MULTISCALE STATISTICAL HYPOTHESIS TESTING FOR K-SAMPLE GRAPH INFERENCE

by

Vivek Gopalakrishnan

A thesis submitted to The Johns Hopkins University

in conformity with the requirements for the degree of

Master of Science and Engineering

Baltimore, Maryland

May, 2021

Abstract

A connectome is a map of the structural and/or functional connections in the brain. This information-rich representation has the potential to transform our understanding of the relationship between patterns in brain connectivity and neurological processes, disorders, and diseases. However, existing computational techniques used to analyze connectomes are often insufficient for interrogating multi-subject connectomics datasets. Several methods are either solely designed to analyze single connectomes, or leverage heuristic graph invariants that ignore the complete topology of connections between brain regions. To enable more rigorous comparative connectomics analysis, we introduce robust and interpretable statistical methods motivated by recent theoretical advances in random graph models. These methods enable simultaneous analysis of multiple connectomes across different scales of network topology, facilitating the discovery of hierarchical brain structures that vary in relation with phenotypic profiles. We validated these methods through extensive simulation studies, as well as synthetic and real-data experiments. Using a set of high-resolution connectomes obtained from genetically distinct mouse strains (including the BTBR mouse—a standard model of autism—and three behavioral wild-types), we show that these methods uncover valuable latent information in multi-subject connectomics data and yield novel insights into the connective correlates of neurological phenotypes. The documentation and code for all analyses in this thesis are available at https://github.com/neurodata/MCC.

Primary Reader and Advisor: Joshua T. Vogelstein

Reader: Carey E. Priebe

Reader: Avanti Athreya

Acknowledgments

This work would not be possible without: the transformational mentorship of Dr. Joshua T. Vogelstein; the teachings of Dr. Carey E. Priebe and Dr. Avanti Athreya; the support from Ben Pedigo, Jaewon Chung, Ronak Mehta, Ronan Perry, Alex Loftus, Eric Bridgeford, Mike Powell, Jayanta Dey, and Sambit Panda; and the NeuroData Lab.

I am eternally grateful to the Department of Biomedical Engineering and the Whiting School of Engineering at Johns Hopkins University for creating an environment wherein my academic curiosity and research interests could flourish. I will never forget the privilege of leading an Undergraduate BME Design Team or my amazement when, during a dissection in *Computational Cardiology Lab*, a decapitated bullfrog autonomically flipped over off its back and started breathing. Over these past four years, I have lost track of how many times I have felt inspired by the knowledge surrounding me.

I will forever cherish the friendships I have made with my peers and professors at JHU. Thank you Natasha, John, Stephen, Lucas, Mitchell, Maggie, Chester, Bolaji, Angela, Christine, Dr. Lu, Dr. Miller, Dr. Fishkind, and Dr. Logsdon for supporting me as I pursue my dreams.

And finally, to my parents, I owe the deepest debt of gratitude. My accomplishments would not be possible without the foundation of love and support you have provided me my entire life. Thank you for always believing in me, even when I did not believe in myself.

Table of Contents

List of Tables

List of Figures

1 Introduction

Understanding how patterns in brain connectivity correlate with observable phenotypes is an active area of research in neuroscience. Computational investigations into large volumes of neuroimaging data from multiple modalities have already yielded insights into the links between brain connectivity and neurological phenomena such as cognition [1, 2], neurodevelopment [3–5], the heritability of brain structures [6–8], and the pathology of neurological diseases [9–12]. Derived from neuroimaging data, a connectome, or map of the functional and/or structural connections between distinct brain regions, has recently become an invaluable tool for such analyses, enabling researchers to model the brain as a network and understand its organization with graph theoretical methods [13–17]. Successfully associating phenotypic profiles with variation in the connectome will enable the elucidation of identifiable neurological structures that underlie phenotypes, thereby transforming our understanding of the human brain [18]. However, to fully realize the promise of the connectome, novel statistical approaches that are principled, robust, and reproducible are required for the analysis of this nascent and highly-complex data type [19].

From a mathematical perspective, connectomes can be modeled as a network (or graph) of the interactions between brain regions [20]. In this network, vertices represent discrete parcellations of the brain, and edges represent the functional or structural connections between these brain regions [21]. From a graph theory perspective, the connectome can be further described at multiple hierarchical scales of network topology [22]: at the extremes are the *local*

scale, characterized by the features of individual edges and vertices in the connectome, and the *global scale*, characterized by global patterns in brain connectivity and often quantified by a variety of graph-level statistics [23]; the intermediate *regional scale* focuses on the interactions between distinct subsets of brain regions (known as communities or blocks), comprising subgraphs of the connectome [24–28]. Simultaneously considering the local, regional, and global sclaes of network topology provides a multiscale overview of patterns in brain connectivity at the level of brain regions of interest.

The fundamental goal of comparative connectome analysis is to identify differences in network architecture across multiple groups [29]. However, most of the current methods for analyzing connectomes solely operate on a single topological scale, ignoring potentially useful biological information encoded in other scales. This is detrimental as the complementary analyses provided by multiscale connectomics approaches make scientific insights into neurological phenotypes more robust [30, 31]. A handful of multiscale connectomics methods have previously been described; however, they are either only designed to analyze a single connectome [32, 33], or leverage graph features that ignore or misrepresent connections between brain regions [34, 35]. The latter approach is particularly problematic: recent studies have shown that no set of graph invariants can comprehensively describe network topology [36, 37]. Therefore, claims that a particular graph invariant explains a given phenotype are suspect.

To overcome these limitations and enable the rigorous interrogation of

multi-subject connectomics datasets, we present inferential statistical methods that provide insight across topological scales of the connectome (Figure 1). Predicated on recent theoretical advances in random graph models, our methods identify components of the connectome that are connectively different across multiple categorical or dimensional phenotypes, which we term *signal* components. Specifically, these algorithms can be used to discover signal edges, vertices, and communities within populations of connectomes defined on the same vertex set and are appropriate for analyzing connectomes estimated from either structural or functional neuroimaging data. We formulate these methods as k-sample hypothesis tests, enabling comparisons of connectomes from more than two distinct phenotypic groups.

We demonstrate the efficacy and utility of our multiscale connectomics methods by applying them to an open access dataset of ultrahigh-resolution structural mouse connectomes (derived at a spatial resolution 20,000 times greater than typical human connectomes) [38]. Additionally, we establish the superiority of our proposed methods over prevailing connectomics analysis strategies via information-theoretic comparisons and extensive simulation studies.

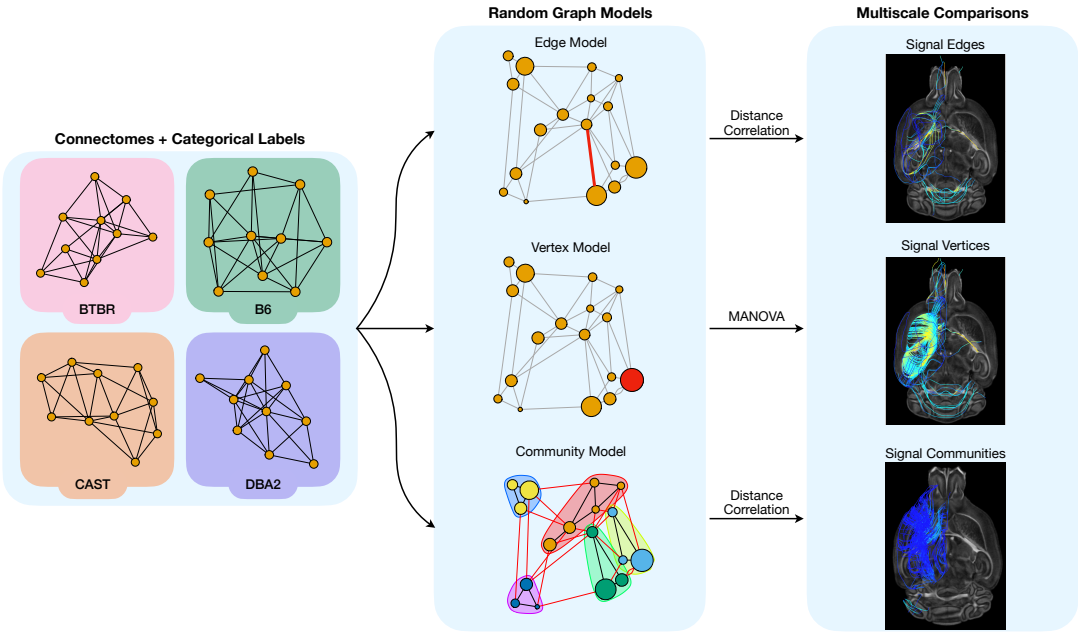


Figure 1: Overview of the statistical framework for multiscale comparative connectomics. (*Left*) All connectomes must be defined on a common set of vertices and have an associated categorical label. Labelled representative networks from four mouse lines whose connectomes we analyze in this work are shown. (*Center*) Connectomes are fit to random graph models that are specifically chosen to model the variation in a network at a given topological scale. (*Right*) Using the categorical labels, k-sample statistical hypothesis tests are applied to the fit parameters from each model, yielding a set of the signal edges, vertices, and communities across the groups in a given multi-subject connectomics dataset. We visualize identified structures using tractography.

2 Results

2.1 A multi-subject mouse connectome dataset for biological algorithmic validation

We assessed our multiscale connectomics methods using an open access dataset of whole-brain diffusion magnetic resonance imaging derived connectomes from four mouse lines: BTBR T+ Itpr3tf/J (BTBR), C57BL/6J (B6), CAST/EiJ (CAST), and DBA/2J (DBA2) [38]. The BTBR mouse strain is a

well-studied model that exhibits the core behavioral deficits that characterize autism spectrum disorders (ASD) in humans [39–41]. Additionally, the BTBR mouse has significant neuroanatomical abnormalities including the complete absence of the corpus callosum, a band of nerve fibers connecting the left and right hemispheres of the brain [42, 43]. The B6, CAST, and DBA2 mice are genetically distinct strains that do not exhibit ASD-like behaviors. They serve as wild type behavioral controls in this analysis.

For each strain, connectomes were generated from eight age-matched mice ($N = 8$ per strain), with a sex distribution of four males and four females. Each connectome was parcellated using a symmetric Waxholm Space [44, 45], yielding a vertex set with a total of 332 regions of interest (ROIs) bilaterally distributed across the left and right hemispheres. Within a given hemisphere, there were seven superstructures consisting up multiple ROIs, resulting in a total of 14 distinct communities in each connectome. Heatmaps of the average log-transformed adjacency matrix for each strain with hierarchical community and hemispheric labels are shown in Figure 2.

2.2 Interrogating the *local scale*

2.2.1 Identifying signal edges

The simplest approach for comparing connectomes is to treat them as a *bag of edges* without considering interactions between the edges [19]. Serially performing univariate statistical analyses at each edge enables the discovery of *signal edges* whose neurological connectivity differs across categorical or

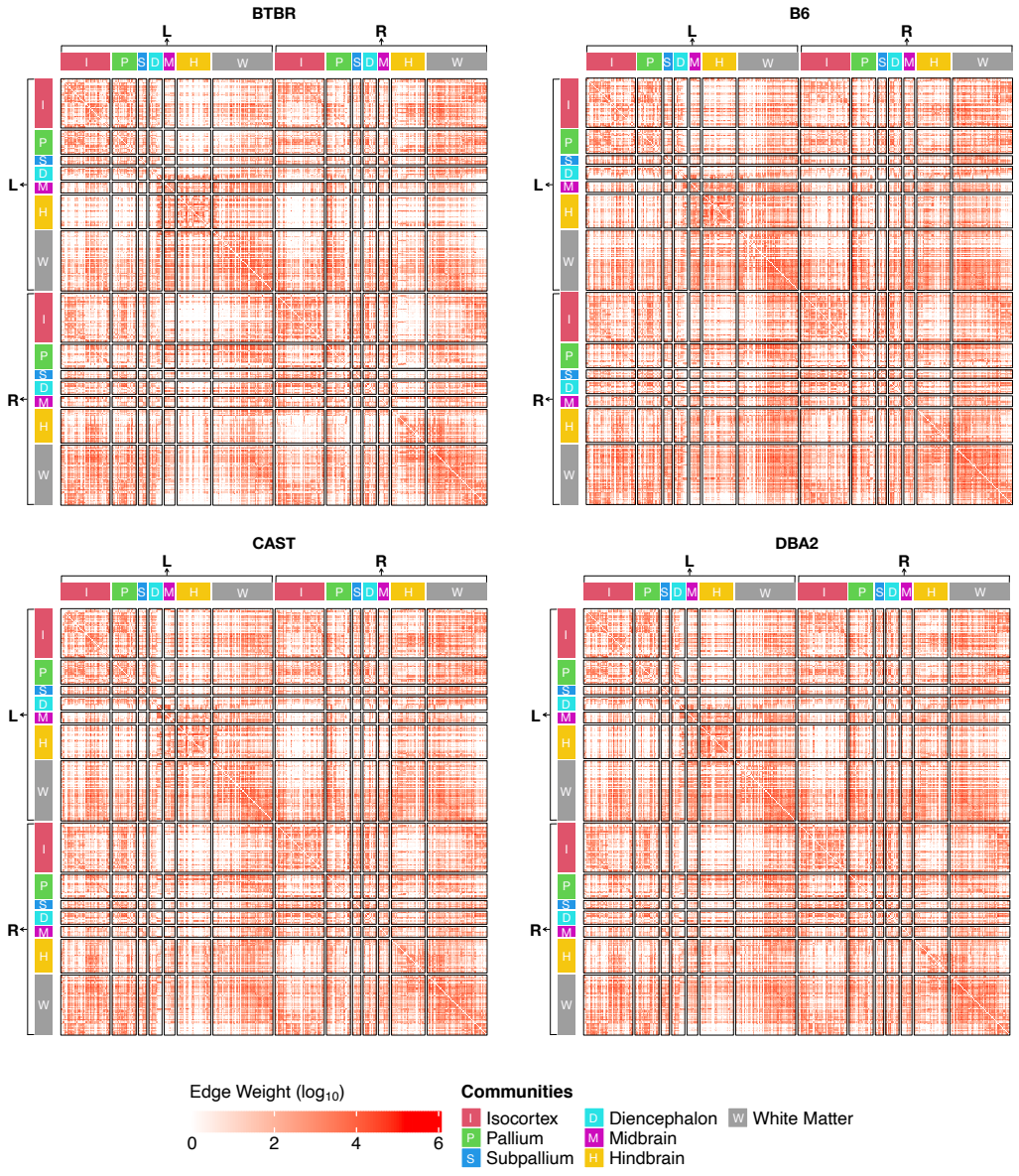


Figure 2: Log-transformed average connectomes for each mouse strain with hierarchical structure labels. Hierarchical labels show the hemispheric and superstructure level. For hemispheric labels, L (R) denotes the left (right) hemisphere. Superstructures are labelled as follows: I) isocortex, P) pallium, S) subpallium, D) diencephalon, M) midbrain, H) hindbrain, and W) white matter. Adapted from [38] with permission.

dimensional phenotypes. We introduce Distance Correlation (DCORR)—a previously established nonparametric universally consistent test [46]—to detect changes in edges. While classical statistical approaches (such as one-way analysis of variance (ANOVA) and the Kruskal–Wallis H test) are widely for edge comparisons [19, 47, 48], we find that DCORR outperforms these standard tests in this application.

