A mapping between structural and functional brain networks: **Supplementary Information**

J. Meier[*], P. Tewarie[†], A. Hillebrand[‡], L. Douw[§,¶], B.W. van Dijk[‡], S.M. Stufflebeam[¶,‖], and P. Van Mieghem[*]

Contents

[*]Faculty of Electrical Engineering, Mathematics and Computer Science, P.O Box 5031, 2600 GA Delft, The Netherlands; *email*: J.M.Meier@tudelft.nl (corresponding author), P.F.A.VanMieghem@tudelft.nl

[†]Department of Neurology, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands; *email*: p.tewarie@vumc.nl

[‡]Department of Clinical Neurophysiology and Magnetoencephalography Center, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands; *email*: a.hillebrand@vumc.nl, bw.dijk@vumc.nl

[§]Department of Anatomy and Neurosciences, VU University Medical Center, Amsterdam, The Netherlands, *email*: douw@nmr.mgh.harvard.edu

[¶]Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging / Massachusetts General Hospital, Boston MA.

[‖]Department of Radiology, Harvard Medical School, Boston MA, *email*: sms@nmr.mgh.harvard.edu

A Theory

We provide here some mathematical background of our method.

A.1 Matrix functions

If $f(z)$ is a function of the complex number z and analytic in a disk with radius R around z_0, then it can be shown (see Section 2.2) that for all matrices A, the matrix function $f(A)$ can be expressed as a polynomial of degree at most $N-1$,

$$f(A) = \sum_{k=0}^{N-1} c_k[f] A^k$$

where the coefficients $c_k[f]$ can be specified as

$$c_k[f] = \frac{1}{k!} \sum_{m=1}^{N} \frac{f(\lambda_m)}{\prod_{j=1;j\neq m}^{n} (\lambda_m - \lambda_j)} \frac{d^k}{dx^k} \prod_{j=1;j\neq m}^{N} (x - \lambda_j) \Bigg|_{x=0}$$

Only if f is a polynomial of degree m at most $N-1$, we find, for all $0 \leq k \leq N-1$, that

$$f_k(0) = c_k[f],$$

where $f_k(0)$ is the k-th coefficient of the Taylor series of f for the development point $z_0 = 0$.

A.2 Role of the diameter

An explanation for the importance of the diameter could be as follows: The matrix powers I, A, A^2, A^3,... are all linearly independent of each other up to A^ρ (Van Mieghem, 2011). For higher powers, we cannot be sure of their dependency. Our analysis shows that using higher powers of the structural connectivity matrix than its diameter ($\rho = 6$) does not improve the goodness of fit of our estimation. Furthermore, in the binary matrix A, the sum $\sum_{k=1}^{m} A^k$ can have zero entries for all $m < \rho$. Reducing the number of zero entries can also be a reason why the goodness of fit increases gradually until adding A^ρ and then converges.

A.3 Analysis in the spectral domain

If a mapping postulated in the previous section is valid, then such a mapping should also hold in the spectral domain (Van Mieghem, 2011). Provided A is symmetric such that $A = X\Lambda X^T$, where the matrix X contains the eigenvectors of A in the columns and $\Lambda = \text{diag}(\lambda_k)$, $1 \leq k \leq N$, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ the real eigenvalues of A, then there exists an alternative to compute $f(A)$. Since eigenvectors are orthogonal, the matrix X is an orthogonal matrix satisfying $X^T X = I$ and $XX^T = I$, where the latter follows from the fact that $X^{-1} = X^T$ and the fact that a matrix and its

inverse commute. We assume here that all eigenvalues are different, as is the case in most real-world networks (Van Mieghem, 2011). Then,

$$f(A) = Xf(\Lambda)X^T = \sum_{k=0}^{N} f(\lambda_k) x_k x_k^T \tag{1}$$

where x_k is the eigenvector of A belonging to the eigenvalue λ_k. Using the spectral form (1) in our assumption (1) reveals that

$$W = Xf(\Lambda)X^T, \tag{2}$$

Since W is symmetric, the spectral decomposition equals

$$W = Y\Upsilon Y^T \tag{3}$$

where Y is the orthogonal matrix containing the eigenvectors $y_1, y_2, \ldots, y_N$ belonging to the eigenvalues $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_N$ and $\Upsilon = \text{diag}(\mu_k), 1 \leq k \leq N$ is the diagonal matrix of the eigenvalues of W. Using the properties of orthogonal matrices, we find from (2) the diagonal matrix $f(\Lambda) = \text{diag}(f(\lambda_k))$ as

$$f(\Lambda) = X^T W X.$$

With (3), we obtain

$$f(\Lambda) = \left(X^T Y\right) \Upsilon \left(Y^T X\right). \tag{4}$$

Since both $f(\Lambda)$ and Υ are diagonal matrices and $\left(X^T Y\right) = \left(Y^T X\right)^{-1}$, there must hold that

$$f(\lambda_k) = \mu_k \tag{5}$$

for each $1 \leq k \leq N$. Furthermore, from (4) follows that

$$\left(Y^T X\right) f(\Lambda) = \Upsilon \left(Y^T X\right). \tag{6}$$

From (5) and the fact that we multiply the matrix $Y^T X$ with the same diagonal matrix from both sides, we can conclude that (under the condition that all eigenvalues are different) $Y^T X = I$ and therefore $X = Y$.

