Loewe et al. BMC Neuroscience 2014, 15:78
http://www.biomedcentral.com/1471-2202/15/78

BMC Neuroscience

METHODOLOGY ARTICLE Open Access

Fast construction of voxel-level functional connectivity graphs

Kristian Loewe[1,2*], Marcus Grueschow[3], Christian M Stoppel[1,4], Rudolf Kruse[2] and Christian Borgelt[5]

Abstract

Background: Graph-based analysis of fMRI data has recently emerged as a promising approach to study brain networks. Based on the assessment of synchronous fMRI activity at separate brain sites, functional connectivity graphs are constructed and analyzed using graph-theoretical concepts. Most previous studies investigated region-level graphs, which are computationally inexpensive, but bring along the problem of choosing sensible regions and involve blurring of more detailed information. In contrast, voxel-level graphs provide the finest granularity attainable from the data, enabling analyses at superior spatial resolution. They are, however, associated with considerable computational demands, which can render high-resolution analyses infeasible. In response, many existing studies investigating functional connectivity at the voxel-level reduced the computational burden by sacrificing spatial resolution.

Methods: Here, a novel, time-efficient method for graph construction is presented that retains the original spatial resolution. Performance gains are instead achieved through data reduction in the temporal domain based on dichotomization of voxel time series combined with tetrachoric correlation estimation and efficient implementation.

Results: By comparison with graph construction based on Pearson's *r*, the technique used by the majority of previous studies, we find that the novel approach produces highly similar results an order of magnitude faster.

Conclusions: Its demonstrated performance makes the proposed approach a sensible and efficient alternative to customary practice. An open source software package containing the created programs is freely available for download.

Keywords: Functional connectivity, Graph theory, Tetrachoric correlation, Resting-state, fMRI

Background

The functioning of the human brain relies on the interplay and integration of numerous individual units in a complex network. Insights into its topology are thus essential to promote our understanding of the brain in general, as well as its maladaptive states associated with dysfunction and disease. An increasingly popular approach to the analysis of functional brain networks is based on the framework of graph theory [1-4]. A graph is a mathematical structure designed for modelling pairwise relationships, known as edges, between an assortment of objects, referred to as nodes. In applications to fMRI, the node set is defined as a collection of brain sites, and edges are established by measuring internodal functional connectivity [5] based on the regions' associated time series. The obtained functional connectivity graph, serving as a simple model of the brain's functional organization in a complex network, is subsequently examined drawing on a rich collection of graph-theoretical metrics that target various aspects of its topology [6]. Several studies indicate, for instance, that the brain's functional network conforms to a small-world architecture [1,2,7,8]. Beyond that, the usefulness of graph-based functional connectivity analyses has been demonstrated in applications to brain development and aging [9-11], gender differences [12], intellectual performance [13], and neurological disorders, such as Alzheimer's disease [14,15] or schizophrenia [16,17].

In most previous studies, functional connectivity graphs have been constructed at the level of regions [1,7,10,14,18], meaning that graphical nodes are defined

*Correspondence: kl@kristianloewe.com
[1]Department of Neurology, Experimental Neurology, Otto-von-Guericke Universität, Leipziger Str. 44, 39120 Magdeburg, Germany
[2]Department of Knowledge and Language Processing, Otto-von-Guericke Universität, Magdeburg, Germany
Full list of author information is available at the end of the article

BioMed Central

© 2014 Loewe et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

based on a parcellation of the brain into regions of interest (ROI), each consisting of several voxels. Due to the limited number of nodes, such analyses are computationally inexpensive and their results are comparatively easy to visualize and interpret. However, region-level nodes involve mixing fMRI time series from the incorporated voxels, thus obliterating more detailed spatial information [4,19]. ROI-based analyses are therefore highly dependent on the quality of the parcellation: If ROI boundaries and actual functional boundaries are inconsistent, the results can be erroneous [20]. Voxel-level analyses, in contrast, are not subject to these limitations, since the parcellation inherent to the original data is used for node definition [8,11,13,15,21-23]. Consequently, voxel-level graphs provide a finer model of the brain's functional network organization, since the original spatial resolution of the fMRI data is preserved [4,24].

Because of the large number of nodes, the construction and analysis of voxel-level graphs can involve considerable computational efforts. In response, the computational burden has often been reduced by sacrificing spatial resolution (using relatively large voxels to begin with or reslicing the data to a lower resolution) thus reducing the number of nodes in the graph [15,25-27]. While reduction of spatial resolution is undesirable in general (given that the main advantage of fMRI as compared to other methods such as EEG/MEG is its superior spatial resolution), it can even render a study infeasible, e.g., when investigating very small brain structures or different regions that lie in close proximity to each other. Efficient algorithms and implementations are therefore required in order to take full advantage of the data's original spatial resolution [24].

Here, we propose a novel approach aiming to increase the computational efficiency of voxel-level graph construction by combining time series dichotomization, tetrachoric correlation estimation, and efficient implementation, while retaining the full spatial resolution of the data. Comparison with conventional graph construction (as carried out in previous studies) shows that the new approach not only produces highly similar results, but also executes an order of magnitude faster.

Methods

This section consists of three parts. We begin with a short introduction to voxel-level functional connectivity graphs and explain their construction from fMRI data. In particular, it is established how time series dichotomization can be combined with tetrachoric correlation estimation to efficiently measure functional connectivity. The second part describes analyses comparing Pearson's r and the tetrachoric correlation coefficient r_t (1) as correlation estimators in the controllable environment of synthetic data, (2) as measures of functional connectivity in the context of graph construction from fMRI data, and (3) with respect to the similarity of graphs resulting from (2). The last part provides implementation details regarding the programs created for this study and assesses their computational performance.

Voxel-level graph construction

Formally, an undirected binary graph is defined as an ordered pair $G_B = (N, E)$, comprised of a set of nodes N and a set of pairwise internodal connections, or edges, E. Individual edges are unordered pairs $\{i, j\}$, where $i, j \in N$. G_B can be represented by a binary adjacency matrix $\mathbf{B}^{|N|\times|N|} = (b_{i,j})$, where $b_{i,j} \in \{0, 1\}$, $i, j \in N$, and $b_{i,j} = 1$ indicates that $\{i, j\} \in E$, i.e., that a connection between the two nodes i and j exists. In order to represent not only the presence or absence of connections but also their strength, a graph can be extended by assigning a weight to each edge.

In applications to fMRI data, graph-based analyses rely on the derivation of a graphical representation of the brain's functional network, which is then examined in terms of graph theory (Figure 1). Voxel-level functional connectivity graphs are constructed based on individual voxels as nodes, that is, the set of nodes N is a collection of voxels. In the literature, N is often defined as all in-brain voxels, or all gray matter (GM) voxels. Functional connectivity is estimated between all pairs of nodes based on their corresponding time series using a measure of association.

To construct a binary graph, edges are established by thresholding the functional connectivity estimates: Two nodes are connected by an edge if their functional connectivity exceeds a given threshold. Alternatively, one can take the pairwise functional connectivity matrix as a basis for a weighted graph, thus conserving the strength of individual functional connections between nodes [28]. For simplicity, and in a manner consistent with the majority of previous work on voxel-level functional connectivity graphs, we presently focus on binary graphs for our analysis.

Measuring functional connectivity

In most previous studies investigating voxel-level functional connectivity graphs, internodal functional connectivity is measured using Pearson's sample correlation coefficient r [2,8,13,15,21-27,29-31]. When using Pearson correlation as a measure of functional connectivity, it seems sensible to assume bivariate normality with respect to the distribution of pairwise observations arising from each pair of voxel time series. This is because Pearson correlation may be a poor measure of association if the data are not normally distributed [32]. Encouragingly, in a recent study employing region-level graphs, data for the most part appeared to meet the assumption of bivariate normality [33].

