Summary

Connectal coding: discovering the structures linking cognitive phenotypes to individual histories by Joshua T Vogelstein, Eric W Bridgeford, et al. 2019

Abstract

Cognitive phenotypes characterize our memories, beliefs, skills and preferences which actually aroused from ancestral and developed in history and have been written into our brain structure. Connectal coding is proposed to study the network structures that link cognitive phenomena to individual history.

In general, there are three parts we should mention when we trying to explain things in scientific way: 1) the constituent part. 2) the properties of those part 3) the interactions among them. In brain science, nowadays technology already done well in constituent part and properties part. The limitation is in interaction part. In order to know more about interactions in brain, we need the connectal coding technology.

Models of brain activity, is referred as neural coding. Models of brain connectivity, is referred as connectal coding. Connectal coding is a mode of brain investigation that is parallel and complementary to neural (activity) coding.

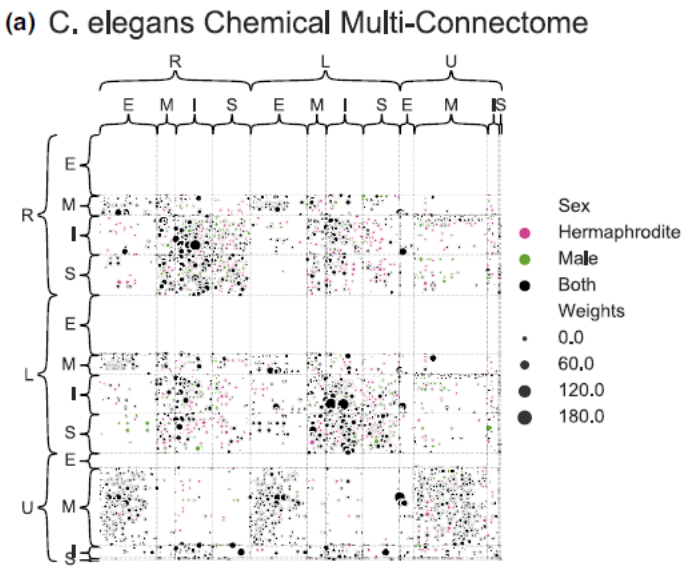
Modeling brains as networks (Definition)

The term “connectome” established by Sporns

Definition: A connectome is an abstract mathematical model of brain structure, denoted G, and is a set of two kinds of objects:

1. Vertices (or nodes), V: a vertex represents a biophysical entity of the brain. A set of constrains must be satisfied:
   a) The spatial extent
   b) The spatial resolution
   c) Type
   d) The development stage
2. Edges (or links), E: an edge between any pair of nodes represents the presence (and lack of edge represents the absence) of a connection or communication between those nodes. A set of constrains must be satisfied:
   a) The kind of communication (transmission of electrical charges, neurotransmitters …)
   b) The temporal duration under which these communications may be present.

We common to represent the connectome via a 2D array. Each row/column pair corresponds to a node, and edges between a pair of nodes u and v are depicted by a non-zero entry in the corresponding element of the array, that is, A(u,v) = 1.



Connectome has an attribute called *edge weights* such that each binary edge is associated with a magnitude that can take any continuous value, which makes the connectome can link to past and ongoing events.

Connectome actually have more attributes. For example, when edges correspond to synapses, they might be attributed with weights, locations, directions of transmission, neurotransmitters, and so on. Similarly, for connectomes, it is natural for nodes to have attributes, such as location, volume, and shape. Finally, when studying populations of connectomes, the entire graph may be endowed with attributes, such as a weight.

An implication of this definition is that one could simultaneously model a given brain with many different connectomes at different times, or at different resolutions, or of different types, and so on.

Example estimated connectomes (several different species connectomes)

We indicate the strength of connection as either the size or contrast of the corresponding matrix element. Showing an adjacency matrix requires first sorting all the nodes in some order; we choose to sort by region or type, and within that by degree (total weight of connections per node).

A. Caenorhabditis elegans(worm).
B. Drosophila melanogaster(fly).
C. Mus musculus(mouse).
D. Homo sapiens(human).

The purpose of brain codes

Neural activity coding can be thought of as converting information from past and ongoing events (stimuli and behavior) to neural activity, and from neural activity to future events (predictions and behaviors). In other words, neural activity codes correspond to the brain's representation of information.

In contrast, connectal coding can be thought of as converting information from past and ongoing events (ancestral, developmental, and experiential) to brain connectivity, and from brain connectivity to future events (behavioral tendencies). In other words, connectal codes correspond to the brain's storage of information.
Therefore, brain activity and connectal codes serve complementary roles in understanding the relationship between genetics, body, world, and brain.

