

METHODS

Kernel machine tests of association using extrinsic and intrinsic cluster evaluation metrics

Alexandria M. Jensen¹, Peter DeWitt², Brianne M. Bettcher³, Julia Wrobel⁴, Katerina Kechris⁴, Debashis Ghosh^{4*}

1 Quantitative Sciences Unit, Stanford School of Medicine, Palo Alto, California, United States of America, **2** Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora, Colorado, United States of America, **3** Behavioral Neurology Section, Department of Neurology, University of Colorado Alzheimer's and Cognition Center, Aurora, Colorado, United States of America, **4** Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, Colorado, United States of America

* debashis.ghosh@cuanschutz.edu



OPEN ACCESS

Citation: Jensen AM, DeWitt P, Bettcher BM, Wrobel J, Kechris K, Ghosh D (2024) Kernel machine tests of association using extrinsic and intrinsic cluster evaluation metrics. *PLoS Comput Biol* 20(11): e1012524. <https://doi.org/10.1371/journal.pcbi.1012524>

Editor: Thomas Serre, Brown University, UNITED STATES OF AMERICA

Received: July 21, 2023

Accepted: September 30, 2024

Published: November 11, 2024

Copyright: © 2024 Jensen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available via a public Amazon S3 bucket from the Autism Brain Imaging Data Exchange (ABIDE): <http://preprocessed-connectomes-project.org/abide/download.html>. The ABIDE dataset is a third party dataset not owned or collected by any of the manuscript authors. Access to this, and other, datasets within ABIDE, can be easily obtained by anyone who follows the link provided. All code associated with the manuscript can be found via an R package published on CRAN, CommKern (<https://cran.r-project.org/web/packages/>

Abstract

Modeling the network topology of the human brain within the mesoscale has become an increasing focus within the neuroscientific community due to its variation across diverse cognitive processes, in the presence of neuropsychiatric disease or injury, and over the lifespan. Much research has been done on the creation of algorithms to detect these mesoscopic structures, called communities or modules, but less has been done to conduct inference on these structures. The literature on analysis of these community detection algorithms has focused on comparing them within the same subject. These approaches, however, either do not accommodate a more general association between community structure and an outcome or cannot accommodate additional covariates that may confound the association of interest. We propose a semiparametric kernel machine regression model for either a continuous or binary outcome, where covariate effects are modeled parametrically and brain connectivity measures are measured nonparametrically. By incorporating notions of similarity between network community structures into a kernel distance function, the high-dimensional feature space of brain networks, defined on input pairs, can be generalized to non-linear spaces, allowing for a wider class of distance-based algorithms. We evaluate our proposed methodology on both simulated and real datasets.

Author summary

Parcellating the brain into clusters, which can be characterized as a way to describe the balance between dense relationships among areas highly engaged in the same processing tasks as well as sparser relationships between regions with different processing assignments, has been an area of recent focus within the neuroscientific community. While many algorithms exist to discover and characterize these clusters, there is a paucity of literature seeking to conduct inference on these parcellations between subjects. We have proposed a semiparametric kernel machine regression framework that can accommodate

[CommKern/index.html](#)). A thorough vignette accompanies the package, walking users through the process from either simulating or using real data, to using the accompanying community detection algorithm to obtain communities, to applying the kernel machine tests of association.

Funding: The author(s) received no specific funding for this work.

Competing interests: The author(s) have declared that no competing interests exist.

either a binary or continuous outcome, where brain connectivity measures are modeled nonparametrically and any additional covariates of interest (e.g., age, sex, etc.) are modeled parametrically. Evaluating our proposed methodology on both simulated and real datasets, we provide evidence of the robustness of this method within a single layer of clusters, thus showing the potential utility in the rapidly-changing field of network neuroscience.

Introduction

In the years since Bullmore and Sporns' landmark 2009 paper that first applied network analysis techniques to the question of brain connectivity [1], research in this field has exploded, due to the parallel maturations of the neuroimaging and network science fields [2]. Broadly, connectivity research is concerned with the anatomical links, statistical dependencies, or causal interactions between distinct units in a nervous system. While initial studies focused on local (i.e., groups of neurons) and global (i.e., the entire brain) characteristics [3, 4], the spatiotemporal patterns of connectivity can be seen as a hierarchy based on the ideas of segregation and integration [2]. These intermediate characteristics, known as the mesoscale, span the entire range of organization, from small, functionally specialized areas to larger regions for general brain function and cognition [5]. This organization displays what Herbert Simon deemed as a "nearly-decomposable system," which is a system where elements have most of their interactions with a subset of elements close to them and much less interaction outside of this subset [6, 7]. Simon argued that a network with near-decomposability posed an evolutionary advantage as modules allow for a stable middle ground as complex systems emerge from more simplistic formations [7]. Similarly, many complex networks are also characterized by small-world topology, where the network is composed of dense clustering connections amongst neighboring nodes; this allows for short path lengths between any two nodes and the minimization of long-range connections [8]. Like most man-made and natural information processing systems, the modular (or community) organization of the brain, and within it the characterizations of near-decomposability and small-world topology, is thought to be a method of minimizing energy consumption in the development and maintenance of connections [9]. The composition of these modules underpins the communication patterns of the brain and promotes a well-balanced and efficient mechanism for integration and segregation [3, 10]. The architecture of mesoscale brain networks has been shown to vary over the lifespan [11–14], in the presence of neuropsychiatric disease or injury [15–18], and across diverse cognitive processes [19].

There exist an assortment of methods and algorithms for detecting network communities. Fortunato [20] and later with Newman [21] provide excellent summaries of the various approaches to community detection, as well as Yang, Algesheimer, and Tessone in 2016 [22] and Mukerjee in 2021 [23]. A recent interest in the literature has been topological data analysis, which uses persistent homology to represent the shape of and cluster data [24]. The utility of topological data analysis in neuroscience is beginning to be explored [25, 26], but, much like any other community detection algorithm, is not without its pitfalls and challenges. In the context of brain connectivity networks, once a community detection algorithm has been run on the data, understanding whether the pattern of assigning nodes to communities varies based on some outcome of interest is the next logical step in the analytic pipeline. However, much of the neuroscience literature on analysis of these community detection algorithms has focused on comparing them within the same subject. Taya et al. used a non-parametric permutation test based on the normalized mutual information to determine which algorithm or atlas provided the best representative structure [27]. Within multilayer or multiplex networks, measuring

layer similarity has been explored by [28] as well as [29]. Onnela et al. compared networks by creating profiles of summary statistics that characterized the community structure of each network at different mesoscopic scales [30]. Bródka et al. quantified relative differences by comparing aggregations of layer property vectors, with a distance between observed distributions of these properties based on histograms [31]. Betzel et al. used a similar resolution parameter approach, but then quantified the average dissimilarity of an individual's community assignments across layers to all other subjects, creating entropy vectors that were then used as an input for a principal component analysis [5]. Lohse et al. created a summary of graph summary measures (bipartivity, community laterality, community radius, number of singletons) to create mesoscopic response functions to describe multi-resolution topological structure in various networks [32]. Finally, Alexander-Bloch et al. compared summary measures of community structures (e.g., normalized mutual information) between groups using a permutation test and then used Pearson's phi to assess regionally-specific differences in community structure [33]. These approaches, however, may not be flexible enough to model the linear and nonlinear effects of potential confounders and community structures on an outcome of interest.

One way to accomplish this is through the use of kernels, which are weighting functions used to estimate the conditional expectation of a random variable [34]. Kernels are a powerful tool within machine learning for multidimensional data, with uses ranging from support vector machines [35] to spatial smoothing for functional MRI analyses [36]. Unlike spline-based models, which require the specification of a smoothness condition of an unknown function, the kernel function in a kernel machine method implicitly determines the smoothness property of the unknown function [34]. When working in terms of a kernel function, one can avoid the explicit introduction of a feature vector, therefore allowing for implicit use of feature spaces of high dimensionality, something linear statistical models lack. Kernel-based methods do not make assumptions on the functional form of the association, assuming a more general relationship based on a Hilbert space of functions.

This study proposes a semiparametric kernel machine regression model for covariate and brain connectivity measures on either a continuous or binary outcome, where covariate effects are modeled parametrically and brain connectivity measures are measured nonparametrically. As defining a measure of network similarity is a non-trivial task, our approach has the flexibility to incorporate either extrinsic or intrinsic cluster evaluation metrics [37]. By incorporating these notions of similarity into a kernel distance function, the high-dimensional feature space of brain networks, defined on input pairs, can be generalized to non-linear spaces; this allows for a wider class of distance-based algorithms, to be used for modeling the similarity between two network community structures.

The remainder of this paper is organized as follows. A description of the cluster evaluation metrics, the general framework for distance-based kernels, and the kernel machine tests of association can be found in the **Materials and Methods** section. In **Results**, we apply these methods under a variety of simulation paradigms using a novel simulation technique and to a subsampled dataset from the publicly available, pre-processed Autism Brain Imaging Data Exchange (ABIDE) study. Finally, we conclude the paper with a discussion and proposal of future directions. All code has been made available in the form of an R package published on CRAN, CommKern (<https://cran.r-project.org/web/packages/CommKern/index.html>).

Materials and methods

Cluster evaluation metrics

Within community detection, the methods for evaluating the performance of a particular algorithm can be classified as either (1) extrinsic, requiring ground truth labels or (2)

intrinsic, not requiring ground truth labels. Focusing on extrinsic evaluation, metrics can be further grouped into four families: (1) metrics based on counting pairs; (2) based on set matching; (3) based on entropy; and (4) based on edit distance. All external clustering measures rely on a $r \times k$ contingency matrix N that is induced by a system-clustering \mathcal{C} and a ground truth partitioning \mathcal{T} , where the rows represent the system-clusters $\mathcal{C}_i \in \mathcal{C}$ for $i = 1, \dots, r$ and the columns represent the ground truth partitions $\mathcal{T}_j \in \mathcal{T}$ for $j = 1, \dots, k$. A cell at the intersection of row i and column j contains the count n_{ij} of points that are common to cluster \mathcal{C}_i and ground truth partition \mathcal{T}_j :

$$N(i, j) = n_{ij} = |\mathcal{C}_i \cap \mathcal{T}_j| \tag{1}$$

[28]. In many cases, however, a ground truth labelling does not exist, such as in the case of brain connectivity. To overcome this issue in the context of between layer communities, Ghawi and Pfeffer proposed a two-way matching be performed, where the matching is performed twice, each time with one of the layer’s clusterings playing the role of “ground truth” and the results combined using a harmonic mean to produce a final matching measure [28]. We have extended this logic to comparing clusterings between two graphs, \mathcal{G}_i and \mathcal{G}_j . This approach overcomes the lack of a ground truth clustering commonly seen in real-world applications while preserving the interpretation of such extrinsic clustering metrics.

There are several options for external cluster evaluation metrics. For all, assume $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$ and $\mathbb{C} = \{c_1, c_2, \dots, c_j\}$ are two sets of clusters.

