Evaluating Network Threshold Selection for Structural and Functional Brain Connectomes

Nicholas Theis MS[1], Jonathan Rubin PhD[2], Joshua Cape PhD[3], Satish Iyengar PhD[3], Raquel E. Gur MD PhD[4], Ruben C. Gur PhD[4], David R. Roalf PhD[4], Michael F. Pogue-Geile PhD[5], Laura Almasy PhD[4], Vishwajit L. Nimgaonkar MD PhD[1,6], Konasale M. Prasad MD[1,6,7*]

[1] Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh PA 15213

[2] Department of Mathematics, University of Pittsburgh, Pittsburgh PA 15213

[3] Department of Statistics, University of Pittsburgh, Pittsburgh PA 15213

[4] Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA

[5] Department of Psychology, University of Pittsburgh, Pittsburgh, Pittsburgh PA, 15213

[6] Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, PA 15240

[7] Department of Bioengineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh PA 15213

*Corresponding author:

3811 O'Hara St
Pittsburgh, PA 15213
Email: kmp8@pitt.edu

Abstract

Structural and functional brain connectomes represent macroscale neurophysical data collected through methods such as magnetic resonance imaging (MRI). Such data may contain noise that contribute to false positive edges, which motivates the use of edge-wise thresholding. Thresholding procedures are useful for reducing network density in graphs to retain only the most informative, non-noisy edges. Nevertheless, limited consensus exists on selecting appropriate threshold levels. We compare existing thresholding methods and introduce a novel thresholding approach in the context of MRI-derived and simulated brain connectomes. Performance is measured using normalized mutual information (NMI), a quantity robust to arbitrary changes in partition labeling, and describes the similarity of community structure between two node-matched networks. We found that the percolation-based threshold and our newly proposed “objective function”-based threshold exhibited the best performance in terms of NMI. We show an application of these two thresholding methods to real data that showed that both percolation-based and objective function-based thresholding provide statistically similar NMI values between real world FC networks and structural connectivity (SC) counterparts, where shared modular structure is assumed. The two thresholding methods, however, achieve these NMI values at significantly different thresholds ($p<0.0001$) in both simulated and real networks. Moreover, the threshold obtained from the objective function gives a more accurate estimate of the number of modules present in the network and includes more flexibility in threshold selection, suggesting that this method may represent a useful option for graph thresholding.

Introduction

Network representations of biological data often produce graphs that are weighted but fully connected (van den Heuvel et al 2017), or sparse but contaminated with false positives (Drakesmith et al 2015; Pu et al 2015). For instance, functional connectomes (FC) generated from functional MRI (fMRI) data consist of edges that are defined by the correlation coefficient between nodal time series of blood oxygenation level dependent (BOLD) signals. Similarly, morphometric measures are used to build structural covariance networks (SCNs), where the edge weight represents the strength of correlation between pairs of regions. Wherever correlation is used to define nodal connectivity, as in FCs or SCNs, no edges will have a value of zero in the weighted network, and the connectivity patterns can include weakly correlated regions that may represent noise. In the structural connectome (SC), derived from anisotropy-based fiber tracking measurements of diffusion streams (a proxy measure of white matter fiber bundles), false positives are less prevalent, but still possible (Sotiropoulos and Zalesky 2017).

A threshold can be applied to create sparser binarized/weighted graphs from fully connected networks (Basset et al 2008), or to remove false positives (Drakesmith et al 2015). But thresholding procedures are not without caveats. Aggressive thresholding could result in unusable overly sparse graph whereas not thresholding at all could results in having several false positive edges. Despite the importance of this analytic step, few comparisons have been made among the existing methods to evaluate and directly compare thresholding techniques. Arriving at a consensus is difficult partly due to the challenge in defining a universal performance measure for thresholding outcome. We apply a commonly used method of comparing performance of different thresholding methods, based on measuring the normalized mutual information (NMI) calculated between the community structures of binary networks resulting from thresholding and the true community structure, if it is known (Alexander-Bloch et al 2012).

We introduce a novel thresholding method, termed the “objective function”, that uses an optimization approach to compute an edge weight threshold for any given network, calculated on any graph measure such as characteristic path length. This “objective function” maximizes a graph metric over a range of reasonable thresholds, heuristically meaning solutions with graphs of density less than one and sufficiently connected and formalized in the objective function. We demonstrate the use of the best performing thresholding approaches on real-world structural and functional brain connectomes, where the “ground truth” of the FC is assumed to be the SC. While structure is not expected to fully explain function, this SC-FC pairing approach is consistent with data that suggests the existence of a relationship between these networks (Zhu et al 2013; Li et al 2016; Battiston et al 2017; Domenico et al 2016 and 2017).

Earlier threshold selection schemes such as the small-worldness range, also called “sigma” range (Basset et al 2008), provide a range of thresholds, but not a specific “optimal” threshold without a measure of quality or meaning for the network. The small-worldness approach requires that the networks themselves have small-world properties, somewhat limiting its application. The statistical thresholding (Chenet al., 2008; Ferrarini et al., 2009) requires that edge weights have associated metadata, such as *p*-values for correlations, that are not attributed to all biological networks. For example, SC networks constructed on the basis of fiber counting do not have an associated *p*-value for the significance of the streamlines *per se*.

More recently published methods rely on downstream classifier accuracy to detect the “correct” network threshold (Zanin et al 2012; Drakesmith et al 2015). For example, the “multi-threshold permutation correction” method (Drakesmith et al 2015) suggests that the optimal threshold can be determined by clustering thresholded networks into groups and evaluating which threshold yields the most accurate classification. While these methods are applicable to a variety of networks (i.e., not just small world) they cannot be performed on a single graph, or a collection of graphs from a single - or multiple unknown - groups.

Another thresholding method (Bordier et al 2017) is the percolation threshold, which is the highest threshold for which a network’s giant connected component (GCC) includes all nodes. This is a straightforward and reasonable selection method. The only drawback with this method is that it could be considered too rigid. In the FC, some nodes may not be expected to be activated in the resting state fMRI or participate in processing a given computational load, indicating that some FCs may not have GCCs containing all nodes. This would indicate that a threshold higher than the percolation threshold is desired. Conversely, a lower-than-percolation threshold may be more appropriate for networks with many suspected redundancies. The objective thresholding method evaluates a range of possible threshold choices including the percolation threshold, identifying the threshold at which a given graph metric is most different from both endpoints of the range.

Methods

MRI Data

Subject Recruitment

The sample consisted of 181 healthy participants who were part of a two-site Multiplex Genetics Investigation (MGI), identified in the communities in Pennsylvania and surrounding states. Details of recruitment and ascertainment are published (Gur et al 2007). Mean age was 39±16 years and 47% were males. The sample is a subset of a multigenerational schizophrenia study, the details of which have been previously published (Roalf et al 2015).

MRI Acquisition

MRI data were acquired on Siemens Tim Trio 3T MRI systems at both sites. The inter-scanner reliabilities were tracked using phantoms and test subjects with no significant differences (Roalf et al 2015). DWI parameters were: b=1000s/mm^2, slices=64, thickness=2.4 mm, TE=90ms, TR=6300 ms, flip angle=90°, matrix=128×128. In-plane resolution was 1.9 mm with a slice thickness of 2 mm. Four regularly interspersed B_0 reference images were also acquired. T1W imaging protocols had a TR=1680ms, TI=1100ms, TE=4.67ms, and a flip angle=15°. The resolution was 0.94 mm in-plane with a 1 mm slice thickness. For fMRI, TR=3000ms, TE=30ms, and the flip angle=70°. Image spatial resolution was 3.44 mm in plane with a 3 mm slice thickness. An emotion recognition task was administered during the fMRI acquisition and BOLD signals at 90 time-points were acquired.

