

Online Methods

Pallares V*, Insabato A*, Sanjuan A, Mantini D, Kühn S, Deco G, Gilson M

September 22, 2017

1 Description of fMRI datasets

Three datasets acquired at different locations were used in this work. The first one was used to study day-to-day variability of resting-state sessions from the same subject along a 6-months period (dataset A). The uniqueness of this data can provide information about how stable the functional connectome of a person is during long periods of time and strengthens the robustness of the results. The second dataset (dataset B) is publicly available and is part of the Consortium for Reliability and Reproducibility (CoRR) (REF: <http://www.nature.com/articles/sdata201449>). We used this dataset to generalize the results of the discriminability to a larger number of subjects. The third dataset (dataset C) was used to include not only the classification of subjects but also of conditions (resting-state vs. movie viewing). We provide next details about the acquisition of the blood-oxygen-level dependent (BOLD) signals for the three datasets. It is important to remark that the results of comparing the different measures and estimates of connectivity hold in spite of the fact that the data was collected in different scanners and preprocessed with different pipelines, which contributes to generalization.

1.1 Dataset A

This dataset has two parts. The first part (A1) is longitudinal and consists of resting-state fMRI sessions from 8 subjects (age 24-32, 6 female). 2 subjects (one male, one female) were not able to continue the study and were discarded. The other 6 subjects underwent scanning between 40 and 50 times along a period of 6 months. The second part of the dataset (A2) was acquired during the same period of time. A total 50 subjects (age 18-32, all female) were scanned during one rsMRI session using the same MRI sequences. Participants were free of psychiatric disorder according to personal interview (Mini-International Neuropsychiatric Interview REF) and had never suffered from a mental disease. The study was approved by the local ethics committee (Charité University Clinic, Berlin).

Images were acquired on a 3 T Magnetom Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany) using a 12-channel radiofrequency head coil. Structural images were obtained using a three-dimensional T1-weighted

magnetization-prepared gradient-echo sequence (MPRAGE) based on the ADNI protocol (www.adni-info.org) (repetition time (TR) = 2500 ms; echo time (TE) = 4.77 ms; TI = 1100 ms, acquisition matrix = $256 \times 256 \times 192$ mm³, flip angle = 7°; bandwidth=140 Hz/pixel, $1 \times 1 \times 1$ mm³ voxel size). Functional images were collected using a T2*-weighted echo planar imaging (EPI) sequence sensitive to blood oxygen level dependent (BOLD) contrast (TR = 2000 ms, TE = 30 ms, image matrix = 64×64 , FOV = $216 \times 216 \times 129$ mm³, flip angle = 80°, bandwidth=2042 Hz/pixel, voxel size= $3 \times 3 \times 3$ mm³, 36 axial slices using GRAPPA acceleration factor, 5:08 min duration).

1.1.1 Pre-processing

The data was preprocessed using SPM5 (REF) and DPARSF (REF) after discarding the first 10 volumes of each session, and included: slice timing and head motion correction (6 parameters spatial transformation), spatial normalization to the Montreal Neurological Institute (MNI) template, and spatial filtering of 4 mm FWHM. The data was parcellated using the Automated Anatomical Labeling (AAL) atlas (REF) into 116 regions of interest (ROI). Linear trends were removed from the fMRI time courses before band-pass filtering (0.01 - 0.08 Hz). Participants were instructed to remain with their eyes closed and data acquisition had to be constrained to 5 min per scan due to experimental limitations. We applied the automated anatomical labeling (AAL) parcellation with 116 ROIs covering the whole brain, including the cerebellum [4].

1.2 Dataset B

This dataset consists of 10 fMRI resting-state sessions acquired from 30 healthy participants every three days for one month (10-minutes length).