To illustrate this, in the Methods (§4.5), we consider a two-population simulation setting where edges are sampled from truncated normal distributions. Supplementary Figure 1 shows that DCORR, ANOVA, and Kruskal–Wallis all identify signal edges that are different in means, and that no particular test is superior to the others in this setting; however, when the mean edge weight is constant but variance is different across groups, only DCORR consistently detects the signal edge. This is because ANOVA and Kruskal–Wallis both test for differences in location (the mean and median, respectively) between groups, whereas DCORR tests for differences in *distribution*.

Using this framework, we identified the signal edges in across our four mouse strains. Application of DCORR showed that no edges were significant at $\alpha = 0.05$ following Holm–Bonferroni correction. However, this does not imply that there were no connections between brain regions whose wiring patterns were heterogeneous across different genotypes: because the number of edges we compared across strains (54,946 edges) was very large, the strongest signal edge would have had to achieve an uncorrected p-value $< 10^{-6}$ to be deemed statistically significant. Due to this limitation imposed by the high-dimensional nature of connectomics data, we use the ranking of the p-value

for each edge to identify the strongest signal edges, not statistical significance. Given this set, we can look for patterns among groups of edges that could be jointly significant.

A list of the 20 strongest signal edges, along with their corresponding test statistics and p-values, are given in Supplementary Table 1. The strongest signal edge is from the left hemisphere corpus callosum to the right hemisphere striatum. In fact, 13 of the 20 strongest signal edges are adjacent to these two ROIs, demonstrating that many connections emanating from these regions are highly heterogeneous across genotypes. This suggests that instead of looking at signal *edges*, we might have more luck looking for signal *vertices*.

2.2.2 Identifying signal vertices

The ability to discover brain regions that are topologically dissimilar across phenotypes (i.e., *signal vertices*) is critical for scientific and clinical analyses of connectomes [49]. Here, we leverage recent advances in the theory of random graph models to propose a principled statistical method for identifying signal vertices. We also demonstrate that this method recovers far more information about ROIs than node-level graph-theoretic features (also known as graph invariants), the predominant connectomics method for analyzing vertices [19, 50, 51].

We use the omnibus embedding (OMNI) [52] — a graph embedding technique — to jointly represent all connectomes in a dataset in a common Euclidean subspace. For each connectome, OMNI jointly maps each vertex to a real-valued vector that corresponds to the vertex's latent position (see the

Methods (§4.2)). For each vertex, we then apply multivariate statistical tests to determine whether the embedding of a specific brain region is different across groups. Previous statistical findings have shown that the latent position vectors produced by OMNI are asymptotically Gaussian [53], motivating our use of multivariate analysis of variance (MANOVA) to identify the signal vertices.

To illustrate the value of this approach, in the Methods (§4.6), we compare this approach to two other graph embedding methods, including (i) the exceedingly popular approach of representing each vertex with a vector of node-level graph-theoretic features, and (ii) Multivariate Distance Matrix Regression(MDMR) [54]. In a two-population simulation setting, we sample graphs from a distribution in which the true number of signal vertices is determined *a priori*. For each vertex, we compute a p-value using each of the three vertex embeddings and set the significance level at $\alpha = 0.05$ following Holm–Bonferroni correction. We measure the performance of each approach via a Receiver Operating Characteristic (ROC) curve, a performance metric which holistically describes the performance of a binary classifier by characterizing the trade-off between sensitivity and specificity as the discriminant threshold of the classifier is varied. Supplementary Figure 2 shows that OMNI is a far superior vertex representation method than graph-theoretic features or MDMR. Even in the most challenging simulation settings where the true number of signal vertices is very small, OMNI correctly identifies the vertices of interest with much higher accuracy than the other two methods. The graph-theoretic approach in particular performed very poorly, achieving accuracy on par with random change in the most challenging settings (Supplementary

Figure 2 *bottom right panel*).

As it has been previously reported that the BTBR mouse has significant neuroanatomical abnormalities in the corpus callosum [42, 43], we expect our methods to identify this brain region as a strong signal vertex. Our results corroborate this hypothesis: across all mouse strains, the left hemisphere corpus callosum is the strongest signal vertex, and its right hemisphere counterpart is the second strongest (Supplementary Table 2). In Figure 3, we plot the vertex embedding of the corpus callosum obtained by OMNI using a pairs plot [55]. Because OMNI embeds each vertex of the graph in d-dimensional space, the pairs plot allows us to visualize the high-dimensional relationships in this Euclidean representation of network connectivity. Each dot represents the embedded corpus callosum of an individual mouse. These plots show scatter plot matrices on the off-diagonal panels, with kernel density estimates (KDEs) of the marginal distributions (smooth approximations of the underlying distribution of the data) on the diagonal. In the lower triangle of each pairs plot, we show 95% prediction ellipses of the distribution of each strain's embedding. Together with the KDEs, these figures show a high degree of separability in the embeddings, highlighting the heterogeneity of these vertices across strains. Thus, OMNI successfully recovers a distinct representation of the corpus callosum in BTBR mice, and the corroboration of known neurobiological results adds further validation to this approach. For comparison, in Supplementary Figure 5, we also show pairs plots of two weak signal vertices: the left hemisphere cingulate cortex area 29c (rank 173 of 332) and the right hemisphere fasciculus retroflexus (rank 260 of 332). Mice from distinct genotypes are much

less separable in embeddings of these vertices compared to the embeddings of the corpus callosum.

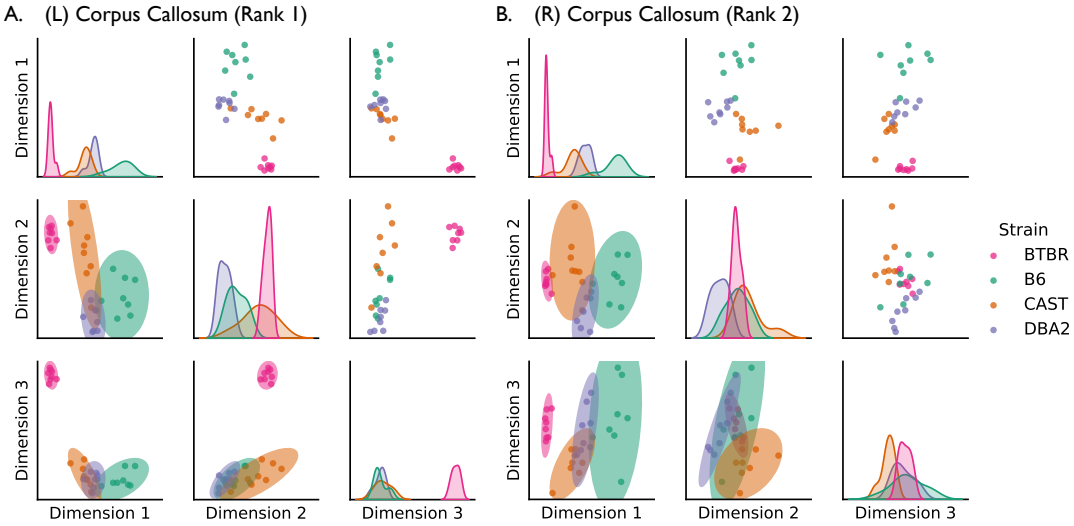


Figure 3: Pairs plots of vertex embedding of the left and right hemisphere corpus callosum produced by the omnibus embedding. The corpus callosum is the bridge between the left and right hemispheres of the brain. In the BTBR mouse, the corpus callosum has highly penetrant neuroanatomical defects; therefore, we expect to differentiate the BTBR mouse from the other strains by examining the embedding of the corpus callosum. We can do this in the embedding of the left hemisphere corpus callosum (the strongest signal vertex) and slightly less clearly in the embedding for the right hemisphere corpus callosum (the second strongest signal vertex). Interestingly, corpus callosum embeddings of wild-type mice are also separable (particularly the B6 strain), suggesting a diversity in corpus callosum architecture across the behavioral controls.

Connectomes in this dataset are bilateral (that is, for each region of interest, there is a corresponding structure in both the left and right hemispheres). Therefore, we can aggregate the connective abnormality of a structure across hemispheres, enabling the identification of pairs of vertices that are significant in both the left and right hemisphere. In Supplementary Table 3, we provide a list of the 10 strongest signal vertex pairs, which we term bilateral signal vertices. For most bilateral signal vertices, the ROI in one hemisphere is

usually a much stronger signal vertex than the other hemisphere (for example, the right hemisphere cerebral peduncle is the 9th ranked signal vertex, while its right hemisphere counterpart is ranked 63rd).

2.3 Interrogating the *regional scale*

2.3.1 Identifying signal communities

Communities of highly interconnected vertices are important structures within connectomes that underlie diverse neurological functions [24, 29]. Therefore, communities are an important topological scale at which connectomes can be analyzed. For analysis of connectomes where vertices are organized using an *a priori* community grouping, we compare four approaches for modelling the connectivity information encoded within a community (see the Methods (§4.4.3)). Each successive approach yields a more holistic summary of the community. The first two approaches are univariate, comparing either (1) the probability of connectivity in a community or (2) the average edge weight in a community across populations. While these are fundamental properties of a community, summarizing the behavior of a community with a single scalar loses information. The last two approaches are multivariate, comparing either (3) the indices of nonzero edges in a community or (4) the vector of edge weights in a community. Note that the operations performed by approaches (1) and (3) require the connectomes to be binarized via Otsu's method [56], whereas approaches (2) and (4) operate on weighted connectomes. To test if the summarized information in a block is different across phenotypes, we again use DCORR.

We compare these approaches in a two-population simulation setting (see the Methods (§4.7)). All methods were robust to false positives (Supplementary Figure 3 *first panel*), however, they differed in their ability to successfully identify signal communities. In settings where edge weights in a community have the same mean but different variances, univariate and binary approaches struggle to identify signal communities, achieving a maximum TPR of 60%. However, comparing multivariate weighted representations of communities proved much more successful with a TPR of 80% for sample sizes $N > 30$ (Supplementary Figure 3 *middle panel*). When edge weights in a community have different means, all algorithms are able to successfully identify signal communities with a high TPR. However, only the weighted approaches can do this with small sample sizes $N < 25$ (Supplementary Figure 3 *right panel*). Therefore, we propose using DCORR to compare communities across subjects in *regional scale* analyses.

In Supplementary Figure 4, we show log-transformed p-values obtained by the four approaches described above. Regions in blue are significant at the Holm–Bonferroni correction. Consistent with the simulations in the Methods (§4.7), we see that univariate tests find less signal communities than the multivariate tests. p-values for all communities are given in Supplementary Table 4. The strongest signal community (i.e., the one with the most heterogeneous topology across genotypes) determined by DCORR is the intra-connection within the right hemisphere white matter. In fact, the majority of the 10 strongest signal communities involve connections to the white matter in both hemispheres.

2.4 Visualizing multiscale differences in network architecture across genotypes via tractography

Using the statistical methods described above, we discover differences in brain connectivity between the four mouse genotypes at each topological scale of the connectome. Figure 4 visualizes the strongest signal edge as detected by DCORR, the strongest signal vertex as detected by OMNI and MANOVA, and the strongest signal community as detected by DCORR using tractograms, renderings of nerve tracts measured by Diffusion Magnetic Resonance Imaging (MRI). For an edge, the tractogram represents all the tracts between its two incident vertices; for a vertex, the tractogram represents all the tracts originating from that vertex; and for a community, the tractogram represents all the tracts interconnecting the vertices in a superstructure.A full list of the tracking parameters using to generate these tractograms is available in the Methods (§4.9).

Tractograms allow us to visualize the heterogeneity in brain connectivity identified by our multiscale algorithms. For example, a tractogram of the left corpus callosum (*middle column*) in the BTBR mice reveals a near absence of cross-hemispheric connections, while all control strains display much more cross-hemispheric connections at this vertex. To contrast with these heterogeneous tractograms, Supplementary Figure 6 shows tractograms for the weakest signal edge, vertex, and community. In comparison to the tractograms shown in Figure 4, the tractograms of weak signal components are much more homogeneous across strains. Note, we ignore edges that are zero for all connectomes.

In addition to tractograms, we also plot the distribution of neurotopological features (edge weights and embeddings) used to determine the strongest signal structure at each scale (Figure 4 *bottom row*). These distributions highlight the differences in connectome structure that each algorithm used to quantify the signal strength of that specific edge, vertex, or community.

