Supplementary Methods

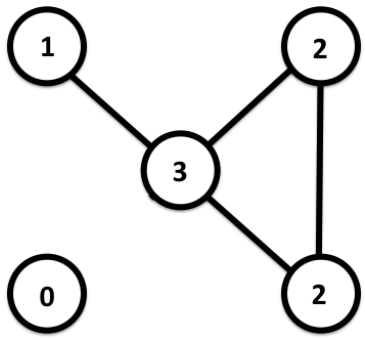
Graph Metrics

Degree

Perhaps the most basic and broadly cited graph metric is the 'degree', which is simply the number of neighbours connected to a given node, which for the ith node can be written as:

$$D_i = \sum_j C_{ij}$$

where C_{ij} represents the connection status between nodes i and j. C_{ij} = 1 when there exists a connection and C_{ij} = 0 when there is no connection (Sporns, 2003; Göttlich et al., 2013; Jan et al., 2013). This measure is only applicable to binary graphs, as only the presence or absence of a connection is taken into consideration, and any weight information is discarded.



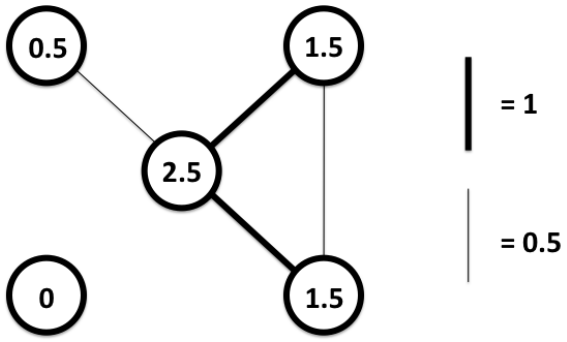
With reference to the diagram above, the number within the nodes denotes the degree of the node, i.e. the number of other nodes to which it is connected.

Strength

The ‘strength’ is closely related to the degree, in that it also counts the number of neighbours of a given node, but with the weightings (here mean FA values) of the connections taken into account (Rubinov and Sporns, 2010; Hwang et al., 2012; Goñi et al., 2013). The strength of the i^{th} node is defined as the sum of the weights of all edges incident upon it.

$$S_i = \sum_j W_{ij}$$

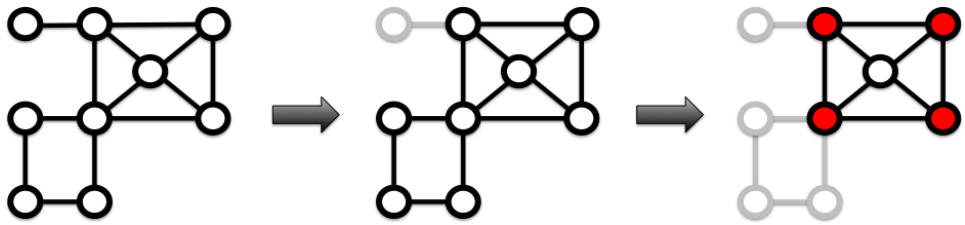
where W_{ij} is the weight of the edge connecting nodes i and j.



With reference to the diagram above, the numbers within the nodes represent their strength, as determined by the thickness of their inbound edges.

Node coreness

If we consider an iterative process whereby nodes with successively higher degree are stripped from the network, we arrive at a sub-network with nodes of minimum degree k – this is known as the k^{th} degree core. The 'node coreness' of a given node is equal to k if it belongs to the k^{th} degree core, but not the $(k+1)^{th}$ degree core (Hagmann et al., 2008).



With reference to the diagram above, the first step removes all nodes of degree 1. In the second step all nodes of degree 2 are removed, leaving a network with minimum node degree 3. The coreness of the 4 external nodes (in red) is equal to 3 as they belong to the 3^{rd} degree core, but not the 4^{th} degree core.

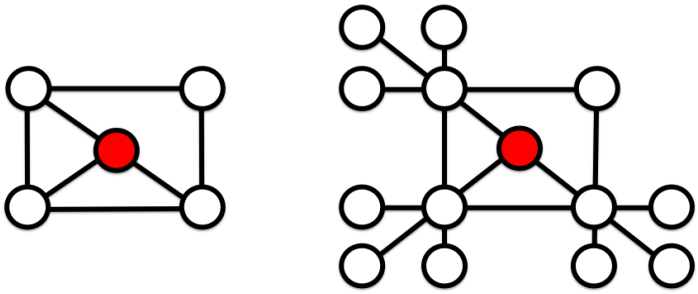
Rich club coefficient

The 'rich club coefficient' also splits a network into nodes of equal degree and is defined as the ratio of the number of connections between nodes of degree k to the number of all possible connections between nodes of degree k (Colizza et al., 2006; van den Heuvel and Sporns, 2011; Grayson et al., 2014).