Definition 1. Purity: a measure that quantifies the extent to which a cluster c_i contains elements from only one partition, or measures how “pure” a cluster is [28]. The formal definition of purity for a clustering \mathbb{C} is given by

$$purity(\Omega, \mathbb{C}) = \frac{1}{N} \sum_k \max_j |\omega_k \cap c_j| = \frac{1}{N} \sum_{i=1}^r \max_{j=1}^k \{n_{ij}\} \tag{2}$$

[28]. Purity is bounded in $[0, 1]$ but is not symmetric due to the need to have a ground truth labeling. To get around this issue, the harmonic mean of the purity between two clusterings is taken, where one partition is taken to be the ground truth labeling, then the other.

Definition 2. Normalized Mutual Information (NMI): a measure that quantifies the “amount of information” obtained about one random variable by observing the other and is linked to the concept of entropy, which attempts to characterize the “unpredictability” of a random variable. Mutual information is both non-negative and symmetric. Normalized variants of mutual information allow for comparisons between different clusterings that have different numbers of clusters (but not a different number of nodes, meaning the underlying networks must be of the same size). The formal definition of the normalized mutual information is given by

$$NMI(\Omega, \mathbb{C}) = \frac{I(\Omega, \mathbb{C})}{[H(\Omega) + H(\mathbb{C})]/2} \tag{3}$$

$I(\Omega, \mathbb{C})$ is the mutual information, defined as

$$I(\Omega, \mathbb{C}) = \sum_k \sum_j P(\omega_k \cap c_j) \ln \frac{P(\omega_k \cap c_j)}{P(\omega_k) P(c_j)} \tag{4}$$

where $P(\omega_k)$, $P(c_j)$, and $P(\omega_k \cap c_j)$ are the probabilities of a node being in cluster ω_k , cluster c_j ,

and in the intersection of ω_k and c_j , respectively. H is the entropy, defined as

$$H(\Omega) = -\sum_k P(\omega_k) \ln P(\omega_k) \tag{5}$$

[38]. $I(\Omega, \mathbb{C})$ in Eq (4) measures the amount of information by which our knowledge of the one clustering increases when we know the other. The normalization by the denominator $[H(\Omega) + H(\mathbb{C})]/2$ in Eq 3 ensures that the NMI is always in $[0, 1]$.

Definition 3. Adjusted Rand Index (ARI): An alternative to an information-theoretic interpretation of clustering evaluation (e.g., NMI, purity) is to view clusterings as a series of decisions, one for each of the $N(N - 1)/2$ pairs of nodes in the network. If we define the following:

- a : the number of pairs of nodes in a network that are in the same cluster in Ω and in the same cluster in \mathbb{C}
- b : the number of pairs of nodes in a network that are in different clusters in Ω and in different clusters in \mathbb{C}
- c : the number of pairs of nodes in a network that are in the same cluster in Ω and in different clusters in \mathbb{C}
- d : the number of pairs of nodes in a network that are in different clusters in Ω and the same cluster in \mathbb{C} ,

then the Rand Index [39] is

$$RI = \frac{a + b}{a + b + c + d} = \frac{a + b}{N(N + 1)/2}. \tag{6}$$

Intuitively, $a + b$ can be considered the number of agreements between Ω and \mathbb{C} , while $c + d$ the number of disagreements between Ω and \mathbb{C} . The adjusted Rand Index (ARI) is the corrected-for-chance version of the Rand Index, which establishes a baseline by using the expected similarity of all pairwise comparisons between clusterings specified by a random model.

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}, \tag{7}$$

where for the contingency table detailed in Eq (1), n_{ij} is the diagonal (when $i = j$), a_i are the row sums, and b_j are the column sums. Both the RI and ARI are symmetric, but while the RI can only yield values in $[0, 1]$, the ARI can yield negative values if the index is less than the expected index.

Each of these extrinsic cluster evaluation metrics can be commonly found within the literature but have their drawbacks. Purity generally increases as the number of clusters increases, thus it cannot be used as a tradeoff between the number of clusters and clustering quality. NMI has the potential to be biased because the symmetric normalization can introduce spurious dependence on the labeling. Finally, ARI assumes a hypergeometric distribution, which is not always appropriate if the clusterings are dependent and can be biased in the presence of heavy cluster size imbalance.

Rather than use extrinsic metrics, which rely either on a ground truth labeling or on a harmonic mean between two clusterings, intrinsic evaluation metrics do not rely on the existence of a ground truth. One of the most commonly used intrinsic metrics is the optimized value of the modularity function; this is based on the fact that networks with very similar community

structure should have very similar modularity values. However, not all community detection algorithms find optimal partitions through use of a modularity function; for example, the spinglass algorithm, first proposed by [40], is optimized by finding the ground state of a spinglass model using a Potts Hamiltonian. Use of the optimized modularity or Hamiltonian value from a community detection algorithm as an intrinsic evaluation metric serve the same purpose, as was first proposed by [30]. In it, they created mesoscopic response functions of the spinglass Hamiltonian across a range of resolution parameters and looked at the distance between the two functions in a more qualitative manner [30]. Because the optimized Hamiltonian or modularity values are not bounded in the same manner as the extrinsic cluster evaluation metrics, a p-norm of the differences between these values for two clusterings was used such that

$$d(\mathcal{H}_\Omega, \mathcal{H}_C) = \|\mathcal{H}_\Omega - \mathcal{H}_C\|_p \text{ for } 1 \leq p < \infty. \quad (8)$$

It should be noted, however, that Emmons et al. argues that the lack of a rigorous definition of a community, even within domain, “is the root of discrepancies amongst stand-alone quality metrics and information recovery metrics [41].” Therefore, it is left to the reader to determine which cluster evaluation metric best suits their dataset and carefully interpret the results of such metrics through this lens.

Kernels

Most approaches used in the analysis of graph theoretic measures are based on linear methods, ranging from t-tests to analysis of variance (ANOVA) to multivariable regression. While these methods have thoroughly studied properties and simple mathematical forms, many times analysis on real data requires a nonlinear methodological approach. The brain is an especially complex system to model, whose dynamics are non-linear at multiple levels [42]. Kernels provide a computationally efficient and mathematically tractable method of extending linear methods into nonlinear ones. This work will provide a brief overview of kernels; for a more comprehensive examination, please refer to [43–45].

Broadly speaking, rather than assuming a linear functional relationship, kernel-based methods assume a more general (i.e., non-linear) relationship. A kernel is a function that takes as its inputs vectors in some original space and returns the dot product of vectors in the feature space. More formally, if we have $\mathbf{x}, \mathbf{z} \in \mathcal{X}$ and a map $\phi : \mathcal{X} \rightarrow \mathbb{R}^N$, then $k(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$ is a kernel function. An important concept relating to kernel methods is the reproducing kernel Hilbert space (RKHS). Because all kernels are positive definite, there exists one or more feature spaces for which the kernel defines the inner product, without having to explicitly define such feature space. Using the Moore-Aronszajn theorem, it can be shown that for each kernel k on a set \mathcal{X} , there is a unique space of functions (known as the Hilbert space) on \mathcal{X} for which k is a reproducing kernel [34].

Letting Z be a multidimensional array of variables and i, j be two subjects, then $k(Z_i, Z_j)$ can be used as a measure of similarity between the pair of subjects i, j since $k(\cdot, \cdot)$ is positive definite. This similarity measure can be then be incorporated into a statistical inference framework to test what extent variation in Z between individuals can explain variation in some outcome of interest Y . A range of kernel functions are used in statistics, where the choice of kernel determines the function space used to approximate the relationship between two variables [46]. Within the context of this work, we will be focusing on distance-based kernels, which are denoted as

$$K_d(x_1, x_2) = \exp \left\{ \frac{-d^2(x_1, x_2)}{\rho} \right\}, \quad (9)$$

where $d^2(x_1, x_2)$ is a distance function and ρ an unknown bandwidth or scaling parameter [34]. This will be a proper kernel for the extrinsic cluster evaluation metrics but will not be for the intrinsic metrics unless conditioned on the optimized values as computing Eq (9), where ρ is based on optimizing some objective function, does not guarantee that $K_d(x_1, x_2)$ will be positive definite.

The kernel machine model for binary outcomes

Suppose a dataset consists of n subjects, where for subject $i = (1, \dots, n)$, y_i is a binary outcome taking values in $\{0, 1\}$, \mathbf{x}_i is a $q \times 1$ vector of clinical covariates and \mathbf{z}_i is a $p \times 1$ vector of cluster evaluation metrics. The outcome y_i depends on \mathbf{x}_i and \mathbf{z}_i through the following semiparametric linear model

$$G(E(y_i)) = \mathbf{x}_i^T \boldsymbol{\beta} + h(\mathbf{z}_i), \tag{10}$$

where $G(\cdot)$ is a link function, $\boldsymbol{\beta}$ is a $q \times 1$ vector of regression coefficients, and $h(\mathbf{z}_i)$ is an unknown, centered and smooth function. Here and in the sequel, we will focus on logistic regression, although the ideas and arguments can be adapted to linear regression with some modifications. Eq (10) models covariate effects parametrically and the brain connectivity metric either parametrically or non-parametrically. There are several special cases of Eq (10): when $h(\cdot) = 0$, the model reduces to a standard logistic regression model. When $\mathbf{x}_i = 1$, the model reduces to a least squares kernel machine regression. A hypothesis test can be conducted to determine whether the multidimensional variable set \mathbf{z}_i is associated with y_i , controlling for \mathbf{x}_i , of the form

$$\begin{aligned} H_0 : h(\cdot) &= 0 \\ H_A : h(\cdot) &\neq 0 \end{aligned}$$

[34, 46]. Assuming that $h(\cdot)$ lies within the function space generated by a kernel function $K(\cdot, \cdot)$, $h(\cdot) \in \mathcal{H}_k$, $\boldsymbol{\beta}$ and $h(\cdot)$ can be simultaneously estimated by maximizing the penalized log likelihood function

$$\begin{aligned} J[\boldsymbol{\beta}, k(\cdot)] &= \sum_{i=1}^n \left[y_i \ln \left(\frac{\mu_i}{1 - \mu_i} \right) + \ln(1 - \mu_i) \right] - \frac{\lambda}{2} \|h\|_{\mathcal{H}_k}^2 \\ &= \sum_{i=1}^n \left[Y_i (X_i^T \boldsymbol{\beta}_i + h(Z_i)) - \ln(1 + \exp \{X_i^T \boldsymbol{\beta}_i + h(Z_i)\}) \right] - \frac{\lambda}{2} \|h\|_{\mathcal{H}_k}^2, \end{aligned} \tag{11}$$

where λ is a regularization parameter that reflects the trade off between model complexity and goodness of fit [47]. At its boundaries, $\lambda = 0$ reflects a saturated model, while $\lambda = \infty$ reduces the model to a fully parametric logistic regression model. However, it should be noted that there are two unknown parameters within $J[\boldsymbol{\beta}, h(\cdot)]$: the regularization parameter λ and bandwidth parameter ρ . Intuitively, λ controls the magnitude of the unknown function $h(\cdot)$ while ρ controls the smoothness of $h(\cdot)$ [46]. The choice of ρ has a strong influence on the resulting estimate and choosing the data-driven, minimally optimal value of ρ is crucial. Using the representer theorem, which states that a solution to the penalized log likelihood function

$$\min_{h(\cdot) \in \mathcal{H}_k} [\ell_y(h(x_1), \dots, h(x_n)) + \Omega \|h\|_{\mathcal{H}_k}^2] \tag{12}$$

takes the form

$$h(\mathbf{z}_i) = \sum_{j=1}^n \alpha_j K(\mathbf{z}_i, \mathbf{z}_j) = \mathbf{k}_i^T \boldsymbol{\alpha} \tag{13}$$