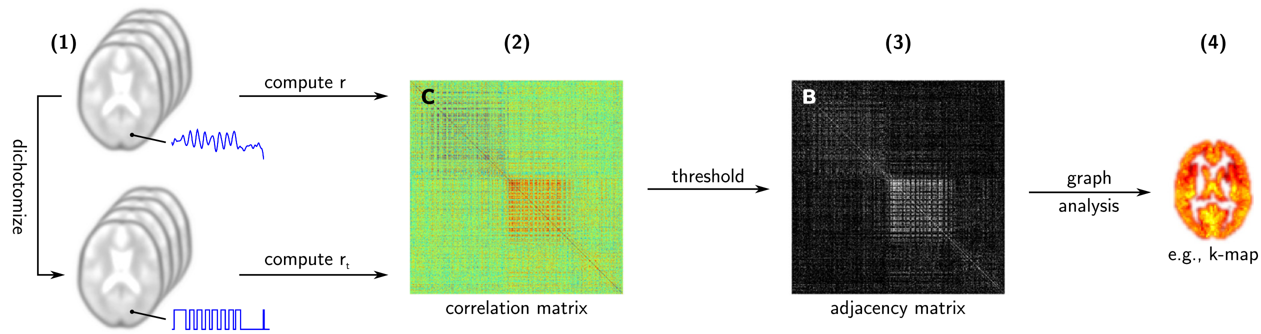

Figure 1 Construction and analysis of voxel-level functional connectivity graphs. Starting with the preprocessed fMRI data, all gray matter voxels are defined as graphical nodes **(1)**. Using their associated time series, pairwise internodal functional connectivity is measured in terms of linear correlation. Typically, this is done using Pearson's r. Alternatively, one can derive binary time series via median-based dichotomization and employ tetrachoric correlation estimation (r_t). In both cases, the result is a correlation matrix **(2)** representing the pairwise functional connectivity between nodes. A binary undirected graph, represented by a binary adjacency matrix **(3)**, is obtained via thresholding. Based on the adjacency matrix, graph-theoretical metrics, such as the node degree k, are computed **(4)**.

Assuming bivariate normality between pairs of voxels, an alternative correlation estimator, the tetrachoric correlation coefficient r_t [34], can be used instead of r. Given two dichotomous variables x_d and y_d, r_t estimates the correlation of the latent continuous-valued variables x_c and y_c associated with x_d and y_d, under the assumption that x_c and y_c follow a bivariate normal distribution. Thus, if we dichotomize, i.e., binarize, the voxel time series data, r_t can be used to estimate the pairwise correlation of the original continuous-valued time series from the dichotomized ones.

Consider two voxels, v and w, and their corresponding time series, $\boldsymbol{s}_v$ and $\boldsymbol{s}_w$. Using the medians of $\boldsymbol{s}_v$ and $\boldsymbol{s}_w$, i.e., $\tilde{\boldsymbol{s}}_v$ and $\tilde{\boldsymbol{s}}_w$, as dichotomization thresholds, we obtain the binary time series $\boldsymbol{d}_v$ and $\boldsymbol{d}_w$. Formally, $d_{v,k} = 1$ if the signal intensity value $s_{v,k}$ amounts at least to $\tilde{\boldsymbol{s}}_v$ and $d_{v,k} = 0$ otherwise, where $k \in \{1, \ldots, T\}$ and T is the number of acquired fMRI volumes [35]. See Section S.1 for details.

By virtue of $\tilde{\boldsymbol{s}}_v$ and $\tilde{\boldsymbol{s}}_w$, the pairs $(s_{v,k}, s_{w,k})$ are divided into four partitions corresponding to four quadrants in the x-y-plane of a Cartesian coordinate system (Figure 2A, bottom). Thus, by counting the number of points falling into each quadrant, a pair of voxels gives rise to a 2×2 contingency table, the (relative) frequencies of which can be expressed in terms of $\boldsymbol{d}_v$ and $\boldsymbol{d}_w$. For example, n_{11}, the frequency of points in time where $s_{v,k}$ and $s_{w,k}$ amount at least to $\tilde{\boldsymbol{s}}_v$ and $\tilde{\boldsymbol{s}}_w$, respectively, is given by $n_{11} = \sum_{k=1}^{T} d_{v,k} \cdot d_{w,k}$. In other words, n_{11} is the number of points where both $\boldsymbol{d}_v$ and $\boldsymbol{d}_w$ are equal to 1, yielding the associated relative frequency p_{11} through $p_{11} = \frac{n_{11}}{T}$.

The probability masses corresponding to the table's relative frequencies are equal to the respective partial volumes belonging to the four quadrants in the x-y-plane under the curve representing the bivariate normal distribution (Figure 2A). The correlation coefficient r_t, for which these partial volumes resemble the relative frequencies in a given table, is an estimate of the population correlation ρ belonging to the underlying distribution. Since a 2×2 contingency table is uniquely defined by the marginal probabilities and one joint probability, r_t can be found by solving, e.g., $p_{11} = \int_{\Phi^{-1}(p_{\bullet 0})}^{\infty} \int_{\Phi^{-1}(p_{0\bullet})}^{\infty} f(z_x, z_y) \mathrm{d}z_x \mathrm{d}z_y$, where Φ is the standard normal distribution function, Φ^{-1} is its inverse, and f is the probability density function of the bivariate normal distribution. While this would typically be solved using numerical techniques, an analytical solution, $r_t = -\cos(2\pi p_{11})$, exists for the case under consideration (Figure 2B). See Section S.2 for details.

Given a pair of voxels, we can determine p_{11} from the dichotomized time series and use the relationship $r_t = -\cos(2\pi p_{11})$ in order to obtain r_t. As a consequence, r_t can be used instead of r to estimate pairwise functional connectivity in the process of graph construction from fMRI data (Figure 1).

Simulations

Building upon the theoretical considerations presented above, we analyzed the characteristics of r_t in the controllable environment of synthetic data. More specifically, we assessed its usefulness relative to r in estimating the correlation parameter ρ of bivariate normal populations of known properties.

Data

Bivariate normal populations were generated, such that each of them exhibited a predefined population correlation ρ, where $\rho \in \{-0.99, -0.98, \ldots, 0, \ldots, 0.98, 0.99\}$. Then, 10000 bivariate samples of size T (one sample represents a pair of time series of length T) were randomly

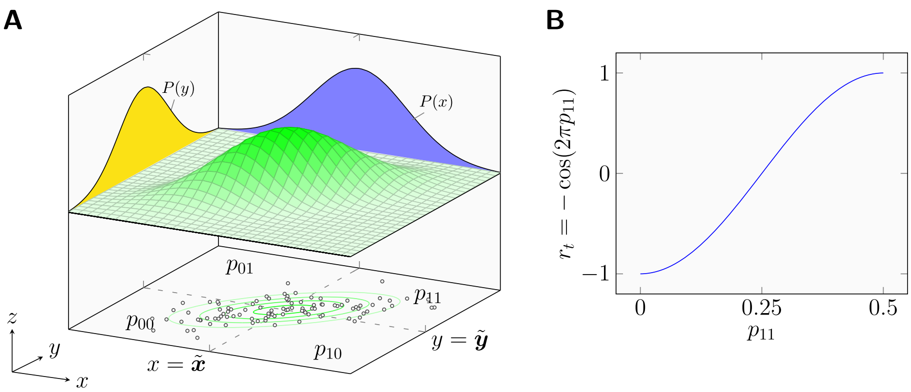

Figure 2 Dichotomization and tetrachoric correlation estimation. Consider a sample $((x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T))$ of size T where (x, y) is distributed according to bivariate normality. Further, let $\boldsymbol{x} = x_1, x_2, \ldots, x_T$ and $\boldsymbol{y} = y_1, y_2, \ldots, y_T$ denote the samples from x and y, respectively. Using $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$, as thresholds, $\boldsymbol{x}$ and $\boldsymbol{y}$ can be dichotomized resulting in the binary samples $\boldsymbol{x}_d$ and $\boldsymbol{y}_d$. **A**: As an example, the density of a bivariate normal distribution ($\rho = 0.7$) is shown (top, 3D curve) along with a sample (bottom, points in the *x*-*y*-plane) drawn from that distribution. By virtue of the two lines $x = \tilde{\boldsymbol{x}}$ and $y = \tilde{\boldsymbol{y}}$, the *x*-*y*-plane is divided into four quadrants, such that the counts of sample points per quadrant form a 2×2 contingency table. The (relative) frequencies in the contingency table can also be expressed in terms of $\boldsymbol{x}_d$ and $\boldsymbol{y}_d$ (e.g., n_{11} is the number of indices where both $\boldsymbol{x}_d$ and $\boldsymbol{y}_d$ are equal to 1, and $p_{11} = n_{11}T^{-1}$). The probability masses corresponding to the table's relative frequencies are equal to the respective partial volumes belonging to the four quadrants in the *x*-*y*-plane under the bivariate normal's curve. The tetrachoric correlation coefficient r_t, for which these partial volumes resemble the relative frequencies in a given table, is an estimate of the population correlation parameter ρ belonging to the underlying distribution. **B**: Relationship between p_{11} and r_t. Given $\boldsymbol{x}_d$ and $\boldsymbol{y}_d$, r_t can be found using $r_t = -\cos(2\pi p_{11})$. For details see text.

drawn from each of the populations. For each sample, r and r_t were calculated. Prior to calculating the latter, the data were dichotomized as described above. The entire procedure was conducted separately for two different sample sizes, $T = 100$ and $T = 300$, resulting in two data sets, where the choice of these numbers was guided by the parameters of the real fMRI data we analyzed.