In summary, equation (1) implies that both the structural matrix A and the functional matrix W must have the same eigenvectors and that the function f maps the eigenvalues (ordered) of A onto those of W (also ordered, see (5)). Moreover, (5) shows that $f(x)$ is non-decreasing in x.

If the empirical matrices have different eigenvectors, then that difference may be due to noise of the measurement. If the difference cannot be explained by noise perturbation, our assumption in (1) that there exists an analytic function f needs to be revisited and a more general form of our fitting function can be considered with matrix coefficients instead of scalar ones

$$F(A) = \sum_{k=0}^{\infty} F_k (A - z_0 I)^k \tag{7}$$

where F_k are $N \times N$ matrices, which reduces when $N = 1$ again to our previous assumption (3).

B Data Analysis

B.1 Visualization of best fits

In Figures 1, 2 and 3, we visualized the best fitted matrices of the group-averaged data set on a bigger scale to better localize the biggest discrepancies between our best fits and the empirical findings. The numbers on the axes of the matrices, 1 to 78 refer to specific brain regions. You can find on the next page a complete list of the regions of interest (ROIs) that we display mostly with numbers:

1	Rectus-L	40	Rectus-R
2	Olfactory-L	41	Olfactory-R
3	Frontal-Sup-Orb-L	42	Frontal-Sup-Orb-R
4	Frontal-Med-Orb-L	43	Frontal-Med-Orb-R
5	Frontal-Mid-Orb-L	44	Frontal-Mid-Orb-R
6	Frontal-Inf-Orb-L	45	Frontal-Inf-Orb-R
7	Frontal-Sup-L	46	Frontal-Sup-R
8	Frontal-Mid-L	47	Frontal-Mid-R
9	Frontal-Inf-Oper-L	48	Frontal-Inf-Oper-R
10	Frontal-Inf-Tri-L	49	Frontal-Inf-Tri-R
11	Frontal-Sup-Medial-L	50	Frontal-Sup-Medial-R
12	Supp-Motor-Area-L	51	Supp-Motor-Area-R
13	Paracentral-Lobule-L	52	Paracentral-Lobule-R
14	Precentral-L	53	Precentral-R
15	Rolandic-Oper-L	54	Rolandic-Oper-R
16	Postcentral-L	55	Postcentral-R
17	Parietal-Sup-L	56	Parietal-Sup-R
18	Parietal-Inf-L	57	Parietal-Inf-R
19	SupraMarginal-L	58	SupraMarginal-R
20	Angular-L	59	Angular-R
21	Precuneus-L	60	Precuneus-R
22	Occipital-Sup-L	61	Occipital-Sup-R
23	Occipital-Mid-L	62	Occipital-Mid-R
24	Occipital-Inf-L	63	Occipital-Inf-R
25	Calcarine-L	64	Calcarine-R
26	Cuneus-L	65	Cuneus-R
27	Lingual-L	66	Lingual-R
28	Fusiform-L	67	Fusiform-R
29	Heschl-L	68	Heschl-R
30	Temporal-Sup-L	69	Temporal-Sup-R
31	Temporal-Mid-L	70	Temporal-Mid-R
32	Temporal-Inf-L	71	Temporal-Inf-R
33	Temporal-Pole-Sup-L	72	Temporal-Pole-Sup-R
34	Temporal-Pole-Mid-L	73	Temporal-Pole-Mid-R
35	ParaHippocampal-L	74	ParaHippocampal-R
36	Cingulum-Ant-L	75	Cingulum-Ant-R
37	Cingulum-Mid-L	76	Cingulum-Mid-R
38	Cingulum-Post-L	77	Cingulum-Post-R
39	Insula-L	78	Insula-R

We can see that the patterns of the fitted matrices seem to be similar to the empirical ones and that the value range is overlapping. Therefore, we can conclude that also only from visual inspections of the fitted matrices our mapping seems to be convincingly accurate.

B.2 Spectral data analysis

After plotting the eigenvalue couples (λ_k, μ_k) in a scatter plot, we can obtain the function f in the spectral domain (see (5)). Polynomial functions were fitted to all possible combinations of scatter plots (i.e. for the combinations structure-function (MEG/fMRI) and vice versa) by minimizing the sum of squared errors. An example of such a fit is depicted in Figure 4.