[48], then the penalized log likelihood function can be rewritten as

$$J[\boldsymbol{\beta}, \boldsymbol{\alpha}] = \sum_{i=1}^n [y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{k}_i^T \boldsymbol{\alpha}) - \ln(1 + \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{k}_i^T \boldsymbol{\alpha}\})] - \frac{\lambda}{2} \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha}, \tag{14}$$

where $\mathbf{k}_i = \{K(\mathbf{z}_i, \mathbf{z}_1), \dots, K(\mathbf{z}_i, \mathbf{z}_n)\}$ and $\boldsymbol{\alpha}$ is a $n \times 1$ vector of unknown parameters. Solving for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ gives the equations

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \frac{1}{\lambda} \left(\mathbf{I} + \frac{\mathbf{K}}{\lambda} \right)^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) \\ \hat{\boldsymbol{\beta}} &= \left[\mathbf{X}^T \left(\mathbf{I} + \frac{\mathbf{K}}{\lambda} \right)^{-1} \mathbf{X} \right]^{-1} \mathbf{X}^T \left(\mathbf{I} + \frac{\mathbf{K}}{\lambda} \right)^{-1} \mathbf{y} \end{aligned} \tag{15}$$

and then, plugging in $\hat{\boldsymbol{\alpha}}$ from (15) into (13)

$$\hat{h}(\mathbf{z}) = \frac{1}{\lambda} \{K(\mathbf{z}, \mathbf{z}_1), \dots, K(\mathbf{z}, \mathbf{z}_n)\} \left(\mathbf{I} + \frac{\mathbf{K}}{\lambda} \right)^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) \tag{16}$$

[46]. However, it is possible to approach $J[\boldsymbol{\beta}, h(\cdot)]$ from a generalized linear mixed models (GzLMM) perspective. As logistic regression is a special case of GzLMM, the kernel estimator within the semiparametric logistic regression model parallels the penalized quasi-likelihood function from a logistic mixed model, letting $\tau = 1/\lambda$ denote the regularization parameter and ρ remaining the bandwidth parameter [46]. These parameters can be treated as variance components, where $h(\cdot) \sim N(0, \tau \mathbf{K}(\rho))$ can be treated as a subject-specific random effect and the covariance matrix $\mathbf{K}(\rho)$ is an $n \times n$ kernel matrix as previously defined [47]. This means that estimating $\boldsymbol{\beta}$ and $h(\cdot)$ can be done by maximizing the penalized log likelihood

$$J[\boldsymbol{\beta}, h(\cdot)] = \sum_{i=1}^n [y_i(\mathbf{x}_i^T \boldsymbol{\beta} + h(\mathbf{z}_i)) - \ln(1 + \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + h(\mathbf{z}_i)\})] - \frac{1}{2\tau} \mathbf{h}^T \mathbf{K} \mathbf{h}, \tag{17}$$

where $\mathbf{h} = \mathbf{K} \boldsymbol{\alpha}$ and $\tau = 1/\lambda$ [47]. This provides an advantage as it allows for testing of the null hypothesis $H_0 : \tau = 1/\lambda = 0$ without explicit specification of basis functions.

However, under the null hypothesis, the kernel matrix K disappears, which makes ρ a nuisance parameter that is inestimable under the null hypothesis. Davies studied the issue of a nuisance parameter disappearing under the null hypothesis [49], and proposed a score test be used, where the score statistic is treated like a nuisance parameter-indexed Gaussian process. As Eq (14) is a nonlinear function of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, a Newton-Raphson algorithm needs to be implemented to maximize Eq (14) in terms of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. If (j) is the j^{th} iteration of the algorithm, then the $(j + 1)$ step solves

$$\begin{bmatrix} \mathbf{X}^T \mathbf{D}^{(j)} \mathbf{X} & \mathbf{X}^T \mathbf{D}^{(j)} \mathbf{K} \\ \mathbf{D}^{(j)} \mathbf{X} & \tau^{-1} \mathbf{I} + \mathbf{D}^{(j)} \mathbf{K} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^{(j+1)} \\ \boldsymbol{\alpha}^{(j+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{D}^{(j)} \tilde{\mathbf{y}}^{(j)} \\ \mathbf{D}^{(j)} \tilde{\mathbf{y}}^{(j)} \end{bmatrix} \tag{18}$$

where $\tilde{\mathbf{y}}^{(j)} = \mathbf{X} \boldsymbol{\beta}^{(j)} + \mathbf{K} \boldsymbol{\alpha}^{(j)} + (\mathbf{D}^{(j)})^{-1} (\mathbf{y} - \mathbf{X}_i^T \boldsymbol{\beta}^{(j)})$, $\mathbf{D}^{(j)} = \text{diag}[\mathbf{X}_i^T \boldsymbol{\beta}^{(j)} - (1 - \mathbf{X}_i^T \boldsymbol{\beta}^{(j)})]$, and $\mathbf{h}^{(j)} = \mathbf{K} \boldsymbol{\alpha}^{(j)}$. Also noting that $\boldsymbol{\beta}$ and $k(\cdot)$, which depend on τ and ρ , can be estimated using penalized

quasi-likelihood under a logistic mixed model paradigm, then we can rewrite Eq (17)

$$J(\beta(v), v) \approx -\frac{1}{2} \log|V| - \frac{1}{2} \log|X^T V^{-1} X| - \frac{1}{2}(\tilde{y} - X\beta)^T V^{-1}(\tilde{y} - X\beta), \tag{19}$$

where $v = (\tau, \rho)$ and $V = D^{-1} + \tau K$. Then \hat{v} can be solved for in the usual way [47]. However, if the derivative of Eq 13 is taken with respect to τ , then the score test for $H_0 : \tau = 1/\lambda = 0$ can be written as

$$S(\rho) = \frac{Q_\tau(\hat{\beta}_0, \rho) - \mu_Q}{\sigma_Q} \tag{20}$$

where $Q_\tau(\hat{\beta}_0, \rho) = (\tilde{y} - X\hat{\beta}_0)^T DK(\rho)D(\tilde{y} - X\hat{\beta}_0) = (\tilde{y} - \mu_0)^T K(\tilde{y} - \mu_0)$, $\hat{\beta}_0$ is the maximum likelihood estimate of β under the null hypothesis, $\hat{\mu}_0 = \text{logit}^{-1}(X\hat{\beta}_0)$, $\mu_Q = \text{trace}[P_0 K(\rho)]$, $\sigma_Q^2 = 2 * \text{trace}[P_0 K(\rho)P_0 K(\rho)]$, $P_0 = D_0 - D_0 X(X^T D_0 X)^{-1} X^T D_0$, and $D_0 = \text{diag}[\mu_0 - (1 - \mu_0)]$ [47].

$S(\rho)$ under the null hypothesis is an approximate, ρ -indexed Gaussian process, which requires application of Davies' results [49] to get the upper bound for the score test's p-value. It can be seen that large values of $Q_\tau(\hat{\beta}_0, \rho)$ would result in a rejection of H_0 and the upper bound of the p-value is

$$\Phi(-M) + \frac{W \exp\left\{-\frac{1}{2}M^2\right\}}{\sqrt{8\pi}} \tag{21}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function, M is the maximum of $S(\rho)$ over all of the searched range of ρ , $W = |S(\rho_1) - S(L)| + |S(\rho_2) - S(\rho_1)| + \dots + |S(U) - S(\rho_m)|$, L and U are the lower and upper bounds, respectively, of the search area for ρ , and ρ_m are the search points between L and U [47, 49]. Liu et al. suggest setting the lower and upper bounds of the ρ search to be $L = 0.1 \min_{i \neq j} \sum_{l=1}^p (z_{il} - z_{jl})^2$ and $U = 100 \max_{i \neq j} \sum_{l=1}^p (z_{il} - z_{jl})^2$ [47].

The kernel machine model for continuous outcomes

The proposed semiparametric kernel machine model (10) can easily be extended to accommodate other types of outcomes. In this section, we will briefly discuss how to extend the estimation and testing procedures detailed for a binary outcome to a continuous data type.

Suppose a dataset consists of n subjects, where for subject $i = (1, \dots, n)$, x_i is a $q \times 1$ vector of clinical covariates and z_i is a $p \times 1$ vector of cluster evaluation metrics, and y_i is a continuous outcome following a distribution within the exponential family density,

$$p(x, \theta) = h(x) \exp \{ \eta \cdot T(X) - A(\eta) \}, \tag{22}$$

where $T(x)$ is the sufficient statistic of the distribution, or a function of the data that holds all information the data x provides with regards to the unknown parameter values, η is the natural parameter, and $A(\eta)$ is the cumulant function. The mean of y_i satisfies $\mu_i = E(y_i) = \partial A / \partial \eta$ while $Var(y_i) = \partial^2 A / \partial \eta^2$. The outcome y_i depends on x_i and z_i through the following generalized semiparametric linear model

$$g(\mu_i) = x_i^T \beta + h(z_i) + e_i, \tag{23}$$

where $g(\cdot)$ is a known monotone link function and $h(\cdot)$ has the same interpretation as for the

logistic case. For a Gaussian-distributed outcome, $g(\mu) = \mu$ gives us a linear kernel machine model. Thus, the regression coefficients β and nonparametric function $h(\cdot)$ can be obtained by maximizing the penalized log-likelihood function

$$J(h, \beta) = \sum_{i=1}^n \ell\{y_i, \mathbf{x}_i, \mathbf{z}_i; \beta, h(\cdot)\} - \frac{\lambda}{2} \|h\|_{\gamma_k}^2, \quad (24)$$

where $\ell(\cdot) = \log(p)$, p is the density given in Eq (22), and λ is a tuning parameter. Using the kernel expression of $h(\cdot)$ specified in Eq (13), we can write

$$g(\mu_i) = \mathbf{x}_i^T \beta + \mathbf{k}_i^T \alpha, \quad (25)$$

and the penalized likelihood can be rewritten as

$$J(\beta, \alpha) = \sum_{i=1}^n \ell(y_i, \mathbf{x}_i, \mathbf{z}_i; \beta, \alpha) - \frac{1}{2} \lambda \alpha^T \mathbf{K} \alpha, \quad (26)$$

where \mathbf{K} is a $n \times n$ matrix whose i, j^{th} element is $K(\mathbf{z}_i, \mathbf{z}_j)$. From here, the procedure is essentially the same as for the binary outcome case, with μ_i specified under 10 and $\mathbf{D} = \text{diag}\{\text{var}(y_i)\}$ under 22. The score test has a straightforward alteration with the only change occurring in \mathbf{D} , where the variance needs to be replaced with the appropriate variance function for y_i .