2.5 Whole-brain *global scale* comparisons via OMNI

To complete a multiscale analysis of multi-subject connectomics data, we demonstrate how results from our previous methods can be aggregated to enable comparisons of patterns in whole-brain connectivity across subjects. OMNI was used to jointly embed every graph in our sample into a low-dimensional subspace. Figure 5A shows the pairwise dissimilarity between each pair of connectomes in our dataset. Pairwise dissimilarity between a pair of connectomes G_1 and G_2 is calculated as the Frobenius norm of the difference between the corresponding embedding of each graph. Note that the dissimilarities in Figure 5A are standardized by dividing by the largest pairwise dissimilarity. The average intra-strain dissimilarity (27%) is smaller than the average inter-strain dissimilarity (67%), providing evidence that the connectomes from mice of the same genotype are globally most similar to one another. Additionally, the BTBR mice are very dissimilar to all other strains (the average inter-strain dissimilarity for the BTBR mice is 79%) while the three control strains are all fairly similar to each other (the average inter-strain dissimilarity for B6, CAST, and DBA2 mice is 55%).

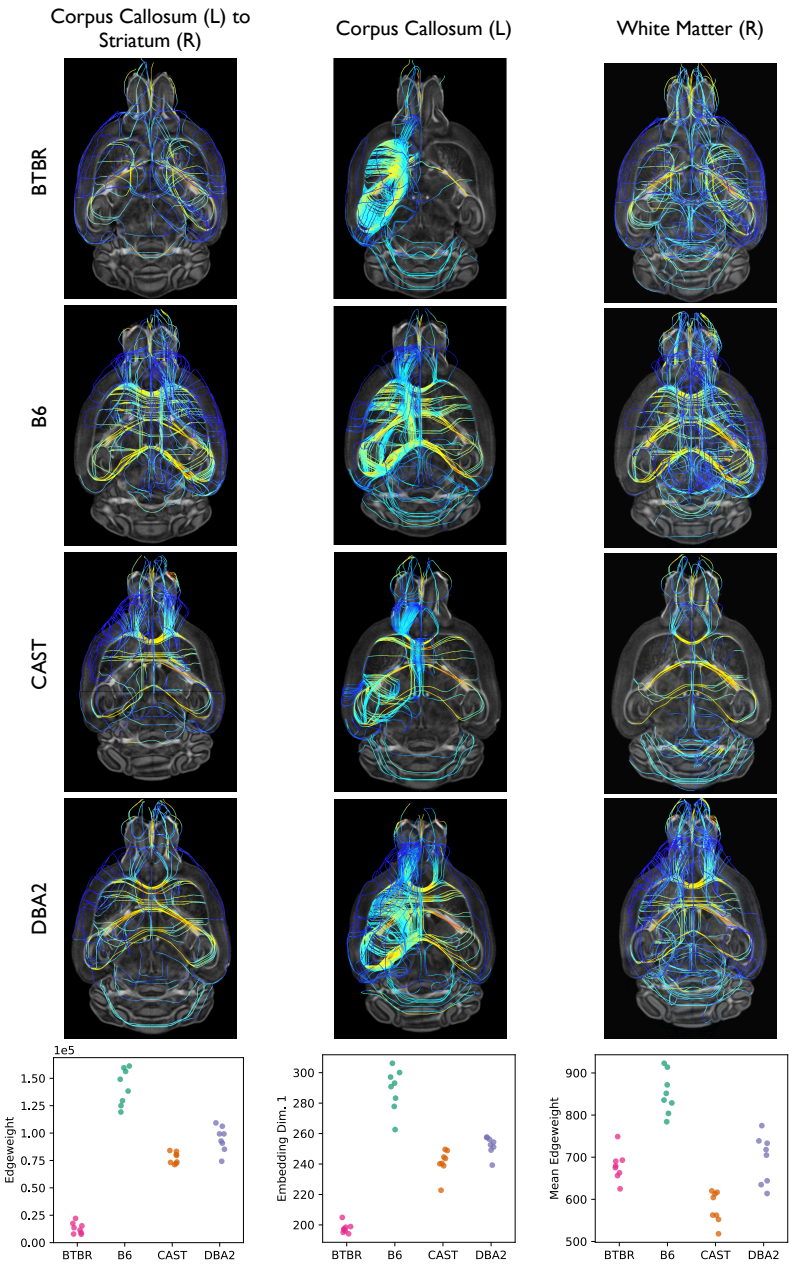


Figure 4: Visualization of the strongest signal edge (left corpus callosum to right hemisphere striatum), vertex (left hemisphere corpus callosum), and community (right white matter) across all mouse strains. At each topological scale, tractograms of these neurological structures are shown for each mouse strain. (*Bottom row*) The distribution of edge weights for the strongest signal edge (*Column 1*); the distribution of the first embedding dimension for the strongest signal vertex (*Column 2*); and the distribution of average edge weight for the most strongest signal community (*Column 3*). Each dot represents data from an individual mouse. Connective differences in the left hemisphere corpus callosum are apparent, with BTBR mice displaying a uniformly small vertex embedding. The average edge weight of the most significant community also shows pronounced variation across strains.

We further reduce the dimensionality of these embeddings by using classical Multidimensional Scaling (CMDS) to embed this distance matrix into a two-dimensional space [57]. This yields a collection of 32 points in $\mathbb{R}^2$, where each point represents a connectome. The BTBR mice are highly separated from the other three control strains (Figure 5B); all wild-type strains are also clearly distinct from each other (Figure 5C). Thus, we can leverage information from OMNI to successfully differentiate all connectomes based on genotype, enabling comparisons of brain connectivity at the *global scale*.

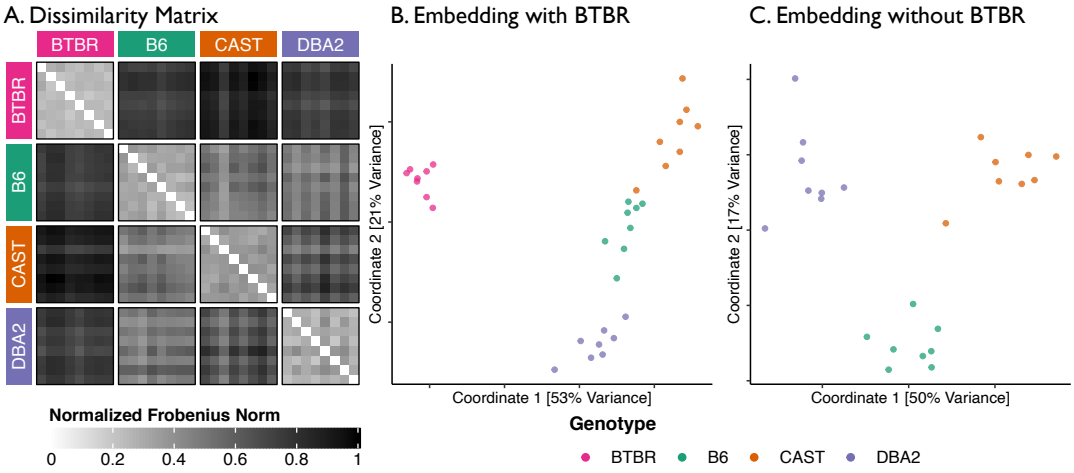


Figure 5: Pairwise distance between each mouse connectome, organized by mouse strain (*left*) and the joint embeddings of each sample with and without connectomes from BTBR mice in a two-dimensional space (*right* and *center*, respectively). Joint embeddings of every connectome were obtained using the omnibus embedding. (*Left*) The pairwise distance between connectomes is calculated as the Frobenius norm of the difference between the embeddings of a pair of connectomes. (*Center*) Two-dimensional representations of each connectome were obtained by using Classical Multidimensional Scaling (CMDS) to reduce the dimensionality of the embeddings obtained by OMNI. (*Right*) Same as center, but without data from BTBR mice. Adapted from [38] with permission.

2.6 The topology of vertices encodes information beyond the anatomy of those vertices

Here, we demonstrate that the characterization of vertex-level brain connectivity provided by the omnibus embedding contains topological information that is not available in commonly-used anatomical features. Following registration of all diffusion imaging data, the following anatomical features were derived for each vertex in every mouse brain: volume, apparent diffusion coefficient (ADC), fractional anisotropy (FA), and radial diffusivity (RD). For every vertex, we tested if the genotype labels and the low-dimensional embedding of the vertex produced by OMNI were conditionally independent given any of the anatomical features of that vertex [58]. The implication of this test is that if genotype and the embedding are conditionally independent given a particular feature, then that feature is a confounder [59]. If the alternative is true, then OMNI captures a characterization of brain connectivity (topology) not explained by that anatomical feature. This conditional independence testing procedure is described in further detail in the Methods (§4.8).

We find that for 282 out of 332 vertices, the genotype labels and OMNI embedding were conditionally dependent given each of the anatomical feature (Figure 6). All four anatomical features were confounders for only 28 vertices, and the majority of these vertices were not in the top 100 strongest signal vertices. Thus, for most vertices, the omnibus embedding provides additional insight into connectivity beyond what is explained by the anatomical features, demonstrating the value of this means of *local scale* analysis. Our findings demonstrate that, in most cases, anatomical features do not encode additional

information about mouse genotype than the connectomes.

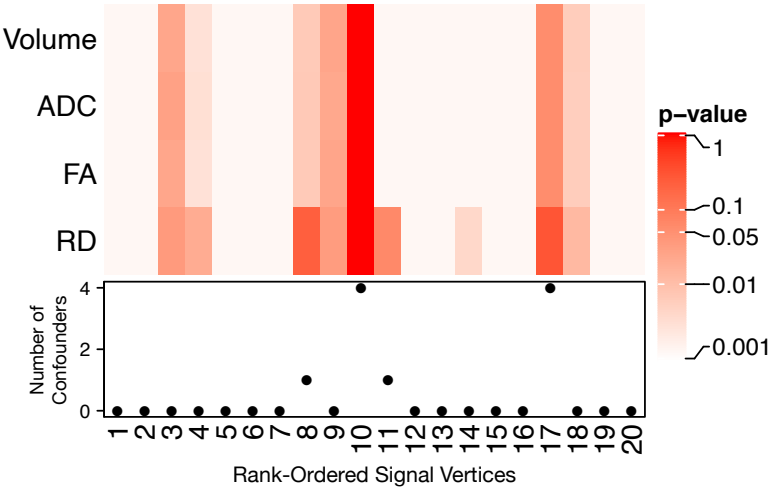


Figure 6: The omnibus embedding provides a novel understanding of brain connectivity at the scale of individual vertices. Using diffusion imaging, the following metrics were measured for each vertex: volume, apparent diffusion coefficient (ADC), fractional anisotropy (FA), and radial diffusivity (RD). To determine if a given feature provided more information than the embedding, these four anatomical features were compared to vertex embeddings produced by OMNI via a conditional independence test. Holm–Bonferroni corrected p-value for each conditional independence test for the top 20 strongest signal vertices (Table 2) are shown on the heatmap above. A cell is red if genotype and the embedding are conditionally independent given the feature, implying that OMNI is confounded by the anatomical feature. The total number of confounding anatomical features for each vertex, sorted by signal strength are shown below the heatmap.

3 Discussion

We introduce multiscale statistical methods for analyzing high-dimensional multi-subject connectomics datasets. Leveraging recent advances in the theory of random graph models, these methods deliver statistically principled and interpretable analysis of brain networks. Specifically, these methods can be used to identify signal edges, vertices, and communities—that is, components of the connectome across multiple scales that characterize the differences in

network architecture observed between populations from distinct phenotypic backgrounds. Using connectomes derived from the BTBR mouse, a well-studied model of autism, and behavioral control strains, we show the utility of these methods and demonstrate that they successfully recover known neurobiological information.

While we demonstrate the superiority of these methods to existing approaches through a variety of simulation studies and synthetic data experiments, a number of future extensions remain. First, the magnitude of the p-values obtained by multiple hypothesis testing of cannot always be trusted. Multiple testing requires p-value corrections to control the false positive rate, all of which make inappropriate assumptions for connectomics data. Even the straightforward Holm–Bonferroni correction can be erroneous, resulting in overly-conservative corrections for edge-wise testing and overly-liberal corrections for vertex-wise testing (Supplementary Figure 2). However, our simulations show that the components with the smallest p-values are the true signal edges, vertices, and communities, demonstrating that the ranking of the p-values can be trusted even if their magnitudes cannot be. While this validates our multiscale approach, it remains difficult to interrogate a single edge, vertex, or community in isolation without comparing it to all other components at the same topological scale.

Second, while the use of distance correlation provides powerful edge-wise testing, it is very computationally expensive. Performing k-sample testing with distance correlation typically requires a costly permutation test to estimate the null distribution and subsequent p-value. While there is a good

chi-square approximation for the null distribution of unbiased distance correlation with comparable finite-sample power [60], this test generally requires a sample size ≥ 20 to be statistically valid. In connectomics, it can sometimes be difficult to achieve a sample this large, particularly if one is studying a rare neurological disorder or using a very time-intensive process to estimate the connectome (such as manual tracing).

Third, these methods are designed for the comparison of populations of connectomes from distinct categorical or dimensional phenotypes. Statistical modelling of the connectome in relation to a continuous phenotypic variable of interest (such as age) is a fundamental challenge for the analysis of dynamic connectomes. While these methods can be applied if the continuous variable is discretized into categorical bins, extensions are required to enable true regression analysis of the connectome in this statistical modelling framework.

The network-level view of brain organization provided by the connectome will enable a transformative understanding of the brain [61]. For an organ system whose function—both in disease and in health—remains so poorly understood, the promise of this new data type is immense. As the wealth of neuroimaging and connectomics data continues to grow, new mathematical and statistical techniques will be required to discover the brain circuits that underlie neurological processes, disorders, and diseases. We anticipate that the multiscale algorithms and techniques presented in this work will be widely used by future researchers to uncover the neurobiological correlates of different phenotypes in multi-subject connectomics datasets.