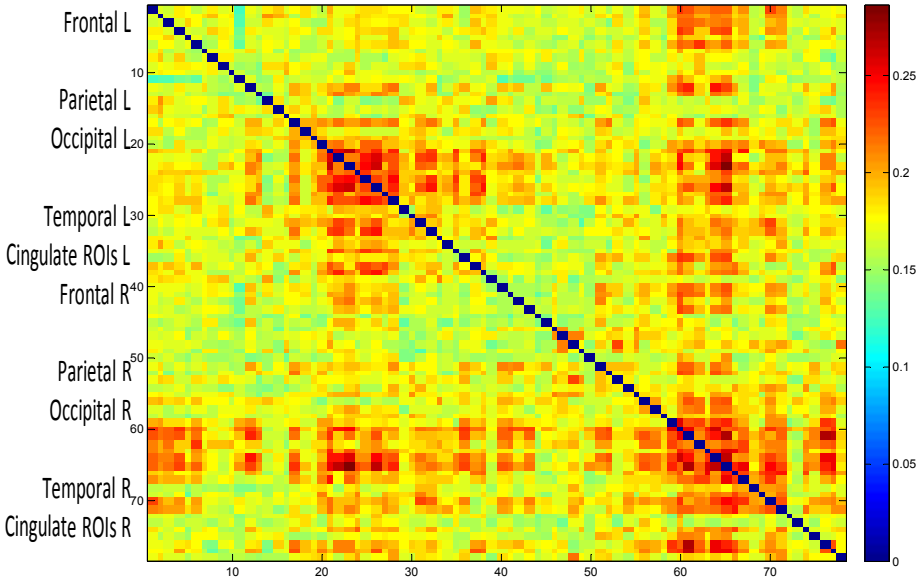
For the goodness of fit in the spectral domain, we computed the adjusted R^2 value for the different mappings (see Figure 5). Overall, we reached already for $K \geq 4$ with all combinations of matrices an adjusted R^2 value of higher than 0.9 indicating a good fit of our mapping. We followed the same approach here as we did for the topology domain, and reversed A and W to repeat the spectral analysis for the function $f^{-1} : W \to A$. Results of this analysis can also be found in Figure 5. For this spectral approach, the adjusted R^2 value did not improve much after adding the same number of terms as was used for the mathematical function in the topology domain (compare Figures 5 with Figure 2). Thus, a functional expression with $K = 5$ was again sufficient for the analyzed mappings. This conclusion held for both modalities (fMRI and MEG) and for both functions $f : A \to W$ and $f^{-1} : W \to A$.

Since we conducted a similar analysis in the topology and in the spectral domain respectively for the functions f and f^{-1}, we compared the estimated coefficient values for the spectral and topology domain. For a correct comparison, we must omit the error matrix E in (5) in the topology domain. A plot of the estimated coefficient values and confidence intervals (obtained by the least squared parameter estimation in MATLAB) is illustrated in SI Figures 6 and 7 for the mapping of structural to functional matrices and in SI Figures 8 and 9 for the other direction. For the mapping f, we only faced small differences between the coefficients (see SI Figures 6 and 7) and in most cases their confidence intervals were overlapping. But for the other direction, function f^{-1}, we observed quite different estimated coefficients. In the case of $f^{-1} : W_{MEG} \to A$ we obtained large confidence intervals when many (> 5) coefficients were fitted (see SI Figure 8), implying insecure estimations of their exact value. This result is in agreement with the finding that 5 coefficients were sufficient to describe the mapping between W and A (Figure 2), and that these extra coefficients did not contribute relevant information to the mapping. The discrepancies between the estimated coefficient values in the topology and spectral domain could be originating from the different eigenvectors of the 3 analyzed empirical matrices. These eigenvector perturbations can potentially be caused by noise in the different measurement techniques.

C Comparison of fitted coefficient values for different modalities

In Table 1, we displayed the different estimated values for the coefficients of our mapping for a maximal fitted exponent of $K = 6$ and $z_0 = 0$ for the group-averaged data set. In addition, in Figures 10 and 11 we show the different coefficient values (without the offset estimates) for the two modalities for both mappings together with their 95% confidence intervals, from structure to function and from function