Simulated datasets

To understand the statistical properties of the kernel machine tests using either extrinsic or intrinsic cluster evaluation metrics, simulated networks with community structure were constructed. In order to simulate similar community structures amongst a group of networks with a corresponding level of inter-subject variability, Beta distributions were utilized. As functional connectivity within real brain imaging data is typically assessed using some form of correlation and then bounded to be non-negative, the support of the Beta distribution, $[0, 1]$, fits nicely within these restrictions. Two different Beta distributions were simulated for each group of networks: one representing between-community edges and one representing within-community edges. Fig 1 provides four examples of beta distributions simulated to reflect either between-community (bottom panel) or within-community (top panel) edge weights that would be sampled for a simulated group of networks. Well-defined community structure, as well as ill-defined community structure, can be simulated through the shape parameters (α and β).

To create the simulated networks, a data structure is first created with node labels and community memberships of these nodes is randomly assigned. “Edge weights” are assigned to all dyadic relationships. If the dyad’s nodes belong the same community, the within-community beta distribution is sampled; if the dyad’s nodes belong to different communities, the between-community beta distribution is sampled. Fig 2 provides an example of four simulated networks, all with 50 nodes and four communities. For all four simulated networks, the between-group edge weights were sampled from $Beta(1, 8)$. For panel A and C, the within-group edge weights were sampled from $Beta(8, 8)$ while for panel B and D they were sampled from $Beta(4, 8)$. As modeled in the figure, panel A and C are very similar, with C having slightly higher within-community edge weights. Panels B and D have comparatively weaker within-group edge weights compared to their panel A and C counterparts.

Extensions of these simulations allow for perturbation of the community assignment for nodes to create additional variability within networks. A “master” community assignment is created and then, for each network, a pre-specified proportion of nodes are randomly selected

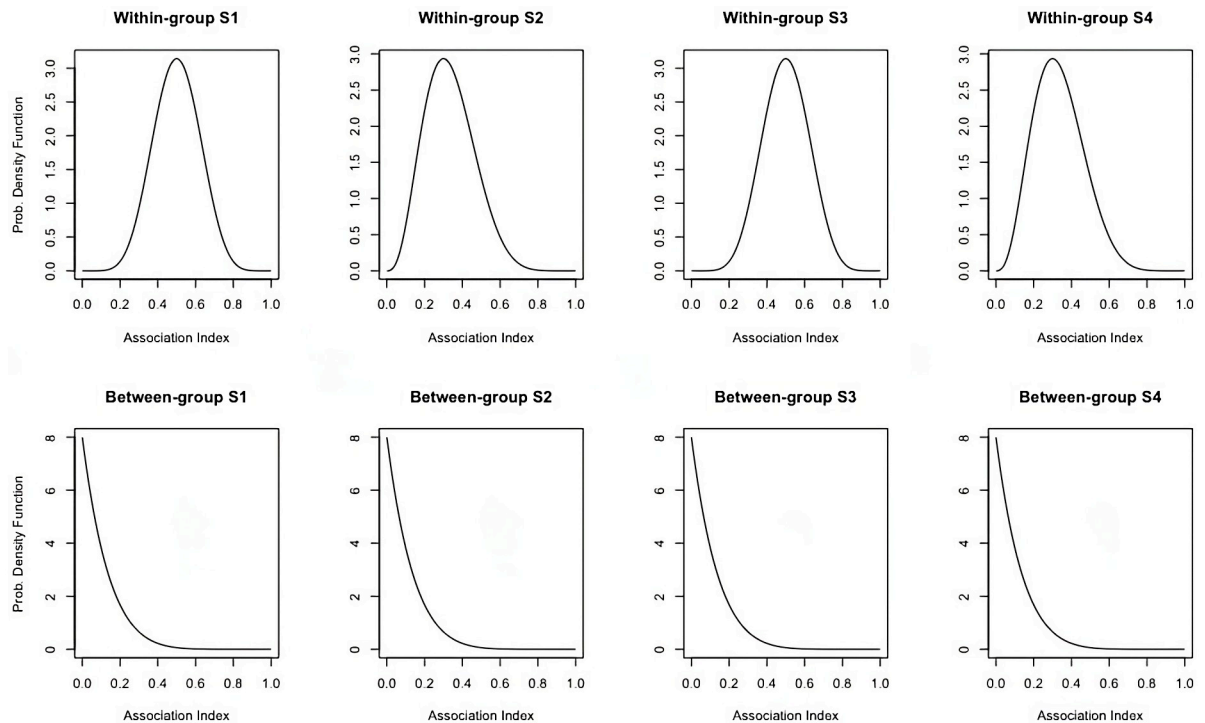


Fig 1. Examples of between- and within-community Beta distributions. Four examples of Beta distributions simulated to reflect the distribution function from which either within-community (top panel) or between-community (bottom panel) edge weights that would be sampled in the simulated network.

<https://doi.org/10.1371/journal.pcbi.1012524.g001>

and reassigned to a new community. From there, the edge weights are assigned to the dyads in the same manner as previously described.

ABIDE dataset

De-identified and pre-processed data from the Autism Brain Imaging Data Exchange (ABIDE) [50] was provided as part of the International Neuroimaging Datasharing Initiative (INDI) consortium's work with the Pre-Processed Connectomes Project (PCP) (<http://preprocessed-connectomes-project.org/abide/index.html>). A total of 1112 datasets, 539 from individuals with Autism Spectrum Disorder (ASD) and 573 typical controls, were aggregated from 16 international imaging sites. For each subject included in ABIDE, a high resolution, 1mm^3 isotropic T1-weighted scan was obtained using a MPRAGE sequence and a resting-state functional MRI (fMRI) was obtained using a BOLD contrast.

The Configurable Pipeline for the Analysis of the Connectomes (C-PAC) pipeline was chosen for the functional pre-processing. The fMRI data was slice time corrected, motion corrected, and the voxel intensity was normalized. Nuisance signal removal was performed using 24 motion parameters, CompCor with five components [51] and low-frequency drifts (linear and quadratic trends), without any global signal regression. Functional data was band-pass filtered (0.01-0.1Hz) and spatially registered using a nonlinear technique to the MNI152 template space. The mean time series for regions of interest were extracted for each subject. The Craddock 200 functional parcellation atlas of the brain [52] was used to reduce the feature vector size. This atlas was generated using a data-driven parcellation of the whole brain into

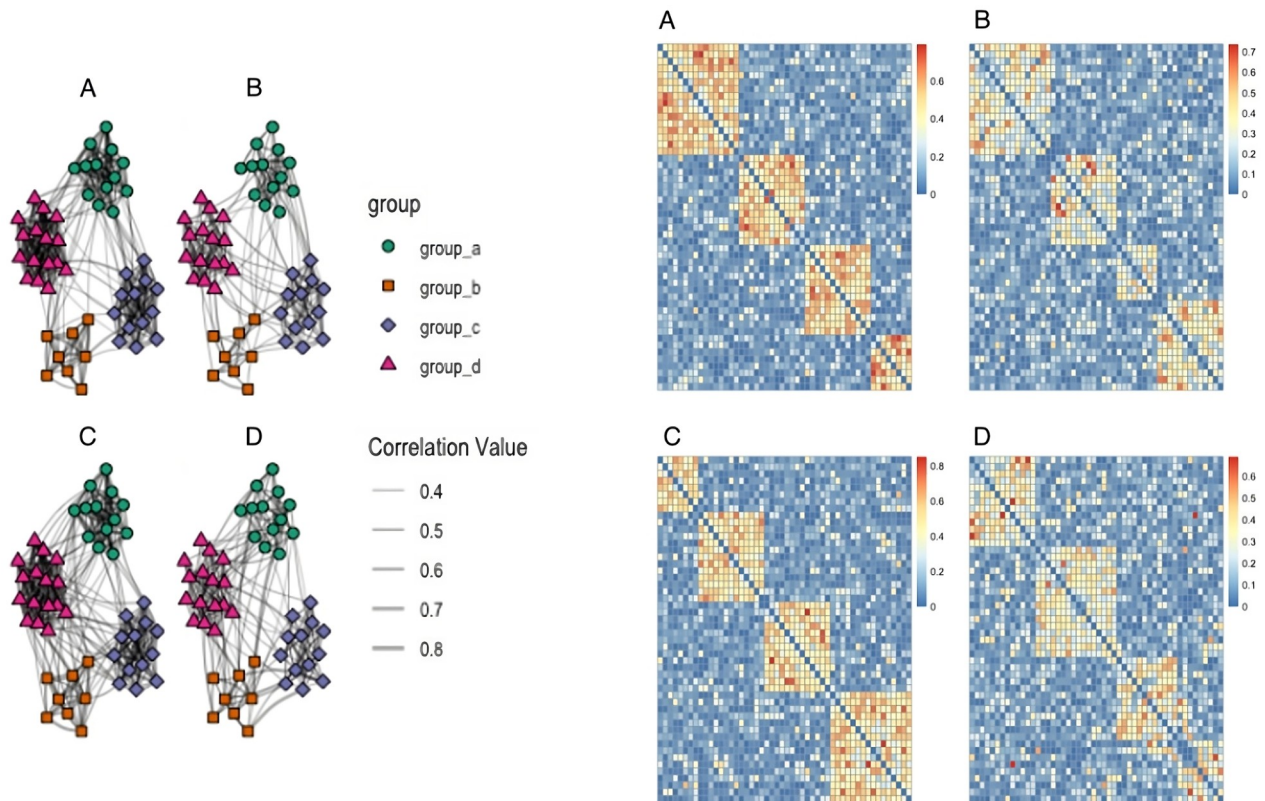


Fig 2. Examples of four simulated networks and adjacency matrices. Four examples of simulated networks of 50 nodes each and four communities. For all simulated networks, the between-group edge weights were sampled from $Beta(1, 8)$. For panels A and C, the within-group edge weights were sampled from $Beta(8, 8)$ while for panels B and D they were sampled from $Beta(4, 8)$.

<https://doi.org/10.1371/journal.pcbi.1012524.g002>

spatially close regions of homogeneous functional activity, resulting in 200 regions. Labels were generated for each of the resulting ROIs from their overlap with the Automated Anatomical Labelling, Eickhoff-Ziles, Harvard-Oxford, and Talarach and Tournoux atlases. Phenotypic information, including age at time of scan, sex, handedness, body mass index (BMI), and an array of neurocognitive testing results were also made available for all subjects.

A stratified random sample was taken from the larger ABIDE pre-processed database. Because every subject in the pre-processed ABIDE database was registered to the same MNI152 template, subjects under 18 years of age or over 45 years of age were excluded due to potential biasing effects during the registration step. The remaining pool of potential subjects were then categorized into one of three buckets: 18-24, 25-34, or 35-45 years of age. This age categorization and the diagnosis group (diagnosed ASD or control) were used in the stratified randomization. Functional connectivity was assessed using partial correlation coefficients. This coefficient is an index of the functional connection of two regions of the brain, along the $[-1, 1]$ range. Negative correlations were zeroed out due their inability to be included in the vast majority of community detection algorithms. Communities were determined using two different algorithms: the Louvain method first proposed by [53] and the spinglass algorithm first proposed by [40]. This was done to understand the robustness of the kernel machine model to choice of community detection algorithm. Resulting community structures as well as the optimized modularity (for the Louvain algorithm) or Hamiltonian (for the spinglass algorithm) values were extracted for each subject's run through the community detection method.

