NeuroImage 152 (2017) 437–449

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations

Martijn P. van den Heuvel[a,*], Siemon C. de Lange[a], Andrew Zalesky[b], Caio Seguin[b], B.T. Thomas Yeo[c], Ruben Schmidt[d]

[a] *Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Heidelberglaan 100 PO Box 85500, Room: A01.126, Utrecht, GA 3508, The Netherlands*

[b] *Melbourne Neuropsychiatry Centre & Melbourne School of Engineering, The University of Melbourne, Australia*

[c] *Dept of Electrical and Computer Engineering, Clinical Imaging Research Center, Singapore Institute for Neurotechnology, Memory Network Program, National University of Singapore, Singapore*

[d] *Brain Center Rudolf Magnus, Department of Neurology, University Medical Center Utrecht, The Netherlands*

ABSTRACT

Graph theoretical analysis has become an important tool in the examination of brain dysconnectivity in neurological and psychiatric brain disorders. A common analysis step in the construction of the functional graph or network involves "thresholding" of the connectivity matrix, selecting the set of edges that together form the graph on which network organization is evaluated. To avoid systematic differences in absolute number of edges, studies have argued against the use of an "absolute threshold" in case-control studies and have proposed the use of "proportional thresholding" instead, in which a pre-defined number of strongest connections are selected as network edges, ensuring equal network density across datasets. Here, we systematically studied the effect of proportional thresholding on the construction of functional matrices and subsequent graph analysis in patient-control functional connectome studies. In a few simple experiments we show that differences in overall strength of functional connectivity (FC) – as often observed between patients and controls – can have predictable consequences for between-group differences in network organization. In individual networks with lower overall FC the proportional thresholding algorithm has to select more edges based on lower correlations, which have (on average) a higher probability of being spurious, and thus introduces a higher degree of randomness in the resulting network. We show across both empirical and artificial patient-control datasets that lower levels of overall FC in either the patient or control group will most often lead to differences in network efficiency and clustering, suggesting that differences in FC across subjects will be artificially inflated or translated into differences in network organization. Based on the presented case-control findings we inform about the caveats of proportional thresholding in patient-control studies in which groups show a between-group difference in overall FC. We make recommendations on how to examine, report and to take into account overall FC effects in future patient-control functional connectome studies.

Introduction

The measurement and investigation of functional connectivity has become an important approach in the field of connectomics, the study of the topological organization of the structural and functional wiring of nervous systems (Bullmore and Sporns, 2009; Damoiseaux et al., 2006; Fox and Raichle, 2007; Smith et al., 2009; Smith et al., 2011; van den Heuvel and Hulshoff Pol, 2010). Furthermore, the examination of the topological aspects of functional brain networks and the possibility of examining possible disruptions in network organization in disease has become an invaluable tool for studying brain dysconnectivity in a wide range of psychiatric and neurological disorders (Filippi et al., 2013; Fornito and Bullmore, 2012; Stam and Reijneveld, 2007).

A typical experimental setting to examine differences in functional brain network organization is the acquisition of resting-state fMRI data (or equivalent EEG/MEG), followed by the computation of functional connectivity by means of correlation analysis between the measured time-series. Performing correlation analysis for all possible pairs of

[*] Corresponding author.
E-mail address: M.P.vandenheuvel@umcutrecht.nl (M.P. van den Heuvel).

http://dx.doi.org/10.1016/j.neuroimage.2017.02.005
Received 8 December 2016; Accepted 2 February 2017
Available online 03 February 2017

brain regions results in a functional connectivity matrix for each of the individual subjects, with the obtained case and control matrices often "thresholded", meaning the selection of those connections that reach a certain absolute or relative threshold. Although studies have suggested that this operation may ignore potentially valuable information during functional network construction (Gallos et al., 2012; Goulas et al., 2015; Santarnecchi et al., 2014), thresholding is a commonly applied approach in functional connectomics to remove spurious connections and to obtain sparsely connected matrices, a prerequisite for the computation of many graph theoretical metrics.

Two of the most commonly applied approaches to perform this thresholding include the "absolute threshold" and the "proportional threshold" approach. The absolute threshold approach describes the selection of those network edges that exceed an absolute threshold T, for example all correlations higher than 0.3, with (in the binary case) all surviving connections set to 1 and all other network connections set to 0. Although a simple and potentially powerful approach to reconstruct functional networks, setting an absolute threshold can lead to different numbers of network edges across datasets, and -importantly for disease studies- different levels of network density between control and patient cases. Network density, expressing the proportion of all possible connections that are present in the network (also commonly referred to as "graph density" or "connection density") has been shown to have a direct effect on the computation of many graph metrics (see in particular the study of Van Wijk et al. (2010) for a detailed theoretical and experimental overview), potentially leading to statistical differences in network metrics between patient and control populations, effects that should be attributed to underlying differences in number of network connections and not directly to disease related differences in network topology. As such, this approach has been suggested to be less favorable for case-control studies (Nicols et al., 2016).

To overcome this issue, studies have proposed an alternative approach of using a proportional threshold (Achard and Bullmore, 2007; Bassett et al., 2009; Van den Heuvel et al., 2008), aiming to keep the number of connections fixed across all individuals to rule out the influence of network density on the computation and comparison of graph metrics across groups. The proportional thresholding approach includes the selection of the strongest PT% of connections in each individual network, setting all (in the binary case) surviving connections to 1 and other connections to 0. This selection procedure is often referred to in literature as an analysis in which the "density" (Jalili, 2016; Van den Heuvel et al., 2008) or "network cost" (Achard and Bullmore, 2007; Bassett et al., 2008; Ginestet et al., 2011) is set fixed across patient and healthy control cases, with potential between-group differences in graph metrics (e.g. clustering, path length) assumed to result from differences in the topological organization of edges and not due to differences in number of edges. Compared to absolute thresholding, proportional thresholding has been argued to reliably separate density from topological effects (Braun et al., 2012; Ginestet et al., 2011) and to result in more stable network metrics (Garrison et al., 2015), making it a commonly used approach for network construction and analysis in disease connectome studies.

However, as discussed in the graph theoretical studies of Van Wijk et al. (2010) and others (e.g. (Alexander-Bloch et al., 2010; Fornito et al., 2013; van den Heuvel and Fornito, 2014) the inclusion of lower and thus potentially less reliable correlations as functional network edges can have an effect on the organization of the constructed functional network, and thus an effect on subsequently derived graph metrics. The effect of including potentially less reliable connections has been discussed in the theoretical setting of artificially generated toy networks (van Wijk et al. 2010). Here, we take an empirical and practical approach on this matter. We studied the effect of proportional thresholding on the formation of functional networks and the subsequent computation and comparison of graph metrics across groups, in particular in the case of studying patient-control differences in functional network organization.

To be more specific about our study aim, we set out to investigate how the use of proportional thresholding can introduce (artifactual) topological differences in network structure in a patient - control brain network study, differences that perhaps should be attributed to underlying between-group differences in functional connectivity and not directly to network architecture. We write this report to caution against the use of this approach in disease network studies in which there is a widespread between-group difference in overall functional connectivity strength (FC). Patient populations often show different levels of FC as compared to controls (be it the result of disturbed brain communication, changes in neural activity and/or of increased noise, global signal or motion), and in the methods and results section of this report we show that this can have a pronounced effect on the computation and between-group comparison of network metrics when using proportional thresholding.

The theoretical background of this effect can be understood as follows (see also (Fornito et al., 2013; van den Heuvel and Fornito, 2014; van Wijk et al., 2010)): When setting a proportional threshold, the number of connections across patient and healthy control subjects is set to the same fixed number, leading to a fixed network cost / density across all included participants. In the case of a dataset in which the edges show lower levels of FC as compared to other datasets in the sample, this network density can only be reached by including more low correlations to reach the required number of network edges.[1] Due to the nature of the computation of the correlation coefficient, lower correlations based on the same number of time-point samples are less reliable, which will increase the chance of including a random noisy connection into the reconstructed network, an effect detrimental for the computation of network metrics (see also Zalesky et al. (2016) and discussion).

While this effect may average out when averaging functional networks, for example in studying the healthy functional connectome, in the setting of a patient-control study this can have severe consequences. Having lower overall functional connectivity in one of the groups could lead to significant differences in network structure due to the inclusion of more random connections, making the network as a whole more comparable to randomly connected networks. The small-world model of Watts and Strogatz (1998) shows that random edges can act as shortcuts in the network, reducing the overall shortest path length and lowering the chance of finding topologically closed local circuits. Moreover, the Watts and Strogatz model illustrates that the inclusion of even a few random edges can strongly reduce the overall shortest path length (and therewith increase global efficiency) in the network, illustrating that graph properties can rapidly change with respect to small changes in network wiring. Following this line of thought, in the case of a patient population showing lower overall connectivity, setting a proportional threshold may introduce additional random shortcuts, which can in turn have a pronounced effect on the creation of shortest paths. This will be reflected in an increase in network global efficiency, reduction in overall network clustering, and a network topology more comparable to that of random networks. Conversely, if patients show increased levels of functional connectivity as compared to healthy controls, this can lead to lower global efficiency and increased local clustering, and thus a -perhaps incorrectly concluded- less efficient and more locally clustered network organization in patients.