Correlation estimation

For each data set and estimator $\hat{\rho} \in \{r, r_t\}$, joint histograms $(\hat{\rho}, \rho)$ with associated marginal histograms were calculated. For the joint histograms a linearly spaced 199×199 grid was used, such that bin centers in both dimensions corresponded to correlations exhibited by the generated bivariate populations. For each estimator, means and standard deviations, as well as mean signed differences $\mathrm{MSD}(\hat{\rho}, \rho)$ were calculated per ρ-bin. Mean signed differences are defined as $\mathrm{MSD}(\hat{\rho}, \rho) = n^{-1}\sum_i^n(\hat{\rho}_i - \rho)$, where n is the number of samples per ρ-bin, i.e., $n = 10000$, and $\hat{\rho}_i$ is the correlation estimate for sample i. Since ρ is not known in the case of real data, additional joint histograms (r_t, r) were calculated in order to facilitate comparability with respect to real data applications. As both estimators exhibit errors with respect to ρ, Deming regression[a] was used in order to fit a linear relationship to the (r_t, r) data.

Application to fMRI data

Comparative graph-based analyses of resting-state fMRI data were carried out based on r vs. r_t as measures of functional connectivity.

Data

MRI data were obtained from the "1000 Functional Connectomes Project" repository [36,37]: We used the "Cambridge" and "Pittsburgh" data sets, contributed by R.L. Buckner and G. Siegle, respectively. These data sets contain resting-state fMRI data from 198 subjects (75 males/123 females, ages: 18–30 years; imaging parameters: TR = 3s, voxel size $= 3 \times 3 \times 3$ mm^3, number of slices $= 47$; number of volumes $= 119$) and 17 subjects (10 males/7 females, ages: 25–54; imaging parameters: TR = 1.5s, voxel size $= 3.125 \times 3.125 \times 3.2$ mm^3, number of slices $= 29$; number of volumes $= 275$), respectively. Both data sets also include anatomical scans for each subject.

Preprocessing

Using SPM8 [38] functional data were motion-corrected by alignment to the mean functional image, then anatomical scans were coregistered to the mean functional image and segmented. In order to account for low frequency intensity drifts and high frequency noise, frequencies below 0.01Hz and above 0.1Hz were removed from the voxels time series by band-pass filtering, as is customary for resting-state data [39]. In order to minimize the impact of preprocessing on the data's correlation structure, we refrained from spatial smoothing and spatial normalization [8,27].

Correlation estimation

Based on r and r_t as a measure of functional connectivity, two voxel-level graphs were constructed for each subject from the two data sets. Nodes were defined as supra-threshold voxels in the subject-specific GM probability maps obtained from the segmentation (threshold $\theta_{GM} = 0.2$). To measure internodal functional connectivity, two correlation matrices, $\mathbf{C}_r$ and $\mathbf{C}_{r_t}$, were calculated based on all pairwise correlations between nodes. $\mathbf{C}_r$ was obtained by calculating Pearson correlations based on the voxels' associated continuous-valued time series from the preprocessed fMRI data, and $\mathbf{C}_{r_t}$ was obtained analogously, except that tetrachoric correlations were calculated instead of Pearson correlations, and binary voxel time series were used instead of continuous-valued ones. Again, binary voxel time series were derived from the continuous-valued ones through median-based dichotomization. In order to compare $\mathbf{C}_r$ and $\mathbf{C}_{r_t}$, their entries were used to calculate joint histograms (r_t, r) in the same fashion as for the synthetic data.

Functional connectivity graphs

Subject-specific binary functional connectivity graphs, $\mathbf{B}_r$ and $\mathbf{B}_{r_t}$, were derived from $\mathbf{C}_r$ and $\mathbf{C}_{r_t}$, respectively, via density-based thresholding: The density κ of a binary undirected graph $\mathbf{B}$ is the proportion of potential edges that actually exist, i.e., $\kappa = \frac{2 \cdot |E|}{|N| \cdot (|N|-1)}$. In order to facilitate comparability across graphs, an individual correlation threshold θ was determined for each correlation matrix, such that the resulting binary graphs exhibited the same density κ. Given $\mathbf{C}$, where $\mathbf{C} \in \{\mathbf{C}_r, \mathbf{C}_{r_t}\}$, and θ, the entries of $\mathbf{B}$ are given by $b_{i,j} = 1$, if $c_{i,j} > \theta$, and $b_{i,j} = 0$ otherwise, where $1 \leq i, j \leq |N|$.

Node degree maps

In graph-based fMRI functional connectivity analyses, one of the most popular graph-theoretical metrics is the node degree, or degree centrality, a measure aiming to characterize the importance of individual nodes in a binary graph. Given a binary graph $\mathbf{B}$, the degree k_i of a node i is defined as the number of nodes that are connected to i via an edge, or, more formally, $k_i = \sum_j^{|N|} b_{i,j}$, where $i, j \in N$ and $i \neq j$ [40]. The node degree has recently been employed in several neuroimaging studies aiming to identify potential hub regions in the human brain [27,31].

Here, node degrees k were calculated for all subject-specific functional connectivity graphs $\mathbf{B}_r$ and $\mathbf{B}_{r_t}$. Degrees were standardized in order to afford comparable scaling across subjects [15]. The spatial distribution of degrees was analyzed by constructing k-maps in individual brain space for each subject. In order to generate group average k-maps for each data set (Cambridge and Pittsburgh), the subject-specific k-maps were spatially normalized to ICBM[b] -template space based on transformation parameters estimated with respect to the mean functional image using SPM8 [41-43]. Since the normalized k-maps have different overlaps due to anatomical differences and differing GM masks, a varying number of subjects "supports" each standard space voxel. Thus, group-level k-maps were derived by voxel-wise averaging of the k-values from the supporting subjects. For enhanced reliability, k-values of voxels supported by less than 20% of all subjects were discarded.

Implementations

The most time-consuming step when constructing a graph from fMRI data consists in the computation of a functional connectivity matrix, which here corresponds to the computation of a correlation matrix based on r or r_t. In the following, the programs created for calculating the voxel-level pairwise correlation matrices $\mathbf{C}_r$ and $\mathbf{C}_{r_t}$ will be referred to by `pcc` and `tetracc`, respectively. For both programs we opted to store only the upper triangular part of $\mathbf{C}$, in order to save memory. In doing so, no information is lost, since $\mathbf{C}$ is symmetric. Because efficient implementation plays an important role when aiming to accelerate large-scale analyses, implementation was conducted using the C programming language, providing Matlab integration via its C interface MEX. A Matlab toolbox and C sources are available for download [44].

Calculation of C_r

Pearson's sample correlation coefficient r is calculated for a pair of voxels v and w using

$$r(s_v, s_w) = \frac{\sum_{k=1}^{T} (s_{v,k} - \bar{\mathbf{s}}_v)(s_{w,k} - \bar{\mathbf{s}}_w)}{\sqrt{\sum_{k=1}^{T} (s_{v,k} - \bar{\mathbf{s}}_v)^2} \cdot \sqrt{\sum_{k=1}^{T} (s_{w,k} - \bar{\mathbf{s}}_w)^2}} .$$

To avoid redundant operations, subexpressions depending on one voxel only are precalculated for all voxels before computing $\mathbf{C}_r$.

In order to take advantage of the processor's cache without the need for explicit knowledge about its size, we adopted a so-called *cache-oblivious algorithm* [45,46] to compute the correlation matrix, rather than explicit *blocking* (with a predetermined block size that optimally fits the cache). The core idea is to recursively divide the problem so that the computations are carried out on smaller and smaller blocks of data. Given that the minimum block size is small enough, there is a division step from which on all computations use only data that fits into the processor cache (regardless of its size), thus making optimal use of the cache by localizing the computations. The division scheme we implemented is illustrated in Additional file 1: Figure S1, which shows the first three steps of dividing the upper triangle of the correlation matrix.

In addition, we exploited SSE2 (Streaming SIMD Extensions version 2, where SIMD stands for Single Instruction Multiple Data) and AVX (Advanced Vector eXtensions) instructions (on processors that support them[c]), which allow for parallelization on a single core by carrying out the same operation on multiple data elements in parallel (also known as vectorized computations). Using SSE2 (AVX), the computation of the numerator of the correlation coefficient can be parallelized by computing four (eight) sums in parallel (if the `float` data type is used; for `double`, two and four sums, respectively, can be computed in parallel). The procedure is illustrated in Additional file 1: Figure S2 for SSE2 using `float` or AVX using `double` (four sums in parallel).