All statistical methods described in this paper are implemented in `graspologic`,

an open-source Python package for statistical network analysis (https://github.com/microsoft/graspologic). Jupyter Notebooks demonstrating the use of these algorithms on real-world connectomics data is also available in the package's documentation.

4 Methods

4.1 Preliminaries

Graphs are convenient mathematical objects for representing connectomes. A graph G consists of the tuple (V, E) where V is the set of vertices and E is the set of edges. The set of vertices can be represented as $V = \{1, 2, \ldots, n\}$, where $|V| = n$ is the number of vertices. The set of edges is a subset of all possible connections between vertices (i.e., $E \subset V \times V$). We say the tuple $(i, j) \in E$ if there exists an connection from vertex i to vertex j. In many connectomics datasets, edges have associated edge weights: real-valued numbers that encode quantitative information about a connection between two vertices. The interpretation of the edge weight is dependent on imaging modality used to estimate the connectome. For example, the edge weight in human structural connectomes are non-negative integers that represent the number of estimated neuronal fibers that traverse from one region of the brain to another [62]. Every connectome has an associated weighted adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ where $\mathbf{A}_{ij}$ denotes the weight of the edge $(i, j) \in E$.

4.2 Statistical models

Statistical modeling of connectomics data enables the principled analysis of these high-dimensional, graph-valued data. Random graph models treat individual connectomes as random variables, enabling mathematical characterization of network structure and accounting for noise within and across observed samples. Below, we present three such models: 1) the Independent Edge (IE) model; 2) the Random Dot Product Graph (RDPG); and 3) the Stochastic Block Model (SBM). Each model is designed to characterize a particular topological scale of a network, either its edges, vertices, or communities. This network modelling approach enables the formulation of inferential statistical hypothesis tests that can be used to identify connective differences at specific scales across multiple phenotypes.

4.2.1 Edge model

In the Independent Edge (IE) model, every possible edge in the connectome $(i, j) \in V \times V$ occurs with some probability $p_{ij} \in [0, 1]$. We construct a matrix of probabilities $\mathbf{P} \in [0, 1]^{n \times n}$, where $\mathbf{P}_{ij} = p_{ij}$. Thus, a graph is given by the model $G \sim \text{IE}(\mathbf{P})$ if its adjacency matrix $\mathbf{A}$ has entries $\mathbf{A}_{ij} \sim \text{Bernoulli}(p_{ij})$ for all $(i, j) \in V \times V$. In the analytical methods presented in the Methods (§4.4), we consider the weighted IE model to account for a network with weighted edges. For this variant, instead of sampling a binary edge from $\text{Bernoulli}(p_{ij})$, we sample each edge weight from a distribution F_{ij}. In this setting, $\mathbf{P}$ represents a matrix of univariate probability distribution functions for the weight of each edge in the connectome. When estimating the $\mathbf{P}$ matrix

for a weighted IE graph, it is typical to assume that all elements of $\mathbf{P}$ are from the same family of distributions.

4.2.2 Vertex model

The Random Dot Product Graph (RDPG) is a generalization of the IE model that belongs to the family of latent position random graphs [63]. In certain scenarios, the probability of a connection p_{ij} may not be directly observable. Instead, each vertex is associated with a *latent position* x_i, which is a member of some space $\mathcal{X}$. The probability of a connection between vertices i and j is given by a link function $\kappa : \mathcal{X} \times \mathcal{X} \mapsto [0,1]$; that is, $p_{ij} = \kappa(x_i, x_j)$. In the RDPG, the latent space $\mathcal{X}$ is a subspace of Euclidean space $\mathbb{R}^d$ and the link function is the dot product [64]. Thus, in a $d-$dimensional RDPG with n vertices, the rows of the matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ encode the latent position of each vertex, and the matrix of connection probabilities is given by $\mathbf{P} = \mathbf{X}\mathbf{X}^\top$. A graph is sampled the model $G \sim \text{RDPG}(\mathbf{X})$ if its adjacency matrix $\mathbf{A}$ has entries $\mathbf{A}_{ij} \sim \text{Bernoulli}(\langle x_i, x_j \rangle)$ for $(i,j) \in V \times V$.

4.2.3 Community model

In the Stochastic Block Model (SBM), every vertex belongs to belongs to one of K communities, which are disjoint subsets of the vertex set V [65]. The SBM is a special case of the RDPG in which all vertices from the same community have identical latent positions; that is, the connection probability is solely determined by community membership. A symmetric $K \times K$ community connectivity probability matrix $\mathbf{B}$ with entries in $[0,1]^{K \times K}$ governs the probability of an edge between vertices given their community memberships. Community

membership is determined by the vertex assignment vector $\vec{\tau} \in \{1, \ldots, K\}^n$, which is either unknown or given *a priori*. Because brain regions in our dataset are hierarchically grouped into superstructures and hemispheres, we assume that $\vec{\tau}$ is given. Thus, a graph is sampled the model $G \sim \text{SBM}(\mathbf{B})$ with $\vec{\tau}$ given if its adjacency matrix $\mathbf{A}$ has entries $\mathbf{A}_{ij} \sim \text{Bernoulli}(\mathbf{B}_{kl})$ where $\tau_i = k, \tau_j = l$, for $(i, j) \in V \times V$, and $k, l \in \{1, \ldots, K\} \times \{1, \ldots, K\}$. As with the IE model, a weighted variant of the SBM can be constructed by replacing the block probability $\mathbf{B}_{ij}$ with a probability distribution function F_{ij} for the entire block.

In our simulations, we will consider two variations of of the two-community Sbm ($K = 2$) with block connectivity matrix $\mathbf{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, abbreviated as $\mathbf{B} = [a, b; c, d]$. These include:

1. Planted Partition: $a = d$ and $b = c$. In this model, the within-community edges share a common probability a, and the between-community edges share a common probability b, where $a \neq b$.

2. Symmetric Heterogeneous: $b = c$. In this model, the between-block edges share a common probability b, but the within-block edges have a disparate probabilities, where $a \neq b \neq d$.

4.3 Correcting for multiple comparisons

Connectomes analyzed in comparative multi-subject studies are typically composed of hundreds of vertices and tens of thousands of edges [66]. Therefore, each statistical method we describe in this work performs multiple comparisons on related structures in the connectome. Many standard methods for

controlling the false discovery rate such as the Benjamini–Hochberg procedure assume independence, but independence within a network is logically impossible for all but the simplest random graph model (weighted Erdős–Rényi), as each edge is defined specifically by virtue of a dependence between pairs of vertices. In this setting, such methods can result in overly liberal statistical corrections and a high false positive rate [67, 68]. Instead, we use the Holm–Bonferroni correction, a simple and stringent multiple comparison correction that controls the false discovery rate by setting the type 1 error rate (commonly referred to as the *significance level*, α) for each test equal to the original significance level divided by the rank of the p-value corresponding to the test [69]. Like the Bonferroni correction [70], the Holm–Bonferroni correction does not make assumptions that are violated by connectomics data. However, the Holm–Bonferroni correction is uniformly more powerful than the Bonferroni correction [69].

4.4 Statistical formulations for multiscale graph hypothesis tests

4.4.1 How to identify signal edges

Univariate edge-wise testing provides an interpretable and simple computational approach for identifying relationships between specific edges and phenotypes. Using the IE model, we assume that each connectome is sampled from a phenotype-conditional probability matrix; that is, we assume that for each phenotype in $\mathcal{Y}$, there is an associated probability matrix in $\{\mathbf{P}^{c_1}, \ldots, \mathbf{P}^{c_k}\}$ from which all connectomes in that phenotype are sampled. For a given edge (i, j), we assume the edge weight for each connectome has been independently

and identically (i.i.d) sampled from the appropriate $\mathbf{P}$ matrix. Specifically, we assume that for every connectome in phenotype $c_i \in \mathcal{Y}$, the edge weight $\mathbf{A}_{ij} \sim \mathbf{P}_{ij}^{c_i}$ i.i.d. Following this assumption, we formulate the following null and alternative hypotheses:

$$H_0 : \forall(y, y') : \quad \mathbf{P}_{ij}^{c_y} = \mathbf{P}_{ij}^{c_{y'}}$$

$$H_A : \exists(y, y') : \quad \mathbf{P}_{ij}^{c_y} \neq \mathbf{P}_{ij}^{c_{y'}}$$

To test the null hypothesis, any univariate statistical test can be employed at each potential edge $e \in E$. We use a one-sided k-sample test of distance correlation (DCORR) to determine if the distribution of an edge weight differs across populations. In the Methods (§4.5), we demonstrate that DCORR is a more powerful test than common non-parametric and Gaussian alternatives. Additionally, serial edge-wise testing requires corrections for an immense number of multiple comparisons. If the sample consists of directed graphs, then the total number of tests is n^2; if the graphs are undirected, then the total number of tests is $\binom{n}{2}$.

4.4.2 How to identify signal vertices

A sample of connectomes can be jointly embedded in a low-dimensional Euclidean space using the omnibus embedding (OMNI). A host of machine learning tasks can be accomplished with this joint embedded representation of the connectome, such as clustering or classification of vertices. Here, we use the embedding to formulate a statistical test that can be used to identify vertices that are strongly associated with given phenotypes.

For each graph in the population, OMNI jointly maps each vertex $v \in V$ to a vector $x_i \in \mathbb{R}^d$ that corresponds to the vertex's latent position in a d-dimensional RDPG. According to a Central Limit Theorem for OMNI, these latent positions are universally consistent and asymptotically normal [53]. This motivates our use of normal-theory inferential statistical tests to determine if the embedding of a given vertex is different across phenotypes. If the number of classes $k = 2$, we use Hotelling's T-squared (HOTELLING'S), a multivariate generalization of the T-TEST; if $k > 2$, we use one-way MANOVA with the Pillai trace as our test statistic. We formulate the following null and alternative hypotheses:

$$H_0 : \forall(y, y') : \quad \mu_{c_y} = \mu_{c_{y'}}$$

$$H_A : \exists(y, y') : \quad \mu_{c_y} \neq \mu_{c_{y'}}$$

where μ_{c_i} is the mean latent position for the phenotype $c_i \in \mathcal{Y}$. This procedure results in n statistical tests (one for each vertex).

4.4.3 How to identify communities

Vertices in a connectome can be hierarchically organized into superstructures such as major brain regions and hemispheres. The interactions within and between these communities of vertices form connective circuits within the brain, and are more correlated with complex behavior and phenotypes than single edges or vertices. Therefore, interrogation of the *regional scale* and identification of signal communities is critical component of multiscale connectomics analysis.

Here, we use the SBM to model the community structure of a connectome. We propose four approaches for describing the connectivity of a community, and an accompanying statistical procedure for each approach. Each subsequent approach provides a more holistic description of a community. For a given community (i, j) and for each phenotype, we do the following:

1. Average Connectivity: Compute the average number of nonzero edges (binarize using Otsu's method [56]), which is equivalent to the community connectivity probability $\mathbf{B}_{ij}$. We then use Pearson's chi-squared to determine if there is a significant difference in the connectivity probability across phenotypes.

2. Average Edge Weight: Compute the average weight of all the edges, essentially treating the community as a single large edge. We then use DCORR to determine if there is a significant difference in the average edge weight across phenotypes.

3. Multivariate Binary: Binarize the community using Otsu's method [56], and vectorize the subgraph of the adjacency matrix corresponding to that community. We binarize to determine if the distribution of nonzero edges is different across phenotypes. Test for differences in this binary representation using DCORR.

4. Multivariate Weighted: Vectorize the subgraph of the adjacency matrix corresponding to that community. Again, test for differences in this weighted representation using DCORR.

In the Methods (§4.7), we demonstrate via simulations that the Multivariate

Weighted method is superior to the other three. If the connectomes in question have b communities and are directed, this procedure results in b^2 comparisons. If the connectomes are undirected, this procedure results in $\binom{b}{2}$ comparisons.