MRI Imaging sessions were performed using a GE MR750 3.0 Tesla scanner (GE Medical Systems, Waukesha, WI) at CCBD, Hangzhou Normal University. T2-weighted echo-planar imaging (EPI: TR = 2000 ms, TE = 30 ms, flip angle = 90°, field of view = 220×220 mm, matrix = 64×64 , voxel size = $3.4 \times 3.4 \times 3.4$ mm, 43 slices) sequence was performed to obtain resting state fMRI images for 10 minutes. A T1-weighted Fast Spoiled Gradient echo (FSPGR: TR = 8.1 ms, TE = 3.1 ms, TI = 450 ms, flip angle = 8°, field of view = 256×256 mm, matrix = 256×256 , voxel size = $1.0 \times 1.0 \times 1.0$ mm, 176 sagittal slices) was carried out to acquire a high-resolution anatomical image of the brain structure. To minimize head movement, straps and foam pads were used to fix the head snugly during each scan. The participants were instructed to relax and remain still with their eyes open, not to fall asleep, and not to think about anything in particular. The screen presented a fixation point and after the scans, all the participants were interviewed, and none of them reported to have fallen asleep in the scanner. The time of day of MRI acquisition was controlled within participants. This data was also parcellated into 116 ROIs using the AAL parcellation.

1.2.1 Pre-processing

Data was preprocessed with DPABI. The first 5 fMRI volumes were discarded in order to let the BOLD signal reach stability. The pre-processing pipeline included: slice-timing correction, motion correction, co-registration of the T1 anatomical image to the mean functional image, detrending, regression of 6 movement parameters, 5 PCA white matter and csf Compcorr, and spatial normalization to MNI coordinates. Scrubbing with Power 0.5 and linear interpolation was applied. Data was parcellated into 116 ROI (AAL), and time courses were band-pass filtered between 0.01 and 0.08 Hz. No global signal regression and spatial smoothing were applied.

1.3 Dataset C

We used a third dataset to study the discrimination between subjects and conditions. In this case, a total of 22 subjects (age 20-31, 15 females) were scanned during rest with eyes opened and natural viewing condition. Volunteers were informed about the experimental procedures, which were approved by the Ethics Committee of the Chieti University, and signed a written informed consent.

Data was acquired with a 3T MR scanner (Achieva; Philips Medical Systems, Best, The Netherlands) at the Institute for Advanced Biomedical Technologies in Chieti, Italy. The functional images were acquired using T2*-weighted echo-planar images (EPI) with BOLD contrast using SENSE imaging. EPIs comprised of 32 axial slices acquired in ascending order and covering the entire brain (230×230 in-plane matrix, TR/TE=2 s/3.5 s, flip angle = 90° , voxel size= $2.875 \times 2.875 \times 3.5$ mm³). For each subject, 2 and 3 scanning runs of 10 minutes duration were acquired for resting state and natural viewing, respectively. Only the first 2 movie scans are used here, to have the same number of time points for the two conditions (i.e., 20 minutes each). Each run included 5 dummy volumes - allowing the MRI signal to reach steady state and an additional 300 functional volumes that were used for analysis. Eye position was monitored during scanning using a pupil-corneal refection system at 120 Hz (Is-can, Burlington, MA, USA). A three-dimensional high-resolution T1-weighted image, for anatomical reference, was acquired using an MP-RAGE sequence (TR/TE=8.1 s/3.7 s, voxel size= $0.938 \times 0.938 \times 1$ mm³) at the end of the scanning session.

In the resting state, participants fixated a red target with a diameter of 0.3 visual degrees on a black screen. In the natural viewing condition, subjects watched and listened to 30 minutes of the movie ‘The Good, the Bad and the Ugly’ in a window of 24×10.2 visual degrees. Visual stimuli were projected on a translucent screen using an LCD projector, and viewed by the participants through a mirror tilted by 45 degrees. Auditory stimuli were delivered using MR-compatible headphones. For each subject, 2 and 3 scanning runs of 10 minutes duration were acquired for resting state and natural viewing, respectively.