(a) W_{MEG}



(b) $\widetilde{W}_{MEG}$

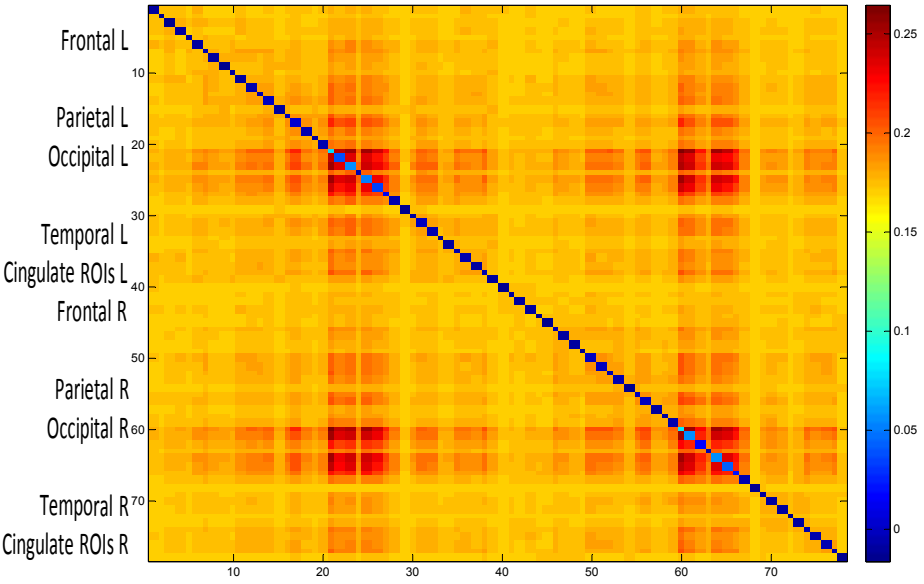
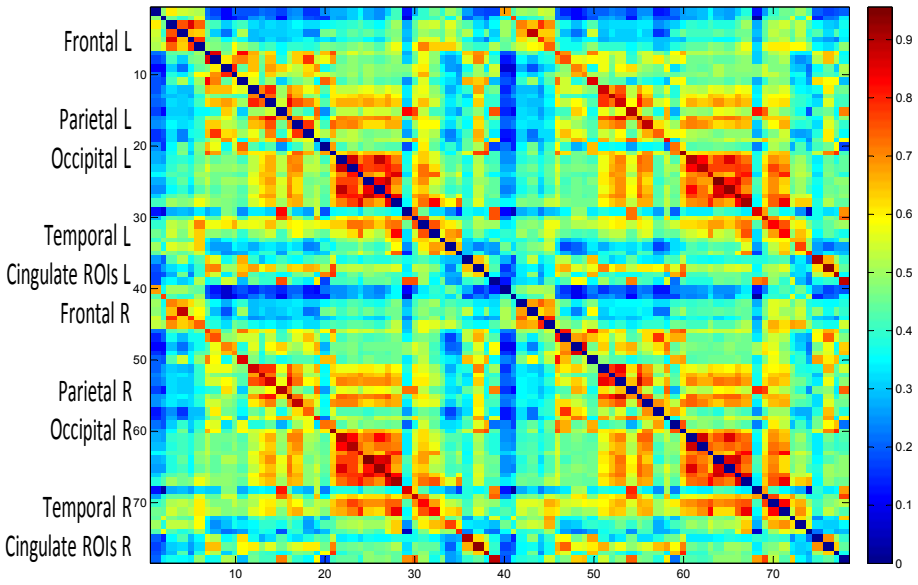


Figure 1: Visualization of the best fits for the function f, which was $\widetilde{W}_{MEG}$ (for $K = 6$ with an error matrix E), under the empirical adjacency matrix W_{MEG} for the group-averaged data set.

(a) W_{fMRI}



(b) $\widetilde{W}_{fMRI}$

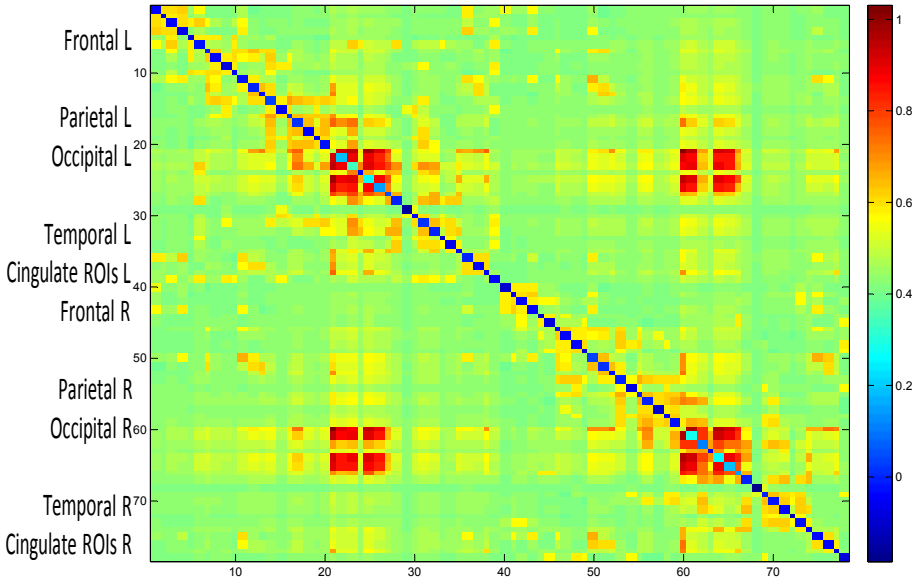
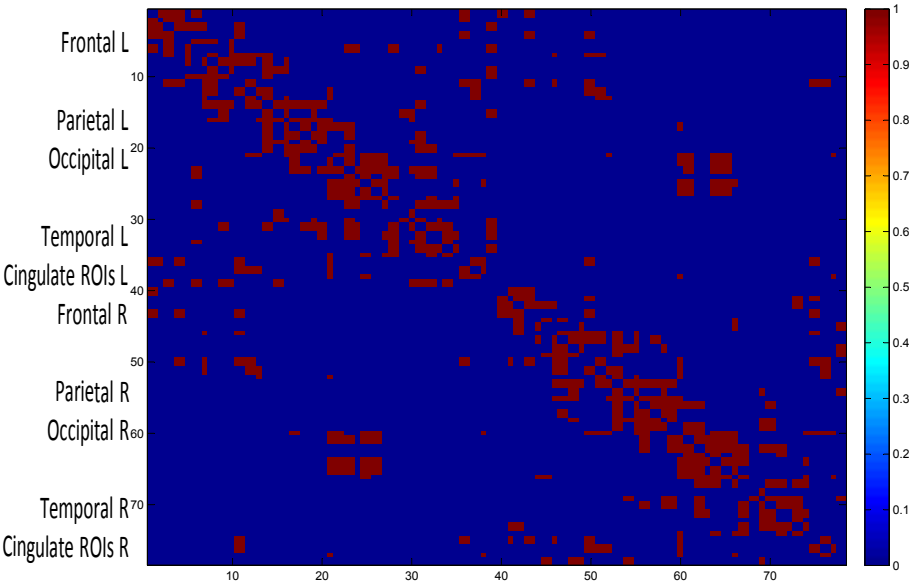