T1W Image Processing

Neuroanatomical scans (T1-weighted images) were processed using the FreeSurfer software version 6.0.0 (Fischl 2012). Automated parcellations of the cortex and automated segmentation of the subcortex were conducted, resulting in 81 non-overlapping regions-of-interest (ROIs) from the Desikan-Killiany-Tourville (DKT) atlas (Klein and Tourville 2012). The image parcellations include 2 cerebellar ROIs (left and right) and a single brain-stem ROI, 16 sub-cortical ROIs (8 on each side) and 62 cortical ROIs.

Diffusion Weighted Imaging (DWI) Processing

DWI were processed using the MRTrix3 software (Tournier et al 2019). Prior to fiber tracking, images were denoised and bias-corrected, and Gibbs ringing artifacts were removed. Brain extraction was performed on the DWI images. MRTrix's Tournier method (Tournier et al 2013) was used to determine the response function (deconvolution kernel) for the preprocessed DWI data. The ROI label masks were then linearly registered (with FSL's linear alignment tool, the FLIRT functionality) to the DWI image space using nearest neighbor interpolation. Fiber tracking was performed on the whole brain by making 10 million random attempts at drawing fibers. Afterward, the "tcksift" function was used to sieve out anatomically improbable tracks, and the track file was queried to determine the subsets of streamlines that connect each possible pair of ROIs.

FMRI Processing

Preprocessing was used to remove motion artifacts and non-brain voxels, as described by the Human Connectome Project (HCP) standards (Glasser et al 2013). Briefly, HCP preprocessing includes spatial artifact removal, region specification and alignment to a template space. The FSL package (version 6) (Jenkinson et al 2012) and AFNI (Cox 1996) were used to preprocess fMRI data. Preprocessing included: 1) rigid-body motion correction using 6 degrees-of-freedom, 2) brain extraction, and 3) spatial smoothing with a 4-mm FWHM (full-width-at-half maximum) smoothing kernel. Next, fMRIs were co-registered to a high-resolution ROI label mask from T1w images using the FLIRT function with trilinear interpolation. Smoothed images were up-sampled for registration to the atlas, and the spatially averaged activation intensity time course for each ROI was then calculated for each subject in MATLAB. This resulted in a time series representing approximate BOLD signal intensity at the ROI.

Network Construction

SC Network Construction

Each entry in the raw SC matrix represents the number of diffusion streams spatially connecting two ROIs after fiber tracking. These methods follow the general recommendations on generating connectome data from DWI (Sotiropoulos and Zalesky 2017; Hasan et al 2011).

FC Network Construction

The functional connectivity between ROI pairs is inferred through Pearson's correlation coefficient (r) of the corresponding time series of BOLD signals in the respective ROIs. The FC can contain negative weights that can be an analytical challenge because many graph properties assume positive weights. While there is some controversy over the source of these negative signals, a common solution is to take the absolute value of the correlation to remove negatives (Andellini et al 2015). In our data, the negative correlations made up 5.61% (mean r=-0.19 ± 0.17) of observed edges at the population level. Thresholding applied to the absolute value of these weights preserves strongly negative edges as connections in the binary network. Negative weights could be treated as a separate network from the positive weight network, but we did not pursue this direction because in our data there were too few negative weights, and they were generally small in magnitude.

LFR Simulations

To evaluate the effectiveness of thresholding techniques, we used a published approach (Bordier et al. 2017), which relies on the generation of benchmark Lancichinetti-Fortunato-Radicchi (LFR) networks so that the true community structure is

known. FMRI-like noise (Welvaert et al 2011) is introduced to these LFR networks and effectiveness of thresholding for noise removal via modularity calculation is evaluated. Effectiveness is quantified by calculating the normalized mutual information (NMI) between the known community assignments of the original LFR network and the community structure detected in the noisy LFR networks after thresholding (Bordier et al. 2017).

We performed two simulations, one termed “LFR-sf” because it has more scale-free qualities and uses parameters similar to that described in Bordier et al. (2017) albeit with fewer nodes, and the other termed “LFR-alt” because it represents an LFR network with an alternative degree-distribution that is closer to that of a regular graph. For all simulations, there were 81 nodes in the final FC networks, and noise added to each of the 100 replicates, or simulated subjects. The SNR value in the simulation environment was set to 35 as described in Bordier et al. (2017). Both LFR-sf and LFR-alt networks had a minimum community size of 5 and maximum of 20, with community mixing and weight mixing parameters both set to 0.2. The power law exponent degree was 2 in both simulations. The only parameters that differed between the two LFR simulations were average degree (5 for the LFR-sf; 9 for the LFR-alt) and maximum degree (20 for LFR-sf; 15 for LFR-alt).

Network Analysis

Binarization (Thresholding)

For a given threshold value, □, each edge in the network is set to 1 if the normalized edge weight is greater than □ (i.e., the edge is present) and 0 if less than or equal to □ (i.e., edge not present) resulting in a binary network of assumed true positives only. We evaluated all networks at a range of thresholds from 0 to 1 inclusive, in step sizes of $\varepsilon = 0.01$, so that 101 equally spaced thresholds were tested in a parameter sweep.

Graph Metric Calculation

Six of the most commonly used network summary measures (Bullmore and Sporns 2009) that span several hypothetical categories of network measures (Fallani et al. 2014) were calculated using the Brain Connectivity Toolbox (BCT) (Rubinov and Sporns 2010). Specifically, density, characteristic path length (λ), efficiency. average clustering coefficient, modularity, and assortativity were tested in the real SC and FC across the full threshold range. Nodal strengths, the sum of weights connecting to a node, was also measured on the weighted networks using the BCT.

For modularity, the community detection algorithm relies on the Newman method (Newman 2006). Briefly, communities are determined by searching for the community structure that maximizes a benefit function called the modularity coefficient (Q). The Newman method for community structure is a spectral partitioning method that expresses this maximization problem in terms of a matrix eigenvector. For characteristic path length, where networks have GCCs with fewer than 100% of the nodes, e.g., where "island networks" appear, the calculation treats the disconnected components as separate networks and averages their results, ignoring pairs of nodes that are not in the same component.

Objective Threshold Method

We developed the objective threshold method under the assumption that the “optimal” threshold for any weighted network lies between the smallest threshold for which the network density is not 1, and the largest threshold for which the GCC includes some percentage of nodes, here 90%. Like many thresholding methods, the objective threshold is determined by first employing a threshold sweep where a range of possible thresholds are evaluated. We propose that for a given graph measure, M - such as density, λ, Q, or any scalar graph summary metric - the optimal threshold should maximize the difference of this metric from the value of the same metric at both end points, termed M_0 and M_1. To formalize this relationship, we introduce the objective function which can be minimized across this range to find the optimal threshold for a given graph metric in any weighted network (**Figure 1**).