The discarded subjects in the present study are 1, 11 and 19, among the 22 subjects available (numbered from 0 to 21).

1.3.1 Pre-processing

Data was preprocessed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK), including slice-timing and head-motion correction, co-registration between anatomical and mean functional image, and spatial normalization to MNI stereotaxic space (Montreal Neurological Institute, MNI) with a voxel size of $3 \times 3 \times 3 \text{ mm}^3$. Mean BOLD time series were extracted from $N = 66$ regions of interest (ROIs) of the brain atlas used in [2] for each recording session.

1.4 Generic structural connectivity (SC)

Structural connectivity is used in the model to determine the skeleton of the effective connectivity. A generic SC for datasets A and B is estimated from... XXX

Structural connectivity for Dataset C was estimated from Diffusion Spectrum Imaging (DSI) data collected in five healthy right-handed male participants [2]. The gray matter was first parcellated into the $N = 66$ ROIs, using the same low-resolution atlas used for the FC analysis. For each subject, we performed white matter tractography between pairs of cortical areas to estimate a neuro-anatomical connectivity matrix. In our method, the DSI values are only used to determine the skeleton: a binary matrix of structural connectivity (SC) obtained by averaging the matrices over subjects and applying a threshold for the existence of connections. The strengths of individual intracortical connections do not come from DSI values, but are optimized as explained below. It is known that DSI underestimates inter-hemispheric connections [2]. Homotopic connections between mirrored left and right ROIs are important in order to model whole-cortex BOLD activity [3]. Here we add all such possible homotopic connections, which are tuned during the optimization as other existing connections. This slightly increases the density of structural connectivity (SC) from 27% to 28%.

2 Connectivity measures and model estimates

2.1 Empirical measures of functional connectivity

For each fMRI session, the BOLD time series is denoted by s_i^t for each region $1 \leq i \leq N$ with time indexed by $1 \leq t \leq T$ (time points separated by a TR=2 seconds). The time series were first centered by removing - for each individual ROI i - the session mean $\bar{s}_i = \frac{1}{T} \sum_t s_i^t$. Following [1], the spatiotemporal FC corresponds to covariances calculated as:

$$\begin{aligned} \hat{Q}_{ij}^0 &= \frac{1}{T-2} \sum_{1 \leq t \leq T-1} (s_i^t - \bar{s}_i)(s_j^t - \bar{s}_j) , \\ \hat{Q}_{ij}^1 &= \frac{1}{T-2} \sum_{1 \leq t \leq T-1} (s_i^t - \bar{s}_i)(s_j^{t+1} - \bar{s}_j) \end{aligned} \tag{1}$$

The classical BOLD correlations (corrFC in the main text) correspond to

$$\hat{K}_{ij}^0 = \frac{\hat{Q}_{ij}^0}{\sqrt{\hat{Q}_{ii}^0 \hat{Q}_{jj}^0}} . \quad (2)$$

For each individual and session, we calculate the time constant τ_x associated with the exponential decay of the autocovariance averaged over all ROIs:

$$\tau_x = \frac{1}{N} \sum_{1 \leq i \leq N} \frac{1}{\log(\hat{Q}_{ii}^0) - \log(\hat{Q}_{ii}^1)} \quad (3)$$

This is used to “calibrate” the model, before its optimization.

2.2 Model of cortical dynamics

The model uses two sets of parameters to generate the spatiotemporal FC:

- The network effective connectivity (EC) between the ROIs (cf. Figure 2 in the main text) is denoted by the matrix C in the following equations; its skeleton is determined by DTI, meaning weights for absent connections are kept equal to 0 at all times.
- The local variability is described by the variances (1 per ROI) on the diagonal of the matrix Σ .

The model FC comes from the propagation of the local variability - inputed to every ROI - that propagates via EC, generating network feedback.

