# Connectal Coding: Discovering the Structures Linking Cognitive Phenotypes to Individual Histories

Joshua T. Vogelstein[1] Eric W. Bridgeford[1], Benjamin D. Pedigo[1], Jaewon Chung[1], Keith Levin[2], Brett Mensh[3], Carey E. Priebe[1]
[1]Johns Hopkins University, [2] University of Michigan, [3]Janelia Research Campus

**Cognitive phenotypes characterize our memories, beliefs, skills, and preferences, and arise from our ancestral, developmental, and experiential histories. These histories are written into our brain structure through the building and modification of various brain circuits. Connectal coding, by way of analogy with neural coding, is the art, study, and practice of identifying the network structures that link cognitive phenomena to individual histories. We propose a formal statistical framework for connectal coding and demonstrate its utility in several applications spanning experimental modalities and phylogeny.**

The premise of connectal coding is that a network, consisting of nodes and edges among them, is a reasonable and useful model of brain structure [1,2]. This model of brain structure is agnostic to scale: the model can be appropriate from the molecular scale to the whole system (and beyond) [3–5]. Although a comprehensive model of the brain includes both brain activity and connectivity, each of activity and connectivity can fruitfully be modeled independently. Much like the goal of brain activity modeling (neural coding) is to link patterns of brain activity to past, ongoing, and future events [6]; so too the goal of brain connectivity modeling (connectal coding) is to link patterns of connectivity to past, ongoing, and future events. The nature of the patterns, events, and links change by virtue of switching focus from activity to connectivity. Moreover, the statistical models one can leverage to learn those links from the data must also change.

The goal of this manuscript is to introduce in clear terms, motivate from first principles, and formalize this emerging approach to studying the brain. While neural activity coding is well established and widely accepted as a (possibly *the*) legitimate framework for studying the brain, connectal coding remains in its infancy. Below we outline our rationale for why connectal codes are not just valuable, but required for a comprehensive understanding of cognitive mechanisms, and how to learn them.

# Modeling Brains as Networks

The idea of the brain as a network dates back to the late 19th century, with Ramon y Cajal's discovery of dendritic spines using Golgi's stain. This marked the founding of the "neuron doctrine", for the first time asserting that the brain is not, in fact, a syncytium [7]. In 1943,

McCulloch and Pitts wrote down a mathematical model of the brain as a network, and proved that such networks could be "Turing-complete" (that is, they can solve any computational problem) [8], thereby founding the field of artificial intelligence [9]. Shortly thereafter, Hebb introduced "cell assemblies" as essentially networks of neurons that are jointly active [10]. A few decades later Little [11], and then Hopfield [12], introduced Little-Hopfield networks (which were the first recurrent neural networks and led to the founding of connectionism [13]). Little and Hopfield leveraged an idea from statistical physics known as the Ising model [14], positing that these assemblies could be used to store information. But in the entire 20th century, it was rare to use the tools of "graph theory" to study brain networks, even though graph theory was founded by Leonhard Euler in 1736 [15].

This all changed with the introduction of the term "connectome" by Sporns *et al.* and Hagmann in 2005 [16,17]. Since then, >3,600 papers have been published using the term. Because many of those papers define connectome differently, we provide our definition below:

**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. In a connectome, one defines a set of constraints, including the spatial extent (e.g., the left mushroom body, a cortical column, or the whole brain), spatial resolution (e.g., a cell, cellular compartment, or a cellular ensemble), type (e.g., a neural, glial, or perivascular cell), and developmental stage (e.g., postnatal day six). The nodes of a connectome are all the nodes satisfying those constraints.
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. In a connectome, that connection/communication must satisfy another set of constraints, including the kind of communication (e.g. transmission of electrical charges, neurotransmitters, physical opposition, or fiber bundles) and the temporal duration under which these communications may be present (e.g., during a particular developmental phase, or during a traumatic injury including brain compression, etc.). The edges of a connectome are all the edges satisfying these constraints between the above described set of nodes. Note that implicit in this definition is that the connections for which edges count in connectomes are direct.

Under this simplest definition of what constitutes a connectome, it is common to represent the connectome via a two-dimensional (2D) array, A (Figure 1). In this representation, 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, i.e., A(u,v)=1. It is tempting to think of this as a matrix, and it certainly is from the computer science perspective. But it is decidedly *not* a matrix from the mathematics perspective, where matrices are linear operators, e.g. y = Ax. Whereas matrix algebra can be applied to study these representations of graphs, we often find it helpful to keep in mind that these "adjacency matrices" are special kinds of 2D arrays. The row identities are inextricably linked to column identities, a property that is not generally true for arbitrary matrices, and changes the kinds of statistical procedures appropriate for these data.

Moreover, in classical graph theory, the graph is simply the tuple G=(V,E), and lacks additional structure. For connectomes to serve as a model for brain structure that can link to past, ongoing, and future events, they often require some additional structure. The most common additional structure is *edge weights*, such that each binary edge is associated with a magnitude that can take any continuous value. More complicated and nuanced edge attributes abound in connectomics. For example, when edges correspond to synapses, they might be attributed with weights, locations, directions of transmission, neurotransmitters, etc. Similarly, for connectomes, it is natural for nodes to have attributes. The most common attribute of a node is a semantic label, such as the Mauthner neuron [18], or Purkinje cell, or primary visual cortex. Nodes in connectomics, like edges, can have other attributes as well, such as location, volume, and shape. Finally, when studying populations of connectomes, the entire graph may be endowed with attributes, such as a weight.

The above definition clarifies what a connectome is for this manuscript, but not what it is not. It is not many things; we illustrate a few. The shape, size, or morphology of a set of brain regions or neurons does not comprise a connectome, nor does the set of all spike trains of a brain. These features can be attributes of nodes, but without also characterizing the edges, one has not modeled a connectome. Similarly, correlation between a pair of nodes cannot be defined as an edge (though it may be used to estimate the presence of an edge). This is because connectomes, as defined above, are models of brain structure and anatomy, and correlation is an emergent property of dynamics on that model, rather than a description of the model itself. While many connectomics papers disagree with this definition, we find it useful, and use it for the remainder of the manuscript.

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, etc. Moreover, the neuron-level model of brain structure does not have an elevated status over, say, compartments of neurons, glia, or brain regions; rather, each scale and type of node can serve as a perfectly adequate model of part of the brain, and those parts may turn out to be the most important parts to explain any particular past or future event.

Given all this, one can measure parts of a brain to estimate a connectome from many different experimental modalities, each of which has its shortcomings. At cellular resolution, as early as 2001, several papers began characterizing how to estimate connectomes (without using the word connectome) from physiology data [19–22]. More recent developments in cellular resolution connectome estimation incorporated unobserved variables [23,24], called “confounders” in the causal inference field [25]. At the millimeter scale, diffusion magnetic resonance imaging data (dMRI) is known to exhibit both false positives and false negatives, as compared to gold standard methods [26]. By the same token, one can use functional MRI (fMRI) data to estimate edges. The most common approach by far is to simply use correlations [27]; these approaches are also known to yield problematic estimates for a number of reasons, including susceptibility to various exogenous variables, such as time of day, week, and month [28]. Methods for addressing confounders have also been proposed in the fMRI literature [29], and while widely cited, are still largely underappreciated.

# Example Estimated Connectomes

We show several different previously published estimated connectomes from different species spanning the phylogenetic tree, with widely disparate definitions of what constitutes nodes and edges (Fig. 1). There are many possible ways to visualize a connectome [30–32]; we choose a relatively simple way, adjacency matrices, as described above. 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), but other sortings could be equally informative. For certain connectomes, experimentalists have measured multiple types of edges. We show these "multi-connectomes" with different colors for the different edge types.

A. **Caenorhabditis elegans (worm).** C. elegans is the only animal for which we have estimated a complete connectome where the nodes represent neurons [1,33,34]. That is, every neuron in the animal is a node, and every edge has been estimated. This connectome has two types of edges corresponding to two types of neural activities: (1) chemical synapses for neurochemical release (Fig. 1A), and (2) gap junctions for electrical activity. Each edge's strength (or weight) corresponds to the approximate total volume of synapses between its parent neurons. C. elegans has two sexes, male and hermaphrodite, with different numbers of neurons (hermaphrodite 302, male 385, with 290 overlapping shown in Fig. 1A) [35,36]. These connectome estimates are derived by cumbersome manual tracing of axons and dendrites and identification of synapses, in nanoscale electron micrographs, updated by Varshney et al. [37], Bentley [38], and most recently by Cook et al. [39].
B. **Drosophila melanogaster (fly).** Eichler et al. [40] published an estimate larval Drosophila connectome of the left (Fig. 1B) and right mushroom body, also derived from serial electron microscopy, using only chemical synapses. These edges are weighted (based on counting the number of synapses between a pair of neurons), and directed (meaning connections can be from u to v and not vice versa) [41].
C. **Mus musculus (mouse).** Calabrese et al. [42] generated a high-resolution connectivity estimate using ex vivo diffusion magnetic resonance imaging (dMRI). This network is undirected, as dMRI lacks directional information, and weights correspond to the number of tracts estimated to go between regions. Because we do not know the mapping from the absolute magnitude of connection weights to any physical connection, we rescale these weights to be between zero and one, and depict them on a log scale (Fig. 1C).
D. **Homo sapiens (human).** The Consortium for Reliability and Reproducibility collects multiple measurements of functional resting-state, anatomical, and/or diffusion magnetic resonance imaging (MRI) per individual [43]. The functional estimated connectomes are Pearson correlation matrices, converted to ranks and then normalized

between zero and one, because such a representation is more reliable than raw or thresholded correlations [44]. The diffusion estimated connectomes are normalized as described above for the mouse. This multi-connectome estimate is derived from averaging the entire dataset of 3,067 diffusion and 1,760 functional MRI connectomes [43] (Fig. 1D).