Assortativity coefficient

The 'assortativity coefficient' builds upon the degree metric, in that it describes the correlation between the degrees of pairs of connected nodes (Newman, 2002; Deuker

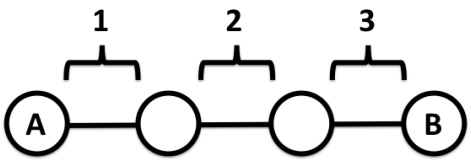
et al., 2009; Foster et al., 2010). The assortativity coefficient is a measure of correlation between the degrees of nodes on either end of a hyperedge. A hyperedge is a connection that exists in a hypergraph, in which a single edge can connect an arbitrary number of vertices. This measure is an indication of the extent to which a given node connects to other nodes with a similar degree. Positive values of assortativity are assigned to nodes that connect to nodes of similar degree, whilst negative values are assigned to nodes to connect to nodes of dissimilar degree. The assortativity coefficient r spans the range $-1 \leq r \leq 1$.



With reference to the diagram above, the assortativity coefficient of the red node in the first diagram is positive as it connects to nodes with the same degree as itself, whereas the assortativity coefficient for the red node in the second diagram is negative as it connects to nodes with dissimilar degree. Note that the connections between the pairs of nodes being compared are not counted in these calculations.

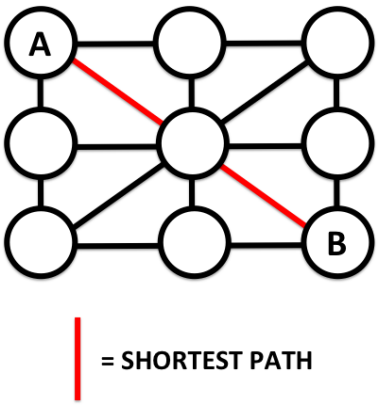
Characteristic path length

Path length is a type of distance used in network analysis, which describes the number of ‘steps’ required to link a pair of nodes (Sporns, 2003). For instance, a path length value of 3 between nodes A and B means that there are 3 edges separating them.



With reference to the diagram above, it may be seen that there is a path length of 3 separating nodes A and B.

There is often more than one possible path between pairs of nodes and it is sometimes of interest to calculate the shortest possible path length – i.e. the one that minimizes the number of edges.



With reference to the diagram above, it may be seen that the shortest path length between nodes A and B is represented by the 2 red edges.

The characteristic path length is the average of all the shortest path lengths between all pairs of nodes in a graph (Sporns, 2003; Dennis et al., 2011; Liang et al., 2013). In

order to obtain an expression for the characteristic path length, one first sums the shortest distances (D) between all pairs of nodes in the network to obtain the total path length C_T.

$$C_T = \sum_{i,j} D(i,j)$$

In order to arrive at the characteristic path length one needs to calculate the average of C_T. One node can make (n-1) connections – where n is the number of nodes in the network. The reason it makes (n-1) connections is that it connects to all nodes except itself. This is repeated n times in order to account for every pair selection. Therefore, the characteristic path length becomes:

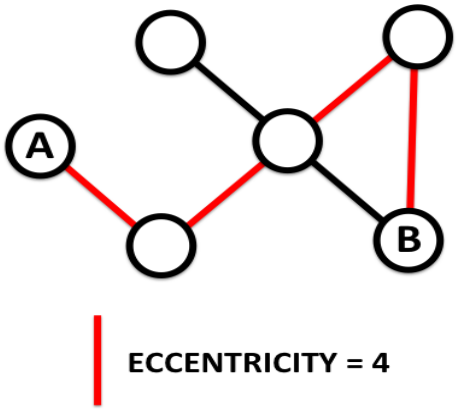
$$C = \frac{C_T}{n(n-1)}$$

Global & local efficiency

The 'global efficiency' is a scalar measure of information flow and is defined as the inverse of all shortest path lengths in a given network (Latora and Marchiori, 2001; Chen et al., 2013; Crone et al., 2014). The 'local efficiency' is a closely related metric that is calculated at the level of individual nodes, rather than at the level of the entire network, as with the global efficiency (Onnela et al., 2005; Klados et al., 2013; Rzucidlo et al., 2013).