The objective function takes the form:

$$F_M(\square) = (M_\square - M_0)^2 + (M_\square - M_1)^2$$
$$\square \ \square \ T = \{\square_0, \square_0 + \varepsilon, \square_0 + 2\varepsilon, \square_1\}$$

where $\square$ is the threshold and the notation M refers to the graph measure in question, with $M_\square$ denoting the value of M evaluated on a graph that has been binarized using threshold $\square$ and the corresponding values of M at the extreme thresholds, denoted M_0, M_1. Each threshold is taken from the set *T* of thresholds tested, ranging in steps of ε = 0.01 from $\square_0 = k*\varepsilon$ to $\square_1 = K*\varepsilon$, for integers $k < K$ such that the boundary thresholds $\square_0$, $\square_1$ are both in the interval [0,1].

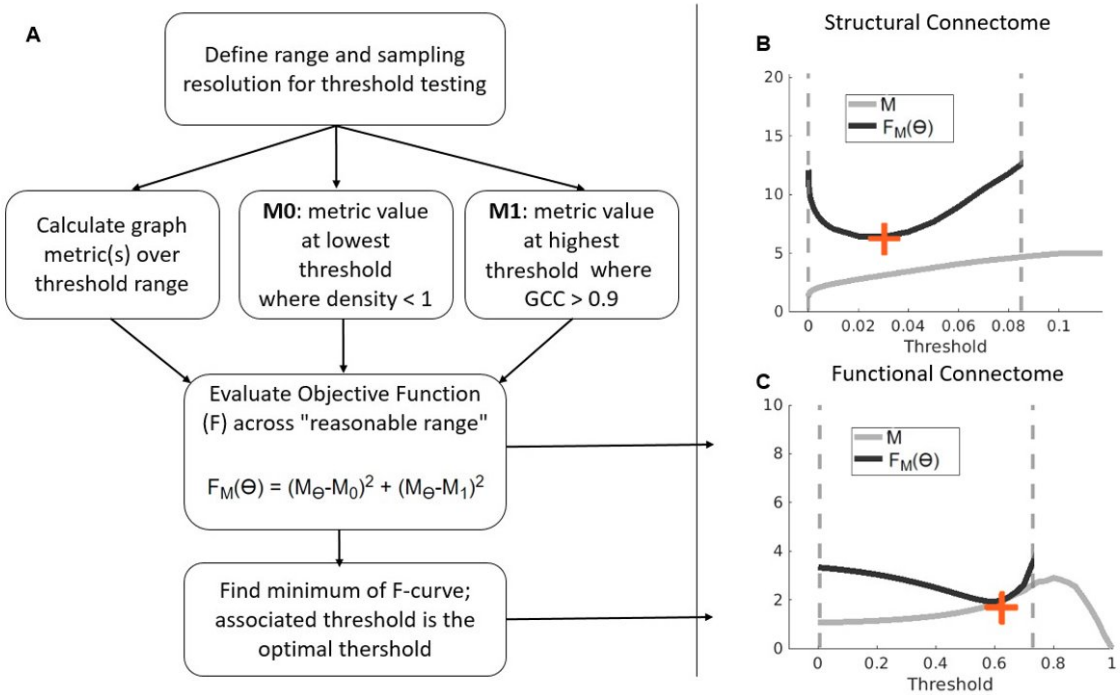

Fig 1. An overview of the objective threshold function is shown in panel A. Briefly, a range of thresholds were examined (here increments of 0.01 from 0 to 1 are used). Next, each subject level network is binarized at every threshold in the range and the density and GCC are calculated, along with the graph metric, M, of interest (gray curves in B and C). The threshold values that yield M_0 and M_1 are taken to be the minimum and maximum of the range. Next, F is calculated for each threshold, producing an F-curve (black curves in B and C). The F-curve has a minimum (red plus sign in B and C) on the interval (specified by vertical dotted lines in B and C). The threshold yielding this minimum for each individual network is that network's "objective threshold."

The determination of M_0 is trivial, and for sparse network it will be found at $\square_0 = 0$. Determining M1 is more difficult; we suggest the $\square_1$ value can be found by assessing the GCC of the graph, and to some extent the modular structure. The GCC refers to the largest collection of nodes within the graph such that there exists a path of nonzero edges connecting each pair of nodes in the collection. For a fully connected graph such as our FC without a threshold, all nodes are in the GCC. From there, if edges are removed uniformly at random, there is typically a rather abrupt transition from a GCC that includes all nodes to a much smaller GCC size (Newman et al., 2001). Hence, the fraction of the network nodes in the GCC provides a measure of the overall connectedness of the graph. We take $\square_1$ to be the smallest threshold such that the proportion of nodes in the GCC is ≥90%. Notably, the "percolation threshold" (Bordier et al. 2017) is given by $\square_1$ when GCC is 100%; however, the assumption that all nodes belong to the GCC may not be suitable for many biological networks, and hence we aimed for a less restrictive choice. As $\square$ increases, the GCC remains constant or monotonically decreases; therefore, setting the GCC cutoff below100% ensures that the

percolation threshold is also evaluated by the objective thresholding function. Furthermore, we find that thresholds above the percolation threshold can yield higher modularity and a greater number of communities, which may be more suitable than the results obtained by requiring that the GCC includes 100% of nodes, suggesting that such higher thresholds should be evaluated for the networks we consider.

Optimization of Objective Threshold Function

The optimal threshold is defined for a given graph measure, M, from within the range of thresholds T, based on the values of the objective function $F_M(\square)$ evaluated for all $\square$ $\square$ T. Specifically, the optimal threshold is the value at which $F_M(\square)$ has an extremum on the interval $[\square_0, \square_1]$. This criterion selects the threshold at which the value of F is farthest from its endpoint values.

In practice, we find that either $M_\square$ rises from a low value at or near 0 at $\square_0$, has an interior maximum, and then declines back to near 0 again at $\square_1$, or else $M_\square$ varies monotonically as $\square$ increases from $\square_0$ to $\square_1$. In the former scenario, the optimal threshold is simply the threshold in $(\square_0, \square_1)$ where the maximum of $F_M(\square)$ occurs. In the latter case, we have $F_M(\square_0) = F_M(\square_1) = (M_0 - M_1)^2$. At the same time, if $F_M(\square)$ is monotonically increasing from $\square_0$ to $\square_1$, then for any $\square$ $\square$ $(\square_0, \square_1)$, $M_0 < M_\square < M_1$, and by the triangle inequality, we have,

$$F_M(\square) = (M_\square - M_0)^2 + (M_\square - M_1)^2 \leq (M_\square - M_0 + M_1 - M_\square)^2 = (M_0 - M_1)^2.$$

That is, $F_M(\square)$ takes its maximal values at $\square_0$ and $\square_1$ and has an interior minimum between these extremes. A similar argument applies and gives an interior local minimum of $F_M(\square)$ if $M_\square$ is monotonically decreasing from $\square_0$ to $\square_1$.

Statistical Threshold

Statistical thresholding refers to the process by which a statistical network such as a correlation matrix, in this case an FC, is made sparse (namely, having a density less than 1) not by removing edges above or below some universal threshold, but by including each edge based on its statistical significance (Bullmore and Basset 2011). Here, we used a nominal uncorrected $p<0.05$ to determine whether a correlational edge should be included in the binary network or not.

Percolation Threshold

The percolation threshold is defined to be the highest intensity threshold at which the network's GCC contains all nodes in the network (Bordier et al 2017). This choice of threshold aims to find the optimal balance between information gained by noise removal

and information lost by over-excessive pruning, assuming that all nodes belong to the same component.