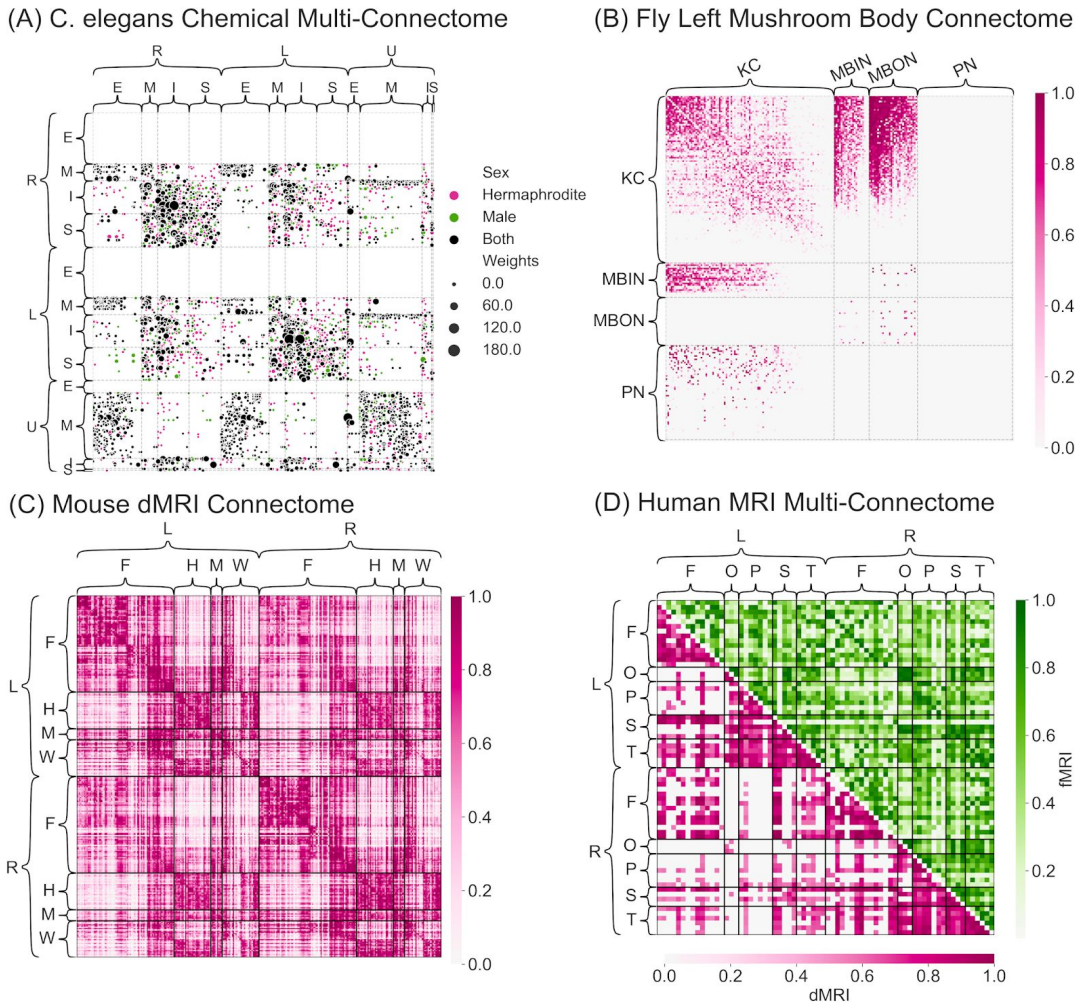


***Figure 1****: Estimated connectomes spanning four levels of the phylogenetic tree, each estimated using different experimental modalities and spatial resolutions, ranging from nanoscale (electron microscopy) to macroscale (MRI regions).* ***(A)*** *C. elegans chemical multi-connectome estimated from male and hermaphrodite connectomes. Size of circles corresponds to the number of synapses between two neurons. Multiscale node labels: left (L) and right (R), which are bilateral pairs, as well as unpaired (U); and four types per side: endorgans (E), interneurons (I), motor neurons (M), and sensory neurons (S).* ***(B)*** *Left mushroom body connectome estimated from Drosophila melanogaster. Nodes represent neurons, and are assigned into kenyon cells (K), input neurons (I), output neurons (O), and projection neurons (P). Color intensity corresponds to the number of synapses between two neurons.* ***(C)*** *Mouse estimated connectomes obtained from dMRI scans. Nodes represent regions of the brain, and are assigned into right (R) and left (L) hemispheres and then further assigned into superstructures such as frontal (F), hindbrain (H), midbrain (M), and white matter (W). Color intensity corresponds to degree of connectivity between two regions.* ***(D)*** *Human multi-connectome estimated from averaging 3,067 dMRI and 1,760 fMRI human connectomes.*

*Since both networks are undirected, only the upper triangle of fMRI connectome and lower triangle of dMRI connectome is shown. Nodes represent regions of the brain, and are assigned into right (R) and left (L) hemispheres and then further assigned into frontal (F), occipital (O), parietal (P), and temporal (T), as well as subcortical structures (S). Color intensity corresponds to degree of connectivity and correlation for dMRI and fMRI, respectively, between two regions.*

# The Purpose of Brain Codes

A code is a (potentially stochastic) system of rules to convert information from one representation into another [45,46].[1] For example, 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.

An important aspect of both activity and connectal codes is that they are stochastic. In other words, 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: activity/connectivity can represent multiple different current/past/future events, respectively. This stochastic property of brain codes is actually required for operating in the world. Even human brains, given finite physical and energetic resources, are incapable of representing and storing the large amount of information impinging on our sensory faculties. In a similar fashion, the connectivity of a human brain is too complex to be explicitly prescribed by their genome. To be concrete, consider that a human brain contains approximately $10^{11}$ neurons [48] and $10^{15}$ connections between pairs of neurons, and yet we only have about $10^{4}$ genes [49]. Thus, for the genome to encode every single synapse would require five to six variants per gene, and literally every variant of every gene would encode the brain's synapses. More likely, the genome encodes the "blueprint", that is, a number of statistical principles governing the probability of connections between nodes across development, as well as all the rules for learning new connections due to activity-dependent plasticity. Those rules are the principles of connectal coding.

# The Role of Connectomes in Connectal Coding

The above definition of connectome sets the stage for understanding the relationship between connectomes and other aspects of an individual or population. By way of analogy,

[1] We acknowledge that it is common in the neuroscience community to ascribe other meanings to the word code, including representational and causal meanings [47]. We use "code" the way Shannon and information theory uses it, purely as a statistical relationship.

recall that 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. In that sense, 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*. 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. We may desire to understand the relationship between connectome and these phenotypes as well. Regardless of which particular kind of phenotype, there may exist one (or many) connectotypes associated with it. Connectal coding is the study of brain structures that encode that information.

Like genotypes, there is not a one-to-one mapping between connectotypes and phenotypes, rather, a given connectotype could stochastically encode in different cognitive phenotypes at different times, and a given cognitive phenotype could be associated with many different connectotypes. For example, the stomatogastric ganglion circuit of a crab can exhibit similar network activity from disparate circuit parameters [50].

In light of this, our view on connectomics is that its primary value is in *generating hypotheses* about connectotypes. This is in contrast to the relatively low-throughput, more classical approach to studying neural circuits. For example, detailed physiological and anatomical characterization of the sound localization circuit of a barn owl [51] effectively confirmed models that explained a certain phenotype. However, the process of hypothesis generation was arduous, taking decades, including numerous models that, in retrospect, were not possible given the underlying neuroanatomy. If one could rapidly collect data about many interwoven neural circuits, and then use those data to screen and/or filter various circuits as implicated (or not) in a given behavior, then subsequent experiments could further refine those results. Thus, 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. Nonetheless, if these data can accelerate the hypothesis-generation process to hone in on a small set of plausible models, they will be extremely valuable.