Eccentricity

The eccentricity is the maximum path length between a given node and any other node in the network (Sporns, 2003).



With reference to the diagram above, it may be seen that the longest path length between nodes A and B is 4. This is defined as the eccentricity:

$$E(i,j) = \max(D(i,j))$$

Diameter & Radius

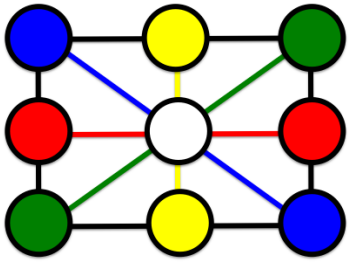
The 'diameter' and 'radius' are graph-level metrics that are defined as the maximum and minimum values of the eccentricity across all nodes in a given network, respectively (Sporns, 2003).

$$D = \max(E)$$

$$R = \min(E)$$

Betweenness centrality

The 'betweenness centrality' is a metric that depends upon the deconstruction of a graph into path lengths and is defined as the ratio of the number of shortest paths passing through a given node to the total number of shortest paths between all other nodes in the network (Brandes, 2001; Fadlallah et al., 2013; van Oort et al., 2013).



For instance, with regard to the diagram above, it can be seen that the central node participates in the shortest path lengths between the pairs of green, blue, yellow and red nodes. Therefore, the middle node is assigned a high value of betweenness centrality. The betweenness centrality of the i^{th} node can calculated by summing the shortest paths between all other pairs of nodes in the network:

$$B_i = \sum_{j,k} \frac{D_i(j,k)}{D(j,k)} \qquad i \neq j \neq k$$

Where $D_i(j,k)$ is the number of shortest paths between the j^{th} and k^{th} node that pass through the i^{th} node, and $D(j,k)$ is the total number of shortest paths between the j^{th} and k^{th} node.

The betweenness centrality may be normalized in order to compare networks of different sizes. This is achieved by dividing the equation above by the number of pairs of nodes excluding the node for which the betweenness centrality is being calculated. For the i^{th} node, a total of (N-1) nodes are able to form a pair with the j^{th} node, as the i^{th} node is excluded. Similarly, a total of (N-2) nodes are able to form a pair with the k^{th} node, as the i^{th} and j^{th} node are excluded. As we are dealing with undirected graphs in the case of fractional anisotropy, inbound and outbound links count as just one link and so one divides by 2 in order to account for this bi-directionality. Therefore, the normalized betweenness centrality becomes:

$$B_i^{NORM} = \frac{2}{(N-1)(N-2)} \sum_{j,k} \frac{D_i(j,k)}{D(j,k)} \qquad i \neq j \neq k$$

Eigenvector centrality

A graph metric that has gained considerable attention in recent years, with the rise of the page-rank algorithms of internet search engines, is 'eigenvector centrality'. This metric is self-referential, in that it assigns a high level of importance to a node if it is connected to other nodes that are themselves important (Binnewijzend et al., 2013; Geerligs et al., 2013; Liang et al., 2013). The eigenvector centrality of the ith node in the network may be calculated through the either a binary or weighted adjacency matrix. In the case of the weighted adjacency matrix a further dimension in the form of strength of connections between nodes of high centrality contribute to classification. The eigenvector centrality of the ith node is equal to the sum of the eigenvector centralities of all of its neighbours:

$$e_i = \frac{1}{\lambda} \sum_j a_{ij} e_j$$

where a_{ij} is the connection status between the ith and jth node and λ is a constant of proportionality. This may be rewritten in the form of the eigenvector equation:

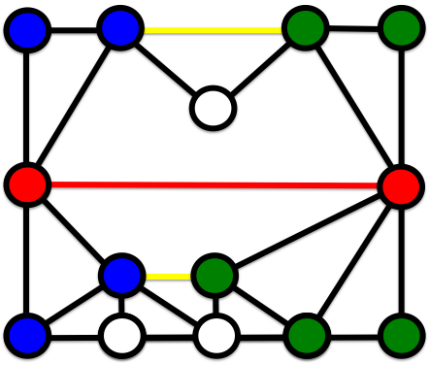
$$Ae = \lambda e$$

There are multiple values of lambda for which a solution to this eigenvector equation exists. However, we may impose the restriction that lambda must be positive, as a negative connection status is not a plausible scenario. The Perron-Frobenius theorem tells us that in the case of a square matrix with strictly positive components, as is the case here, there is a unique solution to the eigenvector equation in the form of the largest eigenvalue. Therefore, the ith component of the eigenvector related to the

largest eigenvalue gives the eigenvector centrality of the i^{th} node.

