

What is heritability?

Read slide

Brain connectivity as connectomes

We model the brain connectivity as connectomes aka networks or graphs.

Connectomes consists of vertices, representing regions of the brain and edges, representing a connectivity measure between pairs of vertices. In structural connectomes, which are what we are interested in, the edges represent the number of estimated neuronal fiber from one region to another. Lastly, we assume our connectomes are undirected, meaning edges have no direction.

Figure on the right shows a connectome overlayed on top of a brain, where dots are the vertices, and red lines are edges. However, this representation isn't particularly useful for analysis so we alternatively represent connectome as an adjacency matrix, which is square and symmetric.

How do we get structural connectomes?

First we obtain magnetic resonance images from human subjects. Then we preprocess the images to remove the skull and fix artifacts like head motion. Then we use linear registration to align the images into a common space. Next, we use tractography which estimates all the neuronal fibers in the brain, and with a brain parcellation that define the regions, we can construct our connectome, represented as an adjacency matrix here. Construction of connectome is simply counting the number of neuronal fibers that go from one region to another.

Heritability as causal problem

We want to pose the heritability analysis as a causal problem. So direct acyclic graphs or DAGs provide a visual representation of causal relationships among a set of variables. Our parents affects our genomes, and our ultimate goal is to estimate the existence of the red arrow. That is we want to know the causal relationship between genomes and connectomes which may also be affected by sex.

Do genomes affect connectomes?

And so we can formulate a hypothesis test, which says does fixing our exposures, in this case genomes, change our outcomes, which are connectomes? This can be alternatively phrased as, does the joint distribution of connectomes and genomes factor into its marginals?

This is known as independence testing, and our test statistic will be distance correlation. The implication if the hypothesis is false is that there exists some associational heritability effect.

I say associational because as we will see later, our DAG doesn't truly capture all our confounding.

What is distance correlation?

Distance correlation measures dependence between two potentially multivariate quantities, and in our case, these are connectomes and genomes. One of the main benefits of distance correlation is that it can detect nonlinear associations.

And so, the distance correlation measures the correlation of all pairwise distances among the observations of one particular quantity to those of another quantity.

The figure below shows our two quantities, connectomes and genomes. Using some measure of distances, we can compute a distance matrix for connectomes and genomes, which will be our input to distance correlation.

How do we compare genomes?

Typically, we do not actually observe genomes in twin studies. Even though we don't observe genomes, we can compare genomes between individuals under some assumptions.

One particular measure is the coefficient of kinship, which is the probability that we find the same gene at a particular location of the genome between subjects.

In the table, I list out the four possible familial relationships and their coefficient of kinship. Given our coefficients, we can define the distance between a pair of genomes as one minus twice the coefficient. This gives the distances zero for identical twins, half for fraternal twins and siblings, and one for unrelated.

How do we compare connectomes?

We leverage a statistical model for graphs called random dot product graph. In this model, each vertex is represented as low dimensional latent vector, or also called latent positions. We can estimate these vectors via adjacency spectral embedding, which is essentially a singular value decomposition.

In the figure, we have a simulated connectome represented as an adjacency matrix. After we apply adjacency spectral embedding, we obtain a two dimensional latent positions, where each dot represents a vertex.

And so, the distance between connectome k and connectome l is given by the Frobenius norm of the difference of latent position matrices after computing some rotation matrix R .

Human connectome project

Alright. So we have our causal model, a hypothesis to test, and a procedure for performing the hypothesis test. Now all we need is the data.

We use the human connectome project, which obtained brain scans from identical, fraternal, and non-twin siblings. The table below shows the demographic information. Using these brain scans, we obtain a connectome per subject using the Glasser parcellation, which divides the brain into 180 regions.

Genome and connectomes are dependent

We first perform a qualitative examination of the distances of connectomes. In the figure, we plot the pairwise connectome distances stratified by the familial relationship, and the diamond represents the median. This is visualizing the values of the connectome distance matrix.

Of course, if the genome does affect connectomes, we should expect the distances between pairs of monozygotic twins should be stochastically smaller than those from dizygotic twins, which should be stochastically smaller than those from non-twin siblings, etc. And the median certainly grows larger as the genomic distances increase. We observe this pattern when we further stratify by sex.

Our independence test are all significant, meaning we can reject the hypothesis that genomes have no affect on connectomes.

Neuroanatomy

Turns out, there is a lot of literature that state that the neuroanatomy is also heritable. That is, measurements like brain volume is highly heritable. So we measure the brain volume per region, as well as diffusivity measure that describe neuroanatomy.

So we want to test another hypothesis, that the genome does not affect neuroanatomy using distance correlation. The implication if the hypothesis is false is that previous result on the effect of genome on connectomes can be explained by neuroanatomy, and it should be included in the causal model. We can compute the distance between neuroanatomy via the Frobenius norm.

The figure below shows our two quantities, neuroanatomy and genomes. And, again the distances matrices will be the input to distance correlation.

Genome and neuroanatomy are dependent

We again perform a qualitative examination of the distances of neuroanatomy. Similar to the figure with connectome distances, we observe that the median increases as we increase genetic distances, which remains even after stratifying for sex.

And we have that all tests are all significant, meaning that we should add neuroanatomy into our causal model and control for neuroanatomy.

DAG including interactions of neuroanatomy

This DAG now contains the effects of neuroanatomy, which is certainly influenced by the genome, and neuroanatomy may also have an affect on the connectomes. In addition, neuroanatomy may be affected by other factors like the environment, but by controlling for neuroanatomy, we can ignore the other effects.

Do genomes affect connectomes given neuroanatomy?

So now we want to test whether the joint distribution of genome and connectomes conditioned on neuroanatomy can be factored into its conditional marginals. This is known as conditional independence test and the test statistic is given by conditional distance correlation. The implication if the hypothesis is false is that there exists causal dependence of connectomes on genomes.

What is conditional distance correlation?

In conditional distance correlation, we augment the procedure with a third distance matrix, given by the distances of neuroanatomy.

Connectomes are still dependent on genome

Again, we observe that all our tests are significant, meaning that we reject the null that there is no causal relationship between genomes and connectomes. That is, even after controlling for the effect of neuroanatomy on connectomes, there still exists a dependence between genome and connectomes.

Summary

We wanted to explore whether genomes affect connectomes. We leveraged recent advances such as statistical models for networks, allowing meaningful comparison of connectomes and the use of distance and conditional distance correlation as test statistic for causal analysis, which is the second of its kind. And lastly, we show that there is heritable effect of connectomes.