Calculation of $\mathbf{C}_{r_t}$

For each pair of voxels, r_t is computed in three steps. First, the bitwise `AND` operator is applied to the voxels' associated binary time series. Second, the set bits in the result are counted to obtain n_{11}. Third, r_t is retrieved from a lookup table of the function $r_t = -\cos(2\pi n_{11} T^{-1})$. The table is indexed by n_{11} and contains the corresponding r_t values for those values that n_{11} can attain. Depending on T being even or odd, these are $0, 1, 2, \ldots, \frac{T}{2}$ or $1, 2, 3, \ldots, \frac{T+1}{2}$, respectively.

Storing the binary time series in integers of, e.g., 32 bit, 32 points in time can be processed in parallel, so that the above three steps need to be executed only $\lceil T/32 \rceil$ times per pair. Hence, it seems conceivable that the computational cost in terms of CPU time could be lower for the calculation of $\mathbf{C}_{r_t}$ than for the calculation of $\mathbf{C}_r$.

Following the procedure outlined above, two programs, `tetracc/32` and `tetracc/128` were created. `tetracc/32` uses 32 bit integers for storing binary time series and a 16 bit lookup table for bit-counting. `tetracc/128` uses `__mm128i` variables (holding 128 bit each) for time series storage. Using the intrinsics `_mm_and_si128` and `_mm_popcnt_u64` for bitwise `AND` and bit-counting, respectively, it is expected to improve over `tetracc/32`, since more data can be processed in parallel and no extra memory access (lookup table) is needed. While `tetracc/32` is platform-independent, `tetracc/128` is only applicable on fairly modern CPUs, because `_mm_and_si128` and `_mm_popcnt_u64` depend on the availability of SSE2 and POPCNT instructions, respectively[d].

Parallel versions

For additional performance gains, parallel versions of `pcc` and `tetracc` have been implemented using multiple threads. This aspect is, however, beyond the focus of this article, since the resulting benefits relative to single-threaded programs are expected to be fairly independent of the choice of r versus r_t as a measure of internodal functional connectivity.

Performance tests

In order to assess the performance of the programs described above, we compared them to three other programs: Matlab's built-in function `corrcoef`, `corr` from Matlab's Statistics Toolbox, and `IPN_fastCorr`, a function from the Matlab toolbox *ipnvoxelgraph* by X.N. Zuo. Experiments were conducted from within Matlab (R2011b) on a desktop computer with an Intel(R) Core(TM) i7-3960X CPU (3.30GHz) and 64GB main memory running Linux (Kernel 3.4). The C/MEX routines that are part of our programs were compiled using the GNU C compiler gcc (version 4.7.1, optimization level 3). In order to prevent programs from making use of multiple cores, Matlab was restricted to one CPU core.

Input data sets ($\mathbf{S}^{V \times T}$) were generated using pseudo-random numbers of type `float`. While the length of time series T was fixed at $T = 200$, the number of voxels V was varied between 10000 and 170000 in steps of 10000. The maximum number of voxels, 170000, follows from the fact that storing the resulting matrix (upper triangular part) requires 53.83 GB ($\frac{V(V-1)}{2}$ `floats`, 4 bytes per `float`). Since the machine used has 64GB of main memory, this seemed a sensible choice in order to leave some memory for other applications and subsequent processing of the matrix. Because `corrcoef`, `corr`, and `IPN_fastCorr` return the complete symmetric matrix, they were only tested using input data sets with up to 120000 voxels corresponding to 53.64 GB of memory required to hold the matrix (V^2 `floats`).

Results

Correlation estimation

Overall, the correlation estimation from dichotomized data using r_t yielded results strongly resembling those obtained through estimation from continuous data using r. For synthetic data, as expected, r and r_t exhibited linear relationships to ρ and also to each other (Figure 3A). Standard deviations (with respect to means per ρ-bin) were greater for r_t than for r, but were reasonably small for both. Peaking at $\rho = 0$ with 0.101 and 0.058 (r), and 0.158 and 0.09 (r_t), for $T = 100$ and $T = 300$, respectively, they exhibited a gradual decrease towards the range limits of ρ. Naturally, deviations from ρ were larger for $T = 100$ than for $T = 300$, because for $T = 100$ each calculated correlation is based on fewer values than for $T = 300$. Close inspection of the mean signed differences $\mathrm{MSD}(r_t, \rho)$ and $\mathrm{MSD}(r, \rho)$ revealed a small systematic bias of both r and r_t as estimators of ρ (Additional file 1: Figure S3). The expected value of r, $E(r)$, can be approximated by $E(r) = \rho - \rho(\frac{1-\rho^2}{2n})$ [47,48]. $E(r) - \rho$ closely matched the empirical results from the simulation represented by $\mathrm{MSD}(r, \rho)$, while $\mathrm{MSD}(r_t, \rho)$ follows a curve of similar shape but larger amplitude. The Pearson correlation between r and ρ, r_t and ρ, and r_t and r, amounted to 0.992 (0.997), 0.978 (0.992), 0.986 (0.995), respectively, for T=100 (T=300).