Edge neighbourhood overlap, matching index and node pair degree

There is a subset of graph metrics that are designed to quantify similarities in connection patterns within a network, which have also found utility in various technological and social sciences settings. For instance, the 'edge neighbourhood overlap' describes the extent to which the immediate neighbours of pairs of nodes overlap (Easley and Kleinberg, 2010). The neighbourhood of an edge consists of the nodes that are removed by exactly one edge from the two adjoining nodes.



With reference to the diagram above, we are interested in calculating the edge neighbourhood overlap of the red edge. The blue circles represent the neighbourhood of the left red node. Similarly, the green circles represent the neighbourhood of the right red node. We are interested in all green and blue nodes that are neighbours of one another – as represented by the yellow edges in the diagram. The edge neighbourhood overlap is defined as the number of nodes on either end of these yellow edges between neighbours of neighbourhoods, divided by the total number of nodes in both neighbourhoods. Therefore, for the network above the edge

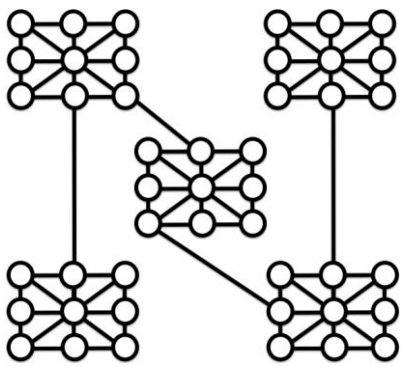
neighbourhood overlap is:

$$\frac{Nodes\ connected\ by\ yellow\ edges}{(Blue\ nodes + Green\ Nodes)} = \frac{(2+2)}{(4+5)} = 4/9$$

In the case of the matching index one is interested in the extent to which two nodes share neighbours (Sporns, 2003). However, in the case of the edge neighbourhood overlap one is interested in the extent to which the neighbours of the two nodes joined by the edge in question are neighbours of one another. The 'node pair degree' is the degree of each pair of nodes considered in the calculation of the edge neighbourhood overlap metric (Sporns, 2003).

Optimal community structure

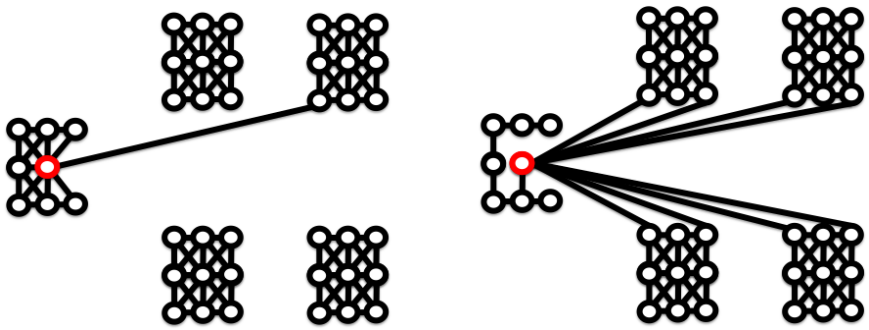
Certain graph metrics depend on the estimation of the 'optimal community structure' - a segregation of a network into densely inter-connected regions (Newman, 2006). A network may divided into modules in which nodes interact more strongly with one another than with those in the rest of the network. These sub-networks may be found by calculating the optimal community structure. The latter produces modules in which the number of inter-modular connections are minimized and the within-modular connections are maximized.



With reference to the diagram above, it may be seen that there are 5 clusters of highly interconnected nodes, with sparse connections between each other. The optimal community structure for this configuration would be split into 5 modules as shown in the diagram above.

Participation coefficient

The participation coefficient is a measure of how deeply an individual node is embedded within its local module, as defined by the optimal community structure (Guimera and Amaral, 2005). It is calculated as the ratio of the number of connections node A has within its module to the number of connections is has within the entire network.