Sigma Range Threshold

Small-worldness is a network characteristic where graphs have high clustering coefficient and low pathlengths even between nodes that are not directly connected (Watts and Strogatz, 1998). Basset et al. (2008) developed the notion that a statistical graph, where edge weights are strictly correlational, can be thresholded at any threshold within a range prescribed by the observed small-worldness quantity. This quantitative measure of small-worldness is often represented by the symbol sigma and hence this range is sometimes called the “sigma range.” The small-worldness quantity sigma is defined as the ratio between the observed clustering normalized by the clustering in a degree-matched randomized graph and the observed characteristic path length normalized by the path length of a degree-matched randomized graph (Basset el al. 2008). For the LFR networks, to avoid excessive removal of edges, we tested the low end of this range, i.e., the lowest threshold for which sigma is at least 1.2.

NMI Calculation

Any network that is sufficiently connected or consists of connected components can be partitioned into communities which represent clusters of relatively highly connected nodes that are relatively isolated from other clusters. These communities represent categorical assignments of the nodes in the network. The amount of information that two such lists share can be quantified using normalized mutual information (NMI), a measure of mutual dependence that arises in information theory (Cover and Thomas 1991), which primarily deals with communication, storage and quantification of digital information but has been widely used in different fields of science including neurobiology (Rieke et al. 1999). NMI is 1 if the nodes with the clusters are the same, and 0 if the clusters share no information. We computed NMI based on methods described in (Alexander-Bloch et al 2012). To summarize, the two community partitions are compared using mutual information, then normalized to the total entropy of the system, so that the uncertainty in the community structure complementarity is no greater than the uncertainty within the entire network.

Other methods for comparing the modular structure between two networks, such as the adjusted Rand index or the Jaccard Index, as well as methods to directly compare network similarity, such as the “DeltaCon” metric (Koutra et al. 2016) were considered for evaluating experimental results to the ground truth. Comparing community partitions through NMI is a widely used approach that provides a reliable measure of network community structure similarity, robust to small changes in edge-wise network content (Lanchichinetti et al 2009; Danon et al. 2005). Therefore, we use comparison of NMI-based community partition comparisons to evaluate existing and novel thresholding methods both *in silico* and for real-world data.

Statistical Tests

To test for significant differences between the unweighted SC, FC, LFR-sf and LFR-alt, 2-way Kolmogorov–Smirnov tests were performed on the non-normally distributed edge-weight distributions of the networks. We used a two-sample T-test to calculate both the statistical significance of differences in NMI results obtained from different thresholding methods, and between the actual threshold values prescribed between pairs of thresholding methods. Samples sizes for LFR comparisons are N=100; SC and FC comparisons have a sample size of N=181.

Results

We considered four types of networks, LFR-sf, LFR-alt, SC and FC. For each, we successively applied each of a collection of evenly spaced thresholds between 0 and 1 and we computed the GCC size for each threshold (**Fig. 2A**). The GCC profiles of LFR-sf and the SC were quite similar. LFR-alt showed a more abrupt drop-off but at a higher range of thresholds. The FC, which started as a fully connected graph at threshold 0, required a higher threshold than the other networks before the size of the GCC started to decrease. Here, the SC and FC results each represent an average of the 181 data-derived connectomes. The original (dotted line) LFR networks are not averages, as only one LFR-sf and one LFR-alt were produced. In the noise addition step, 100 replicates were performed and the average GCC characteristics of these networks (Fig **2A).** Noise shifted the GCC curve towards higher thresholds, with a much larger shift in the LFR-sf case than in the LFR-alt case.

The histograms in **Fig 2B-E** show the differing degree distributions of the 4 networks, which were all significantly different from each other (2-sample Kolmogorov-Smirnov test, all $p<0.0001$). The most similar – but still significantly different – nodal strength distributions were LFR-sf and SC.

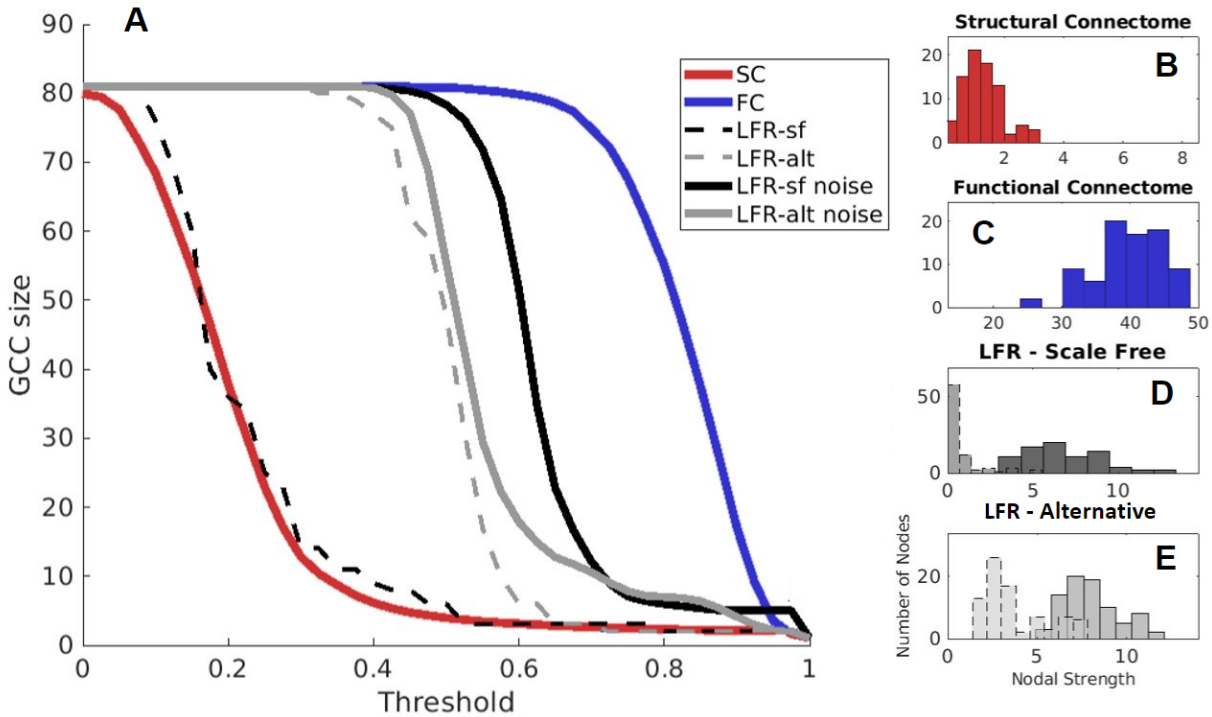

Figure 2. The GCC is shown as a function of threshold in A for the average real structural and functional connectomes (SC and FC, respectively), and the two simulated LFR networks, one with a more scale-free degree distribution, and one with a more regular degree distribution. Histograms of the nodal strength distributions for each of the four networks are shown in figures B-E, where nodal strength of weighted networks is shown on the x-axis and the number of nodes in the network with this strength on the y-axis. In D and E the solid-outlined darker histograms represent the averaged noisy-LFR networks, and the lighter, dotted-outlined histograms represent the noise-free ground-truth LFR.

To better understand how community structure interacts with threshold, we plotted the average number of communities found within different network classes as a function of threshold (**Fig 3**). This figure shows the population-averaged number of communities on the left (**Fig 3A1,3B1,3C1, and 3D1**), and population-averaged community size statistics for maximum, median, and minimum community sizes on the right (**Fig 3A2, 3B2, 3C2, and 3D2**). It is evident that for all graph types considered, the number of communities detected increases as threshold increases from 0 until a maximum number is reached, and the maximal number of communities is found at thresholds where the GCC does not contain all nodes (see **Fig 2A** for GCC curves), suggesting that allowing for thresholds above the percolation threshold, corresponding to a GCC of less than 100%, may be important in some cases where there is the possibility that separate networks are being detected.