Formally, the network dynamics is described by a multivariate Ornstein-Uhlenbeck process, where the activity variable x_i of node i decays exponentially with time constant τ_x - estimated using Eq. (3) - and evolves depending on the activity of other populations:

$$ddx_i = \left(\frac{-x_i}{\tau_x} + \sum_{j \neq i} C_{ij} x_j \right) dt + dB_i , \quad (4)$$

where dB_i is equivalent to white noise with covariance matrix Σ (formally a Wiener process); note that only variances on the diagonal are non zero here.

The simplicity of the model allows for an analytic (feedforward) estimation of the covariances Q_{ij}^τ , where $\tau = 0$ or 1 is the time shift. For $\tau = 0$, it can be calculated by solving the Lyapunov equation (for example using the Bartell-Stewart algorithm):

$$JQ^0 + Q^0 J^\dagger + \Sigma = 0 . \quad (5)$$

For $\tau > 0$, it is simply

$$Q^\tau = Q^0 \expm(J^\dagger \tau) . \quad (6)$$

Here J is the Jacobian of the dynamical system and depends on the time constant τ_x and the network effective connectivity:

$$J_{ij} = \frac{-\delta_{ij}}{\tau_x} + C_{ij} , \quad (7)$$

where δ_{ij} is the Kronecker delta and the superscript \dagger denotes the matrix transpose; **expm** denotes the matrix exponential. In practice, we use the two time shifts $\tau = 0$ and $\tau = 1$ TR, as this is sufficient information to uniquely infer the network parameters.

2.3 Parameter estimation procedure

In order to invert the model (i.e., the model FC reproduces the experimental FC), we iteratively tune the parameters to reduce the model error defined as

$$E = \frac{1}{2} \frac{\sum_{i,j} (\Delta Q_{ij}^0)^2}{\sum_{i,j} (\hat{Q}_{ij}^0)^2} + \frac{1}{2} \frac{\sum_{i,j} (\Delta Q_{ij}^1)^2}{\sum_{i,j} (\hat{Q}_{ij}^1)^2}, \quad (8)$$

where each term - for FC0 and FC1 - is the matrix distance between the model and the data observables, normalized by the norm of the latter; the difference matrices $\Delta Q^0 = \hat{Q}^0 - Q^0$ and $\Delta Q^1 = \hat{Q}^1 - Q^1$. The desired Jacobian update is the matrix

$$\Delta J^\dagger = (Q^0)^{-1} [\Delta Q^0 + \Delta Q^1 \mathbf{expm}(J^\dagger \tau)], \quad (9)$$

which decreases the model error E at each optimization step, similar to a gradient descent. The best fit corresponds to the minimum of E . Finally, the connectivity update is

$$\Delta C_{ij} = \eta_C \Delta J_{ij} \quad (10)$$

for existing connections only; other weights are forced at 0. We also impose non-negativity for the EC values during the optimization. To take properly the effect of cross-correlated inputs into account, we adjust the Σ update from the heuristic update in [1]:

$$\Delta \Sigma = -\eta_\Sigma (J \Delta Q^0 + \Delta Q^0 J^\dagger). \quad (11)$$

As with C for non-existing connections, off-diagonal elements of Σ are kept equal to 0 at all times.

In numerical simulations, we use $\eta_C = 0.0005$ and $\eta_\Sigma = 0.05$.

The optimization code is available with the empirical data on github.com/MatthieuGilson/EC_estimation.

3 Analysis of vectorized EC, corrFC

3.1 Similarity between sessions

We used Pearson correlation coefficient (PCC) as a measure of similarity, both within- and between-subject. To compute the similarity between two sessions, the connectivity matrix C^i for each session i was transformed into a vector v^i by extracting the lower triangle for corrFC, and by applying the SC mask for EC. Each vector v^i had a size of 1x6670 for FC data, and 1×4056 for EC (116 ROIs and 6,670 connectivity links for FC and 4,056 for EC). After vectorization, the

values of all the links were z-scored, using the mean and standard deviation of each v^i :

$$\hat{v}_k^i = \frac{v_k^i - \text{mean}_k(v_k^i)}{\text{std}_k(v_k^i)} ; \quad (12)$$