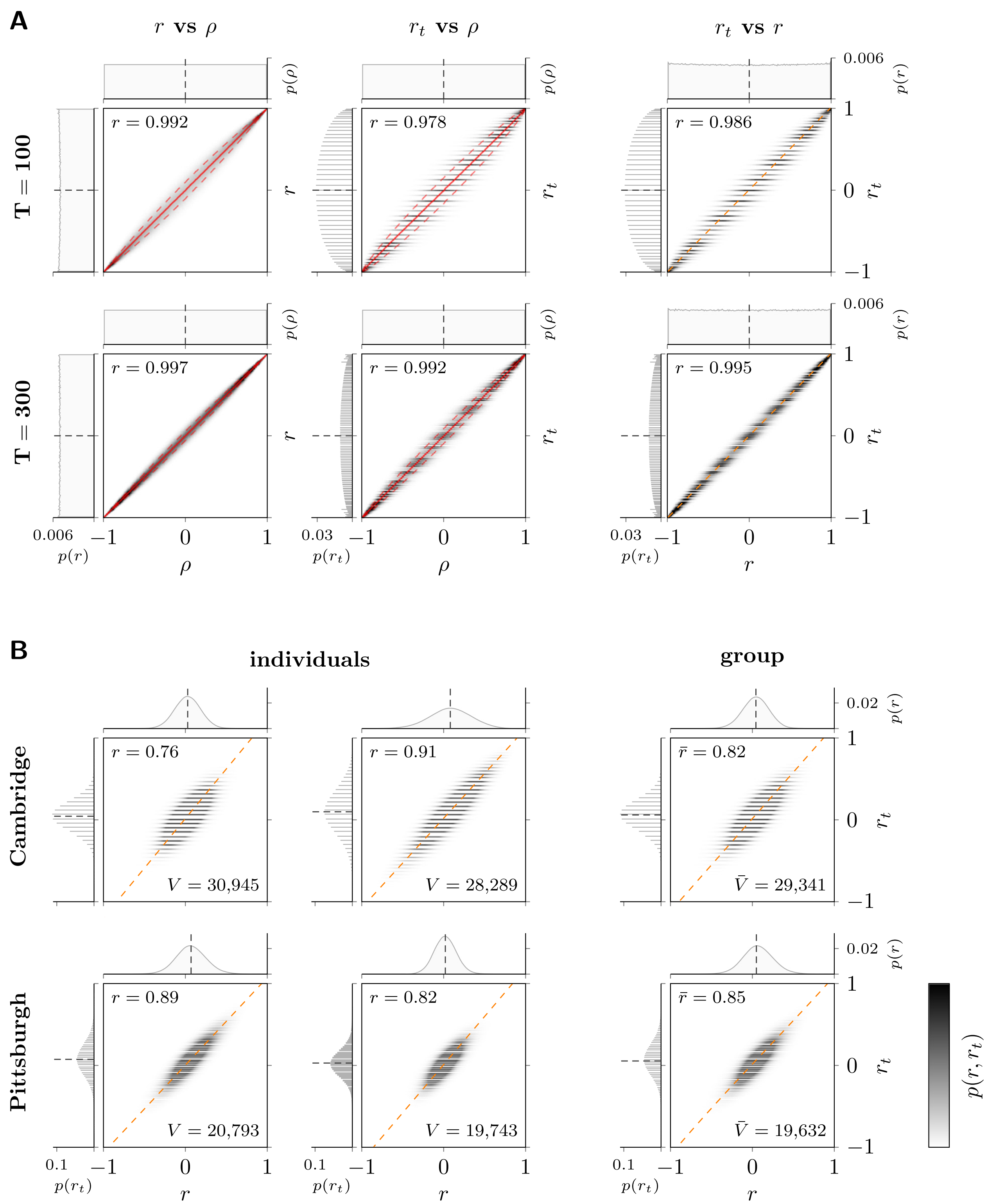

Figure 3 Comparison of correlation estimates r and r_t. Each composite plot consists of a joint histogram and the corresponding marginal histograms. Joint probabilities were mapped to grayscale intensities. **A**: Results for synthetic data. Top row: $T = 100$. Bottom row: $T = 300$. Left and mid column: r vs. ρ and r_t vs. ρ, respectively; mean (solid) and mean $\pm$ STD (dashed) per ρ-bin shown in red. Right column: r_t vs. r; Deming regression line (dashed) shown in orange. **B**: Results for fMRI data (r_t vs. r). Top row: Cambridge data set. Bottom row: Pittsburgh data set. Left and mid column: Based on correlation matrices $\mathbf{C}_r$ and $\mathbf{C}_{r_t}$ from two individual subjects. Right column: Based on *all* subject-specific correlation matrices $\mathbf{C}_r$ and $\mathbf{C}_{r_t}$. Deming regression line (dashed) shown in orange.

For fMRI data, r_t and r followed a linear relationship in both data sets, although there was a slight counterclockwise rotation about the origin as reflected in a slope > 1 as opposed to 1 for a perfect relation $r_t = r$ (Figure 3B). This suggests only limited deviation from the assumption of pairwise bivariate normality, and, moreover, indicates that r_t-based graphs closely resemble r-based graphs. Furthermore, the results for the Cambridge data set ($T = 119$) showed a greater variance than those for the Pittsburgh data set ($T = 275$). Since this feature was also observed in the results for the synthetic data sets, we presume that this was caused by the smaller sample size for the calculation of each sample correlation. The overall correlation between r and r_t amounted to 0.82 (Cambridge) and 0.85 (Pittsburgh).

Note that the vertical gaps in bin occupation in those histograms involving r_t are due to the fact that r_t can attain a distinct set of values only, as explained earlier. The number of attainable values, and hence the (potential) performance of r_t as an estimator, increases with T.

Node degree

Group average node degree maps from r- and r_t-based binary graphs $\mathbf{B}_r$ and $\mathbf{B}_{r_t}$ (derived from $\mathbf{C}_r$ and $\mathbf{C}_{r_t}$, respectively, using a density threshold of $\kappa = 0.01$) are presented in Figure 4. Other thresholds ($\kappa \in \{0.05, 0.1\}$) led to similar results and are hence not shown. In accordance with the strong correlation between r and r_t reported in the previous section, both approaches yielded highly similar spatially distributed node degree maps (Figure 4A). In line with this, k_r and k_{r_t} were very strongly correlated ($r(k_r, k_{r_t}) = 0.95$ for Cambridge and $r(k_r, k_{r_t}) = 0.97$ for Pittsburgh; Figure 4B), although degrees tended to be slightly higher for r_t-based than for r-based graphs. Prominent clusters of high-degree nodes were found within circumscribed regions of the occipital (cuneus, precuneus, fusiform and lingual gyri), parietal (intraparietal sulcus, superior parietal lobe, temporoparietal junction), temporal (superior temporal gyrus, temporal pole, amygdala) and frontal lobes (medial orbitofrontal and rostral ventromedial prefrontal cortex) with a similar distribution pattern as reported in previous work employing r-based node degree mapping [15,27,31].

Performance tests

The most time-consuming step when constructing a graph from fMRI data consists in the computation of a functional connectivity matrix. We compared the performance of our programs on this task to three other programs: Matlab's built-in function `corrcoef`, `corr` from Matlab's Statistics Toolbox, and `IPN_fastCorr`, a function from the Matlab toolbox *ipnvoxelgraph* by X.N. Zuo. Table 1 shows memory requirements, execution times, and speedups relative to `corrcoef` which was selected as reference since it is available to any Matlab user out of the box and we therefore assume that it has a higher prevalence than `corr` or `IPN_fastCorr`. Figure 5 illustrates the performance in terms of data troughput measured in correlation coefficients per second. This measure does not depend on the performance of a reference program and offers more immediate access to the key results. In this sense, it is complementary to Table 1.

In line with expectations, execution times increased quadratically with the number of nodes for all programs (Table 1). While `pcc`'s basic variant (`pcc/naive`) was considerably slower than `corrcoef` (speedup 0.31×), its SSE2- and AVX-based variants achieved speedups around 1.34× and 2.08×, respectively. The performance of `corr` (speedup 1.35×) was comparable to that of `pcc/SSE2`, while `IPN_fastCorr` (speedup 1.63×) ranked between `pcc/SSE2` and `pcc/AVX`. The `tetracc` variants (`32` and `128`) were considerably faster than all programs computing r with speedups (relative to `corrcoef`) around 5.7× and 13.5×, respectively.

As an aside, we note that `IPN_fastCorr` as well as `pcc` and `tetracc` scaled better with the number of cores than `corrcoef` and `corr`. For example, using 6 cores and a data set of 50000 nodes ($T = 200$), the speedups were 1.16× (`corr`), 2.46× (`IPN_fastCorr`), 2.32× (`pcc/SSE2`), 3.42× (`pcc/AVX`), 9.44× (`tetracc/32`), and 20.22× (`tetracc/128`), compared to `corrcoef`'s execution time of 18.5 seconds. Using the same data set but only 1 core the respective speedups were 1.35×, 1.63×, 1.33×, 2.08×, 5.7×, and 13.5× as compared to `corrcoef`'s execution time of 55.9 seconds (Table 1). Thus, the speedup gained by using 6 cores instead of 1 amounts to 4.56× (`IPN_fastCorr`), 5.29× (`pcc/SSE2`), 4.98× (`pcc/AVX`), 4.98× (`tetracc/32`), and 4.49× (`tetracc/128`), while `corr` and `corrcoef` gain only 2.6× and 3×, respectively.

Note that `pcc` and `tetracc` require only half the amount of memory (column 8) that `corrcoef`, `corr`, and `IPN_fastCorr` require (column 2), because they store only half of the symmetric matrix for memory efficiency (Table 1). In addition, `corrcoef`, `corr`, and `IPN_fastCorr` failed with an out-of-memory error for input data sets with 90000 or more nodes. We assume that these programs internally use more memory than we expected, since the resulting matrix of correlation coefficients would require less than half of the available memory ($90000^2 \cdot 4\text{Byte} = 30.17\text{GB}$). Hence, the speedups of the remaining programs compared to `corrcoef` could not be computed for $T \geq 90000$.

Discussion

Graph-based functional connectivity analysis at the level of individual voxels allows for spatially fine-grained characterization of functional networks in the human brain.