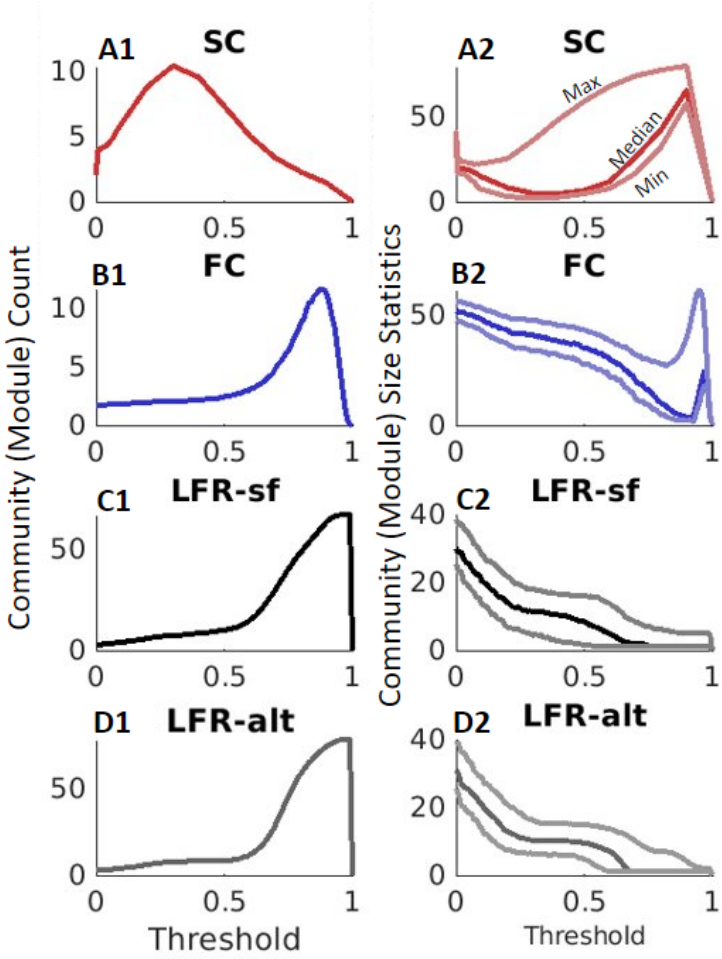

Figure 3. Community structure results for each type of network examined. In the LFR cases, only results for noisy networks are shown. All curves represent group level averages. In A1, B1, C1, and D1 the community count across thresholds is plotted, with curves representing group level means. In A2, B2, C2, and D2, the center curve is the group level average of the median community size detected and the top and bottom curves are the group level average maximum and minimum community sizes detected, respectively. Number of communities and measures of community size are generally, but not entirely, inversely related. Higher thresholds tend to result in more communities, but extremely high thresholds result in a network with no edges and therefore no modularity.

Next, a threshold sweep was performed, and 6 graph measures were computed at each threshold (**Fig. 4**) that provide useful insight into the differing nature of the graphs. For density, the M is bounded between 1 and 0 (**Fig. 4A).** The real FC data is not sparse whereas SC, LFR-sf, and LFR-alt are sparse at a threshold of zero. The λ demonstrates that for all networks, there is a peak in λ at some value in the threshold range (**Fig. 4B).** This trend is also reflected in the efficiency (**Fig 4C)** since λ and efficiency are inversely related**.** Average clustering coefficient (**Fig. 4D**) behaves differently for different types of networks and is the one measure that produces clear differences between the LFR-sf and LFR-alt networks, which give similar results for the other measures. Modularity (**Fig. 4E**) increases with threshold up to a point in the real

networks, while the LFR networks yield increasing Q values all the way up to very high thresholds. These observations are consistent with data shown in **Fig. 3**. Assortativity (**Fig. 4F**) is not defined in the entire threshold range for the real networks but fluctuates near zero where it is defined; SC networks were weakly disassortative, while FC and LFR networks are more assortative. Across all measures shown in **Fig. 4**, the LFR networks behaved most similarly to each other, while the SC and FC differed substantially across all thresholds.

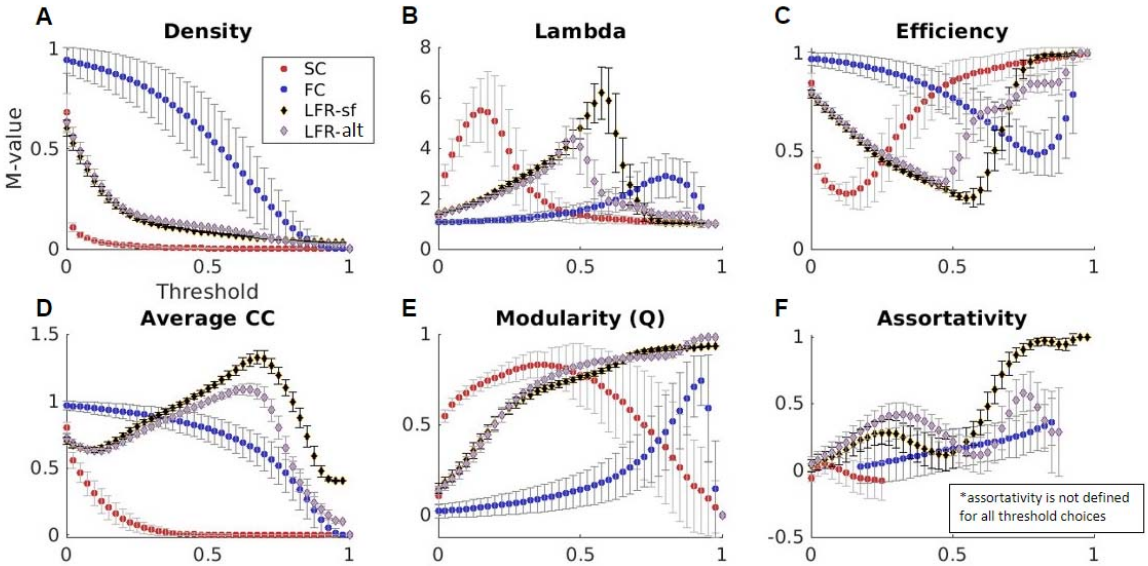

Fig 4. Threshold sweeps for SC, FC, LFR-sf and LFR-alt networks. The M is unitless. A: Graph density; B: characteristic path length (lambda); C: efficiency; D: average clustering coefficient; E: modularity (known as "Q"); F: assortativity. Results for only 40 evenly spaced threshold values between 0 and 1 are plotted for clarity although threshold sweep was conducted at 0.01 intervals resulting in 101 data-points. Points represent group-level means (N=181 for SC and FC; N=100 for LFR) and error bars indicate standard deviations. Legend: SC = structural connectome (red), FC = functional connectome (blue), LFR-sf = simulated scale-free (black), LFR-alt = simulated network with increased regularity (grey).