Results

Simulated datasets

To analyze the robustness of the logistic kernel machine model as previously detailed, several simulations were conducted using the simulated network approach previously described. These simulations were split between the number of groups of functional connectivity matrices that were generated. A simulation under a one generation process presumes that the null hypothesis is true (i.e., all subjects come from the same underlying population) while a two generation process presumes that the null hypothesis is false and subjects come from two distinct populations. To simplify the analyses for these simulations, no covariates were generated.

A series of simulation studies were conducted to evaluate the performance of the kernel machine model under the hypothesis test of $H_0 : h(\cdot) = 0$ versus $H_A : h(\cdot) \neq 0$. As there is no closed-form solution for the test statistic's accompanying p-value, power and Type I error were determined empirically. For each of the 1000 power simulation iterations, 100 different networks were produced, with a 45/55 split between "cases" and "controls." For the cases, the between-community edge weights were sampled from $Beta(1.5, 8)$ and the within-community edge weights were sampled from $Beta(7, 8)$. For the controls, the between- and within-community edge weights were sampled from $Beta(1.5, 8)$ and $Beta(8, 8)$, respectively. Each network contained 50 nodes and three communities, with a community assignment perturbation probability set at 0.1. Bounds of the ρ search were set based on the suggestion from [47]. An indicator function was used to determine whether each simulation's resulting p-value was greater than $\alpha = 0.05$, with the empirical power calculated as the proportion of iterations with a p-value < 0.05. The Louvain and spinglass algorithms were each used to determine community assignment for the simulated datasets. Power simulations were conducted for each of the four different cluster evaluation metrics: optimized modularity/Hamiltonian value, purity, NMI, and ARI.

Similarly, for each of the 1000 Type I error simulation iterations, 100 different networks were produced, all with the same between-community edge weights sampled from $Beta(1.5, 8)$ and within-community edge weight sampled from $Beta(8, 8)$. Each network contained 50 nodes and three communities, with a community assignment perturbation probability set at 0.1. Networks were randomly assigned to either a "case" or "control" group with a 45/55 split. Bounds of the ρ search were again set based on the suggestion from [47]. An indicator function was used to determine whether each simulation's resulting p-value was less than $\alpha = 0.05$, with the empirical Type I error calculated as the proportion of iteration with a p-value < 0.05. The Louvain and spinglass algorithms were each used to determine community assignment and Type I error simulations were conducted for the same four cluster evaluation metrics.

Table 1, below, details the power and Type I error results for all of the cluster evaluation metric simulations. As can be seen, both the intrinsic cluster evaluation metrics (optimized modularity/Hamiltonian value) and the three extrinsic metrics (NMI, ARI, and purity) had comparable power and Type I errors for the nonparametric kernel logistic regression.

Table 1. Logistic kernel: Power and Type I error for cluster evaluation metric simulations.

	Louvain		Spinglass	
	Power	Type I Error	Power	Type I Error
Optimized Algorithmic Value	0.940 (SE: 0.008)	0.038 (SE: 0.006)	0.984 (SE: 0.004)	0.040 (SE: 0.006)
Normalized Mutual Information (NMI)	0.990 (SE: 0.003)	0.030 (SE: 0.005)	0.990 (SE: 0.003)	0.012 (SE: 0.003)
Adjusted Rand Index (ARI)	0.986 (SE: 0.004)	0.018 (SE: 0.004)	0.988 (SE: 0.003)	0.018 (SE: 0.004)
Purity	0.976 (SE: 0.005)	0.024 (SE: 0.005)	0.981 (SE: 0.004)	0.023 (SE: 0.005)

<https://doi.org/10.1371/journal.pcbi.1012524.t001>

Power and Type I error simulations were also conducted for the linear kernel machine method to evaluate its performance under the hypothesis test of $H_0 : h(\cdot) = 0$ versus $H_A : h(\cdot) \neq 0$. Again, simulations were used as there is no closed form solution for the test statistic's p-value. For each of the 1000 power simulation iterations, 100 different networks were produced, with a 45/55 split between “cases” and “controls.” For the cases, the between-community edge weights were sampled from $Beta(1.5, 8)$ and the within-community edge weights were sampled from $Beta(7, 8)$. For the controls, the between- and within-community edge weights were sampled from $Beta(1.5, 8)$ and $Beta(8, 8)$, respectively. Each network contained 50 nodes and three communities, with a community assignment perturbation probability set at 0.1. Two covariates—one continuous and one binary—were randomly simulated using a $N(\mu = 10, \sigma = 5)$ and $Binom(n = 100, p = 0.25)$ distribution, respectively. The outcome variable was distributed under the following statistical model: $y_i = \beta_0 + \beta_1 * (T_{group_i=1}) + \epsilon_i$, where $\beta_0 = 5, \beta_1 = -4, \epsilon \sim N(\mu = 0, \sigma = 2)$. Bounds of the ρ search were set based on the suggestion from [47]. An indicator function was used to determine whether each simulation's resulting p-value was greater than $\alpha = 0.05$, with the empirical power calculated as the proportion of iterations with a p-value < 0.05 . These power simulations were conducted for the four different cluster evaluation metrics: optimized modularity/Hamiltonian value, purity, NMI, and ARI.

Similarly, for each of the 1000 Type I error simulation iterations, 100 different networks were produced, all with the same between-community edge weights sampled from $Beta(1.5, 8)$ and within-community edge weight sampled from $Beta(8, 8)$. Each network contained 50 nodes and three communities, with a community assignment perturbation probability set at 0.1. Networks were randomly assigned to either a “case” or “control” group with a 45/55 split. Two covariates—one continuous and one binary—were randomly simulated using a $N(\mu = 10, \sigma = 5)$ and $Binom(n = 100, p = 0.25)$ distribution, respectively. The outcome variable was similarly randomly generated under a $N(\mu = 100, \sigma = 15)$ distribution. Bounds of the ρ search were again set based on the suggestion from [47]. An indicator function was used to determine whether each simulation's resulting p-value was less than $\alpha = 0.05$, with the empirical Type I error calculated as the proportion of iterations with a p-value < 0.05 . These Type I error simulations were conducted for the four different cluster evaluation metrics: optimized modularity/Hamiltonian value, purity, NMI, and ARI.

Table 2, below, details the power and Type I error results for all of the cluster evaluation metric simulations. As can be seen, both the intrinsic cluster evaluation metric (Hamiltonian value) and the three extrinsic metrics (NMI, ARI, and purity) had comparable power and Type I errors for the semiparametric linear kernel regression, and had comparable test characteristics to the semiparametric logistic kernel regression.

For both the logistic and linear kernel machine methods, the extrinsic cluster evaluation metrics (NMI, ARI, and purity) all displayed higher power and lower Type I error rates as compared to the intrinsic metric. A higher level of inter-subject variability exists for the optimized modularity/Hamiltonian value, due to not only the potential differences in the tolerance stopping criteria of the community detection algorithm implemented but also the influence of

Table 2. Linear kernel: Power and Type I error for cluster evaluation metric simulations.

	Louvain		Spinglass	
	Power	Type I Error	Power	Type I Error
Optimized Algorithmic Value	0.884 (SE: 0.010)	0.060 (SE: 0.008)	0.978 (SE: 0.005)	0.042 (SE: 0.006)
Normalized Mutual Information (NMI)	0.990 (SE: 0.003)	0.012 (SE: 0.003)	0.987 (SE: 0.004)	0.010 (SE: 0.003)
Adjusted Rand Index (ARI)	0.985 (SE: 0.004)	0.006 (SE: 0.002)	0.983 (SE: 0.004)	0.020 (SE: 0.004)
Purity	0.988 (SE: 0.003)	0.018 (SE: 0.004)	0.979 (SE: 0.005)	0.012 (SE: 0.003)

<https://doi.org/10.1371/journal.pcbi.1012524.t002>

the distribution of edge weights within and between communities. These effects may be more pronounced for the Louvain community detection algorithm, which saw a larger difference in power and Type I error for both the logistic and linear kernel machine methods.

ABIDE dataset

Table 3 provides a summary of the phenotypic information on the subsampled ABIDE dataset, both overall and by diagnosis group. Initially, the subsample included 85 subjects, but two had to be excluded due to substantial missingness issues in their fMRI time series data, resulting in the analytic sample size of $n = 83$. As can be seen, the male/female split is quite skewed, but this is not surprising given ASD is identified in females at a substantially lower rate than males, with many studies reporting an approximate 4:1 male to female ratio [54]. Handedness was more evenly distributed for those with ASD as compared to controls. Full-scale IQ scores were equivalent, as was the distribution of subjects to the 10 sites in the subsample. It should be noted that in the larger ABIDE sample, there were a total of 20 sites. However, 10 of these sites only imaged pediatric subjects, excluding them from this subsample.

Quality checks were performed for the fMRI times series data, which included looking at missingness of time series data within regions of interest (ROIs), summary statistics, and time series plots. Three functionally and anatomically diverse ROIs (thalamus, frontal orbital cortex, and middle occipital gyrus) were plotted to check for range of time series values and overall trends. During this quality assurance process, four ROIs had missing time series data for at least one subject:

- Right temporal pole (missing for one subject)
- left temporal pole (missing for eight subjects)
- left cerebellum (missing for one subject)

Table 3. ABIDE dataset subject demographics.

	Overall (n = 83)	Controls (n = 44)	ASD (n = 39)
Age in years, mean (SD)	26.50 (6.63)	27.12 (7.01)	25.81 (6.18)
Sex, n (%)			
Male	75 (90.4%)	37 (84.1%)	38 (97.4%)
Female	8 (9.6%)	7 (15.9%)	1 (2.6%)
Handedness, n(%)			
Ambidextrous	2 (4.1%)	1 (3.6%)	1 (4.8%)
Left	3 (6.1%)	0 (0.0%)	3 (14.3%)
Right	34 (87.8%)	26 (92.9%)	17 (81.0%)
NA	1 (2.0%)	1 (3.6%)	0 (0.0%)
Site, n(%)			
California Inst of Tech	10 (12.0%)	6 (13.6%)	4 (10.3%)
Carnegie Mellon Univ	7 (8.4%)	5 (11.4%)	2 (5.1%)
Univ of Leuven, Sample 1	7 (8.4%)	3 (6.8%)	4 (10.3%)
Ludwig Maximilians Univ Munich	8 (9.6%)	4 (9.1%)	4 (10.3%)
New York Univ	11 (13.3%)	7 (15.9%)	4 (10.3%)
Olin Inst of Living at Hartford Hosp	4 (4.8%)	1 (2.3%)	3 (7.7%)
Univ of Pittsburgh	7 (8.4%)	5 (11.4%)	2 (5.1%)
Social Brain Lab, Netherlands	10 (12.0%)	4 (9.1%)	6 (15.4%)
Trinity Centre for Health Sciences	5 (6.0%)	4 (9.1%)	1 (2.6%)
Univ of Utah	14 (16.9%)	5 (11.4%)	9 (23.1%)

<https://doi.org/10.1371/journal.pcbi.1012524.t003>

- right temporal fusiform cortex, anterior (missing for one subject)

All four of these ROIs are located in the most inferior portion of the brain, where the signal-to-noise ratio is especially poor due to its proximity to the spinal column, skeleton, and sinus cavity. Depending on the field of view of the scan, the inferior part of the brain can be cut off. As the assumption for all cluster evaluation metrics, whether intrinsic or extrinsic, is that the input (the network) is the same, these four ROIs were dropped from all subjects' fMRI datasets. The final analytic fMRI dataset was, therefore, 196 ROIs in size.