4.5 Edge simulations

We consider two populations of networks generated from a two-community SBM. Edge weights are sampled from a truncated normal distribution to emulate correlation matrices. All networks have $n = 20$ vertices with 10 vertices belonging to the first community and 10 vertices belonging to the second community. The community probability matrix for each population is given by

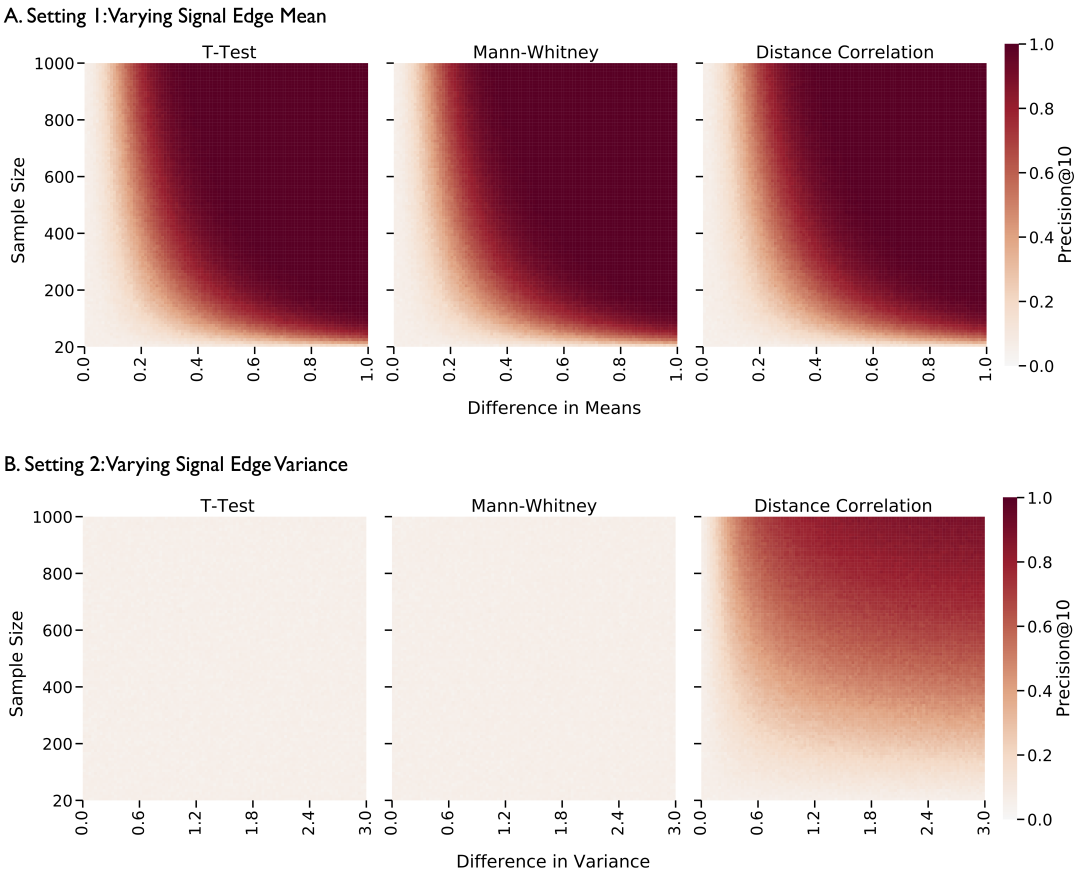
$$\mathbf{B}^1 = \begin{bmatrix} \mathcal{TN}(0, 0.25) & \mathcal{TN}(0, 0.25) \\ \mathcal{TN}(0, 0.25) & \mathcal{TN}(0, 0.25) \end{bmatrix}$$

$$\mathbf{B}^2 = \begin{bmatrix} \mathcal{TN}(0+\delta, 0.25+\phi) & \mathcal{TN}(0, 0.25) \\ \mathcal{TN}(0, 0.25) & \mathcal{TN}(0, 0.25) \end{bmatrix}$$

where $\mathcal{TN}(\mu, \sigma^2)$ denotes a truncated normal distribution with mean μ and variance σ^2 such that all values are bounded within $[-1, 1]$. A total of $m = 20, 40, \ldots, 1000$ networks are sampled ($m/2$ networks per population). In the first population, all edges are sampled from the same edge weight distribution $\mathcal{TN}(0, 0.25)$. In the second population, all edge weights are also sampled from $\mathcal{TN}(0, 0.25)$ except for those in the first community. In this community, the distribution of edge weights has either a different mean, δ, or variance, $0.25 + \phi$, from the first population. Therefore, the edges in the first community of these simulated connectomes are the signal edges that we hope to correctly

identify. For each edge, p-values are computed by three different tests: 1) T-TEST, 2) Mann-Whitney (MW) U test, a non-parametric test of medians, and 3) 2-sample Distance Correlation (DCORR) test, a test of equality in distribution. For each test, the p-values are sorted to find the ten most significant edges, and the performance is evaluated with recall (the number of signal edges that are correctly identified). Supplementary Figure 1 shows the Recall@10 as a function of sample size, mean, and variance. Supplementary Figure 1 top row shows that all three tests can identify signal edges that are different in mean, and that no particular test is superior than another in this setting. Supplementary Figure 1 bottom row shows that only DCORR can detect signal edges with differences in variance when the means are kept the same. This is because T-TEST and MW test for differences in location (e.g. mean or median), whereas DCORR tests for any differences between a pair of observed distributions. Because DCORR achieves about the same precision and recall as the location tests when only location varies, and demonstrates considerably better operating characteristics when the variance varies, we use it in the real-world data.

If the sample size $m \geq 20$, one can use a chi-square test that well approximates the k-sample DCORR test [60]. This can improve the computational efficiency of edge-wise testing by avoiding the costly permutation test that DCORR normally uses to estimate the null distribution.



Supplementary Figure 1: Precision@10 for each edge when comparing two populations of weighted networks using the T-TEST, Mann-Whitney U test, and k-sample DCORR test. The color bar represents precision averaged over 100 trials. *(Top row)* Results for varying the mean δ and sample size while keeping the variance constant ($\phi = 0$). In this setting, all three tests perform similarly, and can detect signal edges when edge distributions differ in means. *(Bottom row)* Results for varying the variance ϕ and sample size wile keeping the mean constant ($\delta = 0$). T-TEST and Mann-Whitney test cannot detect changes in variance regardless of the sample and effect size. k-sample DCORR test is the only test that can detect signal edges with changes in variance.

4.6 Vertex simulations

One goal of connectomics is to identify signal vertices that are different between populations. In this section, we identify signal vertices using different vector-based vertex representations. The first representation we consider is the simplest: it is possible describe a vertex using all its incident edges via the corresponding row (or column) of a vertex in the adjacency matrices, and compare these vectors using Multivariate Distance Matrix Regression (MDMR) [54]. The second representation we consider, which is highly popular in modern connectomics literature, is a set of vertex-level network features. Specifically, we fit an exponential random graph model [71] with the following network features for each vertex: local clustering coefficient (LCC), betweenness centrality (BC), closeness centrality (CC), and number of triangles. The final representation we consider are the low-dimensional latent-space positions estimated by OMNI. Since all vertex representations are multivariate, hypotheses are tested using HOTELLING'S test, a multivariate generalization of the T-TEST.

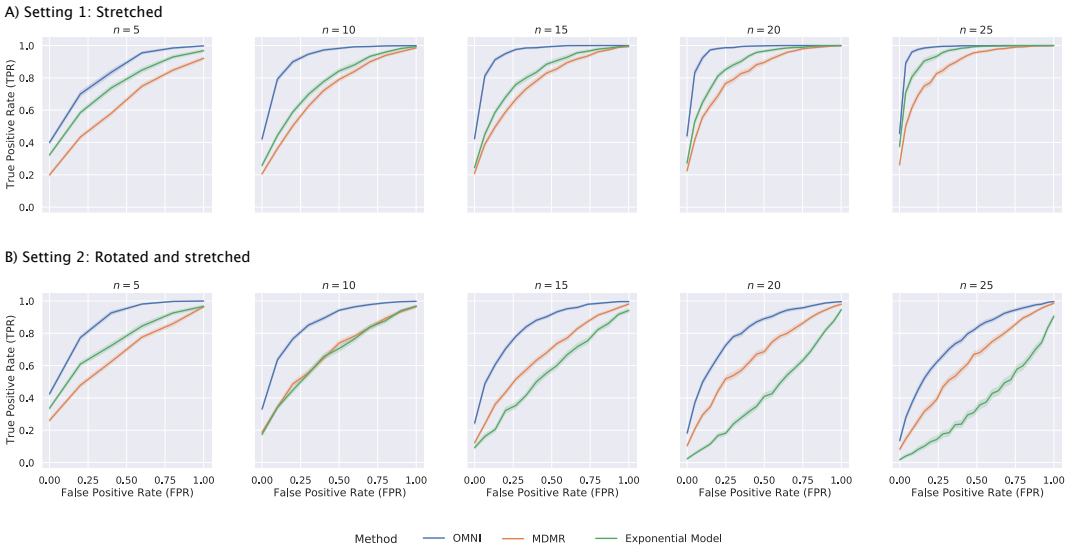
We consider a population of RDPGs in two different settings where the number of signal vertices is varied. In both settings, m null vertices are sampled from the latent position $X_1 = [0.25, 0.25]$ and n signal vertices are sampled from $X_2 = \text{rot}(70^\circ)X_1$, where $\text{rot}(\theta)$ represents a 2-dimensional rotation matrix of angle θ. In setting 1, X_2 is stretched, and a second RDPG is constructed with latent positions X_1 and $X_3 = 0.4X_2$ and an equivalent number of vertices per each position. In setting 2, X_2 is stretched and rotated (i.e., $X_3 = 0.4\text{rot}(10^\circ)X_2$). The vertices sampled from X_2 or X_3 (i.e., from

different latent positions) are considered signal vertices, and we vary the number of these vertices from $n = 0, 5, 10, 15, 20, 25$. The number of null vertices is set to $m = 50 - n$. A total of 100 networks are sampled per population, and the p-values are computed using HOTELLING'S on each of the three vertex representations for each vertex. Vertices with p-values less than $\alpha = 0.05$ after Holm–Bonferroni correction are classified as signal vertices in this simulation. The performance of each algorithm is measured using a Receiver Operator Characteristic (ROC) curve, which shows the trade-off between the False Positive Rate and True Positive Rate for each method.

Supplementary Figure 2 shows that OMNI is uniformly a more accurate method for identifying signal vertices. Across all settings, its accuracy (measured by the Area Under the ROC curve) is higher than the other two methods. In particular, the Exponential Random Graph Model, which uses the widely popular approach of network statistics, performs much worse.

4.7 Community simulations

We consider two populations of networks generated from a three-community block diagonal SBM. As in the edge simulation (Methods §4.5), edge weights



Supplementary Figure 2: Ability of various vertex representations to identify signal vertices. The number of vertices in each network is kept constant (50), but the number of signal vertices is changed ($n = 5, 10, 15, 20, 25$). For each vertex, we compute a p-value from each of the three vertex representations using HOTELLING'S, and set the significance level at $\alpha = 0.05$ following Holm–Bonferroni correction. Colors correspond to the method of vertex representation. *(Top row)* This setting compares two different RDPGs where the perturbed latent position is stretched by a constant. *(Bottom row)* This setting compares RDPGs where the perturbed latent position is rotated and stretched by a constant. Both settings show that OMNI successfully differentiates null and signal vertices with much more accuracy than either MDMR or the exponential random graph model.

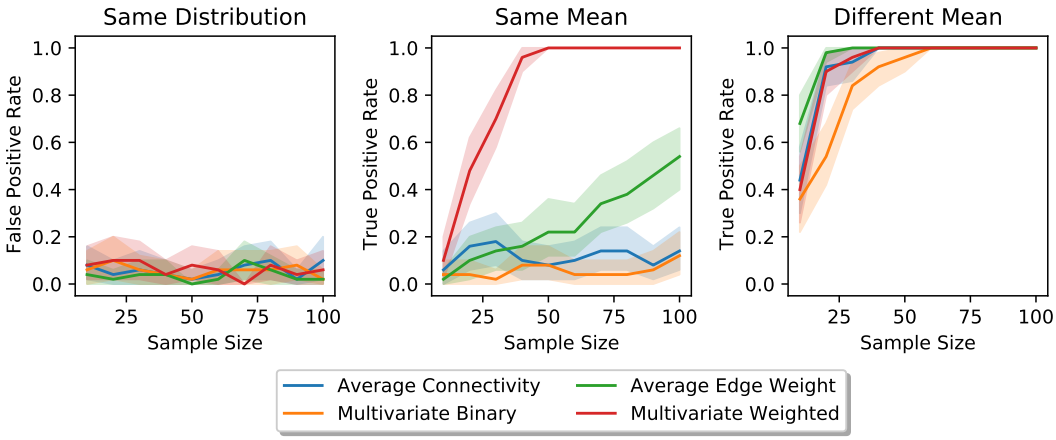
are sampled from truncated normal distributions to emulate correlation matrices. All communities have $n = 10$ vertices each, and the community probability matrix for each population is given by

$$\mathbf{B}^1 = \begin{bmatrix} \mathcal{TN}(0, 0.25) & & \\ & \mathcal{TN}(0, 0.25) & \\ & & \mathcal{TN}(-0.75, 0.25) \end{bmatrix}, \text{ and}$$

$$\mathbf{B}^2 = \begin{bmatrix} \mathcal{TN}(0, 0.25) & & \\ & \mathcal{TN}(0, 0.50) & \\ & & \mathcal{TN}(0.75, 0.25) \end{bmatrix}.$$

A total of $m = 10, 20, \ldots, 100$ networks are sampled ($m/2$ networks per population). Note that the edge weight distributions in the first diagonal block are equal, while the distributions have different variances in the second diagonal block and different means in the third diagonal block. Therefore, data from block one allow us to measure the false positive rate (FPR), while data from blocks two and three allow us to measure the true positive rate (TPR). Each simulation run was conducted 50 times to ensure accurate estimation of the FPR and TPR. All p-values were corrected using Holm–Bonferroni with significance determined at $\alpha = 0.05$.

From these simulations, we see that all four tests have roughly the same FPR (about 5% regardless of sample size) (Supplementary Figure 3 *first panel*). However, they differ in their ability to detect signal communities. When all populations have the same mean edge weight, but different variances, only the Multivariate Weighted method consistently detects signal communities once the sample size is greater than 25 networks (Supplementary Figure 3 *second panel*). All other methods have a lower TPR in this setting. When the



Supplementary Figure 3: The Multivariate Weighted approach for signal community detection is superior to other proposed methods. We consider a two-population simulation where edge weights are sampled from 3-community block diagonal SBM. Distributions in the first block are equal, allowing us to measure the False Positive Rate for each method. We see that all methods achieve a FPR less than or equal to α (*left panel*). In the second block, the two distributions have the same mean, but different variances. The Multivariate Weighted method is the most able to detect community edges in this setting, however it requires a sample size of at least $N = 25$ (*middle panel*). In the third block, the two distributions have different means, and in this setting, the Average Edge Weight method is superior given small sample sizes ($N < 25$) (*right panel*). However, when the samples size is sufficiently large, the Multivariate Weighted method is equivalent. These results demonstrate the superiority of the Multivariate Weighted method over other proposed methods.

edge weights have different means, the TPR of the Multivariate Weighted method lags behind the Average Edge Weight method in low sample size settings ($N < 25$) (Supplementary Figure 3 *third panel*). However, when the sample size is larger than $N > 25$, the Multivariate Weighted and Average Edge Weight methods perform comparably. These simulations demonstrate that the Multivariate Weighted signal community detection method, which includes the most information about connectivity within a block, achieves the highest TPR while maintaining a FPR less than or equal to the significance value of the test.