In the simulated LFR context where the ground truth modular structure is known, the NMI between the recovered modules and the true modules indicates thresholding performance for the LFR-sf and LFR-alt networks (**Fig. 5)**. When 6 M choices are compared (**Fig. 5A1** and **Fig. 5B1),** in both LFR simulations, λ yielded the highest NMI. When the NMI of the objective function using λ is compared with statistical thresholding, sigma-min, percolation, and no thresholding, and to the NMI computed from the weighted networks, all thresholding methods performed statistically differently than no thresholding ($p<0.001$; **Fig. 5A2** and **Fig. 5B2)**. All except the minimum sigma range value provided a meaningful NMI improvement over the no-threshold case. The

objective thresholding method based on characteristic path length yielded the highest NMI for both LFR networks, providing a statistically significant improvement in performance relative to all of the published methods except the percolation method, where the difference was not significant.

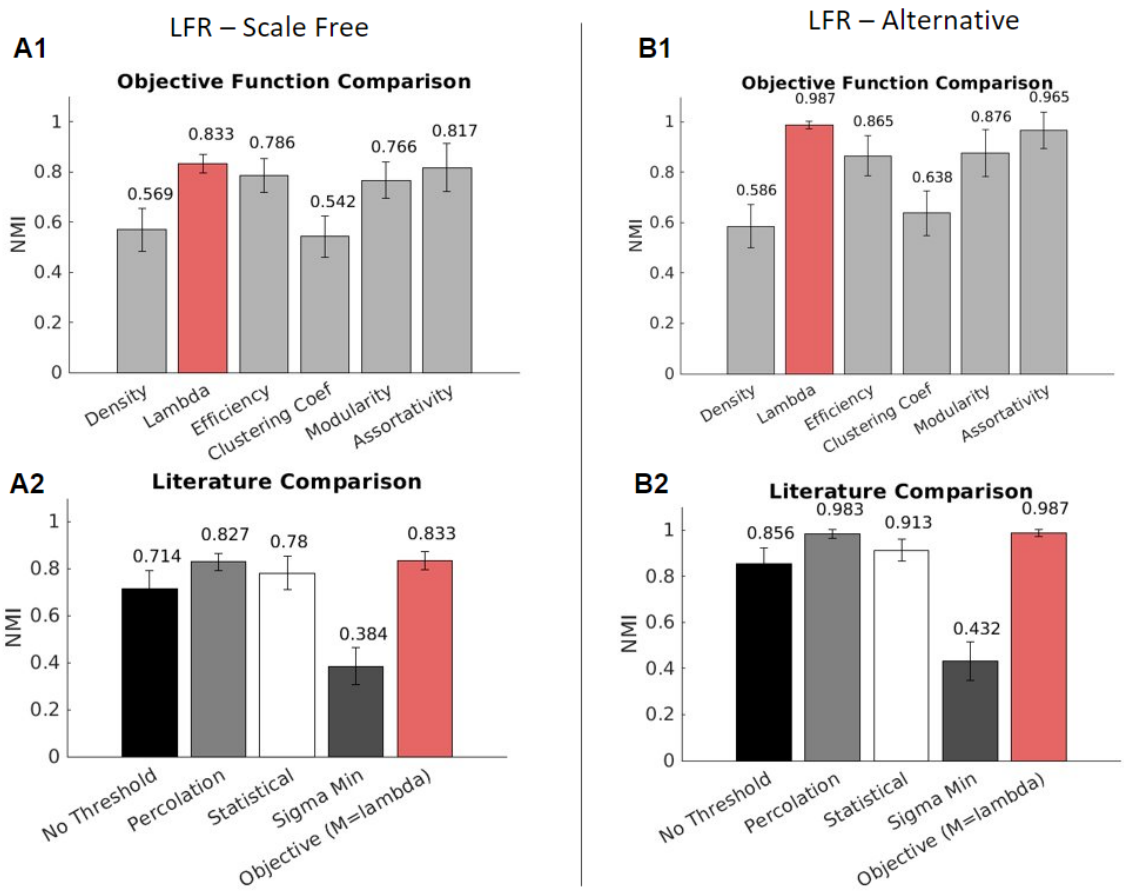

Fig. 5. The objective function threshold based on λ gives the best agreement with ground truth community structure for two LFR networks, **A**: scale free, **B**: alternative. NMI values are calculated between the true community structure of the original LFR network before noise was added, and the modular community structure of each simulated noisy replicant (N=100) after thresholding. Top plots (A1 and B1) represent the NMI calculations from the objective function method only, with the objective function calculated using each of six different metrics. In both the scale-free and alternative LFR simulations, characteristic path length, or lambda, (red bar in each plot) was found to yield the highest NMI among the metrics considered and outperformed other threshold selection methods from the literature (A2 and B2).

Next, we compared these methods with optimal possible performance. Computation of NMI over the full range of possible thresholds showed that the thresholds selected by these methods give near-optimal performance; however, the threshold from the objective function was closer to the optimal possibility (**Fig. 6A**). We also found that the NMI was maximal where the correct number of communities were

found showing the findings are consistent with our comparison of objective function thresholding with other methods (**Fig. 6B-C**).

We next examined the NMI between the SC and FC networks derived from real data. A threshold sweep showed that the NMI between the SC and FC was lower than that in the LFR simulations, consistent with the idea that FC may not be directly determined by SC. As in both LFR cases, in the SC-FC comparison, the percolation threshold and the objective threshold significantly differed in terms of the selected threshold were significantly different (2-sample T-test, $p<0.0001$). Since we do not have a ground truth for the SC-FC case, we compared the number of communities detected in the FC network across the threshold range with the number of communities present in the SC network, which gives a reasonable first estimate for how many FC communities to expect (**Fig. 6D**). The best agreement occurred at an average FC threshold of 0.71± 0.10. The objective threshold method produced a mean FC threshold of 0.64±0.08), which is close to the target value and closer than the threshold from the percolation method, where mean threshold was 0.61±0.10.

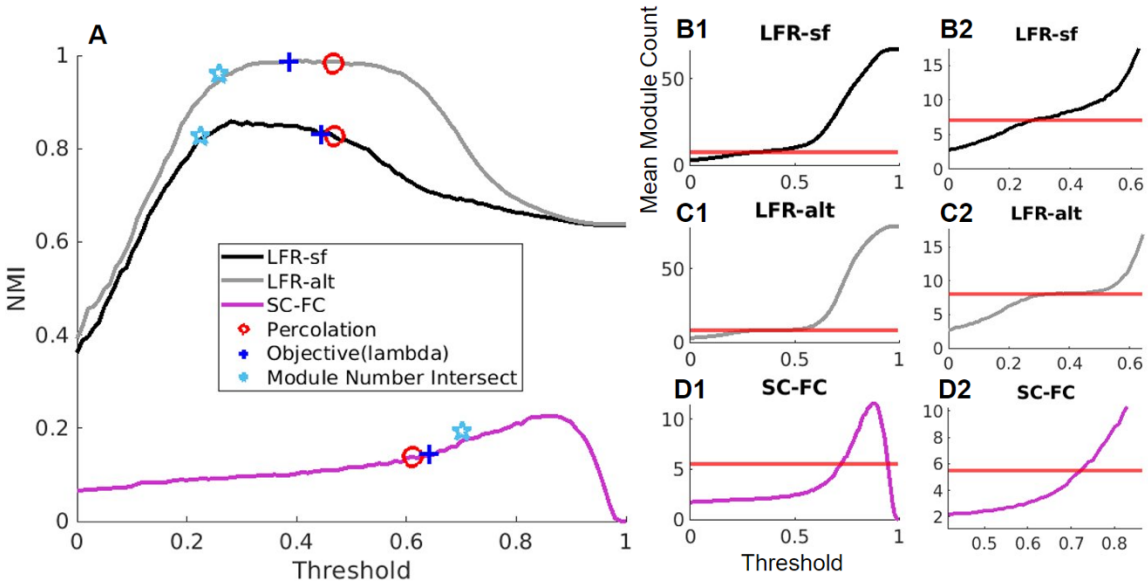

Fig. 6. NMI and number of communities present in multiple network types. **A** The NMI between ground truth modules (for LFR) or SC modules (for real data) and thresholded networks (noisy LFR-sf and LFR-alt, and FC) over the full threshold range are shown. Percolation threshold is identified by the red circle, and objective function threshold with M=lambda and GCC = 0.9 shown as the blue plus symbol. Light blue stars represent the threshold, and corresponding NMI value, of the intersection points shown in **B-D. B-D** The number of communities detected at each threshold is plotted for each network type, and "ground truth" number of modules is show with the red line; the intersection of the red line and the module count curves are shown in **A** as the light blue star. Note that all lines and points are group averages. The right column provides zoomed views of the plots from the left column near the intersection points.