Figure 2: Visualization of the best fits for the function f, which was $\widetilde{W}_{fMRI}$ (for $K = 6$ with an error matrix E), under the empirical adjacency matrix W_{fMRI} for the group-averaged data set.

(a) A



(b) $\widetilde{A}$

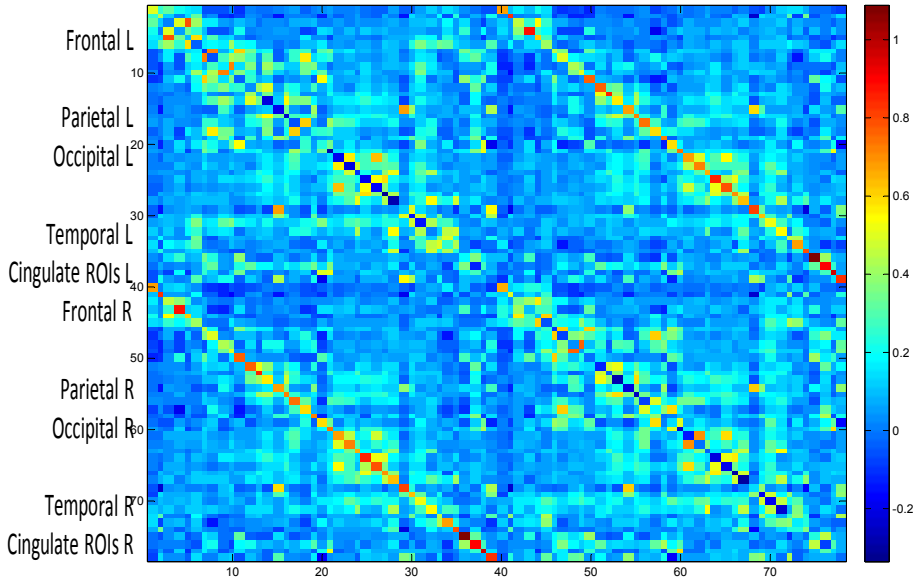


Figure 3: Visualization of the best fits for the function f^{-1}, which was $\widetilde{A}$ (again for $K = 6$ with an error matrix E) under the empirical adjacency matrix A for the group-averaged data set.