4.8 Conditional independence testing

Let $Y \in \mathcal{Y}$ be a random variable representing the genotype of a mouse. For a given vertex i, let $X \in \mathcal{X}$ be a random variable representing the vertex's latent position in a RDPG estimated by OMNI. Additionally, let vertex i be described by a vector of anatomical features (denoted $A \in \mathcal{A}$) measured from raw neuroimaging data. We are interested in relatively how much signal about genotype is encoded in X and A, respectively.

We can do this using an conditional independence testing framework. If genotype and the embedding are conditionally independent given the anatomical features, then the following is true:

$$(Y \perp\!\!\!\perp X) \mid A \implies \Pr(Y \mid X, A) = \Pr(Y \mid A)\,.$$

That is, the information about connectivity encoded in the vertex's latent position is redundant given the anatomical features. However, if null hypothesis of conditional independence is rejected, the latent position contains information about connectivity not represented in the anatomy.

4.9 Tractography

A group average template was constructed from the 4 male mice per strain. A DTI diffusion scheme was used, and a total of 46 diffusion sampling directions were acquired. The b-value was 4000 s/mm2. The in-plane resolution was 0.045 mm. The slice thickness was 0.045 mm. The diffusion data were reconstructed in the MNI space using q-space diffeomorphic reconstruction [72] to obtain the spin distribution function [73]. A diffusion sampling length

ratio of 1.25 was used. The restricted diffusion was quantified using restricted diffusion imaging [74].

5 Data availability

Mouse connectomes were derived by Wang et al. and are described in further detail in the original publication [38]. These data are freely available in `graspologic` (https://github.com/microsoft/graspologic/) [75], an open-source Python package for statistical graph analysis.

6 Code availability

All graph-related simulations and analyses were performed using `graspologic` (https://github.com/microsoft/graspologic/) [75] and all multivariate hypothesis testing was performed using `hyppo` (https://github.com/neurodata/hyppo) [76]. Heatmaps were generated using `ComplexHeatmap` [77]. The documentation and code for all analyses in this work are available at: https://github.com/neurodata/MCC.git.

7 Funding Acknowledgements

This work was supported by funding from Microsoft Research, NIH 1RF1MH121539-01, and R01 AG066184/AG/NIA NIH HHS/United States.

References

1. Poldrack, R. A. Can Cognitive Processes Be Inferred from Neuroimaging Data? en. *Trends in Cognitive Sciences* **10,** 59–63. ISSN: 13646613 (2006).
2. Petersson, K. M., Reis, A. & Ingvar, M. Cognitive Processing in Literate and Illiterate Subjects: A Review of Some Recent Behavioral and Functional Neuroimaging Data. en. *Scandinavian Journal of Psychology* **42,** 251–267. ISSN: 0036-5564, 1467-9450 (2001).
3. Hintz, S. R., Barnes, P. D., Bulas, D., Slovis, T. L., Finer, N. N., Wrage, L. A., Das, A., Tyson, J. E., Stevenson, D. K., Carlo, W. A., Walsh, M. C., Laptook, A. R., Yoder, B. A., Van Meurs, K. P., Faix, R. G., Rich, W., Newman, N. S., Cheng, H., Heyne, R. J., Vohr, B. R., Acarregui, M. J., Vaucher, Y. E., Pappas, A., Peralta-Carcelen, M., Wilson-Costello, D. E., Evans, P. W., Goldstein, R. F., Myers, G. J., Poindexter, B. B., McGowan, E. C., Adams-Chapman, I., Fuller, J., Higgins, R. D. & for the SUPPORT Study Group of the Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Neuroimaging and Neurodevelopmental Outcome in Extremely Preterm Infants. en. *Pediatrics* **135,** e32–e42. ISSN: 0031-4005, 1098-4275 (2015).
4. Huyser, C., Veltman, D. J., de Haan, E. & Boer, F. Paediatric Obsessive–Compulsive Disorder, a Neurodevelopmental Disorder? en. *Neuroscience & Biobehavioral Reviews* **33,** 818–830. ISSN: 01497634 (2009).
5. Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M. E., Hopson, R., Jackson, C., Keefe, J., Riley, M., Mentch, F. D., Sleiman, P., Verma, R., Davatzikos, C., Hakonarson, H., Gur, R. C. & Gur, R. E. Neuroimaging of the Philadelphia Neurodevelopmental Cohort. en. *NeuroImage* **86,** 544–553. ISSN: 10538119 (2014).
6. Ganjgahi, H., Winkler, A. M., Glahn, D. C., Blangero, J., Kochunov, P. & Nichols, T. E. Fast and Powerful Heritability Inference for Family-Based Neuroimaging Studies. en. *NeuroImage* **115,** 256–268. ISSN: 10538119 (2015).
7. Shen, K.-K., Rose, S., Fripp, J., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., Thompson, P. M., Wright, M. J. & Salvado, O. Investigating Brain Connectivity Heritability in a Twin Study Using Diffusion Imaging Data. en. *NeuroImage* **100,** 628–641. ISSN: 10538119 (2014).

8. Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sprooten, E., Nichols, T. E., Wright, S. N., Hong, L. E., Patel, B., Behrens, T., Jbabdi, S., Andersson, J., Lenglet, C., Yacoub, E., Moeller, S., Auerbach, E., Ugurbil, K., Sotiropoulos, S. N., Brouwer, R. M., Landman, B., Lemaitre, H., den Braber, A., Zwiers, M. P., Ritchie, S., van Hulzen, K., Almasy, L., Curran, J., deZubicaray, G. I., Duggirala, R., Fox, P., Martin, N. G., McMahon, K. L., Mitchell, B., Olvera, R. L., Peterson, C., Starr, J., Sussmann, J., Wardlaw, J., Wright, M., Boomsma, D. I., Kahn, R., de Geus, E. J., Williamson, D. E., Hariri, A., van 't Ent, D., Bastin, M. E., McIntosh, A., Deary, I. J., Hulshoff pol, H. E., Blangero, J., Thompson, P. M., Glahn, D. C. & Van Essen, D. C. Heritability of Fractional Anisotropy in Human White Matter: A Comparison of Human Connectome Project and ENIGMA-DTI Data. en. *NeuroImage* **111,** 300–311. ISSN: 10538119 (2015).

9. Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D. & Ji, S. en. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014* (eds Golland, P., Hata, N., Barillot, C., Hornegger, J. & Howe, R.) 305–312 (Springer International Publishing, Cham, 2014). ISBN: 978-3-319-10442-3 978-3-319-10443-0.

10. Spetsieris, P., Ma, Y., Peng, S., Ko, J. H., Dhawan, V., Tang, C. C. & Eidelberg, D. Identification of Disease-Related Spatial Covariance Patterns Using Neuroimaging Data. en. *Journal of Visualized Experiments* **76,** 50319. ISSN: 1940-087X (2013).

11. Sarica, A., Cerasa, A. & Quattrone, A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Frontiers in Aging Neuroscience* **9,** 329. ISSN: 1663-4365 (2017).

12. the Alzheimer's Disease Neuroimaging Initiative, Habeck, C. & Stern, Y. Multivariate Data Analysis for Neuroimaging Data: Overview and Application to Alzheimer's Disease. en. *Cell Biochemistry and Biophysics* **58,** 53–67. ISSN: 1085-9195, 1559-0283 (2010).

13. Bullmore, E. & Sporns, O. Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems. en. *Nature Reviews Neuroscience* **10,** 186–198. ISSN: 1471-003X, 1471-0048 (2009).

14. Galantucci, S., Agosta, F., Stefanova, E., Basaia, S., van den Heuvel, M. P., Stojković, T., Canu, E., Stanković, I., Spica, V., Copetti, M., Gagliardi, D., Kostić, V. S. & Filippi, M. Structural Brain Connectome and Cognitive

Impairment in Parkinson Disease. en. *Radiology* **283,** 515–525. ISSN: 0033-8419, 1527-1315 (2017).

15. Aerts, H., Fias, W., Caeyenberghs, K. & Marinazzo, D. Brain Networks under Attack: Robustness Properties and the Impact of Lesions. en. *Brain* **139,** 3063–3083. ISSN: 0006-8950, 1460-2156 (2016).

16. Hwang, G., Hermann, B., Nair, V. A., Conant, L. L., Dabbs, K., Mathis, J., Cook, C. J., Rivera-Bonet, C. N., Mohanty, R., Zhao, G., Almane, D. N., Nencka, A., Felton, E., Struck, A. F., Birn, R., Maganti, R., Humphries, C. J., Raghavan, M., DeYoe, E. A., Bendlin, B. B., Prabhakaran, V., Binder, J. R. & Meyerand, M. E. Brain Aging in Temporal Lobe Epilepsy: Chronological, Structural, and Functional. en. *NeuroImage: Clinical* **25,** 102183. ISSN: 22131582 (2020).

17. Miranda-Dominguez, O., Feczko, E., Grayson, D. S., Walum, H., Nigg, J. T. & Fair, D. A. Heritability of the Human Connectome: A Connectotyping Study. en. *Network Neuroscience* **2,** 175–199. ISSN: 2472-1751 (2018).

18. Bullmore, E. T. & Bassett, D. S. Brain Graphs: Graphical Models of the Human Brain Connectome. en. *Annual Review of Clinical Psychology* **7,** 113–140. ISSN: 1548-5943, 1548-5951 (2011).

19. Craddock, R. C., Jbabdi, S., Yan, C.-G., Vogelstein, J. T., Castellanos, F. X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S. & Milham, M. P. Imaging Human Connectomes at the Macroscale. en. *Nature Methods* **10,** 524–539. ISSN: 1548-7091, 1548-7105 (2013).

20. Sporns, O., Tononi, G. & Kötter, R. The Human Connectome: A Structural Description of the Human Brain. en. *PLoS Computational Biology* **1,** e42. ISSN: 1553-734X, 1553-7358 (2005).

21. Vogelstein, J. T., Bridgeford, E. W., Pedigo, B. D., Chung, J., Levin, K., Mensh, B. & Priebe, C. E. Connectal Coding: Discovering the Structures Linking Cognitive Phenotypes to Individual Histories. en. *Current Opinion in Neurobiology* **55,** 199–212. ISSN: 09594388 (2019).

22. Kaiser, M. A Tutorial in Connectome Analysis: Topological and Spatial Features of Brain Networks. en. *NeuroImage* **57,** 892–907. ISSN: 10538119 (2011).

23. Rubinov, M. & Sporns, O. Complex Network Measures of Brain Connectivity: Uses and Interpretations. en. *NeuroImage* **52,** 1059–1069. ISSN: 10538119 (2010).

24. Betzel, R. F., Medaglia, J. D. & Bassett, D. S. Diversity of Meso-Scale Architecture in Human and Non-Human Connectomes. en. *Nature Communications* **9,** 346. ISSN: 2041-1723 (2018).

25. Arroyo, J. & Levina, E. Simultaneous Prediction and Community Detection for Networks with Application to Neuroimaging. *arXiv:2002.01645 [stat].* arXiv: 2002.01645 [stat] (2020).

26. Wang, L., Zhang, Z. & Dunson, D. Symmetric Bilinear Regression for Signal Subgraph Estimation. *IEEE Transactions on Signal Processing* **67,** 1929–1940. ISSN: 1053-587X, 1941-0476 (2019).

27. Wang, S., Shen, C., Badea, A., Priebe, C. E. & Vogelstein, J. T. Signal Subgraph Estimation Via Vertex Screening. *arXiv:1801.07683 [stat].* arXiv: 1801.07683 [stat] (2018).

28. Vogelstein, J. T., Roncal, W. G., Vogelstein, R. J. & Priebe, C. E. Graph Classification Using Signal-Subgraphs: Applications in Statistical Connectomics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35,** 1539–1551. ISSN: 0162-8828, 2160-9292 (2013).

29. van den Heuvel, M. P., Bullmore, E. T. & Sporns, O. Comparative Connectomics. en. *Trends in Cognitive Sciences* **20,** 345–361. ISSN: 13646613 (2016).

30. Wang, J., Zuo, X., Dai, Z., Xia, M., Zhao, Z., Zhao, X., Jia, J., Han, Y. & He, Y. Disrupted Functional Brain Connectome in Individuals at Risk for Alzheimer's Disease. en. *Biological Psychiatry* **73,** 472–481. ISSN: 00063223 (2013).

31. Wu, Z., Xu, D., Potter, T., Zhang, Y. & The Alzheimer's Disease Neuroimaging Initiative. Effects of Brain Parcellation on the Characterization of Topological Deterioration in Alzheimer's Disease. *Frontiers in Aging Neuroscience* **11,** 113. ISSN: 1663-4365 (2019).

32. Eichler, K., Li, F., Litwin-Kumar, A., Park, Y., Andrade, I., Schneider-Mizell, C. M., Saumweber, T., Huser, A., Eschbach, C., Gerber, B., Fetter, R. D., Truman, J. W., Priebe, C. E., Abbott, L. F., Thum, A. S., Zlatic, M. & Cardona, A. The Complete Connectome of a Learning and Memory Centre in an Insect Brain. en. *Nature* **548,** 175–182. ISSN: 0028-0836, 1476-4687 (2017).

33. Bhattacharya, A., Aghayeva, U., Berghoff, E. G. & Hobert, O. Plasticity of the Electrical Connectome of C. Elegans. en. *Cell* **176,** 1174–1189.e16. ISSN: 00928674 (2019).

34. Rubinov, M, Kötter, R, Hagmann, P & Sporns, O. Brain Connectivity Toolbox: A Collection of Complex Network Measurements and Brain Connectivity Datasets. en. *NeuroImage* **47,** S169. ISSN: 10538119 (2009).