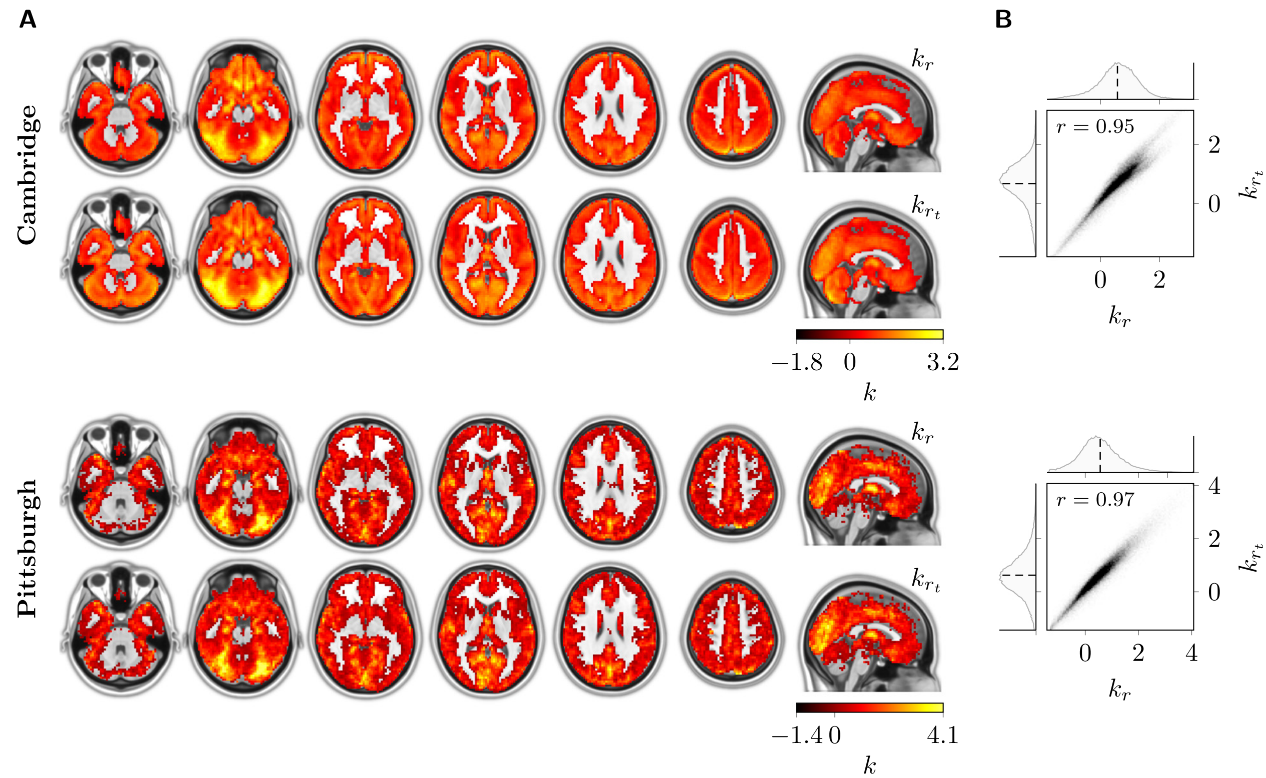

Figure 4 Comparison of node degrees k from r- (k_r) and r_t-based (k_{r_t}) binary graphs. Subject-specific graphs were derived from the correlation matrices $\mathbf{C}_r$ and $\mathbf{C}_{r_t}$ using a density threshold of $\kappa = 0.01$ corresponding to correlation thresholds of $r = 0.48 \pm 0.06$ and $r_t = 0.53 \pm 0.04$ (Cambridge) and $r = 0.46 \pm 0.08$ and $r_t = 0.49 \pm 0.07$ (Pittsburgh), respectively. Individual degree maps were spatially transformed to MNI space to derive group average maps. Note that the degrees were standardized for each subject before averaging, resulting in ranges that one would not commonly expect for degrees. For details see text. **A** Group average degrees on top of axial slices of the MNI brain (shown in neurological convention). Top row: k_r. Bottom row: k_{r_t}. **B** Joint distribution of k_r and k_{r_t}. The composite plots consist of a joint histogram and the corresponding marginal histograms. Joint probabilities were mapped to grayscale intensities.

However, with high-resolution data sets, such analyses can become infeasible due to the computational demands involved. Most previous studies investigating voxel-level functional connectivity graphs relied on Pearson's r for estimating internodal functional connectivity [2,8,13,15,21-27,29-31]. As demonstrated here, the tetrachoric correlation coefficient r_t constitutes a time-efficient alternative to r as a measure of functional connectivity.

In order to reduce the computational costs associated with the analysis of voxel-level graphs, previous studies reduced the data's spatial resolution [15,26,27], spatially restricted the graphical edges incorporated into the analysis [21], or utilized parallel computing [31]. In contrast, efficiency benefits from r_t-based graph construction are achieved without sacrificing spatial resolution, disregarding graphical edges, or exploiting parallel computing. An open source software package containing the created programs is freely available for download [44]. Note that parallel versions of r- and r_t-based graph construction have been implemented in addition to the sequential ones, thus providing additional efficiency gains that depend on the number of available processors. While this aspect is not the main focus of this article, as the resulting benefits (relative to sequential implementations) can be expected to be fairly independent of the choice of r versus r_t as a measure of internodal functional connectivity, the parallel implementations are still included in the software package [44].

Even though the dichotomization procedure (a prerequisite to the computation of r_t) entails discarding information in the time domain, important characteristics of the original data appear to be preserved. In applications to artificially generated as well as real fMRI data the new technique proved capable of closely reproducing results obtained in conventional ways. More specifically, the usefulness of the r_t-based approach was assessed by comparison with r in estimating the correlation parameter ρ of bivariate normal populations of known properties. In this setting, both the bias and standard deviation were greater for r_t than for r, but still reasonably small. Thus, r_t-based correlation estimation yielded results closely resembling

Table 1 Performance comparison for computation of correlation matrices

$\|N\|/10^3$		`corrcoef`	`corr`		`IPN_fastCorr`			`pcc/naive`		`pcc/SSE2`		`pcc/AVX`		`tetracc/32`		`tetracc/128`	
	m [GB]	t [s]	t [s]	s [×]	t [s]	s [×]	m [GB]	t [s]	s [×]	t [s]	s [×]	t [s]	s [×]	t [s]	s [×]	t [s]	s [×]
10	0.4	2.3	1.8	1.29	1.5	1.55	0.2	7.2	0.32	1.7	1.39	1.1	2.16	0.4	5.65	0.2	12.58
20	1.5	9.1	6.8	1.34	5.6	1.63	0.7	28.8	0.32	6.7	1.35	4.3	2.11	1.6	5.71	0.7	13.17
30	3.4	20.2	14.9	1.35	12.4	1.63	1.7	64.9	0.31	15.1	1.34	9.7	2.09	3.5	5.70	1.5	13.33
40	6.0	36.1	26.8	1.35	22.0	1.64	3.0	115.4	0.31	26.9	1.34	17.3	2.09	6.3	5.76	2.7	13.55
50	9.3	55.9	41.5	1.35	34.3	1.63	4.7	180.3	0.31	42.2	1.32	27.0	2.07	9.8	5.72	4.1	13.53
60	13.4	80.3	59.5	1.35	49.3	1.63	6.7	259.4	0.31	60.5	1.33	38.7	2.07	14.0	5.73	5.9	13.57
70	18.3	109.5	80.9	1.35	67.0	1.63	9.1	352.9	0.31	82.4	1.33	52.6	2.08	19.1	5.75	8.0	13.65
80	23.8	180.1	143.4	1.26	87.5	2.06	11.9	461.4	0.39	107.7	1.67	69.1	2.61	24.9	7.24	10.4	17.24
90	30.2						15.1	584.3		137.1		87.9		31.4		13.2	
100	37.3						18.6	721.0		168.7		107.8		38.7		16.3	
110	45.1						22.5	872.1		203.1		130.3		46.8		19.6	
120	53.6						26.8	1037.6		242.0		154.9		55.6		23.4	
130	63.0						31.5	1217.5		284.3		181.6		65.2		27.4	
140	73.0						36.5	1411.7		329.4		210.2		75.6		31.7	
150	83.8						41.9	1620.2		377.9		241.3		86.7		36.4	
160	95.4						47.7	1845.5		430.7		276.4		98.6		41.4	
170	107.7						53.8	2085.0		487.8		313.2		111.3		46.7	

Results are averages from 10 runs on a desktop computer with an Intel(R) Core(TM) i7-3960X CPU (3.3GHz) and 64GB main memory. All programs were restricted to one CPU core. Length of time series T was fixed at $T = 200$. $|N|$: number of nodes. The programs `corrcoef` (Matlab built-in), `corr` (Matlab Statistics Toolbox), `IPN_fastCorr` (X.N. Zuo), and `pcc` (three variants) computed $\mathbf{C}_r$, while `tetracc` (two variants) computed $\mathbf{C}_{r_t}$. m [GB]: memory requirements for result in GB; t [s]: elapsed time in seconds; s [×]: speedup relative to `corrcoef`.