Of note, in connectal coding, the role of estimating connectomes is not about describing the basic anatomical properties of connectomes, or modeling connectomes as a end unto themselves. Rather, in connectal coding, connectomes are interesting insofar as they participate in the understanding of the relationship between brain structure and individual histories or cognitive phenotypes. Part of the rationale of this focus is that most current experimental approaches for estimating connectomes are so error prone, that estimates of the statistical properties of brain networks are difficult to interpret in terms of the underlying biology. In fact, even if the part of the network that is observed is largely correct, if it represents just a subsample of the network of interest, then the resulting network features can be quite different from those features of the entire network [52,53]. Regardless, building statistical model relating connectypes to phenotypes requires models of connectomes.

# Models of Connectomes

Every connectomics study utilizes some mathematical and statistical approach to support the scientific claims. We organize those approaches into three categories, and demonstrate that only one of these frameworks is sufficient for connectal coding, although all three provide complementary insights and perspectives on the connectomes themselves.

The most common approach, dubbed the "bag of edges" (or edgewise statistics) framework. [3,54], treats each edge independently, without taking into account interactions or relationships between them. Such univariate approaches allow researchers to identify easily interpretable relationships between phenotypes and edge weights. However, this approach requires performing many statistical tests, which must be corrected for multiple comparisons to adequately control for the number of false positives. Standard correction techniques such as false discovery rate [55] do not model the dependencies between edges, and therefore may result in overly liberal or conservative corrections [56]. Alternate correction techniques such as network-based statistics [57] or group Benjamini-Hochberg corrections [58] leverage information about the group structure of connectomes to increase statistical power, while attempting to control false positives. Network-based statistics, however, lacks theoretical guarantees identifying the settings in which it successfully controls false positives. Without such an understanding, interpreting its results is problematic. Benjamini-Hochberg has strong theoretical guarantees, but in models that are inappropriate for connectomics data, in that their assumptions are often grossly violated (and rarely checked). Bonferroni corrections are widely believed to be overly conservative [59] and therefore lack the sensitivity desired for connectal coding.

A second popular approach we dub the "bag of features" framework. In this approach, multiple graph-wise or node-wise statistics are calculated [60] and compared. Possible features include degree distribution, degree sequence, clustering coefficient, number of triangles or other motifs, small worldness, efficiency, and modularity [61]. 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. First, for any particular feature, vastly different networks can produce the same value [62]. Second, for a connectome with n nodes, there are $2^{n*n}$ possible subgraphs, each of which could reasonably be considered a feature. Therefore, one cannot reasonably search all features (as $2^{n*n}$ is larger than the number of atoms in the universe for n > sixteen!). It is therefore unclear (and somewhat arbitrary) how one should choose relevant features for a particular dataset. Third, and perhaps most problematic, is that the different features are not typically independent of one another. Therefore, if the question is whether a given phenotype depends on a particular feature being a specific value, it is impossible to determine whether *that* particular feature is responsible. Rather, the phenotype may be dependent upon a subset of the exponentially many features that correlate with both the feature of interest and the phenotype. No experiment could test whether

a feature is uniquely informative with regard to the covariate of interest, even in theory. Thus, studying most network features will fail to yield the principles of connectal coding.

The third framework, which is much less popular in connectomics, but much more popular in network statistics, is statistical modeling of networks [63,64]. The key conceptual herdle associated with using statistical models of networks to model connectomes is that it is a model of the entire network, rather than just the edges or features, as a random variable. This network is a complex high-dimensional random variable, with built-in structure and relationships. The vast majority of work in network modeling focuses on modeling a single network, typically undirected, unweighted, and lacking in network-wise, node-wise, or edge-wise attributes [65]. However, connectomes are typically weighted, sometimes directed, and always include at a minimum node-wise labels (e.g., which cell, cellular compartment, or cell ensemble corresponds to a given node). Moreover, understanding the relationship between connectypes and phenotypes typically requires comparing multiple connectomes. Much of the existing work comparing multiple networks ignores the unique node labels that are often available in connectomics [66]. Therefore, a full accounting of connectotypes will require statistical models of populations of networks with complex attributes. Although a comprehensive theory remains absent, we build connectal coding on the foundational work of network modeling [67].

## Statistical Models of Connectomes

The simplest random network (graph) model is the Erdos-Renyi model, in which each edge is sampled identically and independently [68]. This binary model is the connectal coding homolog to the Poisson process in neural coding, which asserts that a neuron's spikes are sampled identically and independently [69]. Although these models are too simple to explain much, they are excellent starting grounds to build more complex models, such as the inhomogeneous Erdos-Renyi model, which assumes that each edge is an independent coin flip, but each edge can be sampled with a different probability [65]. The next simplest binary models are stochastic block models (SBM) [70–73]. In an SBM, there are K groups, and each node is in one group. Each edge is again sampled independently, but the probability of a connection between a pair of nodes is no longer equal everywhere; rather, it is determined by each node's group (homologous to a heterogeneous Poisson process, where spiking has K different rates, dependent on which of K different states the brain is in). For example, a simple model of brain connectivity would be that contralateral connections have some probability, and ipsilateral connections have a different one. One can further generalize this to a hierarchical block model, where each node is in a given set of nested groups [74–77]. For example, a node might be in a lobe within a hemisphere.

A further generalization asserts that each node is in its own group, and therefore has a "latent position" that characterizes its probability of connecting with other nodes (homologous to latent variable models in neural coding) [78]. A particularly popular version of these models assumes that the probability of connections between a pair of nodes is equal to the dot product between the nodes' latent positions [79–81]. In these models, an extensive set of theoretical investigations have established the kinds of claims we desire when using a statistical model to

make inferences about our data [82,83], as well as a number of extensions, including a generalized random dot product [84], a random dot product with node-wise covariates [85], and a latent structure model [86] (for review, see [87]). However, these models typically only operate on single, unweighted networks lacking attributes.

Some of these single-network models have been generalized to population models. One of the first such models was a non-parametric Bayesian model of populations of networks [88]. This model is essentially a network generalization of mixed effects models popular in biostatistics [89], where the mean network is a fixed effect, and each individual has a unique low-rank distortion relative to the mean. Two extensions of this approach are essentially non-Bayesian variants that enable faster computation [90,91], which was generalized to a mixture of random dot product models [92]. Estimation in each of these models can be thought of as specific tensor factorizations [93]. Although these models still lack much neuroscientific insight or attributes, they establish the statistical foundation for learning connectal codes.

# Statistical Model for Connectal Coding

To formalize connectal coding, we introduce the *connectal coding model*, which is designed to enable investigation of the links among connectypes, individual histories and cognitive phenotypes. Like all models, the connectal coding model makes simplifying assumptions for tractability. Although statisticians like to remind practitioners of George Box's quip, "all models are wrong; some are useful", we find this aphorism to be misleading. Models are maps, designed to get us from one place to another. The question is not "is a given map right or wrong?" but rather, "is the map useful for getting us to our destination?" The connectome coding model is designed to get us to a deeper understanding of the relationship among connectotypes, individual histories, and cognitive phenotypes. Insofar as it does that, it is useful.

In statistics, a code is a conditional distribution characterizing the probability of one random variable taking some value given another random variable taking some value. Let X and Y be random variables; their marginal distributions, P[X] and P[Y], characterize the probability of any particular x or y, and their joint distribution P[X,Y] characterizes the probability of observing x and y. The conditional distribution P[Y|X] characterizes the probability that Y takes a particular value, given that X is some value. In connectal coding, we have the following 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, and
- G = genome of an individual, including epigenetic factors.

Connectal coding concerns estimating the statistical relationships among connectomes and genomes, developmental histories, cognitive phenotypes, and the current environment. More formally, in connectal coding, 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]. We are also interested in P[C | D], P[B | C], and other conditional and joint distributions. In all cases, there exists a random variable that models the connectome, which therefore warrants further study. One random variable which may appear missing to many neuroscientists is brain activity. Much like one does not require modeling a connectome to characterize the relationship between stimuli/behaviors and brain activity, one does not require modeling activity to characterize the relationship between cognitive phenotypes/developmental history/genomes and brain connectivity. Although joint modeling of brain activity and connectivity would be more comprehensive, connectal coding is sufficiently challenging and interesting to warrant its own investigations.

This formalization of connectal coding, is, to our knowledge, novel. That said, this conceptual and formal model can be used to interpret previous connectomics studies, which often have similar conceptual frameworks [94]. More importantly, we hope it will help guide the development and ideation of new connectomic studies, by providing a paradigm within which to ask, formalize, and eventually answer questions about neural circuit mechanisms, how they work, how they fail, and how they can be improved.