Fig 3 provides a visual summary of the functional connectivity for subjects with ASD versus control subjects. As can be seen, average and standard deviation functional connectivity matrices for the two groups look overall very similar, though the histograms show a slightly wider range of possible values, with a higher average correlation value for controls ($\bar{x}_{control} = 0.22$) compared to those with ASD ($\bar{x}_{ASD} = 0.19$) and a higher standard deviation, $s_{control} = 0.17$ and $s_{ASD} = 0.15$, respectively.

Neither community detection method required the pre-specification of the number of communities. The Louvain method is agglomerative, where each node starts as its own community and the method iteratively merges communities if doing so increases the partition's modularity. The process stops if no such merge is possible. Therefore, the number of communities does not need to be specified a priori but rather is determined algorithmically. The spinglass algorithm does require a pre-specification of the maximum number of possible communities ($max(n_c)$) but the optimal solution can include empty communities (or spin states). The choice of $max(n_c)$ is a balance between finding the optimal, data-driven partition and computation time. The preferable number of communities was found through an exploratory analysis, whereby a random subset of 10 subjects from the analytic sample were selected and the spinglass algorithm was run, varying $max(n_c)$ from 3 to 10 communities to determine the consensus number of communities being used by the algorithm across subjects. We recommend such exploratory analysis for any real dataset when using the spinglass algorithm for community detection.

For the binary outcome of diagnosis group (ASD versus control), three different regression models were tested. First, a fully nonparametric logistic model was considered, which did not include any covariates to be parametrically estimated:

$$\text{logit}[\Pr(\mathbf{y} = 1)] = h(\mathbf{Z}), \quad (27)$$

where \mathbf{Z} is the multidimensional distance matrix representing one of the four cluster evaluation metrics: Hamiltonian value, NMI, ARI, or purity. Next, a simple semiparametric logistic model was considered, where the only covariate parametrically estimated was site:

$$\text{logit}[\Pr(\mathbf{y} = 1)] = \beta_0 + \beta_1 * (\text{site}) + h(\mathbf{Z}). \quad (28)$$

Previous studies have shown considerable variability in fMRI measurements due to different scanner manufacturers, acquisition protocols, and even climate effects [55, 56]. Therefore, site was included in the model to control for this potential source of confounding between the cluster evaluation metrics and diagnostic group. Finally, a fuller semiparametric model was fit to control for both site-specific effects and subject-level effects of age and sex:

$$\text{logit}[\Pr(\mathbf{y} = 1)] = \beta_0 + \beta_1 * (\text{site}) + \beta_2 * (\text{age}) + \beta_3 * (\text{sex}) + h(\mathbf{Z}). \quad (29)$$

Table 4 details the results of the score tests for the logistic kernel score tests under the null hypothesis of $H_0 : h(\cdot) = 0$ for all three regression models and the four different cluster evaluation metrics using both the Louvain and spinglass approaches for community detection. In all cases, we failed to reject the null hypothesis. Interestingly, inclusion of potential confounders affected the p-values for the intrinsic and extrinsic metrics in opposing directions when using

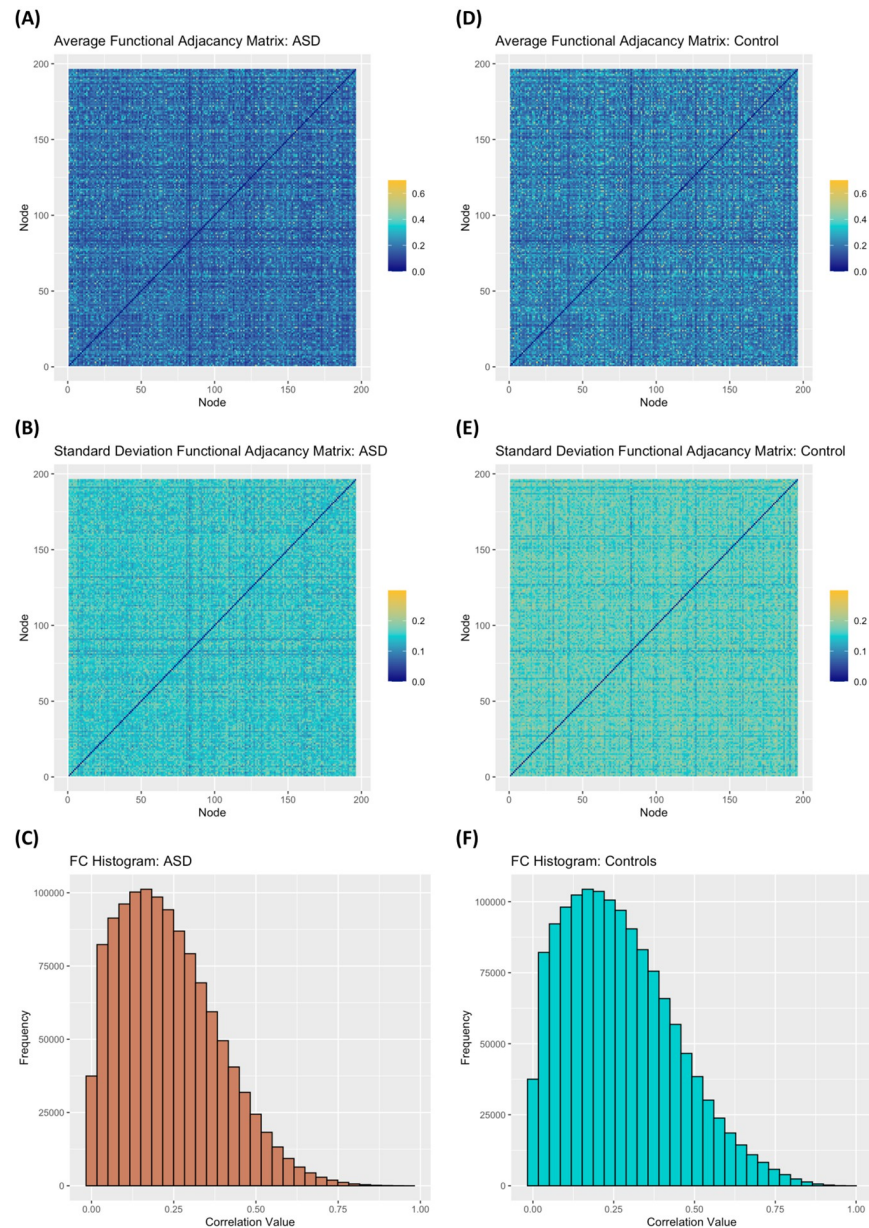


Fig 3. Distribution of cross correlations in functional connectivity in ABIDE sample. For subjects with ASD: (A) Average functional connectivity matrix; (B) Standard deviation of the functional connectivity matrix; (C) Cross correlation distribution of functional connectivity. For control subjects: (D) Average functional connectivity matrix; (E) Standard deviation of the functional connectivity matrix; (F) Cross correlation distribution of functional connectivity.

<https://doi.org/10.1371/journal.pcbi.1012524.g003>

the spinglass community detection algorithm while p-values for all metrics were affected in a similar manner when using the Louvain approach for community detection.

Similarly, the three regression models specified for the binary outcome of diagnostic group, Eqs (27)–(29), were fit for the continuous outcome of verbal IQ score. Intelligence tests, while generally controversial, are an important part of the diagnostic assessment for individuals with suspected ASD. Previous studies have shown that while IQ tests appear to be equivalent and

Table 4. P-values from analysis of ABIDE dataset for binary outcome of diagnostic group.

(a) Louvain algorithm				
Regression	Modularity	NMI	ARI	Purity
Non-parametric	$p = 0.643$	$p = 0.398$	$p = 0.420$	$p = 0.462$
Semiparametric: Site only	$p = 0.325$	$p = 0.262$	$p = 0.265$	$p = 0.309$
Semiparametric: Site, Age, Sex	$p = 0.349$	$p = 0.218$	$p = 0.220$	$p = 0.230$
(b) Spinglass algorithm				
Regression	Hamiltonian	NMI	ARI	Purity
Non-parametric	$p = 0.126$	$p = 0.506$	$p = 0.492$	$p = 0.486$
Semiparametric: Site only	$p = 0.371$	$p = 0.287$	$p = 0.320$	$p = 0.328$
Semiparametric: Site, Age, Sex	$p = 0.317$	$p = 0.220$	$p = 0.273$	$p = 0.274$

<https://doi.org/10.1371/journal.pcbi.1012524.t004>

interchangeable in control subjects, this may not be the case for individuals with ASD [57, 58]. This is especially the case with verbal IQ [57]. Within the ABIDE dataset, full-scale intelligence quotient (FIQ), verbal IQ (VIQ), and performance IQ (PIQ) were derived from the Wechsler Abbreviated Scales of Intelligence (WASI), which has a scale from [50, 160]. However, verbal IQ scores were missing for $n = 19$ subjects in the analytic dataset and had to be dropped. This subset, therefore, included a total of 64 subjects, 29 with ASD and 35 controls.

Table 5 details the results of the score tests for the linear kernel score tests under the null hypothesis of $H_0 : h(\cdot) = 0$ for all three regression models and the four different cluster evaluation metrics. As can be seen, using the spinglass algorithm, several cluster evaluation metrics under the semiparametric paradigms were found to reject the null hypothesis at the $\alpha = 0.05$ significance level; all of these significant p-values came from the extrinsic cluster evaluation metrics. While no such statistical significance was found using the intrinsic metric for either the Louvain or spinglass algorithms, this could be due to a variety of factors. As previously detailed, there is a higher level of inter-subject variability for the optimized modularity/Hamiltonian value, due to not only the potential differences in the tolerance stopping criteria of the algorithms but also the influence of the distribution of edge weights within and between communities.

As a comparison, each subject's optimized algorithmic value was included in conventional regression frameworks for both the binary outcome of diagnostic group and the continuous outcome of verbal IQ score. Table 6 provides the p-values for the regression coefficient associated with the optimized value cluster evaluation metric. In all cases, this measure was not found to be significantly associated with either the binary outcome of diagnostic group or

Table 5. P-values from analysis of ABIDE dataset for continuous outcome of VIQ.