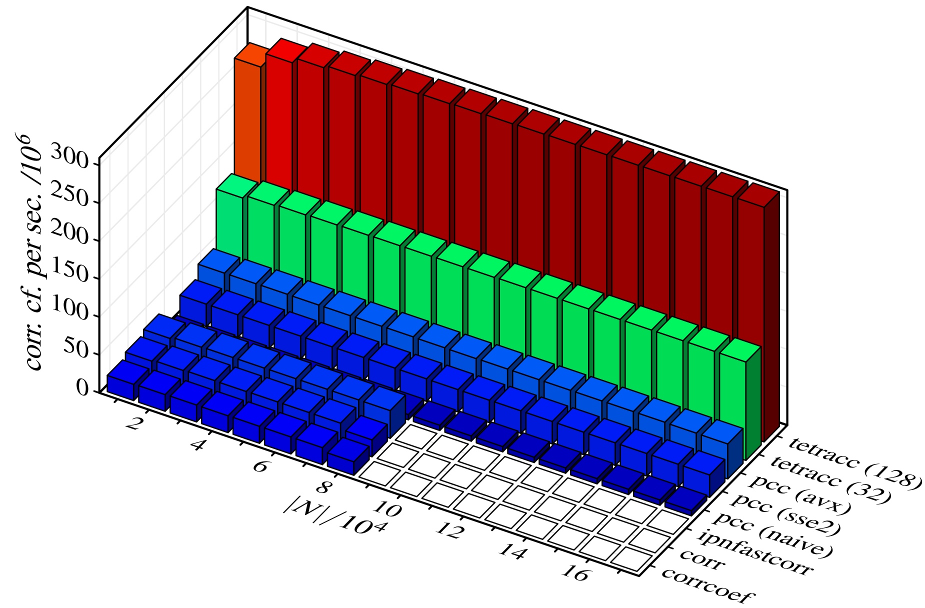

Figure 5 Performance comparison for computation of correlation matrices. Results are averages from 10 runs on a desktop computer with an Intel(R) Core(TM) i7-3960X CPU (3.3GHz) and 64GB main memory. All programs were restricted to one CPU core. Length of time series T was fixed at $T = 200$. $|N|$: number of nodes. The programs `corrcoef` (Matlab built-in), `corr` (Matlab Statistics Toolbox), `IPN_fastCorr` (X.N. Zuo), and `pcc` (three variants) computed $\mathbf{C}_r$, while `tetracc` (two variants) computed $\mathbf{C}_{r_t}$. The performance of each program is given as the number of correlation coefficients computed per second.

those obtained when using r. Beyond that, r- and r_t-based graph construction and node degree computation were carried out for real fMRI data. A strong linear relationship was found between r- and r_t-based correlations indicating that r_t-based graphs closely resemble r-based graphs, since the graphs are derived from the correlation matrices. In line with this, the spatial distribution of node degrees was highly similar for r- and r_t-based graphs and also in good correspondence with previous work [15,27,31].

As data mining approaches are currently gaining momentum in the neuroimaging community [36,37,49,50], the amount of publicly available experimental data is steadily growing. Consequently, development and implementation of efficient exploratory methods, such as the one presented here, are necessary in order to take full advantage of this wealth of data, especially with respect to connectivity analyses [51]. Fast construction and subsequent analysis of graphs may thus open new avenues for applications, including those within a clinical setting, where the voxel-level approach may be of particular importance. It is worth noting in this context that voxel-level graph construction can operate at the original data resolution, thus avoiding the reduction of the analysis' spatial sensitivity [4,24]. For example, disease-related patterns, once identified, may serve as connectivity-based biomarkers that could aid, guide, or facilitate diagnostics and may increase prediction accuracy with respect to disease occurrence, recurrence, severity, or treatment outcome. Here, again, efficient methods are essential to facilitate assessment of individual patients within a narrow time frame [52]. If combined with efficient tools for subsequent analysis, the presented methods for fast graph construction may also be useful for online evaluation of functional connectivity in the context of real-time fMRI. This would allow for connectivity-based adaptation of experimental stimulation and interaction with the subject, for example, in task-based fMRI studies, or neurofeedback-based training. Taken together, we believe that there is a multitude of applications (be them experimental or clinical) that could benefit from the methods presented here, highlighting the growing importance of efficient tools for graph-based analysis of voxel-level connectivity.

Limitations

As illustrated by the results, the accuracy of a correlation estimate naturally increases with the number of data points, i.e., the number of scans. Along the same lines, it has recently been shown that the reliability of functional homogeneity increases with scan duration [53]. For both correlation estimators, it is therefore recommended to avoid a low number of scans (caused, for example, by a short scan duration, or a long TR, or both). Since the deviation from the population correlation ρ is generally higher for r_t than for r, a low number of scans will affect r_t more severely than r.

The main focus of this work lies with the comparison of r and r_t as functional connectivity estimators. To reduce the impact of preprocessing on the data's correlation structure prior to this comparison, we limited the preprocessing of the fMRI data to a minimum. The effect of additional preprocessing steps, or a different preprocessing pipeline altogether, on the robustness of the proposed methods should be subject of future research. Unpublished results from our group indicate, however, that the comparability of r and r_t remains essentially consistent.

Conclusions

Voxel-level graphs allow for spatially fine-grained analyses of functional connectivity networks. In order to reduce the considerable computational demands involved, many previous studies reduced the spatial resolution of the data. Here, a new method for graph construction—exploiting time series dichotomization and tetrachoric correlation estimation—was devised, efficiently implemented, and compared to the conventional approach based on continuous-valued data and Pearson's r. In applications to artificially generated as well as real fMRI data the new technique proved capable of producing highly similar results. Through efficient bit-based implementation adapted to the dichotomized data the novel method runs an order of magnitude faster while the original spatial resolution of the data is retained. Hence, its demonstrated performance, not only in producing consistent results, but in obtaining them substantially faster, makes the new approach a sensible and economical alternative to customary practice. An open source software package containing the created programs is freely available for download [44].

Endnotes

[a]Deming regression is a linear regression method that accounts for errors in both variables.

[b]International Consortium for Brain Mapping.

[c]SSE2 was introduced by Intel with the Pentium 4. It is also supported by AMD CPUs starting with the Athlon 64. AVX is supported starting with the Sandy Bridge (Intel) and Bulldozer (AMD) microarchitectures.

[d]`POPCNT` became available starting with the Nehalem (Intel) and Barcelona (AMD) microarchitectures.

Additional file

Additional file 1: Supporting information.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KL conceived of the study and carried out the data analyses with conceptual input from MG, CMS, and CB. KL and CB created the software package and carried out the performance tests. KL drafted the manuscript. MG, CMS, RK, and CB edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Sarah Donohue, Joseph Harris, Martin von Kurnatowski, Christian Moewes, and Mircea Ariel Schoenfeld for their support. This work was supported by Deutsche Forschungsgemeinschaft SFB 779 (A11) and STSM grant 9059 of COST Action IC0702.

Author details

[1]Department of Neurology, Experimental Neurology, Otto-von-Guericke Universität, Leipziger Str. 44, 39120 Magdeburg, Germany. [2]Department of Knowledge and Language Processing, Otto-von-Guericke Universität, Magdeburg, Germany. [3]Department of Economics, University of Zürich, Zürich, Switzerland. [4]Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Berlin, Germany. [5]European Centre for Soft Computing, Mieres (Asturias), Spain.

Received: 24 November 2013 Accepted: 4 June 2014
Published: 19 June 2014

References

1. Salvador R, Suckling J, Coleman M, Pickard J, Menon D, Bullmore E: **Neurophysiological architecture of functional magnetic resonance images of human brain.** *Cereb Cortex* 2005, **15**(9):1332–1342.
2. Eguiluz V, Chialvo D, Cecchi G, Baliki M, Apkarian A: **Scale-free brain functional networks.** *Physical Rev Lett* 2005, **94:**18102.
3. Bullmore E, Sporns O: **Complex brain networks: graph theoretical analysis of structural and functional systems.** *Nat Rev Neurosci* 2009, **10**(3):186–198.
4. Wang J, Zuo X, He Y: **Graph-based network analysis of resting-state functional MRI.** *Front Syst Neurosci* 2010, **4**(16):1–14.
5. Friston K, Frith C, Liddle P, Frackowiak R: **Functional connectivity: the principal-component analysis of large (PET) data sets.** *J Cereb Blood Flow Metab* 1993, **13:**5–14.
6. Rubinov M, Sporns O: **Complex network measures of brain connectivity: uses and interpretations.** *Neuroimage* 2010, **52**(3):1059–1069.
7. Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E: **A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs.** *J Neurosci* 2006, **26:**63–72.
8. van den Heuvel M, Stam C, Boersma M, Hulshoff Pol H: **Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain.** *Neuroimage* 2008, **43**(3):528–539.
9. Fair D, Cohen A, Power J, Dosenbach N, Church J, Miezin F, Schlaggar B, Petersen S: **Functional brain networks develop from a local to distributed organization.** *PLoS Comput Biol* 2009, **5**(5):e1000381.
10. Supekar K, Musen M, Menon V: **Development of large-scale functional brain networks in children.** *PLoS Biology* 2009, **7**(7):e1000157.
11. Tomasi D, Volkow N: **Aging and functional brain networks.** *Mol Psychiat* 2012, **17**(5):549–558.
12. Tomasi D, Volkow ND: **Gender differences in brain functional connectivity density.** *Hum Brain Mapp* 2012, **33**(4):849–860.
13. van den Heuvel MP, Stam CJ, Kahn RS, Hulshoff Pol HE: **Efficiency of functional brain networks and intellectual performance.** *J Neurosci* 2009, **29**(23):7619–7624.
14. Supekar K, Menon V, Rubin D, Musen M, Greicius MD: **Network analysis of intrinsic functional brain connectivity in Alzheimer's disease.** *PLoS Comput Biol* 2008, **4**(6):e1000100.
15. Buckner R, Sepulcre J, Talukdar T, Krienen F, Liu H, Hedden T, Andrews-Hanna J, Sperling R, Johnson K: **Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease.** *J Neurosci* 2009, **29**(6):1860–1873.
16. Liu Y, Liang M, Zhou Y, He Y, Hao Y, Song M, Yu C, Liu H, Liu Z, Jiang T: **Disrupted small-world networks in schizophrenia.** *Brain* 2008, **131**(4):945–961.