In what follows we show empirical evidence for this phenomenon in

[1] We assume that a large subset of elements in the matrix show reduced correlations. We recognize that this does not always have to be the case: overall FC can be lower while the proportionally thresholded edges across datasets are similar. For example, comparison of the sorted list of edge weights of two toy networks A=[0.9 0.8 0.5 0.4] and B=[0.9 0.8 0.2 0.1] results in network B having a lower overall weight, but a proportional threshold of 50% results in networks with equal strength across selected edges. In the Supplemental Materials we verify that in the empirical datasets examined in this study the overall strength of all as well as of the selected subset of edges is lower.

functional brain networks constructed using proportional thresholding. First, we illustrate the effect in patient and control datasets, derived from both fMRI and EEG data. Second, we explore the consequence of using proportional thresholding in functional networks of a population of healthy control subjects, data taken from the high-quality HCP dataset. We show that by ordering subjects solely on their overall FC we can mimic typically observed patient-control effects of network differences, with the extent of between-group differences dependent on the difference in FC between groups. We conclude by making recommendations for functional network researchers to verify that their reported patient-control effects of disrupted network organization are not a direct result of underlying differences in overall connectivity strength.

Methods

By means of four simple experiments we examined and tested the effect of inter-subject variation in overall FC on the construction of functional networks using proportional thresholding and the subsequent computation of the graph metrics of global efficiency GE and network clustering C, two basic metrics commonly examined in disease connectome studies. We focus our examination on graph metrics of binary versions of the derived functional networks, describing only the presence and absence of connections between cortical regions. We decided to primarily focus on binary networks to show that differences in graph metrics between selected groups are the result of the topological organization of selected network edges and not the result of differences in amount and/or distribution of weights across the set of selected network edges. In the Discussion and Supplemental Materials (page 4–8, section normalized binary and weighted metrics) we discuss and show that the same effect might occur –but with varying degree– in normalized binary and normalized weighted graphs.

In what follows we first describe the fMRI and EEG functional connectivity datasets used in this study, followed by a brief formal description of the examined graph metrics and the procedures used for statistical evaluation of between-group effects. The Results section gives a description of four illustrative experiments that examine the influence of overall FC on graph metrics and between-group effects, as well as strategies to correct for confounding effects of total functional connectivity on graph metrics.

Dataset I: Schizophrenia

The first patient-control dataset was taken from a study on anatomical network connectivity and structural-functional coupling in schizophrenia patients (van den Heuvel et al., 2013), from which we included functional connectivity networks of 48 patients and 44 matched healthy controls. A brief description of the construction of the functional connectivity matrices is given below and for details we refer to previous work of (van den Heuvel et al., 2013). Data was acquired on a 3 T Philips Achieva clinical scanner at the University Medical Center Utrecht, using an eight-element SENSE receiver head-coil. Participants underwent a 45-minute scanning session, in which a resting-state fMRI and an anatomical T1 scan was acquired. Resting-state Blood Oxygenation Level Dependent (BOLD) signals were recorded during a period of 8 min (*parameters:* 3D PRESTOSENSE, TR/TE 22/32 ms using shifted echo, flip-angle 9 degrees; p/s-reduction 2/2; dynamic scan time 502 ms, 4 mm isotropic voxel size, 32 slices covering whole brain). A T1-weighted image was acquired for anatomical reference (*parameters*: 3D FFE using parallel imaging; TR/TE 10 ms/4.6 ms; FOV 240×240 mm, 200 slices, 0.75 mm isotropic voxel size). Data processing of the resting-state fMRI data involved realignment and co-registration to the T1 image, removal of linear trends and first order drifts, removal of global effects (regressing out the white matter, ventricle, and global mean signals, as well as 6 motion parameters) and band-pass filtering (0.02–0.12 Hz). Potential effects of motion were removed by means of ‘scrubbing’ (Power et al., 2012), removing scan frames from the individual time-series in which significant movement was detected (see for details (van den Heuvel et al., 2013)). Next, tissue classification and cortical segmentation was performed on the basis of the T1 scan, followed by parcellation of the cortex into 68 cortical areas using the Desikan-Killiany atlas (Desikan et al., 2006; Hagmann et al., 2008). Functional connectivity between each of the 68 cortical regions (34 left hemisphere, 34 right hemisphere) was assessed by means of correlation analysis, computing the Pearson correlation coefficient between the time-series of region i and region j, for all combinations of regions i and j of the Desikan-Killiany atlas, resulting in a fully filled 68×68 FC matrix.

Dataset II: ADHD and Autism

Functional connectivity matrices of patients with ADHD and healthy controls were downloaded from the open data USC Multimodal Connectivity Database, describing resting-state functional connectivity between 190 brain regions for 190 patients and 330 healthy controls (URL:http://umcd.humanconnectomeproject.org/) (Brown et al., 2012). Functional connectivity matrices of patients with autism and matched healthy controls were taken from the same connectivity database, describing functional connectivity matrices between 264 regions for 42 patients and 37 matched controls.

Dataset III: EEG Autism

To show that the reported effects are not specific to networks derived from resting-state fMRI data, we also examined the effect of proportional thresholding on functional connectivity matrices derived from EEG recordings. FC networks were taken from a previously described EEG study (Boersma et al., 2010), including EEG recordings and subsequent functional connectivity reconstruction in a set of 12 autistic children and 19 matched healthy controls (32 electrodes, 2048 Hz sampling rate, neutral stimuli condition) (Boersma et al., 2010). In this study, functional connectivity was assessed by means of the phase lag index (PLI) (Stam et al., 2007) between the time-series of 32 skull electrodes (a metric ranging from 0 to 1, with 0 indicating no functional coupling and 1 indicating strong coupling, beta-band), resulting in a filled 32×32 functional connectivity matrix for each of the participants.

Dataset IV: Human Connectome Project

Functional connectivity matrices were reconstructed for the Human Connectome Project (Glasser et al., 2013; Van Essen et al., 2012) (Q3 release, resting-state fMRI data of 466 healthy controls included, voxel-size 2mm isotropic, TR/TE 720/33.1 ms, 1200 volumes, 14:33 min, first LR run taken here). fMRI volumes were realigned, co-registered with the T1 image, band-pass filtered (0.01–0.1 Hz), corrected for global effects by regressing out effects of motion (taken as the realignment parameters as described by HCP), global signal mean, ventricle and white matter signal, and scrubbed (FD=0.25, DVARS=1.5) for potential movement artifacts following standard procedures (see (van den Heuvel et al., 2015; van den Heuvel et al., 2016) for all details). T1 scans were used to parcellate the cortex into 68 cortical areas using the Desikan-Killiany atlas (the same as in the schizophrenia dataset), after which a functional connectivity matrix was derived by computing Pearson correlation coefficients between every pair of average regional time-series. In addition, regional signal power was computed over the preprocessed time-series for all regions i, with total signal power over the entire dataset computed as the average power across all regions i.

Within-subject networks

To examine the effect of proportional thresholding on graph metrics of functional networks constructed *within* a single subject we divided the functional time-series (1200 time-points) in half and made a corresponding FC matrix of each of the two parts (i.e. time points 1 to 600 and time points 601 to 1200) in the exact same way as on the entire time-series. We chose to split one time-series in half rather than taking one of the other runs available in the HCP data to rule out any potential difference in physiological state between different runs, assuming that within one run (~15 min) the physiological state of a subject would remain the same. To verify this, we checked across the 466 datasets that the low and high overall FC runs (see next paragraph for the formal definition and computation of overall FC of a matrix) were equally distributed across the two runs to rule out any potential systematic differences between the two runs (for example differences in arousal as one might argue that subjects are potentially more relaxed or more sleepy in the second part of a run). This was indeed the case (242 subjects showed the lowest FC in the first part and the highest in the second part (52% of total group), 224 subjects showed the highest FC in the first and the lowest FC in the second part (48%)). Extending this split-half analysis we also examined dynamical networks creating multiple FC networks by selecting blocks of 100 time-points by means of a moving window across the complete time-series (Results shown in Supplemental Materials, page 15, dynamical networks).

Overall functional connectivity

For each individual matrix, the overall functional connectivity of a matrix (referred to in this paper as overall FC) was taken as the mean of all positive values across all elements of the matrix. Computing overall FC by taking absolute values revealed similar findings.

Proportional thresholding

Proportional thresholding was performed on the FC matrices by selecting the PT% strongest connections (i.e. the strongest PT% correlations) of the derived functional connectivity matrix and setting these connections to 1, with all other connections set to 0. The application of a PT% proportional threshold to a functional connectivity matrix resulted in a binary graph with a density of PT%. In this study we examined a range of levels of PT from 35% to 1% in steps of 1% (see experiments below), with the application of the proportional threshold of PT=15% used to illustrate effects. From now on we refer to setting a proportional threshold of PT% and the subsequent binarization of the matrix to make an unweighted undirected graph as the application of a proportional threshold of PT%.

Graph metrics

After thresholding, topological properties of the reconstructed binary connectivity matrices were quantified by means of graph theoretical analysis. We focus on the commonly used basic metrics of global efficiency GE and clustering C, computed as implemented in the Brain Connectivity Toolbox (Rubinov and Sporns, 2010). Binary global efficiency *GE* was computed as the inverse of the harmonic mean of the shortest path length between all nodes i and j in the network, with higher levels of GE often interpreted as a network topology better suited for efficient network transfer. Binary clustering C was computed as the ratio of the present and total possible number of connected triangles around a network node i, averaged over all nodes i in the network. In the Supplemental Materials we report on a few other commonly used metrics (Supplemental Materials, page 13, other metrics).

