between neural *connectivity* and experience. Indeed, since the days of Hebb [?], neuroscientists have conjectured that memories and experiences are stored in the connectivity patterns of the brain. If the brain stores memories in the physical connections, then it can be free to exhibit whatever pattern it so desires, without fearing a radical reconfiguration of its memories. On the other hand, if experience is somehow stored in ongoing activity, a sudden change of state (for example, by a seizure), could erase all history, like a etch-a-sketch. Thus, we are motivated to build and understand models of brain connectivity.

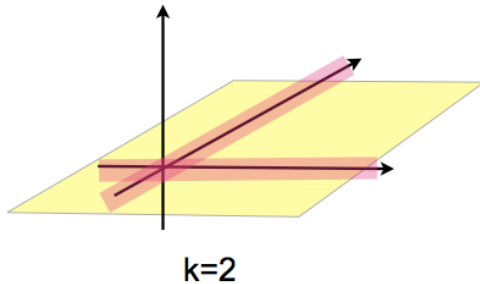


Figure 1: Illustration of two 1-d subspaces and one 2-d subspace; though the two pink lines (two one-dimensional subspaces) are a simpler model than the yellow two-dimensional plane, the set they define is non-convex.

Aims

1. Models of individual connectomes
2. Models of populations of connectomes
3. Models relating populations of connectomes to phenotypes

Subspace Clustering Modeling data as a union of subspaces is a simple and flexible model that fits well to the context of brain-graphs. Algorithmically, the problem of estimating the union of subspaces from data defines the problem of *Subspace Clustering*, or clustering vectors into groups that lie in or near the same subspace.

Let A_G be a $n \times n$ adjacency matrix of graph G , where $A_{uv} = 1$ iff $(u \sim v) \in \mathcal{E}_G$. The goal of subspace clustering is to jointly infer the underlying subspaces from A and to cluster the vertices of A according to their corresponding subspaces. This would in turn give us an understanding of potential clusters of connectivity. From our preliminary results, subspace clustering is a much better tool for clustering connectivity than spectral clustering of the graph or activity, or anatomical clustering, which is the state-of-the-art approach []. There are three open problems that we wish to address.

Let X be a **what size** data matrix, where the rows represent **what** and the columns **what**. The goal of subspace clustering is to infer the underlying subspaces from X and to cluster the columns of X according to the subspaces. This would in turn give us an understanding of potential clusters of connectivity causing activity in the brain, cite here. From our preliminary results, subspace clustering is a much better tool for clustering activity regions than spectral clustering, which is the state-of-the-art approach cite. There are three open problems that we wish to address.

1. Adapting the brain region clustering problem to the framework of subspace clustering, and studying the similarities and differences to the spectral clustering approach.
2. Analyzing the subspace clustering algorithm in this context, which we expect to result in both algorithmic and statistical convergence guarantees.
3. **Something else here, like perhaps dealing realities of actual brain data, applying it, maybe missing data and corruptions, etc.**

1 Spectral Clustering versus Subspace Clustering

This is our first theoretical contribution. The connections are clear in certain problems, give example.

2 Subspace Clustering Analysis

Understanding the theoretical properties of the subspace clustering algorithm is crucial to creating better statistical tests for brain connectivity. The current state-of-the-art algorithms rely heavily on **more here and...**

Algorithmic Analysis The subspace clustering problem is non-convex and finding an optimal solution to the subspace clustering problem is NP-hard [?]. However, it is believed that under mild conditions separating the subspaces, properly initialized simple algorithms like the EM algorithm should provably succeed in identifying the true subspaces underlying the data. A second theoretical contribution of our work will be to pursue this theorem of algorithmic global convergence.

Laura will talk here about avenues we could pursue.

Statistical Analysis Consistency, of course.

3 Missing-Data Subspace Clustering (or general aspects of the reality of data)

Low-rank matrix completion allows one to estimate a single subspace from incomplete data, and this work has recently been extended for the subspace clustering problem [?, ?, ?]. However, the algorithm analyzed in [?] is computationally demanding and intractable in the number of clusters. While the algorithms in [?, ?] are computationally efficient, they have no convergence guarantees.

1. Adapting the brain region clustering problem to the framework of subspace clustering, and studying the similarities and differences to the spectral clustering approach.
2. Analyzing the subspace clustering algorithm in this context, which we expect to result in both algorithmic and statistical convergence guarantees.
3. Something else here, like perhaps dealing realities of actual brain data, applying it, maybe missing data and corruptions, etc.