17. Lynall ME, Bassett DS, Kerwin R, McKenna PJ, Kitzbichler M, Muller U, Bullmore E: **Functional connectivity and brain networks in schizophrenia.** *J Neurosci* 2010, **30**(28):9477–9487.
18. He Y, Wang J, Wang L, Chen ZJ, Yan C, Yang H, Tang H, Zhu C, Gong Q, Zang Y, Evans A: **Uncovering intrinsic modular organization of spontaneous brain activity in humans.** *PLoS One* 2009, **4**(4): e5226.
19. Scheinost D, Benjamin J, Lacadie C, Vohr B, Schneider K, Ment L, Papademetris X, Constable R: **The intrinsic connectivity distribution: a novel contrast measure reflecting voxel level functional connectivity.** *NeuroImage* 2012, **62**(3):1510–1519.
20. Smith S, Miller K, Salimi-Khorshidi G, Webster M, Beckmann C, Nichols T, Ramsey J, Woolrich M: **Network modelling methods for FMRI.** *Neuroimage* 2011, **54**(2):875–891.
21. Tomasi D, Volkow N: **Functional connectivity density mapping.** *Proc Nat Acad Sci* 2010, **107**(21):9885.
22. Tomasi D, Volkow N: **Association between functional connectivity hubs and brain networks.** *Cereb Cortex* 2011, **21**(9):2003–2013.
23. van den Heuvel M, Mandl R, Hulshoff Pol H: **Normalized cut group clustering of resting-state FMRI data.** *PLoS One* 2008, **3**(4):e2001.
24. Hayasaka S, Laurienti P: **Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data.** *Neuroimage* 2010, **50**(2):499–508.
25. Fransson P, Åden U, Blennow M, Lagercrantz H: **The functional architecture of the infant brain as revealed by resting-state fMRI.** *Cereb Cortex* 2011, **21:**145–154.
26. Valencia M, Pastor M, Fernández-Seara M, Artieda J, Martinerie J, Chavez M: **Complex modular structure of large-scale brain networks.** *Chaos* 2009, **19**(2):023119–023119.
27. Zuo X, Ehmke R, Mennes M, Imperati D, Castellanos F, Sporns O, Milham M: **Network centrality in the human functional connectome.** *Cereb Cortex* 2012, **22**(8):1862–1875.
28. Rubinov M, Sporns O: **Weight-conserving characterization of complex functional brain networks.** *Neuroimage* 2011, **56**(4):2068–2079.
29. Sepulcre J, Liu H, Talukdar T, Martincorena I, Yeo BT, Buckner RL: **The Organization of Local and Distant Functional Connectivity in the Human Brain.** *PLoS Comput Biol* 2010, **6**(6):e1000808.
30. Power J, Cohen A, Nelson S, Wig G, Barnes K, Church J, Vogel A, Laumann T, Miezin F, Schlaggar B, Petersen S: **Functional network organization of the human brain.** *Neuron* 2011, **72**(4):665–678.
31. Tomasi D, Volkow N: **Functional connectivity hubs in the human brain.** *Neuroimage* 2011, **57:**908–917.
32. Kowalski CJ: **On the effects of non-normality on the distribution of the sample product-moment correlation coefficient.** *J R Stat Soc Series C (Appl Stat)* 1972, **21:**1–12.
33. Hlinka J, Paluš M, Vejmelka M, Mantini D, Corbetta M: **Functional connectivity in resting-state fMRI: Is linear correlation sufficient?** *Neuroimage* 2011, **54**(3):2218–2225.
34. Pearson K: **Mathematical contributions to the theory of evolution. VII. on the correlation of characters not quantitatively measurable.** *Philos Trans R Soc Lond* 1900, **195:**1–47.
35. Loewe K, Grueschow M, Borgelt C: **Mining local connectivity patterns in fMRI data.** In *Towards Advanced Data Analysis by Combining Soft Computing and Statistics, Volume 285 of Studies in Fuzziness and Soft Computing*. Edited by Borgelt C, Ángeles Gil M, Sousa J, Verleysen M. Berlin Heidelberg: Springer; 2013:305–317.
36. **1000 Functional connectomes project.** [http://www.nitrc.org/projects/fcon_1000]
37. Biswal B, Mennes M, Zuo X, Gohel S, Kelly C, Smith S, Beckmann C, Adelstein J, Buckner R, Colcombe S, Dogonowski A, Ernst M, Fair D, Hampson M, Hoptman M, Hyde J, Kiviniemi V, Kötter R, Li S, Lin C, Lowe M, Mackay C, Madden D, Madsen K, Margulies D, Mayberg H, McMahon K, Monk C, Mostofsky S, Nagel B, et al: **Toward discovery science of human brain function.** *Proc Nat Acad Sci* 2010, **107**(10):4734.
38. **SPM8 - statistical parametric mapping.** [http://www.fil.ion.ucl.ac.uk/spm/software/spm8]
39. van den Heuvel M, Hulshoff Pol H: **Exploring the brain network: a review on resting-state fMRI functional connectivity.** *Eur Neuropsychopharmacol* 2010, **20**(8):519–534.
40. Freeman L: **Centrality in social networks conceptual clarification.** *Soc Netw* 1979, **1**(3):215–239.
41. Friston K, Ashburner J, Frith C, Poline J, Heather J, Frackowiak R: **Spatial registration and normalization of images.** *Hum Brain Mapp* 1995, **3**(3):165–189.
42. Ashburner J, Neelin P, Collins D, Evans A, Friston K: **Incorporating prior knowledge into image registration.** *Neuroimage* 1997, **6**(4):344–352.
43. Ashburner J, Friston K: **Nonlinear spatial normalization using basis functions.** *Hum Brain Mapp* 1999, **7**(4):254–266.
44. **Kristian Löwe's web pages.** [http://www.kristianloewe.com/software.html]
45. Prokop H: **Cache-oblivious algorithms.** *Ma thesis*. Massachusets Institute of Technology, Cambridge, USA 1999.
46. Frigo M, Leiserson C, Prokop H, Ramachandran S: **Cache-oblivious Algorithms.** In *Proc. 40th IEEE Symposium on Foundations of Computer Science (FOCS'99, New York, NY)*. Piscataway, USA: IEEE Press; 1999.
47. Fisher R: **Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population.** *Biometrika* 1915, **10:**507–521.
48. Zimmerman D, Zumbo B, Williams R: **Bias in estimation and hypothesis testing of correlation.** *Psicológica* 2003, **24**(1):133–158.
49. Poldrack R: **The future of fMRI in cognitive neuroscience.** *NeuroImage* 2012, **62**(2):1216–1220.
50. Poline J, Breeze J, Ghosh S, Gorgolewski K, Halchenko Y, Hanke M, Haselgrove C, Helmer K, Keator D, Marcus D, Poldrack R, Schwartz Y, Ashburner J, Kennedy D: **Data sharing in neuroimaging research.** *Front Neuroinform* 2012, **6**(9):1–13.
51. Milham M: **Open neuroscience solutions for the connectome-wide association era.** *Neuron* 2012, **73**(2):214–218.
52. Buckner R: **Human functional connectivity: new tools, unresolved questions.** *Proc Nat Acad Sci* 2010, **107**(24):10769.
53. Zuo XN, Xu T, Jiang L, Yang Z, Cao XY, He Y, Zang YF, Castellanos FX, Milham MP: **Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space.** *Neuroimage* 2013, **65:**374–386.

doi:10.1186/1471-2202-15-78
Cite this article as: Loewe *et al.*: Fast construction of voxel-level functional connectivity graphs. *BMC Neuroscience* 2014 **15**:78.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central