## Connectal Coding Theories

Equipped with the statistical formalization of connectal coding, two questions we can ask are: (1) to what extent is a connectome statistically associated with "X" (e.g., genotype or cognitive phenotype), and (2) where in the connectome is that statistical association (that is, what is the connectotype associated with that genotype/phenotype)? An example of the first question is, "if the genomes of two individuals differ, to what extent are their connectomes different?" An example of the second question is: "where in the connectome are those differences?" Causal questions, such as which connectotype is part of the implementation-level causal mechanism of a given cognitive phenotype, are also possible, but beyond the scope of this article because they are much more difficult to answer convincingly (but see [94] for a wonderful article on this topic focused largely on fMRI based connectome estimates).

Formally asking these questions requires assuming a statistical model. The first kind of question is essentially a hypothesis-testing question. The data required to answer it are two collections of estimated connectomes: $C_1,...C_n$ are estimated connectomes from individuals with one property (e.g., genotype or cognitive phenotype), and $C_{n+1},..., C_{n+m}$ are estimated connectomes from individuals with another property. Assume that $C_1,...C_n$ are all sampled independently and identically from some distribution $P_0 = P[C \mid X = 0]$, and $C_{n+1},...,C_{n+m}$ are sampled independently and identically from another distribution $P_1 = P[C \mid X = 1]$. Then, the formal statement of the hypothesis is:

$$H_0: P_0 = P_1 \qquad \text{vs.} \qquad H_A: P_0 \neq P_1$$

Applying such a test requires defining a test statistic, which quantifies the effect size, and computing the distribution of the test statistic under the null. Such two-sample tests have been

developed for random graphs [95–98]. The theoretical claims associated with each of these two-sample testing results depends on an underlying statistical model of random graphs, such as those described in the previous section. 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.

To answer the second question, global graph features (such as modularity) are inadequate because they cannot indicate *where* the differences are. Moreover, edge-wise statistics are typically underpowered when suitably adjusted for multiple comparisons, given our small sample sizes. Instead, we can search for a "signal subgraph", 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). Signal subgraph searches are akin to feature screening, where one seeks to determine which features are most informative about a specific covariate [99]. The main difference is that signal subgraph methods take advantage of the graph structure to improve their sensitivity and specificity [100]. The first signal subgraph method used a variant of a sparse inhomogeneous Erdos-Renyi random graph model [101]. These methods have been extended to operate under latent variable model assumptions [102], and also to deal with continuous (rather than categorical) covariates [103]. Formally, signal subgraph detection is an estimation, rather than testing, problem: 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

Consider the connectome of the larval *Drosophila* mushroom body [40] (Figure 1B). Priebe et al. [41] conducted an extensive empirical investigation of this connectome, leveraging spectral modeling, and resulting in the development of the latent structure random graph model [86]. Specifically, they discovered that kenyon cells of the mushroom body form a one-dimensional submanifold in a six-dimensional latent space. Figure 2 shows the quality of fit of several of these models to a binary simplification of the left larval mushroom body *Drosophila* data. These models provide a foundation from which to formulate statistical tests to answer the connecal coding questions described above.

Second, consider sex differences in the mouse brain estimated via diffusion magnetic resonance imaging (Fig. 1C). From scans of 55 mouse brains, 32 male and 23 female, connectomes were estimated on 332 nodes, 166 per hemisphere, using the Waxholm atlas [104]. The signal subgraph method of Wang et al. [102] reveals that 10 of the original 332 nodes contain more signal than noise about sex. The top-ranked nodes include a thalamic component and the periaqueductal gray, both of which are important in sexually dimorphic mouse brain development [105,106].

Third, we consider the COBRE data set [107], a collection of 123 functional MRI scans of schizophrenic and healthy human patients. Each scan yields an estimated connectome with 264 nodes, corresponding to 264 brain regions of the Power parcellation [108], with edge weights

given by correlations between BOLD signals measured in those regions. The data set contains scans for 54 schizophrenic patients and 69 healthy controls (like the one shown in Figure 1D). Levin at al. [109] apply their omnibus embedding to jointly estimate a random dot product graph for each connectome, resulting in an estimated three-dimensional latent position for each node of each connectome. For each node, Hotelling's $T^2$ test yields a p-value assessing whether or not the latent positions from healthy connectomes are drawn from the same (normal) distribution as the latent positions from the schizophrenic connectomes. Because the Power parcellation is hierarchical, the nodes can be further organized into 14 parcels. Visualization of the distribution of these p-values within each parcel suggests that the connectomes of schizophrenics differ from those of healthy controls in certain subnetworks but not others (Fig. 3). Specifically, nodes in the default mode network tend to have significantly smaller p-values than nodes in, for example, the visual subnetwork. Independent investigations also implicate the default mode in schizophrenia [110,111], suggesting that our framework can provide statistical rigor to support previous scientific claims.

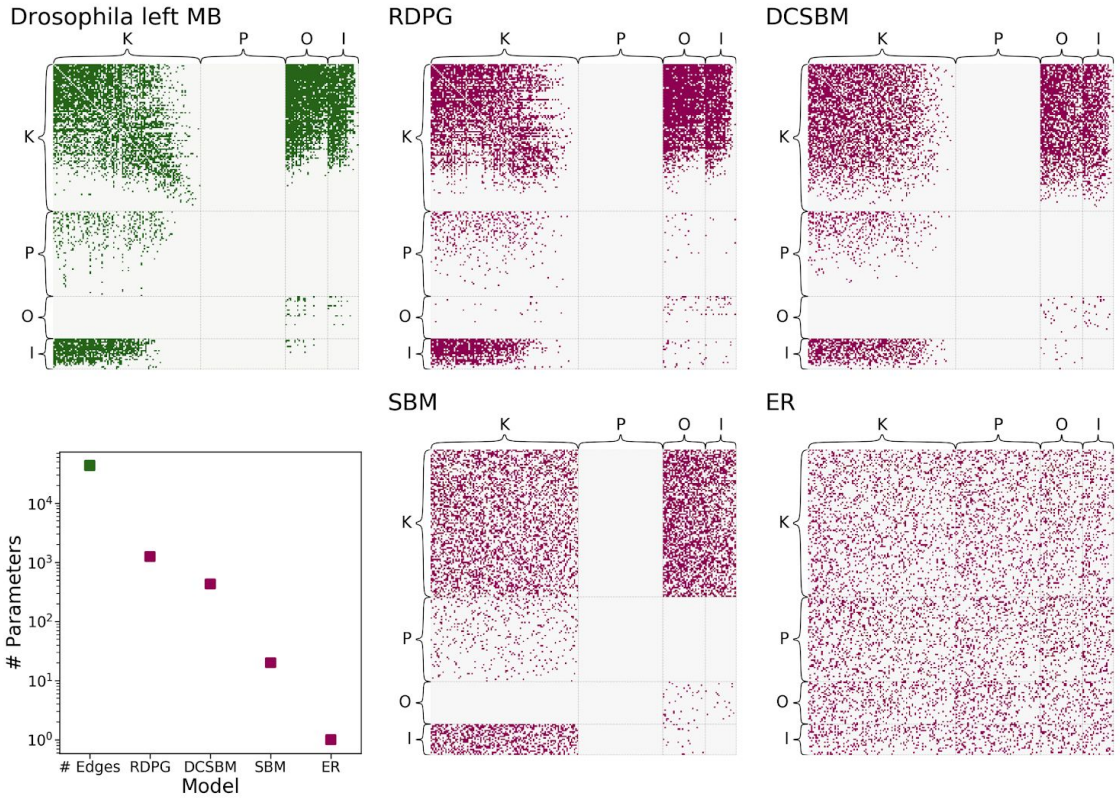


***Figure 2**: Connectome model fitting and complexity. Left larval Drosophila mushroom body adjacency matrix, followed by random samples from four different statistical models of connectomes with decreasing complexity: random dot product graph (RDPG), degree-corrected stochastic block model (DCSBM), stochastic block model (SBM), and Erdos-Renyi (ER). The bottom left shows the number of parameters for each, as compared to the 40,000+ number of edges for the connectome. All graphs are sorted by node degree within each block.*