Then, we computed the similarity between every pair of sessions in data set A1 (Figure 2A):

$$S^{ij} = \text{PCC}(\hat{v}^i, \hat{v}^j) \quad (13)$$

The within-subject similarity was obtained by computing the PCC between every possible pair of vectors, v^i and v^j (for $i \neq j$) from the same subject. For computing the between-subject similarity of datasets A1 and B, the similarity was obtained between all the different combinations of pairs of vectors from different subjects. Figure 2B shows the distributions of the similarity after splitting into within- (in blue) and between-subject (in red) the coefficients for both corrFC and EC. Dataset A2 was used as well to obtain another between-subjects similarity distribution (in green).

3.2 Silhouette coefficient

The silhouette coefficient is defined for each element of a population of z-scored vectors $\hat{v}^{i,s}$ with indices i for the session and s for the subject. Here, each subject s is a cluster and the similarity in Eq. (13) is taken as the metric, but the indices i and j are replaced by doublets of the type (i, s) here. For a given data point i, s , we have the average similarity within his own cluster s defined as

$$a^{i,s} = \text{mean}_{j \neq i}(S^{(i,s)(j,s)}) , \quad (14)$$

and the maximum - over all other clusters - of the same average similarity, but with elements from another cluster:

$$b^{i,s} = \max_{s'}[\text{mean}_k(S^{(i,s)(k,s')})] . \quad (15)$$

The silhouette is then given by the following contrast between the cohesion of the element within its cluster ($a^{i,s}$) and the separation from other clusters ($b^{i,s}$):

$$\sigma^{i,s} = \frac{b^{i,s} - a^{i,s}}{\max(a^{i,s}, b^{i,s})} . \quad (16)$$

Values of silhouette range to 1 for fully separated clusters to -1 for overlapping clusters.

3.3 Dimensionality analysis

To study visually how the data is spread over such a high-dimensional space, we applied PCA to extract the main dimensions reducing the original data. After applying PCA to the whole dataset A1 (6 subjects, 40-50 sessions per subject) the first 6 components (PC1, 2, 3, and PC4, 5, 6) of both FC and EC measures,

are plotted in figure 2C. The silhouette coefficient was computed for each point in these clouds and plotted in figure 2D. The reason for choosing only the first 6 PCs is because the mean silhouette coefficient of EC data reaches a maximum for this value, and then starts to decrease. The silhouette of FC data needs one or two more PCs to reach a maximum (sup.). The same method was applied to the data set B (30 subjects, 10 sessions per subject), and the maximum of the mean silhouette was at 30 PCs (supp). Both distributions of the silhouette, dataset A1 with the first 6 PCs and dataset B with the first 30 PCs, are plotted in Figure 2E.

3.4 Classification of subjects

Figure 3A shows the classification procedure applied to identify subjects using connectivity measures and estimates. First, a fixed number of sessions per subject were selected from each data set, corrFC and EC matrices. These matrices were vectorized as before and individually z-scored, using the mean and standard deviation of each v^i . Then, the corresponding classifier (kNN or MLR) was trained using two different approaches: with and without applying PCA.

The k-NN classifier is a non-linear technique that assigns to new sample the class closest in distance (fig. S4). In our case, we use the same similarity measure used before as a metric. The database has an equal number of sessions per subject (for both corrFC and EC), from 1 to 40. The identity of the remaining sessions from the target group is predicted computing the PCC between each target session and all the sessions in the database and taking the identity with the maximum PCC (Eq. 17). The performance is computed after revealing all the predicted identities.

$$ID_j^* = \operatorname{argmax}(PCC_{j,1}, PCC_{j,2}, \dots, PCC_{j,l}) , \quad (17)$$

where ID^* is the predicted identity for the session j , and l is the number of sessions in the database.