Notice: Both codes are stochastics (a given ongoing stimuli/behavior can stochastically manifest in multiple different patterns of activity, and a given past event can stochastically manifest in multiple different patterns of connectivity.) The inverse is also true.

The role of connectomes in connectal coding

A connectotype is the collection of nodes and edges (and potentially their attributes) associated with a given phenotype. We consider two kinds of phenotypes here: individual histories and cognitive phenotypes. (In order to easily understand, we can use analogy. A genome is the complete genetic sequence of an individual, whereas a genotype is the part of the genetic sequence of an individual that associated with a particular phenotype.)

By individual histories, we mean ancestral, developmental, and experiential histories; we may desire to understand the relationship between connectome and genome, connectome and developmental stage, or connectome and experience.
By cognitive phenotype we mean a set of observable characteristics of an individual related to their cognition, including personality traits, memories, beliefs, skills, preferences, and psychiatric or learning disorders.

Because it's stochastics, our view on connectomics is that its primary value is in *generating hypotheses* about connectoypes. However, the process of hypothesis generation took long time. So connectomes may not on their own provide data sufficient to test hypotheses about how certain genotypes are linked to certain connectotypes and/or how certain connectotypes are linked to certain phenotypes.

Models of connectomes

Three categories of connectomes. (Actually only one of these frameworks is sufficient, although three provide complementary insights.)

A. Bag of edges --- most common approach

a) Treats each edge independently, without taking into account inter- actions or relationships between them.
b) However, this approach requires performing many statistical tests, which must be corrected for multiple comparisons to adequately control for the number of false positives.

B. Bag of features
a) multiple graph-wise or node-wise statistics are calculated and compared.
b) While computing these statistics can be informative about the properties of a given connectome, using them as features to explain differences between genotypes or phenotypes faces serious drawbacks.

C. Statistical modeling of networks
a) it is a model of the entire network, rather than just the edges or features, as a random variable.

Statistical models of connectomes

1. The simplest random network (graph) model is the ErdÅs-Re´nyi model, in which each edge is sampled identically and independently.
2. The next simplest binary models are stochastic block models (SBM).

There are more improved models listed in the paper, go and see if you are interested!

Statistical model for connectal coding

connectal coding model:
1.Let X and Y be random variables; P[X] and P[Y] is their marginal distributions, P[X,Y] is the joint distribution, P[Y|X] is the conditional distribution.
2.Four random variables:

- B = cognitive phenotypes of an individual, including and as measured by behaviors
- C = connectome of an individual, spanning spatial and temporal scales
- D = developmental history of an individual, including past experiences
- E = the current environment acting on individuals
- G = genome of an individual, including epigenetic factors

3. We may seek to estimate the probability of a connectotype, given a genotype and environment, P [C|D, E], and the probability of a cognitive phenotype, given a connectype and environment, P [B|C, E]. And other joint distributions.

Connectal coding theories

Two questions have to be asked:
1. To what extent is a connectome statistically associated with ‘X’ (e.g. geno-type or cognitive phenotype)
2. Where in the connectome is that statistical association (that is, what is the connectotype

associated with that genotype/phenotype)?

The first kind of question is essentially a hypothesis-testing question. The data required to answer it are two collections of estimated connectomes: C1, . . . Cn are estimated connectomes from individuals with one property (e.g. genotype or cognitive phenotype), and Cn+1, . . . , Cn+m are estimated connectomes from individuals with another property.