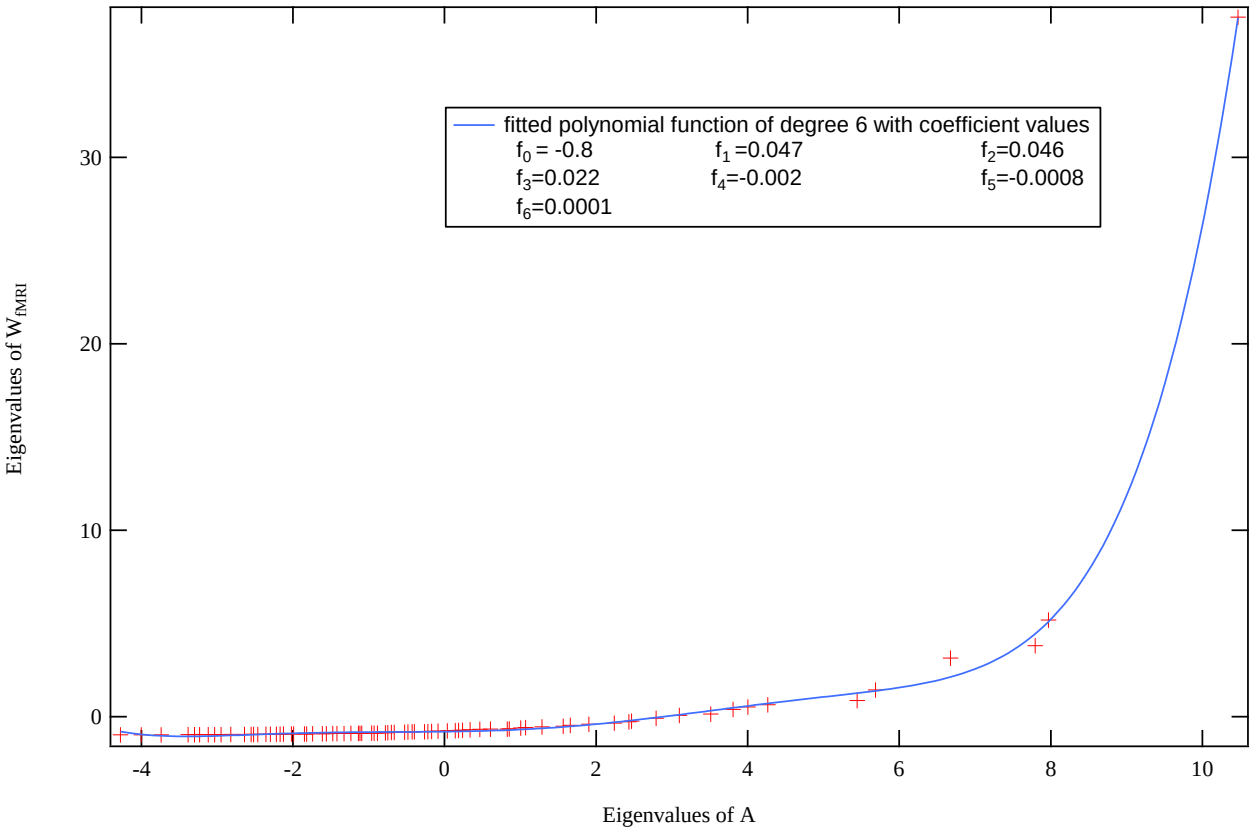


Figure 4: Scatter plot of the eigenvalues of the structural matrix A against the eigenvalues of W_{fMRI} with a least-squared fitted polynomial function with a maximal fitted exponent of $K = 6$ (including the intercept f_0) for the group-averaged data set.