The MLR classifier is a classical tool in machine learning. The parameters (or regressors) of the model are adjusted in order to predict the probabilities of new samples of belonging to each category or class. We used the train set with equal number of sessions per subject to obtain this parameters. The accuracy of the classifier is evaluated by the accuracy of predicting the class of new incoming sessions.

PCA is a preprocessing step applied commonly in machine learning in order to remove noise while keeping the dimensions that capture most of the variance of the data. When applying PCA, the original high-dimensional data is projected into a new space of low dimension. We obtained the performance of the classification with and without applying PCA, increasing the number of PCs or dimensions used to train and test the classifier (Fig. Supp).

We used dataset A1 to study the effect of increasing the samples in the train set, and dataset B to extend the results to a larger group of subjects. The same

method was also applied to dataset C to classify both subjects and condition. Figure 3B shows the accuracy of both classifiers on the two data sets. The curves were obtained after iterating over different sessions and subjects 100 times with cross-validation (mean and standard deviation are plotted).

3.5 Extraction of most discriminative networks

We studied the links that mostly support the classification. For datasets A1 and B these links refer to the classification of subjects, while for dataset C we extracted both subjects and condition support links. In order to extract the supporting links we build upon feature selection methods commonly used in machine learning. In particular we used Recursive Feature Elimination (RFE) to rank links according to their relevance for the classification xxx (Guyon et al. 2002). RFE fits the MLR classifier to the full dataset and removes the link with the smallest weight; the procedure is repeated recursively on shrinking subsets of links until only one is left. This procedure gives a ranking of links according to their relevance for the classification.

We then evaluated the accuracy of MLR on the test set when increasing the number of features following the order given by the RFE ranking. The test set was composed of 10% of randomly chosen samples and train set of the remaining 90%. The fitting/testing was repeated 100 times for different randomly chosen test sets. We selected the number of features for which the mean test set accuracy was maximum. In order to find the maximum we chose the number of features for which the numerical derivative of the mean was less than 10^{-6} . In order to reduce the impact of fluctuations we smoothed the series of means with a rolling average of width 2 features. Since the accuracy is expected to increase initially as a function of the number of features and then either saturate or decrease, this method allow us to find the number of features for which the classifier performance is maximum or adding more features has no practical benefit.

k-NN cannot be used for RFE since it does not estimate weights associated to links. In principle k-NN could be used as a model in a wrapper but given the high amount of features of our setting, wrappers cannot be evaluated on all subsets of features ($\sim 10^{300}$ model fitting/testing would be required for 1000 features). Wrappers may rely on approximate greedy algorithms, for example, eliminating the feature that scores worst. It is well known that greedy algorithms might produce inappropriate solutions if the problem does not have optimal substructure. In addition the computation time almost scales as p^2 , while for RFE it is linear in p .

3.6 Software

Classification and feature extraction analyses were performed in Python together with scikit-learn library for machine-learning routines. Connectivity measures and estimates, as well as similarity analyses were performed in MATLAB

2016 (REF). Network plots in Figure 5 of the main text were done in Gephi 0.9.1.

References

- [1] M. Gilson, R. Moreno-Bote, A. Ponce-Alvarez, P. Ritter, and G. Deco. Estimation of directed effective connectivity from fMRI functional connectivity hints at asymmetries of cortical connectome. *PLoS Comput Biol*, 12:e1004762, 2016.
 - [2] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns. Mapping the structural core of human cerebral cortex. *PLoS Biol*, 6:e159, 2008.
 - [3] A. Messé, D. Rudrauf, H. Benali, and G. Marrelec. Relating structure and function in the human brain: relative contributions of anatomy, stationary dynamics, and non-stationarities. *PLoS CComput Biol*, 10:e1003530, 2014.
 - [4] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15:273–289, 2002.
- ref xxx Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines, *Machine Learning* 46: 389?422.