(a) Louvain algorithm				
Regression	Modularity	NMI	ARI	Purity
Non-parametric	$p = 0.441$	$p = 0.434$	$p = 0.449$	$p = 0.473$
Semiparametric: Site only	$p = 0.091$	$p = 0.133$	$p = 0.145$	$p = 0.169$
Semiparametric: Site, Age, Sex	$p = 0.064$	$p = 0.066$	$p = 0.074$	$p = 0.084$
(b) Spinglass algorithm				
Regression	Hamiltonian	NMI	ARI	Purity
Non-parametric	$p = 0.689$	$p = 0.213$	$p = 0.300$	$p = 0.238$
Semiparametric: Site only	$p = 0.351$	$p = 0.048$	$p = 0.051$	$p = 0.019$
Semiparametric: Site, Age, Sex	$p = 0.227$	$p = 0.020$	$p = 0.022$	$p = 0.005$

<https://doi.org/10.1371/journal.pcbi.1012524.t005>

Table 6. Analysis of ABIDE dataset: Conventional regression frameworks for optimized algorithmic value.

Diagnostic Group	Louvain		Spinglass	
		<i>p</i>		<i>p</i>
	Modularity value only	<i>p</i> = 0.804	Hamiltonian value only	<i>p</i> = 0.095
	Modularity, site	<i>p</i> = 0.473	Hamiltonian, site	<i>p</i> = 0.458
	Modularity, site, age, sex	<i>p</i> = 0.309	Hamiltonian, site, age, sex	<i>p</i> = 0.402
Verbal IQ	Modularity value only	<i>p</i> = 0.351	Hamiltonian value only	<i>p</i> = 0.882
	Modularity, site	<i>p</i> = 0.167	Hamiltonian, site	<i>p</i> = 0.989
	Modularity, site, age, sex	<i>p</i> = 0.581	Hamiltonian, site, age, sex	<i>p</i> = 0.601

<https://doi.org/10.1371/journal.pcbi.1012524.t006>

continuous outcome of verbal IQ. Because of the nature of all the external cluster evaluation metrics—normalized mutual information, adjusted Rand Index, and purity—their incorporation into a conventional regression framework was not possible. However, given the comparable empirical power and Type I error rates for all four cluster evaluation metrics for both the Louvain and spinglass community detection algorithms, we believe the conventional regression frameworks in Table 6 provide a sufficient approach comparison to the kernel machine methods.

Discussion

In this work, we proposed a kernel machine methodological approach for testing the association between cluster evaluation metrics and an outcome of interest, whether binary or continuous. Additional covariates can be modeled parametrically while cluster evaluation metrics are measured nonparametrically, allowing for its relationship to the outcome to be modeled without making any assumption as to the parametric form of its association. Our approach has the flexibility to incorporate not only extrinsic or intrinsic cluster evaluation metrics, but more general distance-based metrics, allowing for broad utilization of this approach. With our application to brain connectivity, by incorporating these notions of similarity into a kernel distance function, the high-dimensional feature space of brain networks, defined on input pairs, can be generalized to non-linear spaces; this allows for a wider class of distance-based algorithms, rather than the restrictive squared distance, to be representative of the similarity between two network community structures. The semiparametric form of these kernel machine methods also allow for potential confounders to be controlled for within a parametric framework, allowing for ease of parameter estimate interpretation, should they be desired. Our simulation studies showed the linear and logistic kernel machine regressions both performed well using either extrinsic or intrinsic cluster evaluation metrics, regardless of the underlying community detection algorithm implemented. Further, despite the high degree of flexibility, our proposed methodology is also computationally efficient. While we did not find significant associations between the cluster evaluation metrics and the outcomes of interest in the ABIDE dataset application, we do not see this as a limitation to our proposed method. Our kernel machine approach serves as a proof of concept for the incorporation of the results of community detection algorithms into questions of association with outcomes of interest, whether continuous or binary, without assuming the functional form of the association. Though we focused on brain functional connectivity in this study, the kernel machine approach can be more broadly applied to any type of network analysis (e.g., social networks, metabolomics, genomics, etc.).

Several methodological considerations are relevant for this work. First, while the kernel machine methods have been shown to have high power and low Type I error rate, they can only be applied to a single layer of communities at a time. When applied to the pre-processed ABIDE subset, no significant associations were found between any of the cluster evaluation metrics and the binary outcome of diagnostic group. However, the linear kernel method showed several significant associations between the extrinsic cluster evaluation metrics and verbal IQ score when using the spinglass algorithm. The simulated results combined with the real data analysis provide evidence of the robustness of this method within a single layer of communities, thus showing the potential utility of multilayer extensions to community detection algorithms. As well, use of partial correlation as the metric for functional connectivity in the ABIDE dataset could be creating interference in the robustness of the connectivity networks themselves. The suppression of the negative partial correlation values could also be masking important information about subjects' functional connectivity; while there exist many algorithms to discover the community structure of networks, there is a limited selection that can account for negative links. As such, we focused on the more widely used community detection algorithms, which do not allow for negative network links, to provide a proof-of-concept and allow for exploration of extensions in future research. Alternatives such as wavelet coherence could be explored, though further research into the frequency band of interest, and whether that differs when applying to subjects with neurodivergence or neuropsychological diseases, needs to be thoroughly investigated. Finally, choice of community detection algorithm on the network data could have a downstream impact on the results of the kernel regressions. For this work, we decided to focus on two community detection algorithms that showed robust performance by Yang, Algesheir, and Tessone [22]: the Louvain (or multilevel) and spinglass algorithms. On a set of Lancichinetti, Fortunato & Radicchi (LFR) benchmark graphs, the accuracy and computation time for the Louvain and spinglass algorithms was found to be comparable for increasing levels of the mixing parameter and network size (number of nodes). As such, we chose to focus on community detection algorithms that showed a high level of accuracy across a variety of LFR benchmark scenarios but differed in their discovery approaches (modularity maximization versus ground state energy). This was done to control for variability due to the community detection algorithm and focus on variability due to either cluster evaluation metric and/or the kernel machine test of association itself. A comparison of community detection algorithm performance is outside the scope of this work, but has been previously examined in depth by [59], [22], and [23].

Potential future directions of this work could extend the semiparametric kernel machine methods to account for nested hierarchical structures, which may be a more accurate representation of brain connectivity's integrated and segregated features. This could be done in one of two ways: (1) through the incorporation of random effects or (2) through a multiple comparisons correction to the score test's p-value. [60] derived a distance-based kernel association test for generalized linear mixed models for use in microbiome data. This derivation could be leveraged for our purposes to control for the nested hierarchical structure of communities in brain connectivity data. Similarly, [61] presented a statistical method to test for the association between a response variable and a fixed tree structure across all levels of the hierarchy while controlling for the overall false positive error rate. While their approach is based on tree structures, rather than nested hierarchical clusterings, their derivation of a multiscale statistic that includes a penalty term for multiple comparisons could be modified for use in the kernel score test of association.

Another future direction of this work would be to investigate other methods for incorporating summaries of the community structure—or the community structure itself—into the kernel machine framework. While we explored both extrinsic and intrinsic cluster evaluation

metrics in this work, there is active work in the literature on ways to assess a community detection algorithm's performance and comparing communities across networks. Understanding how advances in the literature affect the kernel machine methods generally and in the context of brain connectivity would be an important step in the argument in favor of kernel methods.

Author Contributions

Conceptualization: Alexandria M. Jensen.

Formal analysis: Alexandria M. Jensen.

Methodology: Alexandria M. Jensen, Debashis Ghosh.

Software: Alexandria M. Jensen, Peter DeWitt.

Supervision: Brianne M. Bettcher, Julia Wrobel, Katerina Kechris, Debashis Ghosh.

Writing – original draft: Alexandria M. Jensen, Peter DeWitt, Brianne M. Bettcher, Julia Wrobel, Katerina Kechris, Debashis Ghosh.

Writing – review & editing: Alexandria M. Jensen, Debashis Ghosh.

References

1. Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*. 2009; 10(3):186–198. <https://doi.org/10.1038/nrn2575> PMID: 19190637
2. Betzel RF, Bassett DS. Multi-scale brain networks. *NeuroImage*. 2017; 160:73–83. <https://doi.org/10.1016/j.neuroimage.2016.11.006> PMID: 27845257
3. Sporns O. Structure and function of complex brain networks. *Dialogues in Clinical Neuroscience*. 2013; 15(3):247–262. <https://doi.org/10.31887/DCNS.2013.15.3/osporns> PMID: 24174898
4. Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*. 2010; 52(3):1059–1069. <https://doi.org/10.1016/j.neuroimage.2009.10.003> PMID: 19819337
5. Betzel RF, Bertolero MA, Gordon EM, Gratton C, Dosenbach NUF, Bassett DS. The community structure of functional brain networks exhibits scale-specific patterns of inter- and intra-subject variability. *NeuroImage*. 2019; 202:115990. <https://doi.org/10.1016/j.neuroimage.2019.07.003> PMID: 31291606
6. Meunier D, Lambiotte R, Fornito A, Ersche K, Bullmore ET. Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*. 2009; 3. <https://doi.org/10.3389/neuro.11.037.2009> PMID: 19949480
7. Simon HA. The architecture of complexity. *Proceedings of the American Philosophical Society*. 1962; 206(6):467–482.
8. Bassett DS, Bullmore E. Small-world brain networks. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*. 2006; 12(6):512–523. <https://doi.org/10.1177/1073858406293182> PMID: 17079517
9. Bassett DS, Gazzaniga MS. Understanding complexity in the human brain. *Trends in cognitive sciences*. 2011; 15(5):200–209. <https://doi.org/10.1016/j.tics.2011.03.006> PMID: 21497128
10. Betzel RF, Byrge L, He Y, Goñi J, Zuo XN, Sporns O. Changes in structural and functional connectivity among resting-state networks across the human lifespan. *NeuroImage*. 2014; 102:345–357. <https://doi.org/10.1016/j.neuroimage.2014.07.067> PMID: 25109530
11. Gu S, Satterthwaite TD, Medaglia JD, Yang M, Gur RE, Gur RC, et al. Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences of the United States of America*. 2015; 112(44):13681–13686. <https://doi.org/10.1073/pnas.1502829112> PMID: 26483477
12. Chen ZJ, He Y, Rosa-Neto P, Gong G, Evans AC. Age-related alterations in the modular organization of structural cortical network by using cortical thickness from MRI. *NeuroImage*. 2011; 56(1):235–245. <https://doi.org/10.1016/j.neuroimage.2011.01.010> PMID: 21238595
13. Chan MY, Park DC, Savalia NK, Petersen SE, Wig GS. Decreased segregation of brain systems across the healthy adult lifespan. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111(46):E4997–E5006. <https://doi.org/10.1073/pnas.1415122111> PMID: 25368199