35. Xia, C. H., Ma, Z., Cui, Z., Bzdok, D., Bassett, D. S., Satterthwaite, T. D., Shinohara, R. T. & Witten, D. M. *Multi-Scale Network Regression for Brain-Phenotype Associations* en. Preprint (Neuroscience, 2019).

36. Chen, H., Huroyan, V., Soni, U., Lu, Y., Maciejewski, R. & Kobourov, S. Same Stats, Different Graphs: Exploring the Space of Graphs in Terms of Graph Properties. *IEEE Transactions on Visualization and Computer Graphics,* 1–1. ISSN: 1077-2626, 1941-0506, 2160-9306. arXiv: 1911.01527 (2019).

37. Chung, J., Bridgeford, E., Arroyo, J., Pedigo, B. D., Saad-Eldin, A., Gopalakrishnan, V., Xiang, L., Priebe, C. E. & Vogelstein, J. T. *Statistical Connectomics* Preprint (Open Science Framework, 2020).

38. Wang, N., Anderson, R. J., Ashbrook, D. G., Gopalakrishnan, V., Park, Y., Priebe, C. E., Qi, Y., Laoprasert, R., Vogelstein, J. T., Williams, R. W. & Johnson, G. A. Variability and Heritability of Mouse Brain Structure: Microscopic MRI Atlases and Connectomes for Diverse Strains. en. *NeuroImage,* 117274. ISSN: 10538119 (2020).

39. Silverman, J. L., Tolu, S. S., Barkan, C. L. & Crawley, J. N. Repetitive Self-Grooming Behavior in the BTBR Mouse Model of Autism Is Blocked by the mGluR5 Antagonist MPEP. en. *Neuropsychopharmacology* **35,** 976–989. ISSN: 0893-133X, 1740-634X (2010).

40. McFarlane, H. G., Kusek, G. K., Yang, M., Phoenix, J. L., Bolivar, V. J. & Crawley, J. N. Autism-like Behavioral Phenotypes in BTBR T+tf/J Mice. en. *Genes, Brain and Behavior* **7,** 152–163. ISSN: 1601-1848, 1601-183X (2008).

41. Scattoni, M. L., Gandhy, S. U., Ricceri, L. & Crawley, J. N. Unusual Repertoire of Vocalizations in the BTBR T+tf/J Mouse Model of Autism. en. *PLoS ONE* **3** (ed Crusio, W. E.) e3067. ISSN: 1932-6203 (2008).

42. Ellegood, J., Babineau, B. A., Henkelman, R. M., Lerch, J. P. & Crawley, J. N. Neuroanatomical Analysis of the BTBR Mouse Model of Autism Using Magnetic Resonance Imaging and Diffusion Tensor Imaging. en. *NeuroImage* **70,** 288–300. ISSN: 10538119 (2013).

43. Meyza, K. & Blanchard, D. The BTBR Mouse Model of Idiopathic Autism – Current View on Mechanisms. en. *Neuroscience & Biobehavioral Reviews* **76,** 99–110. ISSN: 01497634 (2017).

44. Johnson, G. A., Badea, A., Brandenburg, J., Cofer, G., Fubara, B., Liu, S. & Nissanov, J. Waxholm Space: An Image-Based Reference for Coordinating Mouse Brain Research. en. *NeuroImage* **53,** 365–372. ISSN: 10538119 (2010).

45. Calabrese, E., Badea, A., Cofer, G., Qi, Y. & Johnson, G. A. A Diffusion MRI Tractography Connectome of the Mouse Brain and Comparison with Neuronal Tracer Data. en. *Cerebral Cortex* **25,** 4628–4637. ISSN: 1047-3211, 1460-2199 (2015).

46. Székely, G. J., Rizzo, M. L. & Bakirov, N. K. Measuring and Testing Dependence by Correlation of Distances. en. *The Annals of Statistics* **35,** 2769–2794. ISSN: 0090-5364 (2007).

47. Dagenbach, D. en. in *Connectomics* 27–41 (Elsevier, 2019). ISBN: 978-0-12-813838-0.

48. Varoquaux, G. & Craddock, R. C. Learning and Comparing Functional Connectomes across Subjects. en. *NeuroImage* **80,** 405–415. ISSN: 10538119 (2013).

49. Craddock, R. C., Tungaraza, R. L. & Milham, M. P. Connectomics and New Approaches for Analyzing Human Brain Functional Connectivity. en. *GigaScience* **4,** 13. ISSN: 2047-217X (2015).

50. Fornito, A., Zalesky, A. & Breakspear, M. Graph Analysis of the Human Connectome: Promise, Progress, and Pitfalls. en. *NeuroImage* **80,** 426–444. ISSN: 10538119 (2013).

51. Wang, J., Wang, X., Xia, M., Liao, X., Evans, A. & He, Y. GRETNA: A Graph Theoretical Network Analysis Toolbox for Imaging Connectomics. *Frontiers in Human Neuroscience* **9.** ISSN: 1662-5161 (2015).

52. Athreya, A., Fishkind, D. E., Levin, K., Lyzinski, V., Park, Y., Qin, Y., Sussman, D. L., Tang, M., Vogelstein, J. T. & Priebe, C. E. Statistical Inference on Random Dot Product Graphs: A Survey. *arXiv:1709.05454 [math, stat].* arXiv: 1709.05454 [math, stat] (2017).

53. Levin, K., Athreya, A., Tang, M., Lyzinski, V. & Priebe, C. E. *A Central Limit Theorem for an Omnibus Embedding of Multiple Random Dot Product Graphs* in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (IEEE, New Orleans, LA, 2017), 964–967. ISBN: 978-1-5386-3800-2.

54. Kim, J., Wozniak, J. R., Mueller, B. A., Shen, X. & Pan, W. Comparison of statistical tests for group differences in brain functional networks. *NeuroImage* **101,** 681–694. ISSN: 10538119. https://linkinghub.elsevier.com/retrieve/pii/S1053811914006144 (2021) (2014).

55. Emerson, J. W., Green, W. A., Schloerke, B., Crowley, J., Cook, D., Hofmann, H. & Wickham, H. The Generalized Pairs Plot. en. *Journal of Computational and Graphical Statistics* **22,** 79–91. ISSN: 1061-8600, 1537-2715 (2013).

56. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9,** 62–66. ISSN: 0018-9472, 2168-2909 (1979).

57. Cox, M. A. A. & Cox, T. F. en. in *Handbook of Data Visualization* 315–347 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008). ISBN: 978-3-540-33036-3 978-3-540-33037-0.

58. Wang, X., Wen, C., Pan, W. & Huang, M. Sure Independence Screening Adjusted for Confounding Covariates with Ultrahigh-Dimensional Data. *Statistica Sinica.* ISSN: 10170405 (2017).

59. Imbens, G. W. & Rubin, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* First. ISBN: 978-0-521-88588-1 978-1-139-02575-1 (Cambridge University Press, 2015).

60. Shen, C. & Vogelstein, J. T. The Chi-Square Test of Distance Correlation. *arXiv:1912.12150 [cs, math, stat].* arXiv: 1912.12150 [cs, math, stat] (2020).

61. Morgan, J. L. & Lichtman, J. W. Why Not Connectomics? en. *Nature Methods* **10,** 494–500. ISSN: 1548-7091, 1548-7105 (2013).

62. Kiar, G., Bridgeford, E. W., Gray Roncal, W. R., Consortium for Reliability and Reproducibility (CoRR), Chandrashekhar, V., Mhembere, D., Ryman, S., Zuo, X.-N., Margulies, D. S., Craddock, R. C., Priebe, C. E., Jung, R., Calhoun, V. D., Caffo, B., Burns, R., Milham, M. P. & Vogelstein, J. T. *A High-Throughput Pipeline Identifies Robust Connectomes But Troublesome Variability* en. Preprint (Neuroscience, 2017).

63. Hoff, P. D., Raftery, A. E. & Handcock, M. S. Latent Space Approaches to Social Network Analysis. en. *Journal of the American Statistical Association* **97,** 1090–1098. ISSN: 0162-1459, 1537-274X (2002).

64. Scheinerman, E. R. & Tucker, K. Modeling Graphs Using Dot Product Representations. en. *Computational Statistics* **25,** 1–16. ISSN: 0943-4062, 1613-9658 (2010).

65. Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic Blockmodels: First Steps. en. *Social Networks* **5,** 109–137. ISSN: 03788733 (1983).

66. Myers, P. E., Arvapalli, G. C., Ramachandran, S. C., Pisner, D. A., Frank, P. F., Lemmer, A. D., Bridgeford, E. W., Nikolaidis, A. & Vogelstein, J. T. *Standardizing Human Brain Parcellations* en. Preprint (Neuroscience, 2019).

67. Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. en. *NeuroImage* **15,** 870–878. ISSN: 10538119 (2002).

68. Efron, B. Simultaneous Inference: When Should Hypothesis Testing Problems Be Combined? en. *The Annals of Applied Statistics* **2,** 197–223. ISSN: 1932-6157 (2008).

69. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6,** 65–70. ISSN: 0303-6898 (1979).

70. Weisstein, E. W. *Bonferroni Correction. {From MathWorld—A Wolfram Web Resource}* 2004.

71. Simpson, S. L., Hayasaka, S. & Laurienti, P. J. Exponential Random Graph Modeling for Complex Brain Networks. *PLoS ONE* **6** (ed Sporns, O.) e20039. ISSN: 1932-6203. https://dx.plos.org/10.1371/journal.pone.0020039 (2021) (25, 2011).

72. Yeh, F.-C. & Tseng, W.-Y. I. NTU-90: A High Angular Resolution Brain Atlas Constructed by q-Space Diffeomorphic Reconstruction. en. *NeuroImage* **58,** 91–99. ISSN: 10538119 (2011).

73. Fang-Cheng Yeh, Wedeen, V. J. & Tseng, W.-Y. I. Generalized ${ Q}$-Sampling Imaging. *IEEE Transactions on Medical Imaging* **29,** 1626–1635. ISSN: 0278-0062, 1558-254X (2010).

74. Yeh, F.-C., Liu, L., Hitchens, T. K. & Wu, Y. L. Mapping Immune Cell Infiltration Using Restricted Diffusion MRI: Restricted Diffusion Imaging. en. *Magnetic Resonance in Medicine* **77,** 603–612. ISSN: 07403194 (2017).

75. Chung, J., Pedigo, B. D., Bridgeford, E. W., Varjavand, B. K., Helm, H. S. & Vogelstein, J. T. GraSPy: Graph Statistics in Python. *Journal of Machine Learning Research* **20,** 1–7 (2019).

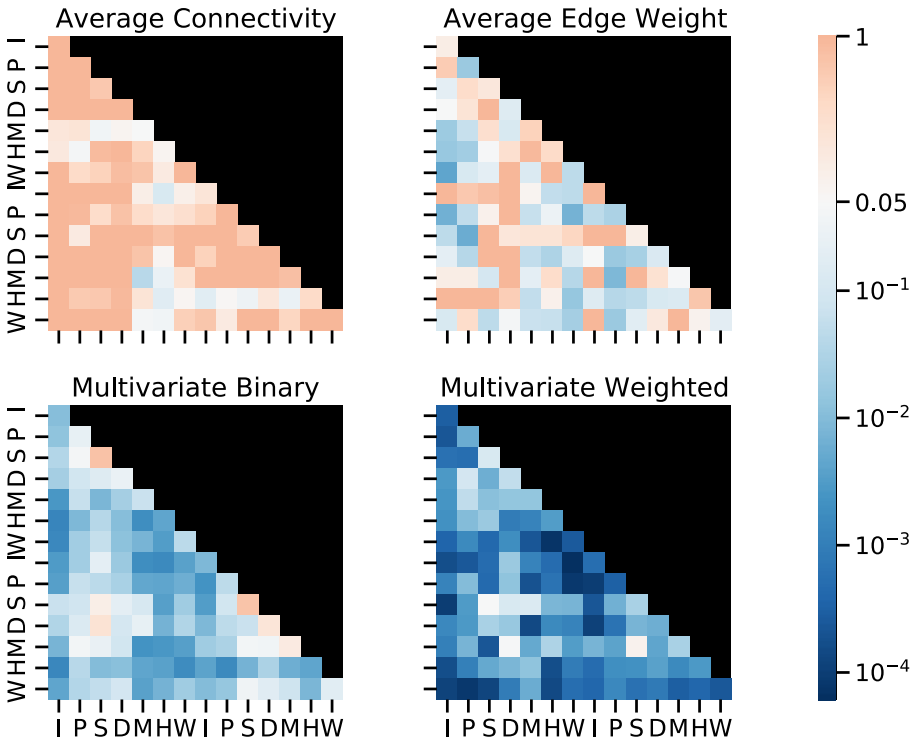
76. Panda, S., Palaniappan, S., Xiong, J., Bridgeford, E. W., Mehta, R., Shen, C. & Vogelstein, J. T. Hyppo: A Comprehensive Multivariate Hypothesis Testing Python Package. *arXiv:1907.02088 [cs, stat].* arXiv: 1907.02088 [cs, stat] (2020).

77. Gu, Z., Eils, R. & Schlesner, M. Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. en. *Bioinformatics* **32,** 2847–2849. ISSN: 1367-4803, 1460-2059 (2016).