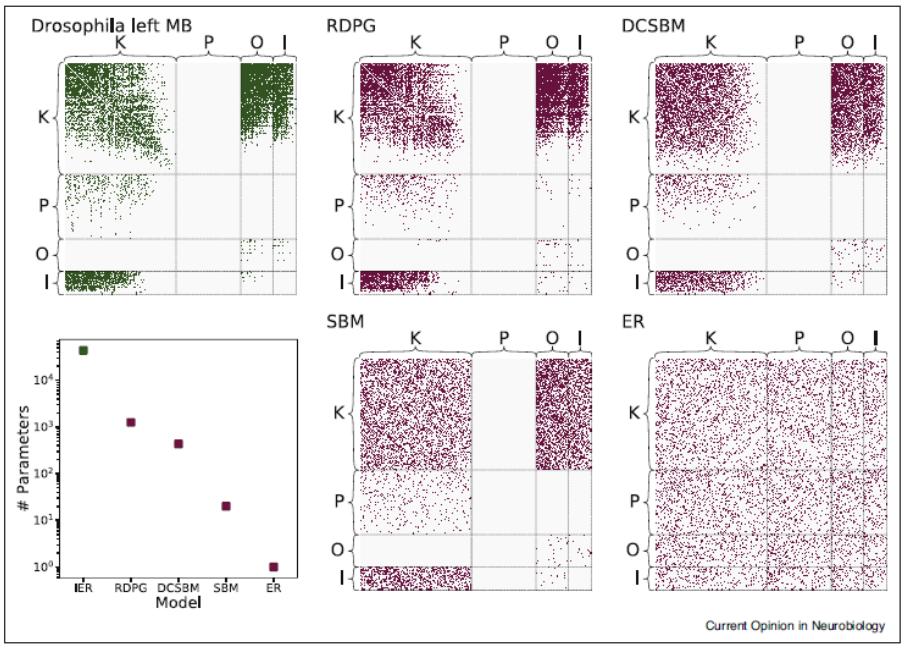
Assume that C1, . . . Cn are all sampled independently and identically from some distribution P0=P[C|X=0], and Cn+1, . . . , Cn+m are sampled independently and identically from another distribution P1 = P[C | X = 1]. Then, the formal statement of the hypothesis is: H0: P0 = P1 versus HA: P0 != P1

1.Evaluation: These tests are holistic: they tell the researchers that there are differences between these populations and the magnitude of those differences (the test statistic), but they do not indicate where those differences are.
2.In order to solve the second question, we introduce the 'signal subgraph' feature, that is, a small set of nodes and edges among them that confer the majority of the signal (the signal subgraph is an estimate of the connectotype). However, signal subgraph detection is an estimation, rather than testing: it seeks to estimate the smallest set of nodes such that the covariate is independent of the remaining nodes, and the signal subgraph is the set of edges among those nodes that carry information about the covariate.

Applications & Discussion

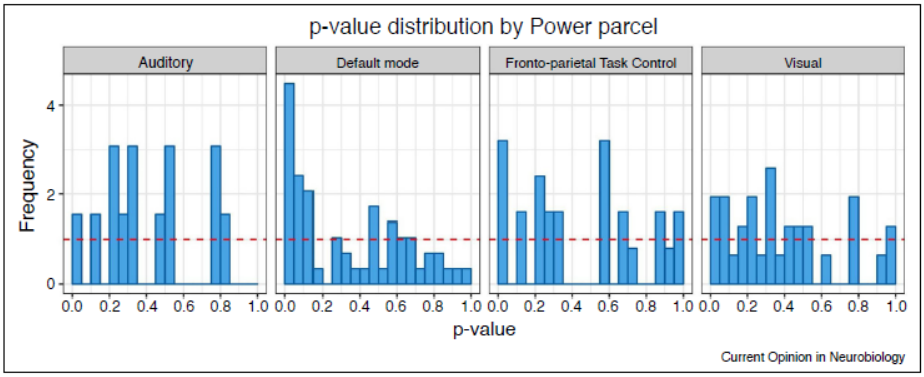
There are three applications. All suggesting that our framework can provide statistical rigor to support previous scientific claims .First, one can use connectal coding to study cognitive disorders, such as schizophrenia, as described above.

1. A widely held belief is that psychiatric illnesses are disorders of neural circuitry, or connectopathies. If true, our ability to develop clinically useful prognostic, diagnostic, and treatment protocols will depend on connectal coding.
2. Could be applied to study healthy brains as well.



Connectome model fitting and complexity. Left larval *Drosophila* mushroom body adjacency matrix, followed by random samples from four different statistical models of connectomes with decreasing complexity: inhomogeneous Erdős-Rényi (IER), random dot product graph (RDPG), degree-corrected stochastic block model (DCSBM), stochastic block model (SBM), and Erdős-Rényi (ER). The bottom left shows the number of parameters for each. All graphs are sorted by node degree within each block.

Figure 3



Normalized histograms of the distribution of *p*-values obtained from applying Hotelling's T^2 test to the omnibus embeddings of brain regions from a few selected Power parcels. The red dashed lines indicate the uniform density, which would be expected to hold if there were no difference between healthy and schizophrenic patients. The default mode network clearly displays a non-uniform *p*-value distribution, suggesting that this parcel differs in schizophrenic patients compared to their healthy counterparts. In contrast, *p*-values in the auditory, visual, and fronto-parietal task control subnetworks appear approximately uniform, providing weak evidence that these systems are not implicated in schizophrenia.

More useful references:

To facilitate using these ideas, we have developed an open source python toolbox for statistical analysis of populations of networks, available at: https:// neurodata.io/graspy/