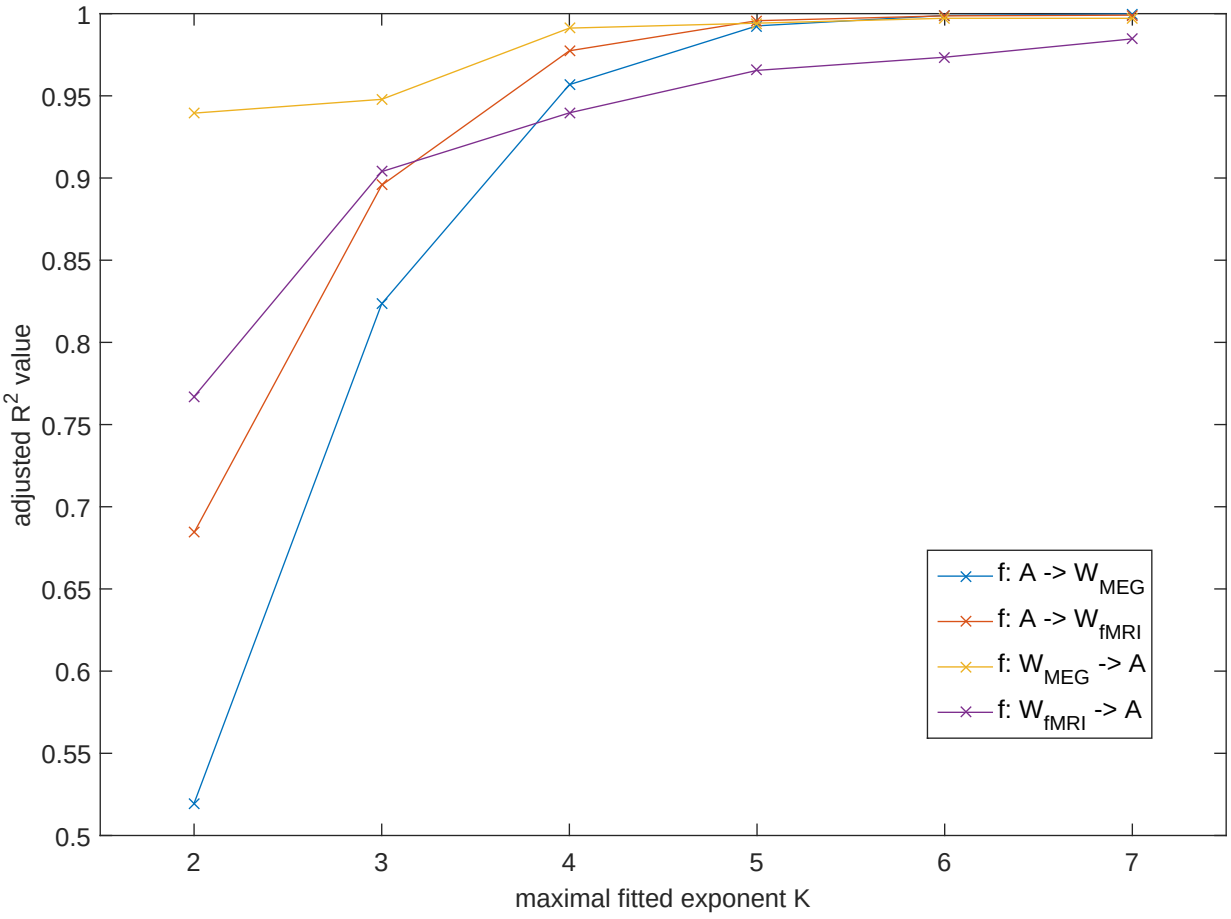


Figure 5: Adjusted R^2 value of the different mappings in the spectral domain for different maximal fitted exponents K for the group-averaged data set for $z_0 = 0$.