Discussion

Thresholding is a critical facet of network examination. We compared several existing thresholding methods and noted some drawbacks in applying these methods to different networks under different contexts. To address these drawbacks, we developed an objective function thresholding method based on brain connectivity measures (Fig 1), and then applied this method to simulated, LFR-based FC networks as well as to SC and FC from human subject data to evaluate their performance *in silico*. Our method is "objective" in the mathematical sense that it involves a function that is optimized to arrive at an optimal threshold value. The function itself quantifies the difference between any scalar graph metric, defined across a range of thresholds (see **Fig 4** for dependence of graph metrics on thresholds), to its values at both endpoints of the threshold range.

Upon evaluating the objective method threshold for a set of graph metrics (density, λ, efficiency, clustering coefficient, modularity, and assortativity), we find λ was the best performing choice of metric (**Fig 5A1 and 5B1**). Performance was defined by the NMI between the network modules detected after binarization to the ground truth modules given by each LFR simulation, which involved two qualitatively different networks; specifically, LFR-sf and LFR-alt were considered along with 100 noisy FC replicates for each LFR simulation. In tandem, we assessed previously proposed methods for threshold selection, including the percolation threshold (Bordier et al 2017), statistical thresholding (Bullmore and Basset 2011), the minimum of the small-worldness range (Basset et al. 2008), with the network considered without any thresholding as a baseline, comparing these to the objective function method using the graph metric, λ **(Fig 5A2 and 5B2**)**.** We show that the objective function threshold and percolation threshold methods produce binary networks that capture the ground truth modular structure of the computationally generated LFR networks better than other methods.

Based on these comparisons, we found that the objective function using λ produces the highest NMI score, in a statistical tie with the NMI obtained from the percolation threshold method, and both were statistically significantly higher ($p < 0.0001$) than using no threshold *or* statistical thresholding in the LFR simulations (**Fig 5A1 and 5B2**). Interestingly, the mean threshold values themselves were statistically different ($p < 0.0001$) between percolation and objective function methods. The objective function is shown to provide a slightly lower threshold than the percolation method in both LFR simulations, and a slightly higher threshold for the real data SC-FC comparison. In both simulations and real data, the objective threshold produced a threshold closer to the threshold where the largest NMI was found **(Fig 6A).** Additionally, the objective function yielded networks for which the number of modules is closer to the baseline number of modules expected (**Fig 6B-D**).

We conclude that the objective function method can be applied to real world brain connectomes with similar results as the percolation method. The objective threshold method also provides flexibility in that different graph metrics can be used and the method applies even for networks with components that are not path connected, such that the GCC does not include the full network even without use of a threshold. Our findings suggest that among those examined, characteristic path length is the most appropriate graph measure to use in the method. Note that we used a quadratic objective function, which is a standard choice that avoids issues of cancellation of terms of opposite signs and provides smoothness. Other functional forms could be considered, however.

Another parameter available to modify according to data priors is the GCC size used in the calculation of M1, the upper bound on the range of thresholds considered. Allowing smaller GCC sizes will increase M1, which changes the range of thresholds to be used in the objective function calculation and optimization, and therefore may change the threshold obtained. The appropriate size of the GCC is not always known, especially in the case of functional brain networks. While the SC is expected to be fully connected, the GCC in the FC could reasonably be smaller than 100% of the node set. Results in **Fig 3** suggest that larger numbers of modules can be detected when thresholds above the percolation threshold are used, suggesting that M1 values in the objective threshold can reasonably be calculated at thresholds higher than the percolation threshold. Insofar as the SC can constitute a ground-truth modular structure for the FC, greater NMI can be calculated at thresholds higher than the percolation threshold (**Fig 5A & 5D**).

More generally, one may wonder whether findings from a (single) network (layer) with comparatively less noise and smaller graph density (e.g., the SC) can be leveraged to help determine the correct level of edge pruning (i.e., thresholding) in a related network (e.g., the FC). One recent review of structure-function observations in neuroimaging found that “direct structural connectivity optimistically explains no more than 50%” of functional connectivity (Suarez et al 2020), and that SC-FC edge-weight correlations range between R=0.3 and R=0.7. As observed in **Fig 4,** measures of SC and FC connectivity such as graph modularity, density, and network clustering coefficient differ strongly across thresholds, and NMI between the modules of these graph never exceeds 0.3 in our sample **(Fig 6A)**. These relatively low SC-FC NMI values corroborate other low structure-function results and provoke interesting questions about the relationship between structure and function at the level of brain MRI.

Network thresholding plays a crucial, early role in the overall process of seeking to understand the complex relationship between structural and functional brain connectomes. Ideally, threshold selection will be done via an automated procedure based on calculable quantities, to allow for reproducibility and reduce bias. The method

that we have proposed provides such a procedure, which can be flexibly adapted to reflect knowledge about the underlying network before it is applied, for example to adjust the criterion used to set the upper bound on the range of thresholds considered. This, and similar methods, can help in exposing structure-function relationships within the brain.

References

Alexander-Bloch A, Lambiotte R, Roberts B, Giedd J, Gogtay N, Bullmore E. 2012. The Discovery of Population Differences in Network Community Structure: New Methods and Applications to Brain Functional Networks in Schizophrenia. *Neuroimage*. 59(4):3889-900.

Anand K, Bianconi G, Severini S. 2011. Shannon and von Neuman entropy of random networks with heterogeneous expected degree. *Physical Review E.* 83. DOI: 10.1103/PhysRevE.83.036109

Andellini M, Cannatà V, Gazzellini S, Bernardi B, Napolitano A. 2015. Test-retest reliability of graph metrics of resting state MRI functional brain networks: A review. *Journal of Neuroscience Methods*. 253: Pages 183-192.

Bassett D, Bullmore E, Verchinski B, Mattay V, Weinberger D, Meyer-Lindenberg A. 2008. Hierarchical organization of human cortical networks in health and schizophrenia. J Neurosci. 28(37): 9239-9248.

Battiston F, Nicosia V, Latora V. 2014. Structural measures for multiplex networks. *Physics Review E.* 89(3).

Broido A, Clauset A. 2019. Scale-free networks are rare. *Nat Commun*. 10(1017).

Bullmore E, Basset B. 2011. Brain Graphs: Graphical Models of the Human Brain Connectome. *Annu. Rev. Clin.* Psychol. 7:113–40.

Bullmore E, Sporns O. 2009. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*. 10: 186-98.