14. Liu K, Yao S, Chen K, Zhang J, Yao L, Li K, et al. Structural Brain Network Changes across the Adult Lifespan. *Frontiers in Aging Neuroscience*. 2017; 9:275. <https://doi.org/10.3389/fnagi.2017.00275> PMID: 28860988
15. de Haan W, van der Flier WM, Koene T, Smits LL, Scheltens P, Stam CJ. Disrupted modular brain dynamics reflect cognitive dysfunction in Alzheimer's disease. *NeuroImage*. 2012; 59(4):3085–3093. <https://doi.org/10.1016/j.neuroimage.2011.11.055> PMID: 22154957
16. Alexander-Bloch AF, Gogtay N, Meunier D, Birn R, Clasen L, Lalonde F, et al. Disrupted Modularity and Local Connectivity of Brain Functional Networks in Childhood-Onset Schizophrenia. *Frontiers in Systems Neuroscience*. 2010; 4:147. <https://doi.org/10.3389/fnsys.2010.00147> PMID: 21031030
17. Fornito A, Bullmore ET. Connectomic intermediate phenotypes for psychiatric disorders. *Frontiers in Psychiatry*. 2012; 3:32. <https://doi.org/10.3389/fpsy.2012.00032> PMID: 22529823
18. Tost H, Bilek E, Meyer-Lindenberg A. Brain connectivity in psychiatric imaging genetics. *NeuroImage*. 2012; 62(4):2250–2260. <https://doi.org/10.1016/j.neuroimage.2011.11.007> PMID: 22100419
19. Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE. Intrinsic and task-evoked network architectures of the human brain. *Neuron*. 2014; 83(1):238–251. <https://doi.org/10.1016/j.neuron.2014.05.014> PMID: 24991964
20. Fortunato S. Community detection in graphs. *Physics Reports*. 2010; 486(3):75–174. <https://doi.org/10.1016/j.physrep.2009.11.002>
21. Fortunato S, Newman MEJ. 20 years of network community detection. *Nature Physics*. 2022. <https://doi.org/10.1038/s41567-022-01716-7>
22. Yang Z, Algesheimer R, Tessone CJ. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific Reports*. 2016; 6(1):30750. <https://doi.org/10.1038/srep30750> PMID: 27476470
23. Mukerjee S. A systematic comparison of community detection algorithms for measuring selective exposure in co-exposure networks. *Scientific Reports*. 2021; 11(1):15218. <https://doi.org/10.1038/s41598-021-94724-1> PMID: 34312444
24. Combs K, Bihl T. Clustering and Topological Data Analysis: Comparison and Application; 2023. Available from: <http://hdl.handle.net/10125/102731>.
25. Sizemore AE, Phillips-Cremens JE, Ghrist R, Bassett DS. The importance of the whole: Topological data analysis for the network neuroscientist. *Network Neuroscience*. 2019; 3(3):656–673. https://doi.org/10.1162/netn_a_00073 PMID: 31410372
26. Caputi L, Pidnebesna A, Hlinka J. Promises and pitfalls of topological data analysis for brain connectivity analysis. *NeuroImage*. 2021; 238:118245. <https://doi.org/10.1016/j.neuroimage.2021.118245> PMID: 34111515
27. Taya F, de Souza J, Thakor NV, Bezerianos A. Comparison method for community detection on brain networks from neuroimaging data. *Applied Network Science*. 2016; 1(1):1–20. <https://doi.org/10.1007/s41109-016-0007-y> PMID: 30533500
28. Ghawi R, Pfeffer J. A community matching based approach to measuring layer similarity in multilayer networks. *Social Networks*. 2022; 68:1–14. <https://doi.org/10.1016/j.socnet.2021.04.004>
29. Zhang RJ, Ye FY. Measuring similarity for clarifying layer difference in multiplex ad hoc duplex information networks. *Journal of Informetrics*. 2020; 14(1):100987. <https://doi.org/10.1016/j.joi.2019.100987>
30. Onnela JP, Fenn DJ, Reid S, Porter MA, Mucha PJ, Fricker MD, et al. Taxonomies of networks from community structure. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*. 2012; 86(3 Pt 2):036104–036104. <https://doi.org/10.1103/PhysRevE.86.036104> PMID: 23030977
31. Bródka P, Chmiel A, Magnani M, Ragozini G. Quantifying layer similarity in multiplex networks: a systematic study. *Royal Society Open Science*. 2018; 5(8):171747. <https://doi.org/10.1098/rsos.171747> PMID: 30224981
32. Lohse C, Bassett DS, Lim KO, Carlson JM. Resolving Anatomical and Functional Structure in Human Brain Organization: Identifying Mesoscale Organization in Weighted Network Representations. *PLoS Computational Biology*. 2014; 10(10):e1003712. <https://doi.org/10.1371/journal.pcbi.1003712> PMID: 25275860
33. Alexander-Bloch A, Lambiotte R, Roberts B, Giedd J, Gogtay N, Bullmore E. The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia. *NeuroImage*. 2012; 59(4):3889–3900. <https://doi.org/10.1016/j.neuroimage.2011.11.035> PMID: 22119652
34. Zhan X, Patterson AD, Ghosh D. Kernel approaches for differential expression analysis of mass spectrometry-based metabolomics data. *BMC Bioinformatics*. 2015; 16(1):77. <https://doi.org/10.1186/s12859-015-0506-3> PMID: 25887233

35. Vapnik V. *Statistical Learning Theory* | Wiley. New York City: Wiley Books; 1998. Available from: <https://www.wiley.com/en-us/Statistical+Learning+Theory-p-9780471030034>.
36. Mikl M, Mareček R, Hlušík P, Pavlicová M, Drastich A, Chlebus P, et al. Effects of spatial smoothing on fMRI group inferences. *Magnetic Resonance Imaging*. 2008; 26(4):490–503. <https://doi.org/10.1016/j.mri.2007.08.006> PMID: 18060720
37. Han J, Kamber M, Pei J. 10—Cluster Analysis: Basic Concepts and Methods. In: Han J, Kamber M, Pei J, editors. *Data Mining (Third Edition)*. Morgan Kaufmann; 2012. p. 443–495.
38. McDaid AF, Greene D, Hurley N. Normalized Mutual Information to evaluate overlapping community finding algorithms; 2013. Available from: <http://arxiv.org/abs/1110.2515>.
39. Prates LdO. A more efficient algorithm to compute the Rand Index for change-point problems; 2021. Available from: <http://arxiv.org/abs/2112.03738>.
40. Reichardt J, Bornholdt S. Statistical mechanics of community detection. *Physical Review E*. 2006; 74(1):016110. <https://doi.org/10.1103/PhysRevE.74.016110> PMID: 16907154
41. Emmons S, Kobourov S, Gallant M, Börner K. Analysis of Network Clustering Algorithms and Cluster Quality Metrics at Scale. *PLoS ONE*. 2016; 11(7):e0159161. <https://doi.org/10.1371/journal.pone.0159161> PMID: 27391786
42. McKenna TM, McMullen TA, Shlesinger MF. The brain as a dynamic physical system. *Neuroscience*. 1994; 60(3):587–605. [https://doi.org/10.1016/0306-4522\(94\)90489-8](https://doi.org/10.1016/0306-4522(94)90489-8) PMID: 7936189
43. Müller KR, Mika S, Rätsch G, Tsuda K, Schölkopf B. An introduction to kernel-based learning algorithms. *IEEE transactions on neural networks*. 2001; 12(2):181–201. <https://doi.org/10.1109/72.914517> PMID: 18244377
44. Hofmann T, Schölkopf B, Smola AJ. Kernel methods in machine learning. *The Annals of Statistics*. 2008; 36(3). <https://doi.org/10.1214/009053607000000677>
45. Schölkopf B, Smola AJ. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; 2018. Available from: <https://direct.mit.edu/books/book/1821/Learning-with-KernelsSupport-Vector-Machines>.
46. Liu D, Lin X, Ghosh D. Semiparametric Regression of Multidimensional Genetic Pathway Data: Least-Squares Kernel Machines and Linear Mixed Models. *Biometrics*. 2007; 63(4):1079–1088. <https://doi.org/10.1111/j.1541-0420.2007.00799.x> PMID: 18078480
47. Liu D, Ghosh D, Lin X. Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. *BMC Bioinformatics*. 2008; 9(1):292. <https://doi.org/10.1186/1471-2105-9-292> PMID: 18577223
48. Kimeldorf G, Wahba G. Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*. 1971; 33(1):82–95. [https://doi.org/10.1016/0022-247X\(71\)90184-3](https://doi.org/10.1016/0022-247X(71)90184-3)
49. Davies RB. Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternatives. *Biometrika*. 1987; 74(1):33–43. <https://doi.org/10.2307/2336019>
50. Cameron C, Yassine B, Carlton C, Francois C, Alan E, Andrés J, et al. The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*. 2013; 7. <https://doi.org/10.3389/conf.fninf.2013.09.00041>
51. Behzadi Y, Restom K, Liao J, Liu TT. A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI. *NeuroImage*. 2007; 37(1):90–101. <https://doi.org/10.1016/j.neuroimage.2007.04.042> PMID: 17560126
52. Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*. 2012; 33(8):1914–1928. <https://doi.org/10.1002/hbm.21333> PMID: 21769991
53. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008; 2008:10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
54. Ratto AB, Kenworthy L, Yerys BE, Bascom J, Wieckowski AT, White SW, et al. What About the Girls? Sex-Based Differences in Autistic Traits and Adaptive Skills. *Journal of autism and developmental disorders*. 2018; 48(5):1698–1711. <https://doi.org/10.1007/s10803-017-3413-9> PMID: 29204929
55. An HS, Moon WJ, Ryu JK, Park JY, Yun WS, Choi JW, et al. Inter-vender and test-retest reliabilities of resting-state functional magnetic resonance imaging: Implications for multi-center imaging studies. *Magnetic Resonance Imaging*. 2017; 44:125–130. <https://doi.org/10.1016/j.mri.2017.09.001> PMID: 28888769
56. Van Horn JD, Toga AW. Multi-Site Neuroimaging Trials. *Current opinion in neurology*. 2009; 22(4):370–378. <https://doi.org/10.1097/WCO.0b013e32832d92de> PMID: 19506479

57. Ribeiro de Oliveira L, Brian J, Kelley E, Beal D, Nicolson R, Georgiades S, et al. Exploring the use of the verbal intelligence quotient as a proxy for language ability in autism spectrum disorder. *Research in Autism Spectrum Disorders*. 2020; 73:101548. <https://doi.org/10.1016/j.rasd.2020.101548>
58. Williams DL, Goldstein G, Kojkowski N, Minshew NJ. Do individuals with high functioning autism have the IQ profile associated with nonverbal learning disability? *Research in autism spectrum disorders*. 2008; 2(2):353–361. <https://doi.org/10.1016/j.rasd.2007.08.005> PMID: 18516234
59. Lancichinetti A, Fortunato S, Kertesz J. Detecting the overlapping and hierarchical community structure of complex networks. 2008.
60. Koh H, Li Y, Zhan X, Chen J, Zhao N. A Distance-Based Kernel Association Test Based on the Generalized Linear Mixed Model for Correlated Microbiome Studies. *Frontiers in Genetics*. 2019; 10. <https://doi.org/10.3389/fgene.2019.00458> PMID: 31156711
61. Behr M, Ansari MA, Munk A, Holmes C. Testing for dependence on tree structures. *Proceedings of the National Academy of Sciences*. 2020; 117(18):9787–9792. <https://doi.org/10.1073/pnas.1912957117> PMID: 32321827