Between-group comparison and statistical evaluation

Statistical evaluation of differences in graph metrics was assessed using t-tests, with differences between two groups (i.e. patient / controls) tested using two-sample t-tests and differences within individual datasets (see experiment 2 evaluating HCP data) tested by means of paired samples t-tests. Non-parametric testing by means of permutation testing (random shuffling group assignment) (Bassett et al., 2008; van den Heuvel et al., 2010)(10,000 permutations examined) revealed similar findings. We examined and statistically tested a wide range of proportional thresholds as well as several different patient and healthy control groups to illustrate that the same effect occurs over a range of thresholds and across a wide range of conditions. To test across a wide range of settings and analysis strategies a two-sided alpha threshold of 0.05 was used.

Results

Experiment 1: Disease datasets

In the first experiment we examined empirical differences in GE and C in the patient-control datasets, examining graph organization in schizophrenia, ADHD and autism, across both fMRI and EEG datasets. For this we followed a standard analysis procedure for disease connectome studies, with individual functional connectivity matrices first proportionally thresholded, binarized and then analyzed with graph theory, followed by statistical evaluation of the derived graph metric values across the patient and control group.

Schizophrenia dataset

As expected, the population of schizophrenia patients showed significantly higher GE and lower C as compared to the population of controls. For example, for an exemplary threshold of 15%, patients showed a significantly higher global efficiency GE ($p=0.0284$, Fig. 1) and trend-level lower clustering C ($p=0.0523$, Fig. 1) as compared to the population of healthy controls, which is commonly interpreted as a more random network organization in patients. We tested overall FC between patients and controls, observing a 4.8% lower overall FC in patients ($p=0.0052$). Fig. 1C shows the effect in GE for the range of examined proportional thresholds.

Next, we examined whether these group differences could be driven by a general relationship between overall FC and graph metrics across the group of subjects. Across the complete group (thus patients and controls combined) overall FC correlated to binary GE (proportional threshold: 15%, $r=-0.87$, $p < 0.001$, Fig. 1D), with networks based on lower FC connections on average showing higher GE. Overall FC and binary C were also correlated ($r=0.82$, $p < 0.001$).

To further illustrate the potential influence of overall FC on graph metrics, in particular in the context of between-group comparison of graph metrics, we ordered the set of 48 patients and 44 controls according to individual overall FC, with the set of controls and patients ordered separately. We then tested, for an exemplary proportional threshold of 15% (see below for an examination of the entire range between 35% and 1%) the difference in GE and C for a subsample of patients and controls that no longer showed a significant difference in overall FC, removing one by one the remaining lowest FC scoring patient and the top highest FC scoring control from the two samples until the overall between-group difference in overall FC showed a $p > 0.05$. The first subsample that reached this criterion involved 46 patient and 42 control datasets, i.e. the removal of 4 datasets in total. Statistical testing of this subpopulation of patients and controls no longer revealed a significant effect in GE ($p=0.134$) nor in C ($p=0.239$). To be more strict on differences in FC (to rule out that small effects in FC could still result in changes in GE and C) we also performed the same analysis but now with a stricter threshold, removing subjects until a t-statistic of < 1 ($p\sim0.15$) was reached. The first subsample that reached this criterion involved 44 patient and 40 control datasets. Statistical testing of this subpopulation of patients and controls no longer revealed any indication of a between-group effect in GE

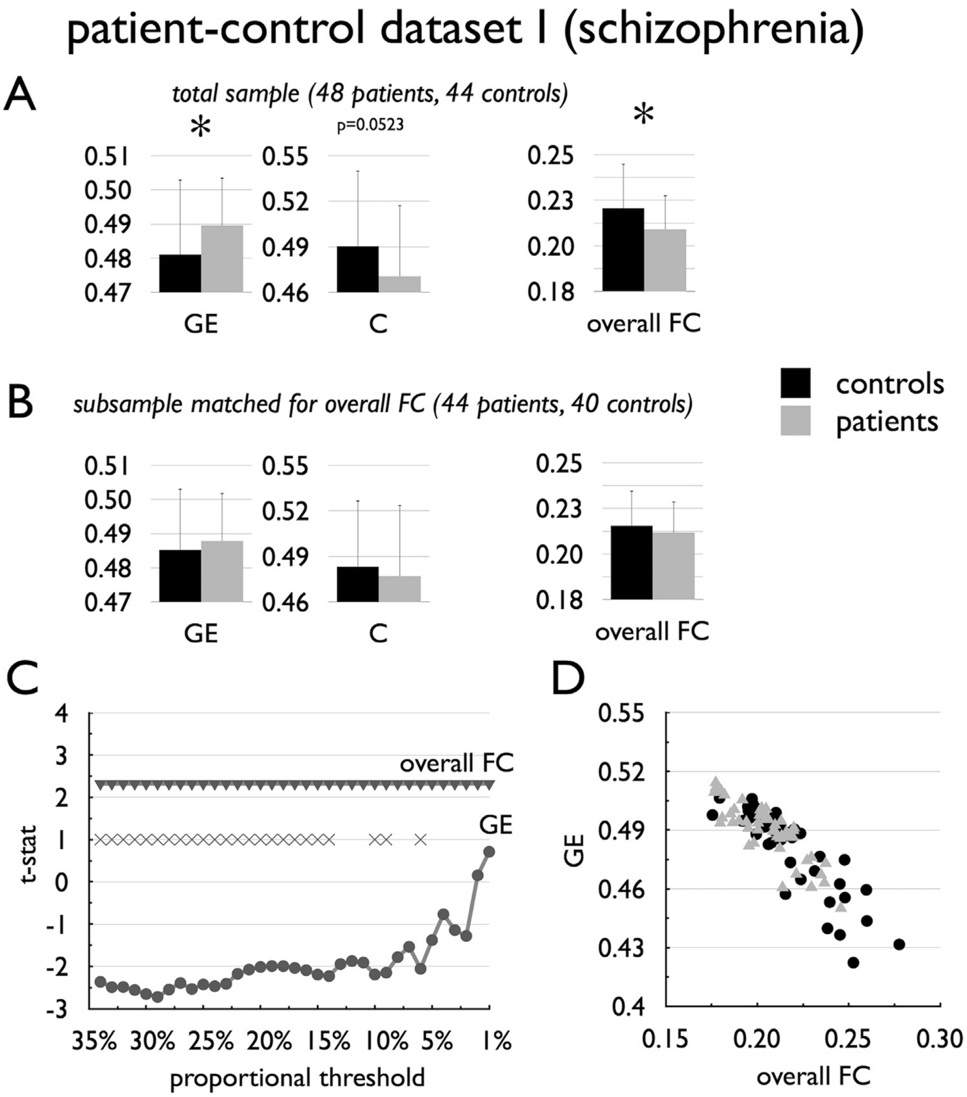

Fig. 1. Effect of proportional thresholding on an exemplary patient-control dataset. **Panel A** reports the results of a typical graph theoretical analysis on functional connectivity (FC) networks of schizophrenia patients (n=48) and matched healthy controls (n=44). FC networks are thresholded with a proportional threshold of 15% and from the binary graphs global efficiency GE and clustering C are computed and tested, resulting in increased GE and reduced C (trend-level) in patients. The right bar plot shows the significant difference in total functional connectivity (FC) between patients and controls. **Panel B** describes the results of the same dataset, but now with the two samples matched for overall FC, by removing the top 4 patients showing the lowest FC and the top 4 controls showing the highest FC removed from the sample (overall FC t-score < 1). GE and C now do not show group-differences. **Panel C** shows the difference (as expressed in t-scores) in GE between controls and patients across a range of proportional thresholds (35% to 1%). X indicates a significantly higher GE in patients as compared and controls. Overall FC shows the t-statistic of testing group differences in overall FC between the control and patient population using a two-sample t-test. Panel D shows the relationship between overall FC and GE across the entire sample. Control samples are plotted as black dots, patient samples as grey triangles.

(p=0.4272, Fig. 1) nor in C (p=0.5217). To further verify that this reduction was not an effect of reduced study power (due to a smaller sample size), we performed a similar test on a subset of 44 patients and 40 controls, randomly selected from the total population of patients and controls, which did again reveal differences in GE (p=0.0272, Fig. 1) and a trend level difference in C (p=0.0681), as well as a difference in overall FC ($p < 0.001$). Performing a 1000 random draws revealed similar findings (e.g. GE: median t-score −2.01 corresponding to p=0.02168). To finally show that the effect was not the result of the removal of the most severely ill patients, we performed the same subsample analysis one more time, now removing only control samples (7 removed until between-group FC showed a t-score < 1). This similarly revealed a diminishing effect on group differences in GE (p=0.551) and C (p=0.406).