A Supplementary investigations of real data

A.1 Comparison of *regional scale* methods

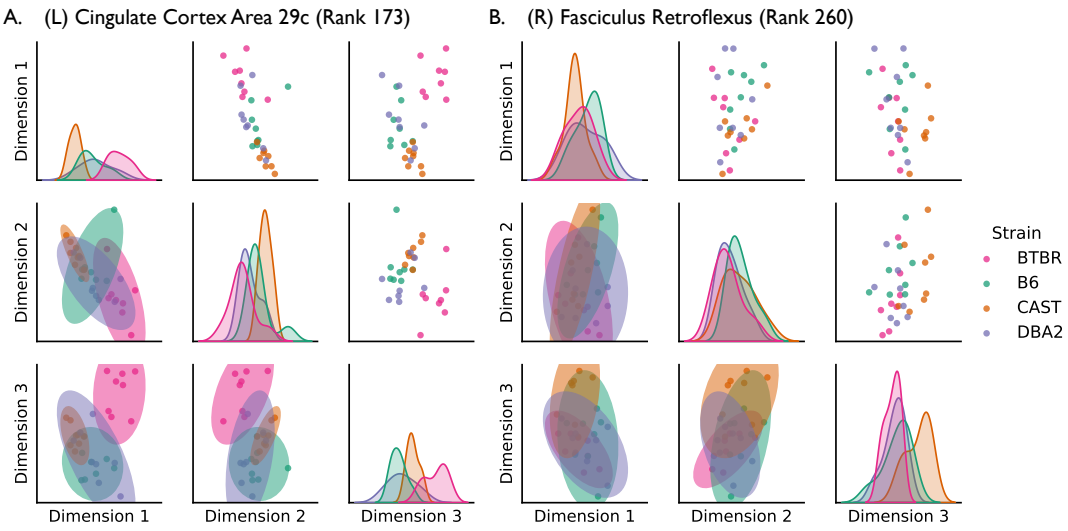
In Supplementary Figure 4, we show the log-transformed p-value for each approach, with communities in blue denoting signal communities (significant after Holm–Bonferroni correction).



Supplementary Figure 4: P-values for each proposed approach for summarizing the information in a communities. More signal communities are found as the amount of information encoded by the summary method increases. Because connectomes in this dataset are undirected, this significance matrix will be symmetric and therefore only the upper triangle of these matrices should be considered. The colorbar represents $\log_{10}$ p-values. Significance is determined at $\alpha = 0.05$ following Holm–Bonferroni correction; therefore, communities in blue are signal communities.

A.2 Embeddings of weak signal vertices

The strongest signal vertex identified by this method in our real-world data was the left hemisphere corpus callosum. In Figure 3, we demonstrate how a *pairs plot*, a d-dimensional scatter plot matrix, can be used to visualize the vertex embedding produced by OMNI. In Supplementary Figure 5, we show pairs plots of two weak signal vertices for comparison.

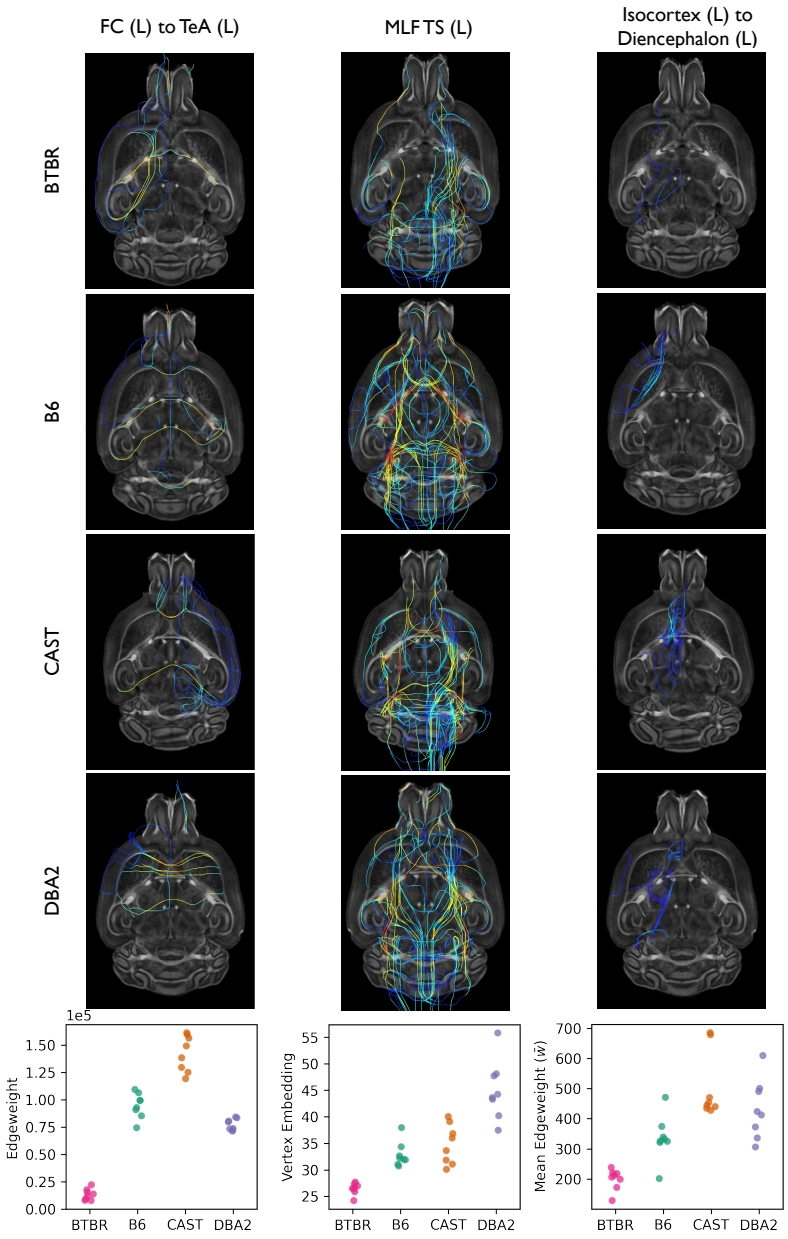


Supplementary Figure 5: Pairs plots of the vertex embeddings of the left cingulate cortex area 29c and the right fasciculus retroflexus, two weak signal vertices identified by OMNI and MANOVA. As shown by the kernel density estimates (*diagonal*) and the 95% prediction ellipses (*lower triangle*), the distribution of these embeddings is not separable. This emphasizes the homogeneity of these vertices across strains relative to the two strongest signal vertices shown in Figure 3.

A.3 Visualizing null edges, vertices, and communities

Using the procedures described above, we identified the weakest signal edge (the left frontal cortex to the right temporal association cortex), vertex (the left medial longitudinal fasciculus and tectospinal tract), and community (left

isocortex to left diencephalon) across all mouse strains. To contrast with the strongest signal structures shown in Figure 4, we plotted tractograms of the weakest components neurological structures (Supplementary Figure 6). These tractograms are much more homogeneous than those plotted in Figure 4, as are the distributions of numerical features for each component (shown in the bottom row). Since this edge, vertex, and community contain no information about the strain of a mouse, we term them *null components*.



Supplementary Figure 6: Visualization of the weakest signal edge (the left frontal cortex to the right temporal association cortex), vertex (the left medial longitudinal fasciculus and tectospinal tract), and community (left isocortex to left diencephalon) across all mouse strains. At each topological scale, tractograms of these neurological structures are shown for each mouse strain. Compared to Figure 4, tractograms here are much more homogeneous across strains, displaying minor differences in connectivity. (*Bottom row*) The distribution of edge weights for the weakest signal edge (*Column 1*); the distribution of the first embedding dimension for the weakest signal vertex (*Column 2*); and the distribution of average edge weight for the most weakest signal community (*Column 3*). Each dot represents data from an individual mouse. Differences in distributions are much less pronounced compared to Figure 4. In these distributions, one strain is typically aberrant while the other three strains are homogeneous.

B Statistical rankings of signal components

Rankings for the signal components in the real-world mouse data set are provided in the following tables.

Vertex 1	Vertex 2	statistic	p-value
Corpus Callosum (L)	Striatum (R)	0.717	0.054
Corpus Callosum (L)	Internal Capsule (R)	0.699	0.073
Corpus Callosum (L)	Reticular Nucleus of Thalamus (R)	0.698	0.074
Corpus Callosum (L)	Zona Incerta (R)	0.686	0.092
Septum (R)	Corpus Callosum (R)	0.671	0.118
Lateral Ventricle (L)	Striatum (R)	0.667	0.125
Striatum (L)	Striatum (R)	0.664	0.132
Corpus Callosum (L)	Ventral Thalamic Nuclei (R)	0.663	0.133
Hippocampus (L)	Middle Cerebellar Peduncle (L)	0.658	0.144
Caudomedial Entorhinal Cortex (R)	Ventral Hippocampal Commissure (R)	0.656	0.150
Corpus Callosum (L)	Midbrain Reticular Nucleus (R)	0.653	0.159
Midbrain Reticular Nucleus (L)	Superior Cerebellar Peduncle (L)	0.648	0.172
Corpus Callosum (L)	Corpus Callosum (R)	0.646	0.179
Spinal Trigeminal Nerve (L)	Middle Cerebellar Peduncle (L)	0.645	0.180
Secondary Visual Cortex L (L)	Striatum (R)	0.641	0.192
Globus Pallidus (R)	Midbrain Reticular Nucleus (R)	0.632	0.225
Striatum (L)	Corpus Callosum (R)	0.632	0.226
Primary Somatosensory Cortex HL (L)	Secondary Visual Cortex MM (L)	0.629	0.236
Corpus Callosum (L)	Primary Somatosensory Cortex J (R)	0.628	0.238
Corpus Callosum (L)	Ventral Orbital Cortex (R)	0.628	0.240

Table 1: The top 20 signal edges (out of 54,946 total edges) ranked by the order of their Holm–Bonferroni corrected p-value. Eleven of the top 20 signal edges are adjacent to either the left or right hemisphere corpus callosum.

Vertex	Pillai	F(15, 78)	p-value
Corpus Callosum (L)	2.591	32.914	5.09e-25
Corpus Callosum (R)	2.556	29.951	1.09e-23
Fimbria (L)	2.440	22.637	7.27e-20
Secondary Motor Cortex (L)	2.438	22.544	8.21e-20
Midbrain Reticular Nucleus (R)	2.430	22.161	1.38e-19
Substantia Nigra (R)	2.305	17.254	2.20e-16
Internal Capsule (R)	2.304	17.229	2.29e-16
Secondary Motor Cortex (R)	2.297	16.989	3.40e-16
Cerebral Peduncle (R)	2.247	15.511	4.34e-15
Internal Capsule (L)	2.238	15.266	6.71e-15
Striatum (L)	2.236	15.230	7.13e-15
Lateral Ventricle (L)	2.218	14.739	1.74e-14
Stria Terminalis (R)	2.202	14.349	3.59e-14
Cerebellar White Matter (R)	2.199	14.278	4.08e-14
Optic Tracts (L)	2.186	13.956	7.52e-14
Subthalamic Nucleus (L)	2.178	13.781	1.05e-13
Hippocampus (R)	2.177	13.764	1.08e-13
Stria Terminalis (L)	2.177	13.747	1.11e-13
Frontal Association Cortex (L)	2.170	13.601	1.47e-13
Rostral Linear Nucleus (R)	2.165	13.473	1.88e-13

Table 2: The top 20 signal vertices (out of 332 total vertices) ranked by the order of their Holm–Bonferroni corrected p-values. Pillai's trace and approximate F statistic (along with degrees of freedom) as calculated by one-way MANOVA are also reported. The corpus callosum in the left and right hemisphere are the top two signal vertices.

	Left Hemisphere		*Right Hemisphere*		
Vertex	**p-value**	**Rank**	**p-value**	**Rank**	**Avg. Rank**
Corpus Callosum	5.09e-25	1	1.09e-23	2	1.5
Secondary Motor Cortex	8.21e-20	4	3.40e-16	8	6.0
Internal Capsule	6.71e-15	10	2.29e-16	7	8.5
Fimbria	7.27e-20	3	1.71e-12	28	15.5
Stria Terminalis	1.11e-13	18	3.59e-14	13	15.5
Lateral Ventricle	1.74e-14	12	1.07e-12	23	17.5
Ventral Tegmental Area	4.18e-12	30	1.04e-12	22	26.0
Hippocampus	6.62e-11	44	1.08e-13	17	30.5
Globus Pallidus	8.49e-11	46	1.12e-12	25	35.5
Cerebral Peduncle	5.79e-10	63	4.34e-15	9	36.0

Table 3: The top 10 bilateral signal vertex pairs (out of 166 total vertex pairs) ranked by the order of their average rank.

Community 1	Community 2	statistic	p-value
White Matter (R)	White Matter (R)	0.888	6.14e-06
Hindbrain (R)	White Matter (R)	0.872	7.87e-06
White Matter (L)	White Matter (L)	0.871	8.02e-06
Subpallium (R)	White Matter (R)	0.852	1.07e-05
Hindbrain (L)	White Matter (L)	0.848	1.14e-05
Isocortex (L)	Isocortex (L)	0.844	1.21e-05
Isocortex (R)	White Matter (R)	0.825	1.64e-05
White Matter (L)	Midbrain (R)	0.809	2.09e-05
Hindbrain (L)	White Matter (R)	0.809	2.09e-05
Isocortex (L)	White Matter (L)	0.810	2.09e-05

Table 4: The top 10 signal communities (out of 105 total communities) ranked by the order of their Holm–Bonferroni corrected p-values as calculated by the Multivariate Weighted method.

Biographical Note

Vivek Gopalakrishnan was born in Bryan, Texas in 1999, the only child of two immigrants from India. He graduated from Lexington High School in 2017 and subsequently enrolled in Johns Hopkins University. While at Hopkins, he studied Biomedical Engineering with a focus in Biomedical Data Science for his Bachelor's of Science degree. He enrolled in the department's combined Bachelor's and Master's program in Biomedical Engineering, where he focused his research on the development of algorithms to extract latent information from large, complex, and noisy biomedical datasets. In May 2021, he earned both his Bachelor's and Master's degrees.

Next year, Vivek will begin pursuing his Ph.D. in Medical Engineering and Medical Physics (MEMP) at the Harvard-MIT Program in Health Science and Technology (HST). Through his research, he hopes to create a future where robust and interpretable machine learning systems are fundamental tools for all biomedical researchers and clinicians, helping transform data into scientific advancements and improved patient outcomes.