With reference to the diagram above, the participation coefficient of the red node in the left network is higher, as it has more links within its local module relative to its total number of links, as compared with the red node in the right network. The ratio R of the i^{th} node's links within module m to its total number of links can be written as follows:

$$\frac{D_i^m}{D_i}$$

where D_i is the degree of node i. We then sum the ratio above over all modules m:

$$R_i = \sum_m \frac{D_i^m}{D_i}$$

We then normalize the ratio of within-module links to total links for node i across all modules and define the result as the participation coefficient P of node i:

$$P_i = 1 - \sum_m \left(\frac{D_i^m}{D_i}\right)^2$$

It can be seen from the above equation that the participation coefficient ranges from $0 \leq P < 1$, where P=0 denotes the case where the i[th] node links only to nodes within its own module, i.e. $D_i^m = D_i$. On the other hand, P approaches a value of 1 when the i[th] node connects as strongly with its local module as with all other modules, i.e. $D_i^m \ll D_i$

Diversity coefficient

There is an inherent uncertainty in assigning a given node to its local community, as quantified by the 'diversity coefficient', a quantity closely related to Shannon's definition of entropy in information theory (Shannon, 1948; Rubinov and Sporns, 2010). We can define the uncertainty in assigning the i[th] node to a module as:

$$H(i) = E[I(i)]$$

where E is the expected value operator and I is the information content. The expected value is the value one would obtain if one were to assign the node in question to a module, based on the optimal community structure algorithm, infinitely many times

and take the average of the values obtained. The information content is the 'quantity' of information associated with the outcome that a node has been assigned to a module. The information content of the i^{th} node being assigned to a particular module is inversely proportional to the probability of this event occurring. In other words, when there is a high probability that a certain node will be placed within a particular module, then the information content associated with this event having taken place is low.

Information content must be both positive and additive in nature. This means that the information content associated with both the i^{th} and j^{th} node being assigned to particular modules is equal to the sum of the information content associated with the i^{th} node being assigned to a particular module, plus the information content associated with the j^{th} node being assigned to a particular module:

$$I(i \cap j) = I(i) + I(j)$$

where $i \cap j$ is the intersection of the two independent events of nodes i and j being assigned to particular modules. This means that the information content associated with the outcome that the i^{th} node was assigned to a particular module can be written:

$$I(i) = log\left(\frac{1}{p(i)}\right) = -\log(p(i))$$

and combining this with the formula above, the entropy can be written:

$$H(i) = E[-\log(p(i))]$$

It may be seen that the entropy H tends to infinity when the probability of the i^{th} node being assigned to a particular module tends to 0. Similarly, the entropy is equal to zero when the probability of the i^{th} node being assigned to a particular module is equal to 1. In other words, the more equally distributed the probability distribution of a given node being assigned to modules in the network, the greater the uncertainty in

which module will be selected as its own.

As we are unable to run the optimal community structure algorithm infinitely many times in order to measure an expected value, we must deal with summations over finite sample sizes. In this case, the expected value becomes the probability of the i^{th} node being assigned to a particular module:

$$E(i) \rightarrow \sum_i p(i)$$

and the equation for the uncertainty (or diversity coefficient) in assigning the i^{th} node to a particular module can be rewritten:

$$H(i) = -\sum_i p(i)\log[p(i)]$$

Clustering coefficient

The extent to which nodes tend to group together is quantified by metrics describing 'clustering'. The 'clustering coefficient' is defined as the number of 'triangles' on the node level (Watts and Strogatz, 1998; Onnela et al., 2005; Dubbelink et al., 2013). For a network consisting of N nodes, the i^{th} node N_i can have a maximum of (N-1) inbound connections, as it can connect to every node in the network except itself. Summing over the network, each of the N nodes can make a maximum of (N-1) connections, giving a total of N(N-1) connections. For a node in an undirected network, an inbound and outbound link to another node counts as only one link, which means it can only make $\frac{(N-1)}{2}$ undirected connections. Therefore, for the entire undirected network, the maximum number of connections are $\frac{N(N-1)}{2}$. The local clustering coefficient is the ratio of connections that actually exist, to the total number

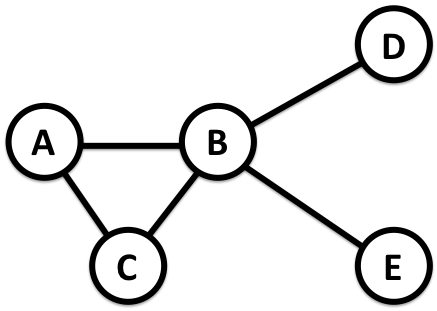
of possible connections:

$$Clustering\ Coefficient = \frac{2C}{N(N-1)}$$

where C is number of present connections.