To further examine the extent of FC differences on the computation and evaluation of graph metrics between groups, we next performed a series of comparisons between a range of subsamples of patients and controls. First, from both the patient and control population a subsample of the top m=20 subjects (i.e. subjects [1,2,.,m]) scoring respectively the lowest (for the patients) and the highest (for the controls) on FC were selected and compared (see Fig. 2A for a schematic overview of this analysis). Next, subsamples with the second highest / lowest overall FC, i.e. subjects [2,3,.,m+1], of both populations were selected and compared, followed by a selection of the subsample [3,4,.,m+2] etcetera, until n − m + 1 ordered subgroups were selected (i.e. up until the set [n-m+1,n-m+2,.,n], with n the sample size of the smallest of the two groups). [To match the size of the two samples we sampled until the size of the smallest of the two groups was reached. Excluding the four lowest FC samples of the largest of the two groups revealed similar results]. As such, for the first test the difference in overall FC between the patient and control sample was maximized (controls having 17% more overall FC than the subsample of patients, $p < 0.0001$), but per subsequent test the total difference in FC between the tested patient and control set was reduced, and with

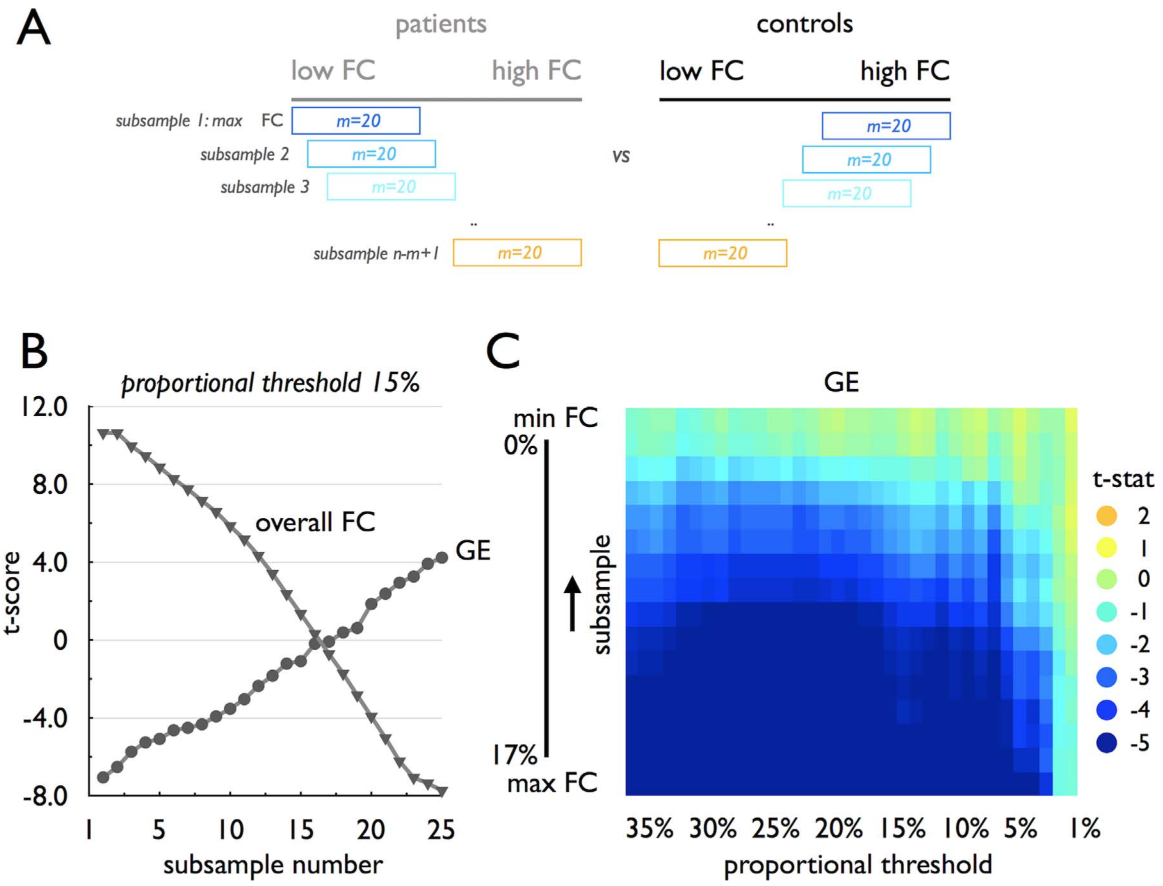

Fig. 2. Panel A provides a schematic overview of the performed subsample analysis, selecting subsamples of maximum and minimum differences in overall FC. Patients and controls were ordered according to their overall FC and starting with two subsamples of m=20 patients with the lowest FC and m=20 controls with the highest FC maximizing the largest group difference in overall FC (lower row). Next, a subsample of m=20 of the top-1 scoring patients and controls was tested (second row), followed by testing the top-2 subsample (third row) etcetera, until a minimum positive difference of overall FC was reached (upper row). **Panel B** shows that the difference in overall FC across tested subsamples declines (a trivial effect given the ordered selection of the subsamples), an effect accompanied by a decrease in the between-group difference in GE, with eventually (at levels where overall FC is matched across groups) effects in GE disappearing. **Panel C** shows the results of the exact same analysis as Panel B, but now for the entire range of proportional thresholds 35% to 1% (left to right). Panel shows that the effect of overall FC is present across almost all proportional thresholds, and particularly disruptive at high to medium network density levels and in subsamples where the difference in overall FC is the strongest. △GE is given as the t-statistic score in GE between the subsamples of patients and controls. △FC is computed as the percentage of difference in overall FC between patients and controls, computed (controls - patients) / controls x 100%.

this the effect in GE diminished (as shown in Fig. 2B). At the level at which the group difference ΔFC was around 0 (subsample 16, showing the minimal ΔFC meaning that patients and controls showed equal levels of overall FC), effects in GE were no longer present. In subsequent subsamples the set of patients demonstrated higher FC than the controls (as we were now selecting the highest FC patients and the lowest FC controls) with GE now lower in the patient population than in the control population. The relationship between effects in GE and effects in FC became more even apparent when we correlated ΔGE to ΔFC across all subsamples, showing a strong association between the two values ($r=-0.96$, $p < 0.001$).

We next tested the effect of △FC on △GE across the range of proportional thresholds (Fig. 2C). Testing across proportional thresholds from 35% to 1% again showed the strongest group effects in network metrics to be present in subsamples of maximal differentiating levels of overall FC and with diminished effects when subsamples of equal levels of FC were tested. The right panel of Fig. 2B shows for the total range of proportional thresholds (left to right: 35% to 1% network density) the computed between-group effect size in GE (in percentage of change between patients and controls, and corresponding t-statistic) for each of the subsequently tested subsamples until the minimal ΔFC of ~0% was reached (bottom to top). The left panel shows the accompanying difference in overall FC between the tested patient and control samples (left to right).

As for an alternative strategy to examine the confounding influence of overall FC on comparisons of graph metrics between two groups, we performed a final analysis in which patient - control status was ignored altogether, comparing differences in overall FC, GE and C between groups randomly selected from the total included population of 48 + 44 datasets. For 10,000 iterations, we drew two random groups (n=48, n=44) and computed the difference in overall FC, binary GE and C (resulting in △FC, △GE and △C respectively) between the two randomized groups. Across the 10,000 random iterations, △FC was strongly correlated to △GE ($r=-0.87$, $p < 0.001$) and △C ($r=0.82$, $p < 0.001$), further confirming a strong influence of overall FC on between-group differences in network organization.

Autism dataset

Similar findings were observed when examining the fMRI functional connectivity dataset of the autism sample. Functional connectivity matrices of autism patients showed significantly lower levels of overall FC as compared to controls (proportional threshold 15%, p=0.0070), as well as higher levels of GE (p=0.0180) and lower C (p=0.0094) (Supplemental Figure 1). Excluding the 14 lowest FC patient datasets and the 14 highest FC controls until the between-group difference in FC level showed a t-test score <1, group effects vanished in GE (p=0.404) and C (p=0.345) (Supplemental Figure 1).

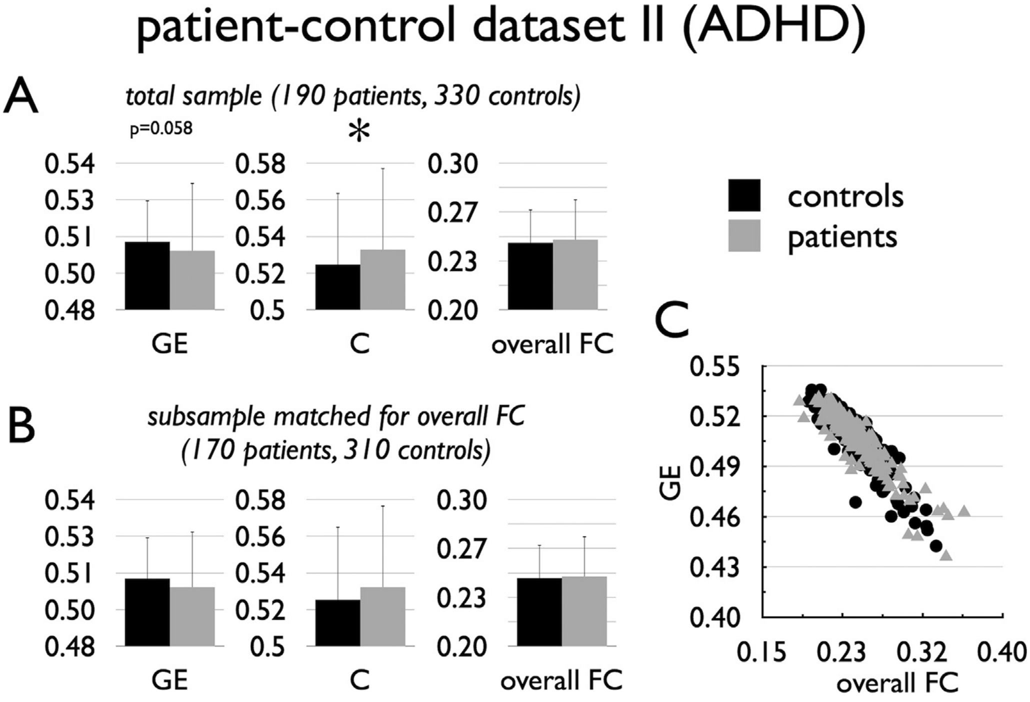

Fig. 3. Data for the exemplary ADHD dataset. **Panel A** shows that, opposite to the schizophrenia dataset, testing the entire sample resulted in ADHD patients revealed a slightly (but non-significant) higher overall FC in the patient population in comparison to the controls. This was accompanied by a borderline lower GE (p=0.058, ns) and higher C in the patient population. **Panel B** shows that matching the samples in terms of overall FC diminished findings in GE and C. **Panel C** shows the relationship between overall FC and GE across the entire sample. Control samples are plotted as black dots, patient samples as grey triangles.