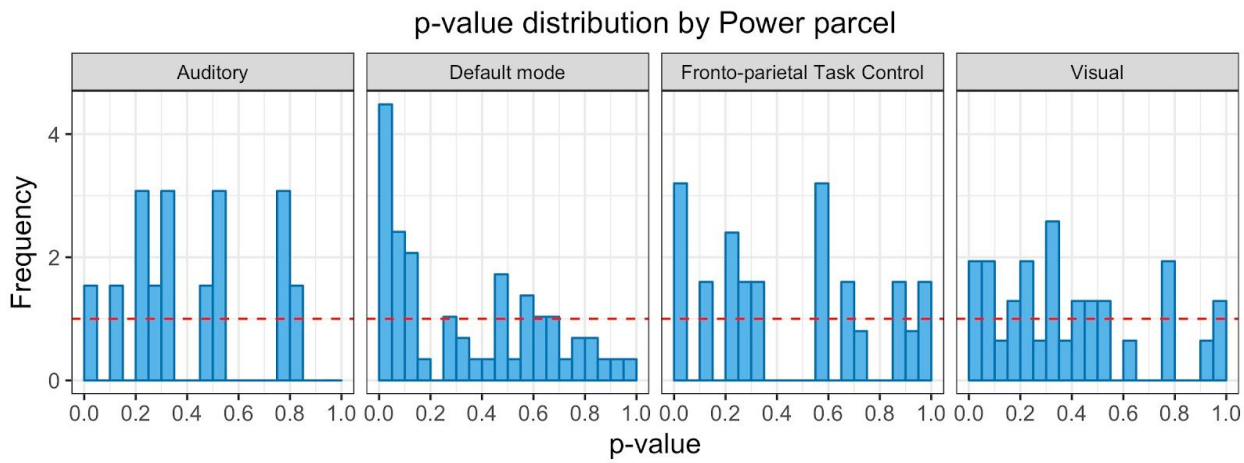


***Figure 3**: Normalized histograms of the distribution of p-values obtained from applying Hotelling's $T^2$ test to the omnibus embeddings of brain regions from a few selected Power parcels. The red dashed lines indicate the uniform density, which would be expected to hold if there were no difference 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.*

# Discussion

The connectome coding framework can be widely applied. Here, we discuss a few additional applications. First, one can use connectal coding to study cognitive disorders, such as schizophrenia, as described above. The consensus report of the American Psychiatry Association's working group on neuroimaging markers of psychiatric disorders concluded that "there are currently no brain imaging biomarkers that are currently clinically useful for any diagnostic category in psychiatry." [112]. This is despite nearly 30 years (at the time of that report) of brain imaging. A widely held belief is that psychiatric illnesses are disorders of neural circuitry, or connectopathies [113–118]. If true, our ability to develop clinically useful prognostic, diagnostic, and treatment protocols will depend on connectal coding. The same strategy could be applied to study healthy brains as well. For example, memories are believed to be stored in "engrams", which are defined as biophysical or biochemical changes in the brain underlying memories [119]. While engrams for specific memories largely remain elusive [120], connectal coding could accelerate this search, by formulating the specific memory as a cognitive phenotype. In these settings, heterogeneity of connectomes may be a significant impediment which would require further methodological developments. Further, the methodologies developed to study contrasts within and across individuals within a species can also be applied across species [121,122], although such studies should also take into account differences in body plan and life cycle.

Second, in other fields of inquiry such as particle physics and geology, simulations play a key role in understanding nonlinear dynamical systems. In brain science, however, simulations remain in their infancy. That said, recently, several high-profile efforts have emerged to simulate the brain of various species, including humans [123–126]. While the precise detailed requirements for biofidelic and useful simulations of a brain remain fiercely debated [127], it is unambiguous that *some* assumptions about nodes, their properties, and connections are required for any such simulation. Therefore, connectal coding could be exploited to learn which connectotypes are required for the simulation to exhibit which cognitive phenotypes.

Third, as alluded to above, the idea that understanding biological intelligence can inform machine intelligence dates back to the early days of computer science and the so-called first wave of artificial intelligence (AI). The first serious artificial neural networks were simple one-layer networks, called perceptrons [128]. The second wave of AI began when a few people realized that artificial neural networks, like biological neural networks, could have multiple layers to increase their expressive capacity. More formally, while perceptrons (one-layer networks) can only represent linear functions [129], multi-layer perceptrons (with only one hidden layer) can represent arbitrarily complex functions [130,131]. The third wave artificial intelligence, brought about by deep learning [132], appreciated that real connectomes do not have huge unstructured single hidden layers, but rather, many relatively small hidden layers [133,134]. Perhaps incorporating more biological constraints in these searches could further improve their efficiency [135]. Indeed, there are some signs that the third wave of artificial neural networks might be waning and/or asymptoting, and some posit that their main hope for continued rapid improvement is to incorporate more ideas from connectomics [136].

Fourth, connectal coding could help explain evolution. Recall that for individuals to thrive in a given environment, they must exhibit certain phenotypes with sufficiently high probability. Further, for parents to increase their offspring's ability to thrive, they must transmit phenotypes to them that tend to increase survival probabilities. The primary strategy parents have for transmitting these proclivities is via a genome whose resulting developmental program tends to result in a body and brain structure that tends to yield the selected phenotypes. For all cognitive phenotypes, this includes connectomes [137]. Thus, for the genome to encode the probabilistic rules of behavior to facilitate the organism winning the evolutionary game, it must act through the connectome. Connectal codes are a key component of the link from genotype to phenotype, and from generation to generation.

In Sydney Brenner's Nobel lecture he quoted his own review article [138], stating: "Behavior is the result of a complex ill-understood set of computations performed by nervous systems and it seems essential to decompose the question into two: once concerned with the question of the genetic specification of nervous systems and the other with the way nervous systems work to produce behavior." Connectal coding is an approach to provide a partial answer to both of those questions by modeling brain structure as a network, and applying the wealth of statistical pattern recognition techniques to answer them. To facilitate using these ideas in practice, we have developed an open source python toolbox for statistical analysis of populations of networks, available at https://neurodata.io/graspy/ [139].

## Acknowledgements

The authors from JHU acknowledge support from NSF 16-569 NeuroNex contract 1707298, and KL's support comes from NSF grant DMS-1646108. The authors are grateful for helpful feedback from Konrad Koerding. The authors declare no conflicts of interest.

# References

1. White JG, Southgate E, Thomson JN, Brenner S. The structure of the ventral nerve cord of Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci. 1976;275: 327–348. Available: https://www.ncbi.nlm.nih.gov/pubmed/8806

2. Morgan JL, Lichtman JW. Why not connectomics? Nat Methods. 2013;10: 494–500. doi:10.1038/nmeth.2480

3. Craddock RC, Jbabdi S, Yan C-G, Vogelstein JT, Castellanos FX, Di Martino A, et al. Imaging human connectomes at the macroscale. Nat Methods. 2013;10: 524–539. doi:10.1038/nmeth.2482

4. Helmstaedter MN. Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. Nat Methods. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013;10: 501–507. doi:10.1038/nmeth.2476

5. Osten P, Margrie TW. Mapping brain circuitry with a light microscope. Nat Methods. 2013;10: 515–523. doi:10.1038/nmeth.2477

6. Johnson KO. Neural coding. Neuron. 2000;26: 563–566. Available: https://www.ncbi.nlm.nih.gov/pubmed/10896153

7. Shepherd GM. Foundations of the Neuron Doctrine (History of Neuroscience) [Internet]. 1 edition. Oxford University Press; 1991. Available: https://www.amazon.com/Foundations-Neuron-Doctrine-History-Neuroscience/dp/0195064917

8. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. Kluwer Academic Publishers; 1943;5: 115–133. doi:10.1007/BF02478259

9. Russell SJ, Norvig P. Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited,; 2016.

10. Hebb DO. The organization of behavior; a neuropsychological theory. Oxford, England: Wiley The organization of behavior; a neuropsychological theory.; 1949;335. Available: https://psycnet.apa.org/fulltext/1950-02200-000.pdf

11. Little WA. The existence of persistent states in the brain. Math Biosci. 1974;19: 101–120. doi:10.1016/0025-5564(74)90031-5

12. Hopfield JJ. Neural Networks and Physical Systems with Emergent Collective

Computational Abilities. Proceedings of the National Academy of Sciences. 1982;79: 2554–2558. doi:10.1073/pnas.79.8.2554

13. Hinton GE, Mcclelland JL, Rumelhart DE. Distributed representations, Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations. MIT Press, Cambridge, MA; 1986.

14. Onsager L. Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition. Phys Rev. American Physical Society; 1944;65: 117–149. doi:10.1103/PhysRev.65.117

15. Euler L. Solutio problematis ad geometriam situs pertinentis. Commentarii academiae scientiarum Petropolitanae. 1741; 128–140. Available: https://scholarlycommons.pacific.edu/cgi/viewcontent.cgi?article=1052&context=euler-works

16. Sporns O, Tononi G, Kötter R. The human connectome: A structural description of the human brain. PLoS Comput Biol. 2005;1: e42. doi:10.1371/journal.pcbi.0010042

17. Hagmann P. From diffusion MRI to brain connectomics. EPFL; 2005; Available: https://infoscience.epfl.ch/record/33696

18. Eaton RC, Lee RK, Foreman MB. The Mauthner cell and other identified neurons of the brainstem escape network of fish. Prog Neurobiol. 2001;63: 467–485. Available: https://www.ncbi.nlm.nih.gov/pubmed/11163687

19. Kamiński M, Ding M, Truccolo WA, Bressler SL. Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance. Biol Cybern. 2001;85: 145–157. doi:10.1007/s004220000235

20. Truccolo WA, Eden UT, Fellows MR, Donoghue JP, Brown N, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J Neurophysiol. 2005;93: 1074–1089. doi:10.1152/jn.00697.2004

21. Pillow J, Paninski L, Shlens J, Simoncelli E, Chichilnisky E. Modeling multi-neuronal responses in primate retinal ganglion cells. Comp Sys Neur. 2005;5.