Transitivity

The 'transitivity' is a scalar descriptor of the clustering coefficient, being defined at the network level, rather than at the individual node level (Onnela et al., 2005; Humphries and Gurney, 2008; Anderson and Cohen, 2013). The transitivity is based on triplet and triangle configurations. Triplets are sub-categorized into open and closed triplets. An open triplet is a set of 3 nodes, only 2 of which are connected. A closed triplet is a set of 3 nodes, all of which are connected.



For example, in the diagram above, the open triplets are as follows: **C-B-E**, **E-B-D**, **A-B-D.** The closed triplets are: **A-B-C**, **B-C-A**, **C-A-B**. The latter collection of 3 closed triplets are collectively known as a triangle. From this we can see that:

$$No.of\ Closed\ Triplets = 3 \times \text{No. of Triangles}$$

The number of triangles in the network can be calculated by summing over the connection statuses between all pairs of nodes:

$$No.\,of\ Triangles = \frac{1}{2}\sum_{i,j,k} a_{ij}a_{ik}a_{jk}$$

where aij is the connection status between the i^{th} and j^{th} node. The connection status equals 1 if there is a connection and zero if there is no connection. Note that the factor of ½ is due to the fact that the network is undirected in the case of fractional anisotropy and that therefore an inbound and outbound link count as only one connection. Transitivity is defined as the ratio of the number of closed triplets to the number of both closed and open triplets.

Minimising overfitting via the elastic net

Overfitting is a common problem in statistical analysis that occurs when a model describes noise rather than meaningful signals in the data. An overfitted model is therefore unable to make accurate predictions, often due to an excess amount of predictors (e.g. graph metrics) relative to observations (e.g. subjects). Various techniques have been developed in order to deal with overfitting in predictive models. These techniques attempt to create efficient models by minimizing the effects of redundant predictors in the models (Hastie et al., 2009).

One recent such technique is known as the elastic net, and is a combination of two previous techniques, known as the least absolute shrinkage and selection operator (LASSO) and ridge regression (Tibshirani, 1996). If a given dataset contains multiple highly correlated predictors (as is the case in our study) ridge regression will decrease the effect of groups of these simultaneously, but will not drive them all the way to zero. This is not useful for our study, as we wish to eliminate redundant graph metrics entirely from the models, in order to be left with only the most valuable. On the other

hand, the LASSO will only eliminate a single redundant predictor at a time. This is also not ideal for our purposes, as the single predictor being removed may be highly correlated with many others, which the LASSO would however retain in the model. This is where the elastic net can be highly advantageous, as it is able to penalize groups of highly correlated predictors simultaneously, whilst eliminating them entirely from the model. Therefore, the elastic net takes advantage of properties of both the LASSO and ridge regression to create a hybrid model that is highly suitable for a study such as ours (Zou and Hastie, 2005).

References

Anderson A, Cohen MS (2013) Decreased small-world functional network connectivity and clustering across resting state networks in schizophrenia: an fMRI classification tutorial. Front Hum Neurosci 7.

Binnewijzend MA, Adriaanse SM, Flier WM, Teunissen CE, Munck JC, Stam CJ, Scheltens P, Berckel BN, Barkhof F, Wink AM (2013) Brain network alterations in Alzheimer's disease measured by Eigenvector centrality in fMRI are related to cognition and CSF biomarkers. Hum Brain Mapp.

Brandes U (2001) A faster algorithm for betweenness centrality. J Math Sociol 25:163-177.

Chen Z, Liu M, Gross DW, Beaulieu C (2013) Graph theoretical analysis of developmental patterns of the white matter network. Front Hum Neurosci 7.

Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. Nat Phys 2:110-115.

Crone JS, Soddu A, Höller Y, Vanhaudenhuyse A, Schurz M, Bergmann J, Schmid E, Trinka E, Laureys S, Kronbichler M (2014) Altered network properties of the fronto-parietal network and the thalamus in impaired consciousness. NeuroImage: Clinical 4:240-248.

Dennis EL, Jahanshad N, Rudie JD, Brown JA, Johnson K, McMahon KL, de Zubicaray GI, Montgomery G, Martin NG, Wright MJ (2011) Altered structural brain connectivity in healthy carriers of the autism risk gene, CNTNAP2. Brain connectivity 1:447-459.