Across the entire population overall FC correlated significantly to both GE (r=−0.93, $p < 0.001$) and C (r=0.92, $p < 0.001$). Similar effects were observed when testing other proportional thresholds.

ADHD dataset: an opposite effect

Similar observations were made for the examined ADHD dataset, but now the bias in GE and C was in the opposite direction. The functional networks of the ADHD patients did not show a significantly higher overall FC as compared to the functional networks of the healthy controls (patients, on average 1% higher FC, p=0.30 ns). We did however observe the proportionally thresholded networks of the ADHD population to show a trend-level effect of lower GE (proportional threshold: 15%, p=0.058) and higher C (p=0.0291, Fig. 3) as compared to the healthy controls. Despite FC not being statistically different across groups, the differences in GE and C could still be influenced by individual variation in FC. Indeed, across the entire population overall FC significantly correlated to both GE (r=−0.90, $p < 0.001$) and C (r=0.89, $p < 0.001$).

Autism EEG dataset

We also examined the same effects in functional connectivity networks derived from EEG recordings. Overall FC was found to be lower in patients as compared to controls (p=0.0182), but proportional thresholding of functional networks did not reveal a significant difference in GE (proportional threshold 25%, p=0.19 ns; proportional threshold 35%, p=0.19 ns) nor in C (p=0.18 ns) as compared to controls. Results were thus less pronounced than in the fMRI examples and mostly present at the higher proportional thresholds. This effect might be due to the small sample size. As in the fMRI experiments, GE and overall FC were still significantly correlated (r=−0.60, p=0.0010), which suggested an effect of overall FC on graph metrics. Indeed, testing small subsamples (here m=6) by ordering all patients and controls and examining subsets of the lowest FC patients and highest FC controls revealed a strong association between the ΔFC between subsamples and ΔGE (r=−0.88, p=0.0181). Alternatively, selecting 10,000 times two random subsamples of equal the size of the patient and control group from the total dataset again revealed a significant correlation between △FC and △GE (r=−0.60, $p < 0.001$).

Experiment summary

Findings show overall FC in combination with proportional thresholding to have pronounced effects on between-group comparison of graph organizational metrics. Across four different patient-control datasets GE and C showed a general relationship with overall FC, with data points with high FC showing low GE and high C. Across the disease datasets, we replicated commonly reported differences in GE and C, with between-group differences diminished when removing the most extreme cases of high/low FC from the patient and control populations. Furthermore, reported between-group differences in graph metrics GE and C were found to gradually decrease when testing subgroups of patients and controls with gradually matching levels of FC.

Experiment 2: HCP data

One possible argument for the relationship between overall FC and network metrics could be that changes in overall FC and changes in network topology are both pathological effects in the patient population, occurring simultaneously and in parallel, but not directly influencing each other. To further show that this effect is also present in a healthy population (thus with no neurological or psychiatric disease pathology) we performed a second experiment in which we examined the same phenomenon in healthy controls of the HCP dataset. After reconstruction and quality control of the FC matrices (466 HCP datasets remained, Q3 release, see methods), datasets were ordered according to overall FC and a subsample of the n=100 top lowest FC and the subsample of n=100 highest FC subjects were selected for

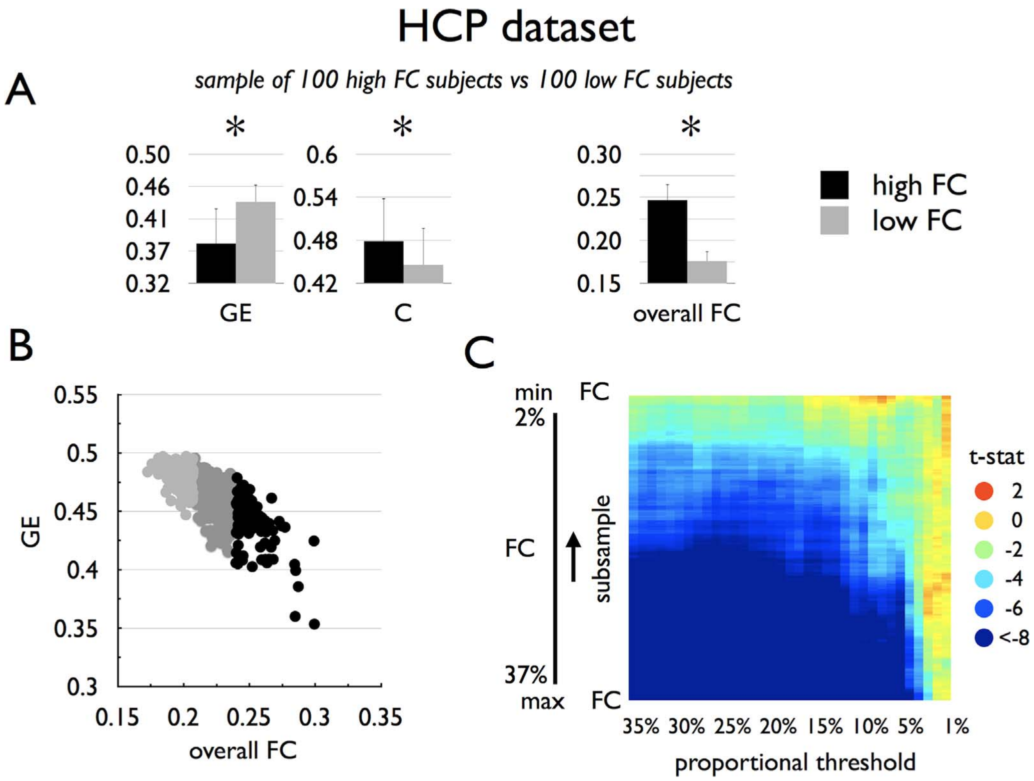

Fig. 4. HCP data. Figure shows the results for experiment 2, examining the effect of proportional thresholding on HCP healthy control data. **Panel A** shows the levels of GE and C of the subsample of the top n=50 highest (black) and top n=50 lowest overall FC subjects out of the entire HCP subset. Right panel illustrates the difference in overall FC between the two groups. **Panel B** shows the direct association between overall FC of the matrix and GE computed on the extracted binary proportionally thresholded functional graph (proportional threshold of 15%), clearly indicating that GE is dependent on overall FC. **Panel C** illustrates the findings in GE as computed by testing subsamples of maximal and minimal differences in FC. Panel shows the results of the same analysis as shown in Fig. 2 on the schizophrenia data. All HCP subjects were ordered according to overall FC and subsamples of m=50 were selected. The first t-test included the top [1,2,.,50] vs last [417,425,.,466] sample (with maximal difference in overall FC between subsamples), the second subsample the top [2,3,.,51] vs [416, 424,.,465], etcetera, until the subsample of minimal difference in overall FC was selected (i.e the two subsamples in the middle of the distribution). Panel shows t-statistic scores in GE computed across a range of proportional thresholds (35–1%), with negative values indicating higher GE in the low FC population versus the high FC population. △GE is given as the t-statistic score in GE between the selected subsamples of HCP subjects. △FC is computed as the percentage of difference in overall FC between the two groups, computed as (A - B) / A x 100%, with A the high FC and B the low FC group. Low FC subjects are plotted as light grey dots, high FC subjects as black dots, and the rest as grey dots.

further examination (selecting the top n=50 or top n=200 revealed similar results). FC matrices were proportionally thresholded, after which binary graph metrics GE and C were computed. First, we examined effects using a proportional threshold of 15% and compared derived graph metrics GE and C across the top n=100 lowest and top n=100 highest FC subjects. We observed the same effects as seen in the patient - control comparisons of experiment 1, namely a significantly higher GE ($p < 0.001$, Fig. 4A) and significantly lower C ($p < 0.001$, Fig. 4A) in the group of low FC subjects as compared to the group of high FC subjects (Fig. 4). Overall FC was (by construction) different between the two groups (22% higher in the high overall FC group, $p < 0.001$).

Examination of the entire set of HCP subjects revealed the same effect. First, across the entire HCP dataset GE ($r=-0.73$, $p < 0.001$, Fig. 2) and C ($r=0.62$, $p < 0.001$) significantly correlated to overall FC. Second, selecting opposite groups of m=50 subjects out of the lowest and highest FC scoring subjects (i.e. comparing groups of lowest and highest [1,2,.,m], [2,3,.,m+1], etcetera, see experiment 1 and Fig. 2A) showed that differences in GE and C go hand in hand with underlying group differences in overall FC, with between-group differences in graph metrics being lower (and eventually disappearing) when subgroups with smaller differences in FC are tested (Fig. 4C).

We continued by examining differences in graph metrics across the entire HCP dataset. Similar as in the schizophrenia dataset, we randomly selected two groups of each m=100 subjects from the HCP dataset and computed △FC, △GE and △C as the differences in respectively overall FC, GE and C between the two selected groups. Across 10,000 iterations, △GE ($r=-0.73$, $p < 0.001$) showed a strong correlation with △FC. △C showed a similar △FC dependency ($r=0.62$, $p < 0.001$).