22. Mishchenko Y, Vogelstein JT, Paninski L. A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. Ann Appl Stat. 2011;5: 129–1261. doi:10.1214/09-AOAS303

23. Kulkarni JE, Paninski L. Common-input models for multiple neural spike-train data. Network. 2007;18: 375–407. doi:10.1080/09548980701625173

24. Yatsenko D, Josić K, Ecker AS, Froudarakis E, Cotton RJ, Tolias AS. Improved estimation and interpretation of correlations in neural circuits. PLoS Comput Biol. 2015;11: e1004083. doi:10.1371/journal.pcbi.1004083

25. Pearl J. Causality: Models, Reasoning, and Inference [Internet]. First Edition. Cambridge

University Press; 2000. Available: https://www.amazon.com/Causality-Reasoning-Inference-Judea-Pearl/dp/0521773628

26. Maier-Hein KH, Neher PF, Houde J-C, Côté M-A, Garyfallidis E, Zhong J, et al. The challenge of mapping the human connectome based on diffusion tractography. Nat Commun. Nature Publishing Group; 2017;8: 1349. doi:10.1038/s41467-017-01285-x

27. Biswal BB, Mennes M, Zuo X-NN, Gohel S, Kelly AMC, Smith SM, et al. Toward discovery science of human brain function. Proceedings of the National Academy of Sciences. National Acad Sciences; 2010;107: 4734–4739. doi:10.1073/pnas.0911855107

28. Yan C-G, Craddock RC, Zuo X-N, Zang Y-F, Milham MP. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. Neuroimage. 2013;80: 246–262. doi:10.1016/j.neuroimage.2013.04.081

29. Smith SM, Miller KL, Salimi-khorshidi G, Webster M, Beckmann CF, Nichols TE, et al. Network modelling methods for FMRI. Neuroimage. Elsevier Inc.; 2011;54: 875–891. doi:10.1016/j.neuroimage.2010.08.063

30. Xia M, Wang J, He Y. BrainNet Viewer: a network visualization tool for human brain connectomics. PLoS One. journals.plos.org; 2013;8: e68910. doi:10.1371/journal.pone.0068910

31. LaPlante RA, Douw L, Tang W, Stufflebeam SM. The Connectome Visualization Utility: Software for Visualization of Human Brain Networks. Marinazzo D, editor. PLoS One. Public Library of Science; 2014;9: e113838. doi:10.1371/journal.pone.0113838

32. Gerhard S, Daducci A, Lemkaddem A, Meuli R, Thiran J-P, Hagmann P. The Connectome Viewer Toolkit: An Open Source Framework to Manage, Analyze, and Visualize Connectomes. Front Neuroinform. 2011;5. doi:10.3389/fninf.2011.00003

33. White JG. Neuronal connectivity in Caenorhabditis elegans. Trends Neurosci. 1985;8: 277–283. doi:10.1016/0166-2236(85)90102-X

34. White JG, Southgate E, Thomson JN, Brenner S. The structure of the nervous system of the nematode Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci. 1986;314: 1–340. Available: https://www.ncbi.nlm.nih.gov/pubmed/22462104

35. Jarrell TA, Wang Y, Bloniarz AE, Brittin CA, Xu M, Thomson JN, et al. The connectome of a decision-making neural network. Science. American Association for the Advancement of Science; 2012;337: 437–444. doi:10.1126/science.1221762

36. Emmons SW. Neural Circuits of Sexual Behavior in Caenorhabditis elegans. Annu Rev Neurosci. 2018;41: 349–369. doi:10.1146/annurev-neuro-070815-014056

37. Varshney LR, Chen BL, Paniagua E, Hall DH, Chklovskii DB. Structural properties of the Caenorhabditis elegans neuronal network. PLoS Comput Biol. 2011;7: e1001066. doi:10.1371/journal.pcbi.1001066

38. Bentley B, Branicky R, Barnes CL, Chew YL, Yemini E, Bullmore ET, et al. The Multilayer

Connectome of Caenorhabditis elegans. PLoS Comput Biol. 2016;12: e1005283. doi:10.1371/journal.pcbi.1005283

39. Cook, S. J., Jarrell, T. A., Brittin, C., Wang, Y., Bloniarz, A. E., Yakovlev, M. A., Nguyen, K. C. Q., Tang, L. T.-H., Bayer, E. A., Duerr, J. S., Buelow, H., Hobert, O., Hall, D. H., and Emmons, S. W. Whole-animal connectomes of both C. elegans sexes. Nature. 2019;

40. Eichler K, Li F, Litwin-Kumar A, Park Y, Andrade I, Schneider-Mizell CM, et al. The complete connectome of a learning and memory centre in an insect brain. Nature. 2017;548: 175–182. doi:10.1038/nature23455

41. Priebe CE, Park Y, Tang M, Athreya A, Lyzinski V, Vogelstein JT, et al. Semiparametric spectral modeling of the Drosophila connectome [Internet]. arXiv [stat.ML]. 2017. Available: http://arxiv.org/abs/1705.03297

42. Calabrese E, Badea A, Cofer G, Qi Y, Johnson GA. A Diffusion MRI Tractography Connectome of the Mouse Brain and Comparison with Neuronal Tracer Data. Cereb Cortex. 2015;25: 4628–4637. doi:10.1093/cercor/bhv121

43. Zuo X-N, Anderson JS, Bellec P, Birn RM, Biswal BB, Blautzik J, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. Sci Data. 2014;1: 140049. doi:10.1038/sdata.2014.49

44. Kiar G, Bridgeford E, Roncal WG, Consortium for Reliability and Reproducibliity (CoRR), Chandrashekhar V, Mhembere D, et al. A High-Throughput Pipeline Identifies Robust Connectomes But Troublesome Variability [Internet]. bioRxiv. bioRxiv; 2018. p. 188706. doi:10.1101/188706

45. Shannon CE. A Mathematical Theory of Communication. Bell System Technical Journal. 1948;27: 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

46. Cover TM, Thomas JA. Elements of Information Theory [Internet]. New York: John Wiley & Sons; 2012. Available: https://market.android.com/details?id=book-VWq5GG6ycxMC

47. Brette R. Is coding a relevant metaphor for the brain? Behav Brain Sci. 2019; 1–44. doi:10.1017/S0140525X19000049

48. Herculano-Houzel S. The human brain in numbers: a linearly scaled-up primate brain. Front Hum Neurosci. 2009;3: 31. doi:10.3389/neuro.09.031.2009

49. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, et al. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum Mol Genet. 2014;23: 5866–5878. doi:10.1093/hmg/ddu309

50. Prinz AA, Bucher D, Marder E. Similar network activity from disparate circuit parameters. Nat Neurosci. 2004;7: 1345–1352. doi:10.1038/nn1352

51. Carr CE, Konishi M. A circuit for detection of interaural time differences in the brain stem of the barn owl. J Neurosci. 1990;10: 3227–3246. Available: http://www.ncbi.nlm.nih.gov/pubmed/2213141

52. Lee SH, Kim P-J, Jeong H. Statistical properties of sampled networks. Physical review E. 2006;73: 16102.

53. Olhede SC, Wolfe PJ. Degree-based network models [Internet]. arXiv [math.ST]. 2012. Available: http://arxiv.org/abs/1211.6537

54. Varoquaux G, Craddock RC. Learning and comparing functional connectomes across subjects. Neuroimage. 2013;80: 405–415. doi:10.1016/j.neuroimage.2013.04.007

55. Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage. 2002;15: 870–878. doi:10.1006/nimg.2001.1037

56. Efron B. Simultaneous inference: When should hypothesis testing problems be combined? Ann Appl Stat. Institute of Mathematical Statistics; 2008;2: 197–223. doi:10.1214/07-AOAS141

57. Zalesky A, Fornito A, Bullmore ET. Network-based statistic: identifying differences in brain networks. Neuroimage. 2010;53: 1197–1207. doi:10.1016/j.neuroimage.2010.06.041

58. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B Methodological. JSTOR; 1995;57: 289–300. doi:10.2307/2346101

59. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika. Oxford University Press; 1986;73: 751–754. doi:10.1093/biomet/73.3.751

60. Mhembere D, Roncal WG, Sussman D, Priebe CE, Jung R, Ryman S, et al. Computing scalable multivariate glocal invariants of large (brain-) graphs. 2013 IEEE Global Conference on Signal and Information Processing. 2013. pp. 297–300. doi:10.1109/GlobalSIP.2013.6736874

61. Bullmore ET, Bassett DS. Brain Graphs: Graphical Models of the Human Brain Connectome. Annu Rev Clin Psychol. Annual Reviews; 2010;7: 113–140. doi:10.1146/annurev-clinpsy-040510-143934

62. Matejka J, Fitzmaurice G. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics Through Simulated Annealing. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2017. pp. 1290–1294. doi:10.1145/3025453.3025912

63. Zheng AX, Fienberg SE, Airoldi EM, Goldenberg A. A Survey of Statistical Network Models. Foundations and Trends in Machine Learning. 2009;2: 129–233. doi:10.1561/2200000008

64. Kolaczyk ED. Statistical Analysis of Network Data [Internet]. Springer; 2014. doi:10.1007/978-0-387-88146-1

65. Bollobas B. Modern Graph Theory [Internet]. 1st ed. Springer-Verlag New York; 1998. doi:10.1007/978-1-4612-0619-4

66. Vishwanathan SVN, Schraudolph NN. Graph Kernels. J Mach Learn Res. 2010;11:

1201–1242.