Deuker L, Bullmore ET, Smith M, Christensen S, Nathan PJ, Rockstroh B, Bassett DS (2009) Reproducibility of graph metrics of human brain functional networks. Neuroimage 47:1460-1468.

Dubbelink KTO, Hillebrand A, Stoffers D, Deijen JB, Twisk JW, Stam CJ, Berendse HW (2013) Disrupted brain network topology in Parkinson's disease: a longitudinal magnetoencephalography study. Brain:awt316.

Easley D, Kleinberg J (2010) Networks, crowds, and markets: Reasoning about a highly connected world: Cambridge University Press.

Fadlallah BH, Keil A, Principe JC (2013) Functional dependence in the human brain: A graph theoretical analysis. In: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp 2948-2951: IEEE.

Foster JG, Foster DV, Grassberger P, Paczuski M (2010) Edge direction and the structure of networks. PNAS 107:10815-10820.

Geerligs L, Saliasi E, Renken RJ, Maurits NM, Lorist MM (2013) Flexible connectivity in the aging brain revealed by task modulations. Hum Brain Mapp.

Goñi J, van den Heuvel MP, Avena-Koenigsberger A, de Mendizabal NV, Betzel RF, Griffa A, Hagmann P, Corominas-Murtra B, Thiran J-P, Sporns O (2013) Resting-brain functional connectivity predicted by analytic measures of network communication. PNAS:201315529.

Göttlich M, Münte TF, Heldmann M, Kasten M, Hagenah J, Krämer UM (2013) Altered Resting State Brain Networks in Parkinson's Disease. PLoS One 8:e77336.

Grayson DS, Ray S, Carpenter S, Iyer S, Dias TGC, Stevens C, Nigg JT, Fair DA (2014) Structural and Functional Rich Club Organization of the Brain in Children and Adults. PLoS One 9:e88297.

Guimera R, Amaral LAN (2005) Functional cartography of complex metabolic networks. Nature 433:895-900.

Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O (2008) Mapping the structural core of human cerebral cortex. PLoS Biol 6:e159.

Hastie T, Tibshirani R, Friedman J, Hastie T, Friedman J, Tibshirani R (2009) The elements of statistical learning: Springer.

Humphries MD, Gurney K (2008) Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence. PLoS One 3.

Hwang K, Hallquist MN, Luna B (2012) The development of hub architecture in the human functional brain network. Cereb Cortex.

Jan H, Chao Y-P, Cho K-H, Kuo L-W (2013) Investigating the effects of streamline-based fiber tractography on matrix scaling in brain connective network. In: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp 523-526: IEEE.

Klados MA, Kanatsouli K, Antoniou I, Babiloni F, Tsirka V, Bamidis PD, Micheloyannis S (2013) A Graph Theoretical Approach to Study the Organization of the Cortical Networks during Different Mathematical Tasks. PLoS One 8:e71800.

Latora V, Marchiori M (2001) Efficient behavior of small-world networks. Phys Rev Lett 87:198701.

Liang X, Connelly A, Calamante F (2013) Graph analysis of resting-state ASL perfusion MRI data: Nonlinear correlations among CBF and network metrics. Neuroimage.

Newman ME (2002) Assortative mixing in networks. Phys Rev Lett 89:208701.

Newman ME (2006) Finding community structure in networks using the eigenvectors of matrices. Physical review E 74:036104.

Onnela JP, Saramaki J, Kertesz J, Kaski K (2005) Intensity and coherence of motifs in weighted complex networks. Phys Rev Lett 71.

Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52:1059-1069.

Rzucidlo JK, Roseman PL, Laurienti PJ, Dagenbach D (2013) Stability of whole brain and regional network topology within and between resting and cognitive states. PLoS One 8:e70275.

Shannon CE (1948) Bell System Tech. J. 27 (1948) 379; CE Shannon. Bell System Tech J 27:623.

Sporns O (2003) Graph theory methods for the analysis of neural connectivity patterns. In: Neuroscience Databases, pp 171-185: Springer.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Series B 267-288.

van den Heuvel MP, Sporns O (2011) Rich-club organization of the human connectome. J Neurosci 31:15775-15786.

van Oort E, van Walsum A, Norris D (2013) An investigation into the functional and structural connectivity of the Default Mode Network. Neuroimage.

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

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B-Statistical Methodology 67:301-320.