HCP within subject variation

We continued by examining the *within-subject* matrices to show that the effect is potentially not due to biological individual variation in overall FC. For each HCP subject, we took the low and high FC matrix (obtained by splitting the time-series in half and computing for each part the overall FC, see methods). Both matrices were proportionally thresholded (exemplary proportional threshold 15%), and graph metrics GE and C were computed for each of the two parts. Testing differences in graph metrics between the low FC and high FC parts revealed significant differences in graph metrics between the two parts, with the proportionally thresholded matrices based on the low FC matrices showing higher GE ($p < 0.001$) and lower C ($p < 0.001$). Examination of dynamical networks revealed similar findings, with GE and C across runs related to overall FC (data shown Supplemental Materials, page 15, dynamical networks).

Experiment summary

The characteristic effects in GE and C metrics seen in patient-control datasets were also observed in subsets of healthy subjects of the HCP dataset that were solely selected on the basis of whether they showed low or high overall FC. Testing sub-groups of HCP subjects with diminishing group differences in overall FC showed a similar diminishing group effect in GE and C as seen in experiment 1. Furthermore, selecting and testing overall FC, GE and C between randomly drawn subsets of HCP data showed a strong relationship between-group differences in FC and between-group differences in GE and C. The effect of overall FC on graph metrics was not only present between selected groups of subjects, but also present within the data of single subjects, as shown by testing network metrics between proportionally threshold graphs derived from the first and second half of the individual fMRI time-series.

Experiment 3: Edge prevalence

How can differences in functional connectivity strength result in differences in graph theoretical metrics when they are computed on binary functional networks? We hypothesize this effect to be the result of lower FC connections to have (on average) a higher probability of being spurious, and therefore to result in the inclusion of more noisy and potentially false-positive edges in the final binary graph. To test this hypothesis, we examined whether the edges in the functional networks derived from lower FC subjects in the HCP data would show a lower edge prevalence, expressing the number of times a network edge is observed across the included population and a metric indicative of potentially less reliably measured network edges (de Reus and van den Heuvel, 2013; Roberts et al., 2016). Across the group of HCP subjects we determined for each observed edge in the network the level of edge prevalence, counting the number of times a binary edge was present across the total examined population. Next, for each individual dataset, we selected the top 100 lowest (as we hypothesized the most varying) and top 100 highest FC connections (hypothesized as the most stable and thus most group prevalent) of the proportionally thresholded graph. From these two groups of edges, per individual dataset, we computed the average group prevalence by taking the mean prevalence of the selected edges, a metric indicative of how often the lowest FC and highest FC connections of a subject's dataset were found to be present across the total group of subjects. Within the HCP dataset, prevalence of the class of low FC connections revealed a positive correlation to overall FC ($r=0.16$, $p < 0.001$), suggesting that networks based on lower overall FC show on average less reliably measured edges across the HCP group. As hypothesized, prevalence of edges based on high FC correlations did not reveal this effect ($p > 0.05$).

Experiment summary

Further testing HCP data shows that network edges resulting from proportional thresholding in low FC datasets are on average less frequently found across the total group of datasets and thus potentially more variable as compared to edges in the functional networks of high FC subjects.

Experiment 4: potential strategies for correction for overall FC

In a fourth experiment, we once again examined the patient and HCP dataset and now aimed to examine potential counter measures to correct or compensate for the effect of overall FC across groups. First, in the schizophrenia dataset, overall FC was regressed out of the graph metrics across the total population (i.e. the total group of both controls and patients) and the corrected graph metrics (i.e. the residuals) were tested between groups. Including overall FC as a covariate diminished between-group effects (threshold 15%, GE: $p=0.822$, C: $p=0.974$). Similarly, regressing out overall FC also diminished between-group effects in the autism (proportional threshold 15%, fMRI: GE: $p=0.0732$, C: $p=0.746$; EEG: GE: $p=0.312$, C: $p=0.877$) and ADHD datasets (exemplary threshold 15%, GE: $p=0.879$). A marginal effect in C remained in the ADHD dataset ($p=0.0354$), which may suggest a potential remaining group-effect in global clustering after correction for overall FC. Regressing out overall FC from GE and C in the entire HCP dataset and testing the n=100 lowest versus n=100 highest FC subjects no longer showed differences in graph metrics (exemplary threshold 15%, GE: $p=0.6015$, C: $p=0.4722$).

As a second alternative we explored the use of permutation testing. In a typical permutation test a null-distribution of between-group differences is obtained by randomly drawing subsamples from the total population, to examine which effect sizes occur irrespective of patient/control status. In addition to a normal permutation approach (in which only group assignment is randomized), here we also took into account the observed difference in overall FC between the two groups. First, group assignment was randomized by randomly drawing two samples of the size of the patient and control population (e.g. n=48 and n=44 in the schizophrenia dataset) from the total set of participants. Next, two random subjects were drawn, one from each sample, and subjects were swapped between groups until the difference in overall FC between the two samples reached the level of the original between-group difference in FC (i.e. 4.8% in the schizophrenia dataset). The patient and control distribution was kept fixed to that of the initially randomly drawn samples, by swapping only control subjects for control subjects and patient subjects for patient subjects between groups. Once the two pseudo-randomly drawn samples were established, matrices were proportionally thresholded, graph metrics were computed and differences in graph metrics GE and C between the two groups were obtained. We performed this procedure for 10,000 permutations, resulting in a null-distribution of expected effects under the null-hypothesis of A) no effect of group assignment and B) a difference in overall FC equal to the observed difference between the patient and control group. Next, similar as in a typical permutation testing approach, the obtained null-distribution was used to assign a p-value to the originally observed effects in GE, by computing the proportion of the null-distribution that exceeded the originally observed group difference in GE.

In the schizophrenia dataset, overlapping with the results of the t-tests, normal permutation testing revealed a difference in GE ($p=0.0152$) and C ($p=0.0259$). However, using the alternative null-distribution in which we controlled for differences in FC, the effects in GE and C were no longer found to be significant (GE:$p= 0.4605$, C:$p=0.344$), indicating that the h0 hypothesis of GE and C being equal in the patient population could no longer be rejected. Similar statistical effects were observed in the autism dataset (GE:$p=0.241$, C:$p=0.414$) with a permutation test controlling for differences in overall FC showing no longer a significant difference in graph metrics. In the ADHD dataset, effects in GE were diminished (GE:$p=0.0890$), but some effects in C remained ($p=0.0218$), which again may suggest a remaining group effect in network clustering in ADHD patients after correcting for group differences in overall FC.

We note that adding a constraint to a permutation test reduces the amount of possible permutations. To quantify this effect, we computed for both the random and pseudo-random permutation test the level of overlap of each permutation with all other permutations, counting the number of samples similarly included across the 10,000 iterations. In the schizophrenia dataset, in the random permutation condition the average overlap was on average 21 and 25 (as expected from respectively the size of the control and patient group), and in the pseudo-random condition the empirical overlap was on average 23 (std:2.3) and 27 (std:2.3).

Experiment summary

We tested two potential strategies for correction of overall FC on between-group comparison of graph metrics. Findings showed that taking overall FC as a covariate may compensate for the influence of FC on graph metrics (see also Discussion on potential drawbacks of this method). Second, we examined the use of a permutation based test in which group differences in overall FC were incorporated in the nullcondition and thus taken into account when testing group differences in derived graph metrics GE and C.

Discussion

In this study we examined the effects of proportional thresholding on the construction of functional connectivity graphs and the computation of graph theoretical metrics. The main conclusion is that proportional thresholding should be used with care when there are differences in total functional connectivity between the examined groups, because a minimal difference in overall FC may introduce potential between-group differences in network metrics. We present a few simple recommendations to examine, report and potentially correct for the effect of differences in total functional connectivity in disease connectome studies.

Our study findings confirm the intuitive notion that the inclusion of lower correlations as functional edges will lead to the inclusion of more noisy and thereby potentially more random connections into the reconstruction of a functional network, an effect reflected in the evaluation of graph metrics. We conclude from our presented findings that low FC connections tend to have a higher probability of being spurious, with their inclusion into network reconstructions leading to the inclusion of more random connections as compared to network reconstructions based on high FC. In turn, the inclusion of these more random connections can give the topology of the reconstructed graph a more random character, most notably reflected in higher global network efficiency and lower network clustering.

Zalesky and colleagues (2016) examined the effects of false-positive (FP) connections and false-negatives (FN) edges to network analysis of healthy human and animal connectomes and showed empirical and theoretical evidence of the inclusion of FPs to be more detrimental to the computation of network metrics as compared to the exclusion of FN connections (Zalesky et al., 2016). Specificity was reported to be at least twice as important as sensitivity with respect to computation and evaluation of graph theoretical metrics of reconstructed brain networks, advocating a 2:1 ratio of FP and FN inclusions in anatomical brain networks. Our current functional results are clearly in line with these findings and extend these findings by showing that proportional thresholding of matrices may further inflate the effect, with profound consequences for between-group comparisons.

We have no intention of arguing that alterations in functional brain connectivity as commonly reported in brain disorders are in any way false effects or effects driven by artifacts. We thus intentionally refrain from naming specific studies that used proportional thresholding in examining for example schizophrenia, ADHD and/or autism.[2] In contrast, we emphasize that disturbances in neuronal activity, accompanying changes in BOLD fluctuations or EEG/MEG recordings, as well as disturbances in region-to-region functional communication are all likely to include key factors in many brain disorders, resulting in the correct observation of changes in interregional functional connectivity. Our main point here is that subsequent proportional thresholding of such functional connectivity matrices in disease studies may translate into or influence differences in graph metrics as measured between groups, effects that may rather be the result of the inclusion of (even a few) spurious connections in one of the two groups. These reported between-group differences in network organization may be driven by, or at least cannot fully be disentangled from, underlying changes in overall functional connectivity. The goal of our study is thus not to name specific studies that used this approach, but rather to inform about the potential consequence of using proportional thresholding in context of between-group differences in overall FC. This to create awareness for future studies to test and put effort in correcting for overall connectivity where needed and where possible.