67. Crane H. Probabilistic foundations of statistical network analysis [Internet]. Chapman and Hall/CRC; 2018. Available: https://www.taylorfrancis.com/books/9781351807333

68. Erdős P, Rényi A. On random graphs, I. Publ Math Debrecen. 1959;6: 290–297. Available: http://www.renyi.hu/~p_erdos/Erdos.html#1959-11

69. Rieke F. Spikes: Exploring the Neural Code [Internet]. Sejnowski TJ, Poggio TA, editors. Cambridge: MIT Press; 1997. p. xvi, 395 p. Available: https://market.android.com/details?id=book-CU6QQgAACAAJ

70. Rohe K, Chatterjee S, Yu B. Spectral Clustering and the High-Dimensional Stochastic Blockmodel. Ann Stat. Institute of Mathematical Statistics; 2011;39: 1878–1915. doi:10.1214/11-AOS887

71. Holland PW, Laskey KB, Leinhardt S. Stochastic blockmodels: First steps. Soc Networks. North-Holland; 1983;5: 109–137. doi:10.1016/0378-8733(83)90021-7

72. Sussman DL, Tang M, Fishkind DE, Priebe CE. A consistent adjacency spectral embedding for stochastic blockmodel graphs [Internet]. arXiv [stat.ML]. 2011. pp. 1119–1128. doi:10.1080/01621459.2012.699795

73. Wasserman S, Anderson C. Stochastic a posteriori blockmodels: Construction and assessment. Soc Networks. Elsevier; 1987;9: 1–36. doi:10.1016/0378-8733(87)90015-3

74. Peixoto TP. Hierarchical Block Structures and High-Resolution Model Selection in Large Networks. Phys Rev X. American Physical Society; 2014;4: 011047. doi:10.1103/PhysRevX.4.011047

75. Lyzinski V, Tang M, Athreya A, Park Y, Priebe CE. Community Detection and Classification in Hierarchical Stochastic Blockmodels. IEEE Transactions on Network Science and Engineering. 2017;4: 13–26. doi:10.1109/TNSE.2016.2634322

76. Betzel RF, Bassett DS. Multi-scale brain networks. Neuroimage. 2017;160: 73–83. doi:10.1016/j.neuroimage.2016.11.006

77. Meunier D, Lambiotte R, Fornito A, Ersche KD, Bullmore ET. Hierarchical modularity in human brain functional networks. Front Neuroinform. Frontiers Research Foundation; 2009;3: 1–12. doi:10.3389/neuro.11.037

78. Hoff PD, Raftery AE, Handcock MS. Latent Space Approaches to Social Network Analysis. J Am Stat Assoc. 2002;97: 1090–1098. doi:10.1198/016214502388618906

79. Young SJ, Scheinerman ER. Random Dot Product Graph Models for Social Networks. Algorithms and Models for the Web-Graph. Springer Berlin Heidelberg; 2007. pp. 138–149. doi:10.1007/978-3-540-77004-6_11

80. Scheinerman ER, Tucker K. Modeling graphs using dot product representations. Comput Stat. Springer; 2009;25: 1–16. doi:10.1007/s00180-009-0158-8

81. Sussman DL, Tang M, Priebe CE. Consistent latent position estimation and vertex classification for random dot product graphs. IEEE Trans Pattern Anal Mach Intell. 2014;36: 48–57. doi:10.1109/TPAMI.2013.135

82. Athreya A, Lyzinski V, Marchette DJ, Priebe CE, Sussman DL, Tang M. A central limit theorem for scaled eigenvectors of random dot product graphs [Internet]. arXiv [math.ST]. 2013. p. 1305.7388. Available: http://arxiv.org/abs/1305.7388

83. Tang M, Priebe CE. Limit theorems for eigenvectors of the normalized Laplacian for random graphs [Internet]. arXiv [stat.ML]. 2016. Available: http://arxiv.org/abs/1607.08601

84. Rubin-Delanchy P, Priebe CE, Tang M, Cape J. A statistical interpretation of spectral embedding: the generalised random dot product graph [Internet]. arXiv [stat.ML]. 2017. Available: http://arxiv.org/abs/1709.05506

85. Binkiewicz N, Vogelstein JT, Rohe K. Covariate-assisted spectral clustering. Biometrika. academic.oup.com; 2017;104: 361–377. doi:10.1093/biomet/asx008

86. Athreya A, Tang M, Park Y, Priebe CE. On estimation and inference in latent structure random graphs [Internet]. arXiv [stat.ME]. 2018. Available: http://arxiv.org/abs/1806.01401

87. Athreya A, Fishkind DE, Tang M, Priebe CE, Park Y, Vogelstein JT, et al. Statistical Inference on Random Dot Product Graphs: a Survey. J Mach Learn Res. 2018;18: 1–92. Available: http://jmlr.org/papers/v18/17-448.html

88. Durante D, Dunson DB, Vogelstein JT. Nonparametric Bayes Modeling of Populations of Networks. J Am Stat Assoc. Taylor & Francis; 2017;112: 1516–1530. doi:10.1080/01621459.2016.1219260

89. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models [Internet]. Cambridge University Press; 2006. Available: https://market.android.com/details?id=book-c9xLKzZWoZ4C

90. Wang S, Vogelstein JT, Priebe CE. Joint Embedding of Graphs [Internet]. arXiv [stat.AP]. 2017. Available: http://arxiv.org/abs/1703.03862

91. Wang L, Zhang Z, Dunson D. Common and Individual Structure of Brain Networks [Internet]. arXiv [stat.CO]. 2017. Available: http://arxiv.org/abs/1707.06360

92. Nielsen AM, Witten D. The Multiple Random Dot Product Graph Model [Internet]. arXiv [stat.ME]. 2018. Available: http://arxiv.org/abs/1811.12172

93. Zhang Z, Allen GI, Zhu H, Dunson D. Tensor network factorizations: Relationships between brain structural connectomes and traits [Internet]. arXiv [stat.AP]. 2018. Available: http://arxiv.org/abs/1806.02905

94. Mill RD, Ito T, Cole MW. From connectome to cognition: The search for mechanism in human functional brain networks. Neuroimage. Elsevier; 2017;160: 124–139. doi:10.1016/j.neuroimage.2017.01.060

95. Tang M, Athreya A, Sussman DL, Lyzinski V, Park Y, Priebe CE. A Semiparametric

Two-Sample Hypothesis Testing Problem for Random Graphs. J Comput Graph Stat. Taylor & Francis; 2017;26: 344–354. doi:10.1080/10618600.2016.1193505

96. Tang M, Athreya A, Sussman DL, Lyzinski V, Priebe CE. A nonparametric two-sample hypothesis testing problem for random graphs. Bernoulli . Bernoulli Society for Mathematical Statistics and Probability; 2017;23: 1599–1630. doi:10.3150/15-BEJ789

97. Ghoshdastidar D, von Luxburg U. Practical methods for graph two-sample testing [Internet]. arXiv [stat.ML]. 2018. Available: http://arxiv.org/abs/1811.12752

98. Ginestet CE, Li J, Balachandran P, Rosenberg S, Kolaczyk ED. Hypothesis testing for network data in functional neuroimaging. Ann Appl Stat. Institute of Mathematical Statistics; 2017;11: 725–750. doi:10.1214/16-AOAS1015

99. Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space. J R Stat Soc Series B Stat Methodol. 2008;70: 849–911. doi:10.1111/j.1467-9868.2008.00674.x

100. Arroyo-Relión JD, Kessler D, Levina E, Taylor SF. Network classification with applications to brain connectomics [Internet]. Annals of Applied Statistics. Available: http://arxiv.org/abs/1701.08140

101. Vogelstein JT, Gray Roncal W, Vogelstein RJ, Priebe CE. Graph classification using signal-subgraphs: applications in statistical connectomics. IEEE Trans Pattern Anal Mach Intell. 2013;35: 1539–1551. doi:10.1109/TPAMI.2012.235