Chen ZJ, He Y, Rosa-Neto P, Germann J, Evans AC. 2008. Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cereb.Cortex.* 18: 2374–2381.

Cox RW. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*. 29:162-173.

Cover T, Thomas J. 1991. Elements of Information Theory. *Wiley*.

Crawford J, Greene C. 2020. Graph Biased Feature Selection of Genes is Better Than Random for Many Genes. *Biorxiv*.: https://doi.org/10.1101/2020.01.17.910703

Danon L, Diaz-Guilera A, Duch J, Arenas A. 2005. Comparing community structure identification. *J Stat Mech*.

Domenico M, Sasai S, Arenas A. 2016. Mapping Multiplex Hubs in Human Functional Brain Networks. *Frontiers in Neuroscience*. 10(326).

Domenico M. 2017. Multilayer modeling and analysis of human brain networks. *Giga Science*. 6: 1-8.

Dumoulin S, Fracasso A, van der Zwaag W, Siero J, Petridou N. 2018. Ultra-high field MRI: Advancing systems neuroscience towards mesoscopic human brain function. *Neuroimage.* 168: 345-357.

Fallani F, Richiardi J, Chavez M, Achard S. 2014. Graph analysis of functional brain networks: practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society B*. 369.

Ferrarini L, Veer IM, Baerends E, van Tol MJ, Renken RJ, van der Wee NJA, Veltman DJ, Aleman A, Zitman FG, Penninx BWJH, van Buchem MA, Reiber JHC, Rombouts SARB, Milles J. 2009. Hierarchical functional modularity in the resting-state human brain. *Hum. Brain Mapp*. 30: 2220–2231.

Fischl B. 2012. FreeSurfer. *NeuroImage*. 62(2): 774-781.

Garrison K, Scheinost D, Finn E, Shen X, Constable T. The (in)stability of functional brain network measures across thresholds. *Neuroimage.* 118: pages 651-661. *https://doi.org/10.1016/j.neuroimage.2015.05.046*

Glasser M, Sotiropoulos S, Wilson A, Coalson T, Fischl B, Anderson J, Xu J, Jbabdi S, Webster M, Polimeni J, Van Essen D, Jenkinson M. 2013. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*. 80: 105-24.

Gur R, Nimgaonkar V, Almasy L, Calkins M, Ragland J, Pogue-Geile M, Kanes S, Blangero J, Gur R. 2007. Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. *Am J Psychiatry*. 165(5): 813-819.

Hasan K, Walimumi S, Abid H, Hahn KR. 2011. A review of diffusion tensor magnetic resonance imaging computational methods and software. *Computers in Biology and Medicine*. 41: 1062-1072.

Van den Heuvel M, Lange S, Zalesky A, Seguin C, Yeo BT, Schmidt R. Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *NeuroImage.* Vol 152, pages 437-449. https://doi.org/10.1016/j.neuroimage.2017.02.005

Klein, A., & Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in neuroscience*, *6*, 171. https://doi.org/10.3389/fnins.2012.00171

Koutra D, Shah N, Vogelstein J, Gallagher B, Faloutsos C. 2016. DeltaCon: A Principled Massive-Graph Similarity Function. *ACM Transactions on Knowledge Discovery from Data*. 10(3).: 28 1–43. https://doi.org/10.1145/2824443

Lancichinetti A, Fortunato S, and Radicchi F. 2008. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E.* 78:46110.

Lancichinetti A, Fortunato S, Kertesz J. 2009. Detecting the overlapping and hierarchical community structure in compelx networks*. New Journal of Physics.* doi:10.1088/1367-2630/11/3/033015

Li R, Liao W, Li Y, Yu Y, Zhang Z, Lu G, Chen H. 2016. Disrupted structural and functional rich club organization of the brain connectome in patients with generalized tonic-clonic seizure. *Human Brain Mapping*. 37(12): 4487-4499.

Newman, M. E., Strogatz, S. H., & Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. Physical review E, 64(2), 026118.

Newman ME. 2006. Finding community structure in networks using the eigenvectors of matrices. *arXiv physics.*

Pu J, Wang J, Yu W, Shen Z, Lv Q, Zejic K, Zhang C, Sun B, Liu G, Wang Z. 2015. Discriminative Structure Feature Engineering for Macroscale Brain Connectomes. *IEEE Transactions on Medical Imaging.* 32(11): 2333 – 2342.

Raichle M. 2009. A brief history of human brain mapping. *Trends in Neuroscience*. 32(2): 118-126.

Roalf, D. R., Vandekar, S. N., Almasy, L., Ruparel, K., Satterthwaite, T. D., Elliott, M. A., Podell, J., Gallagher, S., Jackson, C. T., Prasad, K., Wood, J., Pogue-Geile, M. F., Nimgaonkar, V. L., Gur, R. C., & Gur, R. E. 2015. Heritability of subcortical and limbic brain volume and shape in multiplex-multigenerational families with schizophrenia. *Biological psychiatry*, *77*(2), 137–146. https://doi.org/10.1016/j.biopsych.2014.05.009

Rieke F, Warland D, Van Stevenick R, Bialek W (1999) Spikes: Exploring the Neural Code (Computational Neuroscience) (Reprint), Ed: Sejnowski T & Poggio T, Bradford Books, Cambridge MA

Roe A. 2019 Columnar Connectome: toward a mathematics of brain function. *Network Neuroscience.* 3(3): 779-791 https://doi.org/10.1162/netn_a_00088

Rutt B, Lee D. 1996. The Impact of Field Strength on Image Quality in MRI. *JMRI*. Vol 6(1) 57-62.

Schwarz A, Gozzi A, Bifone A. 2008. Community Structure and Modularity in Networks of Correlated Brain Activity. *Magnetic Resonance Imaging*. 914-920.

Sotiropoulos SN and Zalesky A. 2017. Building connectomes using diffusion MRI: why, how and but. *NMR In Biomedicine*. 32(4).

Suarez LE, Markello RD, Betzel RF, Misic B. 2020. Linking Structure and Function in Macroscale Brain Networks. *Cell Press Review Trends in Cognitive Sciences.* 24(4): 302-315.

Tournier JD, Smith R, Raffelt D, Tabbara R, Dhollander T, Pietsch M, Christiaens D, Jeurissen B, Yeh CH, Connelly A. 2019. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*. 202: 116–37.

Varshney LR, "Bottleneck capacity of random graphs for connectomics," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 6305-6309, doi: 10.1109/ICASSP.2016.7472890.

Watts DJ, Strogatz SH (1998) Collective dynamics of “small-world” networks. Nature 393:440 –442.

Welvaert M, Durnez J, Moerkerke B, Verdoolaege G, and Rosseel Y. 2011. neuRosim: an R package for generating fMRI data. *J. Stat. Softw.* 44, 1–18.

Yook S, Radicchi F, Meyer-Ortmanns H. Self-similar scale-free networks and disassortativity. *Phys Rev E StatNonlin Soft Matter Phys*. 2005;72(4 Pt 2):045105. doi:10.1103/PhysRevE.72.045105

Zanin M, Sousa P, Papo D, Bajo R, Garcia-Prieto J, Pozo F, Menasalvas E, Boccaletti S. Optimizing Functional Network Representation of Multivariate Time Series. *Nature Scientific Reports*. DOI: 10.1038/srep00630

Zhu D, Li K, Terry D, Puente A, Wang L, Shen D, Miller S, Liu T. 2013. Connectome-scale assessments of structural and functional connectivity in MCI. *Human Brain Mapping*. 35(7): 2911-2923.