Binary versus functionally weighted networks

In this study we focused on the examination of binary functional networks with the presence and absence of functional edges forming the main topic of investigation. We focused on binary networks to illustrate the disruptive effect of low FC connections on the topological organization of networks. For this, we thus excluded any additional influence of differences in edge weights between subject groups. Including weights on the edges would potentially introduce a second effect, with between group differences in overall edge weight directly translating into graph metrics. For example, functional networks of low FC subjects would include (on average) lower weighted network edges, resulting now in lower GE values. To examine to what extent potential between-group effects in graph metrics go beyond simple differences in edge weights, studies most often compare normalized GE and C across groups (i.e. taking the ratio of GE and C with their counterparts as computed in comparable random networks). Normalized weighted metrics are often argued in literature to counteract potential between-group differences in network strength, as the total sum of weights in the network of interest and the randomized networks are equal and thus cancel out. A second suggested strength of this type of approach is that connections with a higher strength make a stronger contribution in the computation of graph metrics, with lower weight connections (here argued to be less reliable and thus more random) having less impact (but see also (Drakesmith et al., 2015; Ginestet et al., 2014; Ginestet et al., 2011) for discussion). In the case of evaluating weighted graphs, false positive connections based on lower correlations may thus inherently have a less disruptive impact on network topology. Nevertheless, with many of the graph metrics (and in particular global efficiency and clustering) still dependent on underlying binary patterns, the disruptive effect of including more random edges in low FC networks may -to some extent- remain. In the Supplemental Materials we examined the effect of overall FC on (normalized) weighted graph metrics in a patient-control setting (Supplemental Materials, page 5–6, normalized weighted networks). We report on attenuated, but potentially remaining effects of overall FC on normalized weighted global efficiency and thus between-group evaluation of global network organization. From our simple post-hoc analyses the influence of overall FC on normalized weighted clustering appeared to be less severe, suggesting that weighting and normalization may counteract the influence of overall FC on local graph organization. Future work specifically focused on the use and the development of new (normalized) weighted metrics optimized for the influence of overall FC on graph organization is clearly of great importance to the field.

Anatomical networks

Our main topic of study here is functional networks. Proportional thresholding is less commonly used in the reconstruction and analysis of structural graphs (as the matrix is most often already sparse), but the inclusion of false-positive edges may –in principle– in a similar way influence anatomical network reconstruction and as such introduce influence between-group differences in graph metrics when comparing groups. For example, overall lower number of reconstructed streamlines and/or lower levels of fractional anisotropy of edges in the patient and/or control population potentially lead to the inclusion of more false-positive edges in proportionally thresholded anatomical graphs and as such influence the computation of graph metrics (Zalesky et al., 2016). The examination of structural networks is out of the scope of this study, but future studies examining this effect in more detail in anatomical networks and compare across DWI reconstruction strategies to show or to rule out the influence of overall connectivity strength

[2] We examined schizophrenia, ADHD and autism as exemplary datasets. We argue that overall FC has an effect on group comparison of graph metrics in functional connectome studies in general, thus also including studies that examine neurological conditions.

on the evaluation and between-group comparison of graph metrics would be of interest.

Alternative methods for thresholding

Studies have suggested useful alternative approaches to avoid or reduce the effect of thresholding in the examination of functional networks. One alternative approach includes the evaluation of graph metrics suited for fully weighted networks, avoiding the need for any type of thresholding in the first place. Although the main topic of investigation in this study is the evaluation of proportional thresholding on functional networks and not the evaluation of non-thresholded approaches, we do argue that potentially the same effect of overall FC might influence the computation of such graph metrics. Here too, the inclusion of functional edges based on lower correlations could lead to the inclusion of less accurate estimations of connections and hence potentially more random network connections. Similar to the use of weighted networks (see discussion above), the advantage of a weighted approach is that such edges will have lower weights and are thus argued in literature to have less impact on overall network organization. However, some of the effect of a more randomly organized network may still be present. Indeed, a post-hoc analysis of graph metrics on fully weighted functional connectivity matrices of the patient-control and HCP dataset again revealed a –but less severe– remaining effect of overall FC on graph metrics, with differences in overall FC between groups going hand in hand with between-group differences in network metrics (see Supplemental Materials, page 5–6, weighted normalized metrics).

A second class of proposed alternative strategies does not aim to directly avoid any use of thresholding, but rather aims to avoid the selection of one specific threshold. Examples of these approaches include the computation of a *minimal spanning tree (MST)* or the related *local k-nearest neighbor graph* (k-NNG) of functional matrices (Alexander-Bloch et al., 2010; Jalili, 2016; Tewarie et al., 2015) and approaches that work by integrating effects across a wide range of thresholds, such as so-called Area Under the Curve (AUC) methods (Ginestet et al., 2011; Hosseini et al., 2012) (see also (Langer et al., 2013) for discussion) and multi-threshold permutation correction (MTPC) methods (Drakesmith et al., 2015). The MST describes the tree of minimal number of strongest edges needed to keep the network connected. Related to this, in a *local* k-NNG a local threshold is applied to the functional matrix, selecting the *k* strongest edges of each node, often with the MST used as a starting point to ensure global connectedness of the resulting graph. MST and k-NNG approaches have been successfully applied in several functional connectivity studies and argued to avoid methodological biases for the selection of an arbitrary threshold level, having the strong advantage of ensuring equal network density levels across groups (Tewarie et al., 2015). In mathematical terms, the selection of the MST could be seen as one of the strictest levels of proportional thresholding, namely the application of a (n-1)/(n x (n-1)) = 1/n threshold with the additional selection rule that the network has to remain connected. As such, we could argue that the MST approach may also be subject to the same issues of proportional thresholding as described in this paper, albeit substantially lower as the MST is optimized for including edges corresponding to strong (and thus more reliable) correlations. Similarly, one could argue that k-NGG thresholding may be influenced by variation in overall FC (Alexander-Bloch et al., 2010). Moreover, since the k-nearest neighbor approach mandates a minimal number of edges per node, this may result in the inclusion of more weaker edges as compared to the application of a global proportional threshold, something that may exacerbate the effect of overall FC on graph metrics. Indeed, as expected based on the high-threshold effects as shown in Fig. 2, post-hoc analysis in the HCP data indeed revealed that the effect of overall FC is much less severe on MST graphs, but that adding additional locally thresholded edges in k-NGG graphs may again result in inflated between-group comparisons of graph metrics (data shown in Supplemental Materials, page 8–10, MST and k-NGG).

In contrast, MTPC and AUC methods avoid the selection of a single threshold by alternatively integrating effects across multiple thresholds. With our findings suggesting that overall FC influences between-group comparison of graph metrics across almost the entire range of tested thresholds (see Fig. 1c), one could argue that methods that integrate effects across threshold levels are potentially as sensitive to the influence of overall FC as methods that use one single threshold. Indeed, testing group-differences between high and low FC subjects in the HCP data similarly showed significant differences in both GE (tested thresholds 35% to 1% with steps of 1%, $p < 0.001$) and C ($p < 0.001$) when using proportional thresholding in combination with AUC and MTPC (see Supplemental Materials, page 10–13, AUC and MTPC). Moreover, by merging effects across thresholds, AUC methods may potentially further inflate the described effect of overall FC, with now even smaller group differences in overall FC still resulting in significant between-group differences in graph metrics. Indeed, testing GE and C across sub-groups of HCP subjects with declining levels of between-group difference in overall FC showed more pronounced effects when using AUC as compared to the use of a single proportional threshold (see Supplemental Results, page 10–13, AUC and MTPC).

Strategies to compensate and control for the effect of overall FC on graph metrics

We discuss two potential strategies to correct for the effect of differences in overall FC when using proportional thresholding. First, in case of a difference in overall FC across participant groups, we show that the inclusion of the overall FC as a covariate could be used to compensate for the effect of proportional thresholding on the computation of graph metrics during statistical evaluation. However, we recognize that this involves a rather strict correction, as in the severe case of all patients showing lower FC as compared to the controls this could lead to the removal of a large portion of potentially true differences in network organization between patients and controls. As a potential second strategy, we advise patient-control network studies to include post-hoc control analyses in which one examines subsamples matched on FC levels, for example by removing some of the most severe cases on both sides of the spectrum and show that the same between-group effects in graph organization are also present when samples are matched on FC. In case there is evidence of the highest or lowest FC subjects to include the most severely ill patients, permutation approaches with null-distributions matched on FC could be used as a potential alternative.

We note that both of the above mentioned strategies do not really correct for the potential selection of more random connections in the construction of the functional network, but rather aim to control for the effect during statistical evaluation. We strongly encourage the field to design methods that take care of this potential bias earlier in the analysis, preferably already during network construction (e.g. (Ramsey et al., 2011; Ramsey et al., 2010; Smith et al., 2011)) and/or during the computation or normalization of the graph metrics. Examples of these might include better techniques to distinguish between true and false positives, other types of evaluation null-models, and/or the inclusion of more subtle covariates in the statistical evaluation. Until such normalization and/or correction approaches are proposed we advise that patient-control connectome studies include some of the discussed post-hoc analyses, just to verify that one's reported case-control differences in network organization are not simply the result of underlying differences in overall FC, but reflect true differences in network configuration. Moreover, individual variation in FC could cloud true between-group differences in network organization, which further advocates controlling for the effects of overall FC in connectome studies.