102. Wang S, Shen C, Badea A, Priebe CE, Vogelstein JT. Signal Subgraph Estimation Via Vertex Screening [Internet]. arXiv preprint arXiv. 2018. Available: https://scholar.google.ca/scholar?cluster=6440696796438427662&hl=en&as_sdt=0,5&sciodt=0,5

103. Wang L, Zhang Z, Dunson D. Symmetric Bilinear Regression for Signal Subgraph Estimation [Internet]. arXiv [stat.ME]. 2018. Available: http://arxiv.org/abs/1804.09567

104. Johnson GA, Badea A, Brandenburg J, Cofer G, Fubara B, Liu S, et al. Waxholm space: an image-based reference for coordinating mouse brain research. Neuroimage. 2010;53: 365–372. doi:10.1016/j.neuroimage.2010.06.067

105. Raznahan A, Probst F, Palmert MR, Giedd JN, Lerch JP. High resolution whole brain imaging of anatomical variation in XO, XX, and XY mice. Neuroimage. 2013;83: 962–968. doi:10.1016/j.neuroimage.2013.07.052

106. Spring S, Lerch JP, Henkelman RM. Sexual dimorphism revealed in the structure of the mouse brain using three-dimensional magnetic resonance imaging. Neuroimage. 2007;35: 1424–1433. doi:10.1016/j.neuroimage.2007.02.023

107. Aine CJ, Bockholt HJ, Bustillo JR, Cañive JM, Caprihan A, Gasparovic C, et al. Multimodal Neuroimaging in Schizophrenia: Description and Dissemination. Neuroinformatics. 2017;15: 343–364. doi:10.1007/s12021-017-9338-9

108. Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church J a., et al. Functional

network organization of the human brain. Neuron. Elsevier Inc.; 2011;72: 665–678. doi:10.1016/j.neuron.2011.09.006

109. Levin K, Athreya A, Tang M, Lyzinski V, Park Y, Priebe CE. A central limit theorem for an omnibus embedding of random dot product graphs [Internet]. arXiv [stat.ME]. 2017. Available: http://arxiv.org/abs/1705.09355

110. Hu M-L, Zong X-F, Mann JJ, Zheng J-J, Liao Y-H, Li Z-C, et al. A Review of the Functional and Anatomical Default Mode Network in Schizophrenia. Neurosci Bull. 2017;33: 73–84. doi:10.1007/s12264-016-0090-1

111. Ongür D, Lundy M, Greenhouse I, Shinn AK, Menon V, Cohen BM, et al. Default mode network abnormalities in bipolar disorder and schizophrenia. Psychiatry Res. Elsevier; 2010;183: 59–68. doi:10.1016/j.pscychresns.2010.04.008

112. Michael First, Kelly Botteron, Cameron Carter, Francisco Xavier Castellanos, Daniel P. Dickstein, Wayne Drevets, Kerri L. Kim, Matthew F. Pescosolido Scott Rausch, Karen E. Seymour, Yvette Sheline, Jon-Kar Zubieta. Consensus Report of the APA Work Group on Neuroimaging Markers of Psychiatric Disorders [Internet]. Available: https://www.psychiatry.org/File%20Library/Psychiatrists/Directories/Library-and-Archive/resource_documents/rd2012_Neuroimaging.pdf

113. Castellanos XF, Di Martino A, Craddock RC, Mehta AD, Milham MP. Clinical applications of the functional connectome. Neuroimage. Elsevier; 2013;null: 527–540. doi:10.1016/j.neuroimage.2013.04.083

114. Van Dam NT, O'Connor D, Marcelle ET, Ho EJ, Cameron Craddock R, Tobe RH, et al. Data-Driven Phenotypic Categorization for Neurobiological Analyses: Beyond DSM-5 Labels. Biol Psychiatry. 2017;81: 484–494. doi:10.1016/j.biopsych.2016.06.027

115. Spronk M, Kulkarni K, Ji JL, Keane B, Anticevic A, Cole MW. A whole-brain and cross-diagnostic perspective on functional brain network dysfunction [Internet]. bioRxiv. 2018. p. 326728. doi:10.1101/326728

116. Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature. 2018;562: 210–216. doi:10.1038/s41586-018-0571-7

117. Powell MA, Garcia JO, Yeh F-C, Vettel JM, Verstynen T. Local connectome phenotypes predict social, health, and cognitive factors. Netw Neurosci. 2018;2: 86–105. doi:10.1162/NETN_a_00031

118. Braun U, Schaefer A, Betzel RF, Tost H, Meyer-Lindenberg A, Bassett DS. From Maps to Multi-dimensional Network Mechanisms of Mental Disorders. Neuron. 2018;97: 14–31. doi:10.1016/j.neuron.2017.11.007

119. Lashley KS. In search of the engram. Symp Soc Exp Biol. 1950;4: 30.

120. Berlot E, Popp NJ, Diedrichsen J. In search of the engram, 2017. Current Opinion in

Behavioral Sciences. 2018;20: 56–60. doi:10.1016/j.cobeha.2017.11.003

121. van den Heuvel MP, Bullmore ET, Sporns O. Comparative Connectomics. Trends Cogn Sci. 2016;20: 345–361. doi:10.1016/j.tics.2016.03.001

122. Rilling JK, van den Heuvel MP. Comparative Primate Connectomics. Brain Behav Evol. 2018;91: 170–179. doi:10.1159/000488886

123. Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, et al. Reconstruction and Simulation of Neocortical Microcircuitry. Cell. 2015;163: 456–492. doi:10.1016/j.cell.2015.09.029

124. Sanz Leon P, Knock SA, Woodman MM, Domide L, Mersmann J, McIntosh AR, et al. The Virtual Brain: a simulator of primate brain network dynamics. Front Neuroinform. 2013;7: 10. doi:10.3389/fninf.2013.00010

125. Niebur E, Erdös P. Theory of the locomotion of nematodes: control of the somatic motor neurons by interneurons. Math Biosci. 1993;118: 51–82. doi:10.1016/0025-5564(93)90033-7

126. Arena P, Patané L, Termini PS. An insect brain computational model inspired by Drosophila melanogaster: Simulation results. The 2010 International Joint Conference on Neural Networks (IJCNN). ieeexplore.ieee.org; 2010. pp. 1–8. doi:10.1109/IJCNN.2010.5596513

127. Theil S. Why the Human Brain Project Went Wrong--and How to Fix It. Scientific American. 1 Oct 2015. doi:10.1038/scientificamerican1015-36

128. Rosenblatt F. The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory; 1957.

129. Minksy M, Papert S. Perceptrons: An Introduction to Computation Geometry. MIT press. 1969;200: 355–368.

130. Cybenko G. Approximation by superpositions of a sigmoidal function. Math Control Signals Systems. Springer; 1989;2: 303–314.

131. Hornik K. Approximation capabilities of multilayer feedforward networks. Neural Netw. 1991;4: 251–257. doi:10.1016/0893-6080(91)90009-T

132. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning [Internet]. Bach F, editor. MIT press Cambridge; 2016. Available: https://www.amazon.com/dp/B01MRVFGX4/ref=dp-kindle-redirect?_encoding=UTF8&btkr=1

133. Hassabis D, Kumaran D, Summerfield C, Botvinick M. Neuroscience-Inspired Artificial Intelligence. Neuron. 2017;95: 245–258. doi:10.1016/j.neuron.2017.06.011

134. Cepelewicz J. The U.S. Government Launches a $100-Million “Apollo Project of the Brain.” Scientific American. 8 Mar 2016. Available: https://www.scientificamerican.com/article/the-u-s-government-launches-a-100-million-ap

ollo-project-of-the-brain/. Accessed 16 Nov 2018.

135. Pham H, Guan MY, Zoph B, Le QV, Dean J. Efficient Neural Architecture Search via Parameter Sharing [Internet]. arXiv [cs.LG]. 2018. Available: http://arxiv.org/abs/1802.03268

136. Underwood E. Neuroscience. Barcoding the brain. Science. 2016;351: 799–800. doi:10.1126/science.351.6275.799

137. Seung S. Connectome: How the Brain's Wiring Makes Us Who We Are [Internet]. None edition. Houghton Mifflin Harcourt; 2012. Available: https://www.amazon.com/Connectome-How-Brains-Wiring-Makes/dp/0547508182/ref=sr_1_1?keywords=9780547508177&linkCode=qs&qid=1554693948&s=books&sr=1-1

138. Brenner S. The genetics of Caenorhabditis elegans. Genetics. 1974;77: 71–94. Available: https://www.ncbi.nlm.nih.gov/pubmed/4366476

139. Chung J, Pedigo BD, Bridgeford EW, Varjavand BK, Vogelstein JT. GraSPy: Graph Statistics in Python [Internet]. arXiv [cs.SI]. 2019. Available: http://arxiv.org/abs/1904.05329