Conclusion and recommendations

We show the influence of overall FC on proportional thresholding of functional brain networks and the subsequent computation and comparison of graph metrics across groups. Fixed thresholding of matrices and resulting differences in network density has rightfully been suggested to have an important effect on the computation of graph metrics (van Wijk et al., 2010) and thus to be less suitable for the examination of network organization in patient-control studies. Our current findings now similarly advise against the use of the proposed alternative approach of using density matched networks in situations where a clear difference exists in underlying total functional connectivity strength between groups. We make two recommendations for future patient-control functional connectome studies. First, in a between-group functional connectome study, we advise authors to examine, statistically test and to report overall FC between tested groups. Second, in case of the suspicion of a potential difference in overall FC between groups we advise that these differences are taken into account when graph metrics are statistically tested between patients and controls, for example by including overall FC as a covariate and/or by including post-hoc control analyses in which one verifies that reported between-group differences in graph metrics remain when testing subsamples matched for overall FC. We hope these recommendations will be of use for future functional disease connectome studies.

Acknowledgements

MPvdHeuvel was funded by an ALW open (ALWOP.179) and VIDI (452-16-015) grant from the Netherlands Organization for Scientific Research (NWO) and a Fellowship of MQ. TY was supported by grants from Singapore MOE Tier 2 (MOE2014-T2-2-016), NUS Strategic Research (DPRT/944/09/14), NUS SOM, Aspiration Fund (R185000271720), Singapore NMRC (CBRG14nov007) and NUSYIA. We thank Chantal Kemner for kindly sharing the EEG matrices and Rachel Brouwer and Rene Mandl for inspiring discussions.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuroimage.2017.02.005.

References

Achard, S., Bullmore, E., 2007. Efficiency and cost of economical brain functional networks. PLoS Comput. Biol. 3, e17.

Alexander-Bloch, A.F., Gogtay, N., Meunier, D., Birn, R., Clasen, L., Lalonde, F., Lenroot, R., Giedd, J., Bullmore, E.T., 2010. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. Front Syst. Neurosci. 4, 147.

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

Bassett, D.S., Bullmore, E.T., Meyer-Lindenberg, A., Apud, J.A., Weinberger, D.R., Coppola, R., 2009. Cognitive fitness of cost-efficient brain functional networks. Proc. Natl. Acad. Sci. USA 106, 11747–11752.

Boersma, M., Smit, D.J., de Bie, H.M., Van Baal, G.C., Boomsma, D.I., de Geus, E.J., Delemarre-van de Waal, H.A., Stam, C.J., 2010. Network analysis of resting state EEG in the developing young brain: structure comes with maturation. Hum. Brain Mapp..

Braun, U., Plichta, M.M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., Mier, D., Mohnke, S., Heinz, A., Erk, S., Walter, H., Seiferth, N., Kirsch, P., Meyer-Lindenberg, A., 2012. Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. Neuroimage 59, 1404–1412.

Brown, J.A., Rudie, J.D., Bandrowski, A., Van Horn, J.D., Bookheimer, S.Y., 2012. The UCLA multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis. Front. Neuroinform 6, 28.

Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10, 186–198.

Damoiseaux, J.S., Rombouts, S.A., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks across healthy subjects. Proc. Natl. Acad. Sci. USA 103, 13848–13853.

de Reus, M.A., van den Heuvel, M.P., 2013. Estimating false positives and negatives in brain networks. Neuroimage 70, 402–409.

Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980.

Drakesmith, M., Caeyenberghs, K., Dutt, A., Lewis, G., David, A.S., Jones, D.K., 2015. Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data. Neuroimage 118, 313–333.

Filippi, M., van den Heuvel, M.P., Fornito, A., He, Y., Hulshoff Pol, H.E., Agosta, F., Comi, G., Rocca, M.A., 2013. Assessment of system dysfunction in the brain through MRI-based connectomics. Lancet Neurol. 12, 1189–1199.

Fornito, A., Bullmore, E.T., 2012. Connectomic intermediate phenotypes for psychiatric disorders. Front. Psychiatry 3, 32.

Fornito, A., Zalesky, A., Pantelis, C., Bullmore, E.T., 2013. Schizophrenia, neuroimaging and connectomics. Neuroimage 62, 2296–2314.

Fox, M.D., Raichle, M.E., 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. 8, 700–711.

Gallos, L.K., Makse, H.A., Sigman, M., 2012. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. Proc. Natl. Acad. Sci. USA 109, 2825–2830.

Garrison, K.A., Scheinost, D., Finn, E.S., Shen, X., Constable, R.T., 2015. The (in)stability of functional brain network measures across thresholds. Neuroimage 118, 651–661.

Ginestet, C.E., Fournel, A.P., Simmons, A., 2014. Statistical network analysis for functional MRI: summary networks and group comparisons. Front. Comput. Neurosci. 8, 51.

Ginestet, C.E., Nichols, T.E., Bullmore, E.T., Simmons, A., 2011. Brain network analysis: separating cost from topology using cost-integration. PLoS One 6, e21570.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., Consortium, W.U.-M.H., 2013. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80, 105–124.

Goulas, A., Schaefer, A., Margulies, D.S., 2015. The strength of weak connections in the macaque cortico-cortical network. Brain Struct. Funct. 220, 2939–2951.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. PLoS Biol. 6, e159.

Hosseini, S.M., Hoeft, F., Kesler, S.R., 2012. GAT: a graph-theoretical analysis toolbox for analyzing between-group differences in large-scale structural and functional brain networks. PLoS One 7, e40709.

Jalili, M., 2016. Functional brain networks: does the choice of dependency estimator and binarization method matter? Sci. Rep..

Langer, N., Pedroni, A., Jancke, L., 2013. The problem of thresholding in small-world network analysis. PLoS One 8, e53199.

Nicols, T.E., Das, S., Eickhoff, S.B., Evans, A.C., Glastard, T.G., Hanke, M., Kriegeskorte, N., Milham, M.P., Poldrack, R.A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D.C., White, T., Yeo, B.T.T., 2016. Best practices in data analysis and sharing in neuroimaging using MRI. bioRxiv.

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142–2154.

Ramsey, J.D., Hanson, C., Glymour, C., 2011. Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study. Neuroimage 58, 10.

Ramsey, J.D., Hanson, S.J., Hanson, C., Halchenko, Y.O., Poldrack, R.A., Glymour, C., 2010. Six problems for causal inference from fMRI. Neuroimage 49, 1545–1558.

Roberts, J.A., Perry, A., Roberts, G., Mitchell, P.B., Breakspear, M., 2016. Consistency-based thresholding of the human connectome. Neuroimage..

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 2, 10.

Santarnecchi, E., Galli, G., Polizzotto, N.R., Rossi, A., Rossi, S., 2014. Efficiency of weak brain connections support general cognitive functioning. Hum. Brain Mapp. 35, 4566–4582.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. Proc. Natl. Acad. Sci. USA 106, 13040–13045.

Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for FMRI. Neuroimage 54, 875–891.

Stam, C.J., Nolte, G., Daffertshofer, A., 2007. Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Hum. Brain Mapp. 28, 1178–1193.

Stam, C.J., Reijneveld, J.C., 2007. Graph theoretical analysis of complex networks in the brain. Nonlinear Biomed. Phys. 1, 3.

Tewarie, P., van Dellen, E., Hillebrand, A., Stam, C.J., 2015. The minimum spanning tree: an unbiased method for brain network analysis. Neuroimage 104, 177–188.

van den Heuvel, M.P., Fornito, A., 2014. Brain networks in schizophrenia. Neuropsychol. Rev..

van den Heuvel, M.P., Hulshoff Pol, H.E., 2010. Exploring the brain network: a review on resting-state fMRI functional connectivity. Eur. Neuropsychopharmacol. 20, 519–534.

van den Heuvel, M.P., Mandl, R.C., Stam, C.J., Kahn, R.S., Hulshoff Pol, H.E., 2010. Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. J. Neurosci. 30, 15915–15926.

van den Heuvel, M.P., Scholtens, L.H., Feldman Barrett, L., Hilgetag, C.C., de Reus, M.A., 2015. Bridging cytoarchitectonics and connectomics in human cerebral cortex. J

Neurosci. 35, 13943–13948.

van den Heuvel, M.P., Scholtens, L.H., Turk, E., Mantini, D., Vanduffel, W., Feldman Barrett, L., 2016. Multimodal analysis of cortical chemoarchitecture and macroscale fMRI resting-state functional connectivity. Hum. Brain Mapp. 37, 3103–3113.

van den Heuvel, M.P., Sporns, O., Collin, G., Scheewe, T., Mandl, R.C., Cahn, W., Goni, J., Hulshoff Pol, H.E., Kahn, R.S., 2013. Abnormal rich club organization and functional brain dynamics in schizophrenia. JAMA Psychiatry 70, 783–792.

Van den Heuvel, M.P., Stam, C.J., Boersma, M., Hulshoff Pol, H.E., 2008. Small-world and scale-free organization of voxel based resting-state functional connectivity in the human brain. Neuroimage 43, 11.

Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., Consortium, W.U.-M.H., 2012. The Human Connectome Project: a data acquisition perspective. Neuroimage 62, 2222–2231.

van Wijk, B.C., Stam, C.J., Daffertshofer, A., 2010. Comparing brain networks of different size and connectivity density using graph theory. PLoS One 5, e13701.

Watts, Duncan J., Strogatz, Steven H., 1998. Collective dynamics of ‘small-world’ networks. Nature 393.6684, 440–442.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L.L., van den Heuvel, M.P., Breakspear, M., 2016. Connectome sensitivity or specificity: which is more important? Neuroimage.