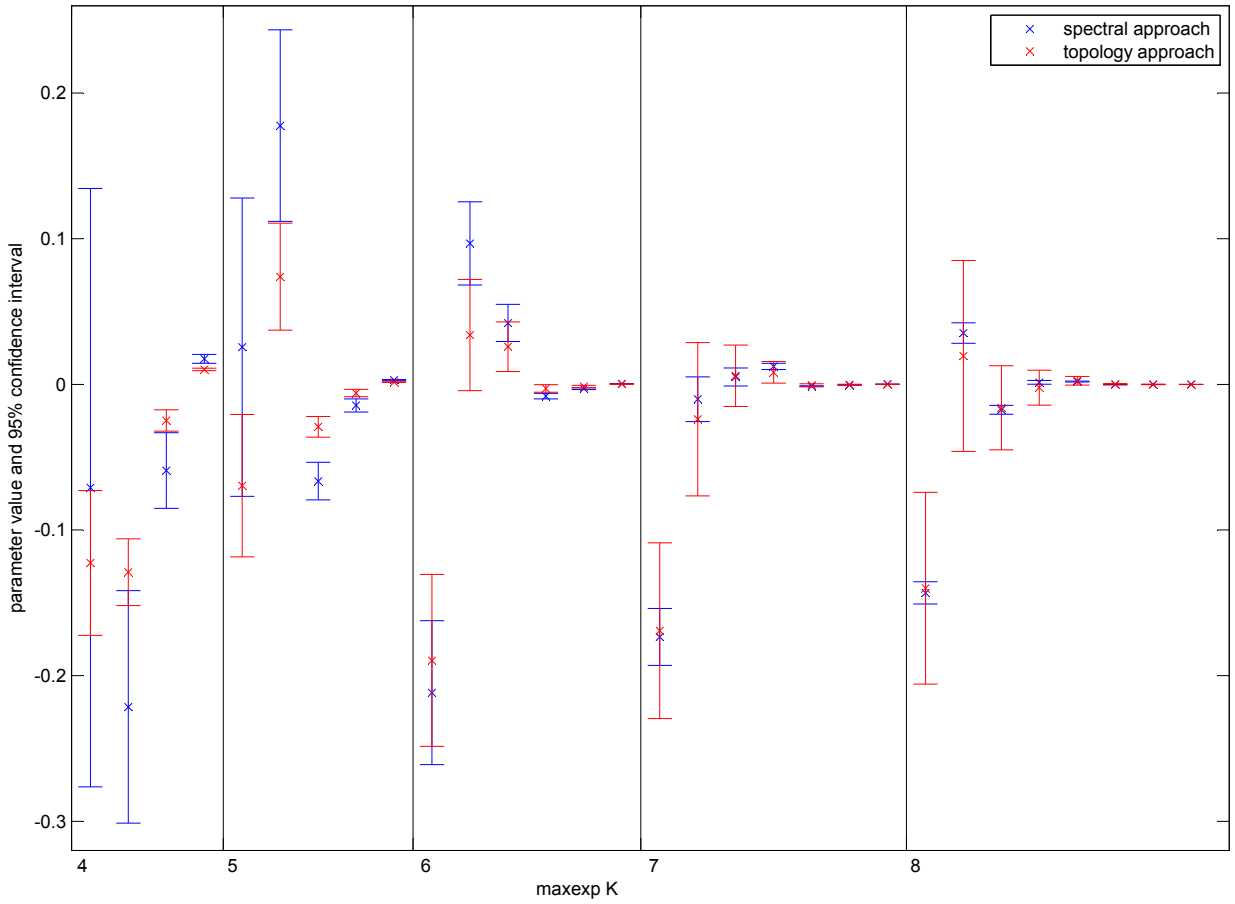


Figure 6: Plot of the estimated coefficient values for different maximal exponents K with their 95% confidence interval as an errorbar for the mapping $f : A \rightarrow W_{MEG}$ for the group-averaged data set. The spectral and topology approach are marked in blue and red, respectively. Note that, the confidence intervals of the coefficient values overlap and the distance between the estimated coefficient values is becoming smaller when more coefficients are used.

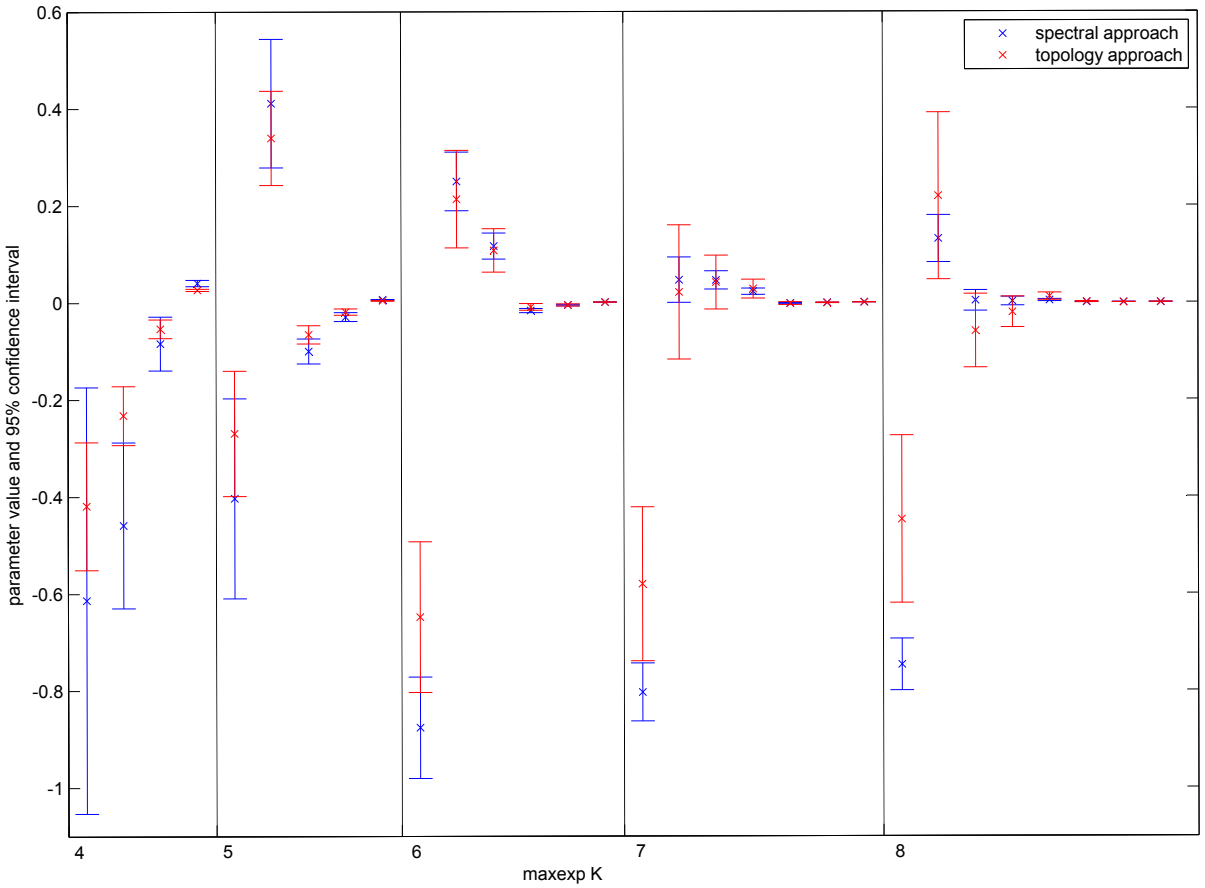


Figure 7: Plot of the estimated coefficient values for different maximal exponents K with their 95% confidence interval as an errorbar for the mapping $f : A \to W_{fMRI}$ for the group-averaged data set. The spectral and topology approach are marked in blue and red, respectively. Note that, in nearly all cases, the confidence intervals of the coefficient values overlap pointing towards similar estimated coefficients for the spectral and topology domain.

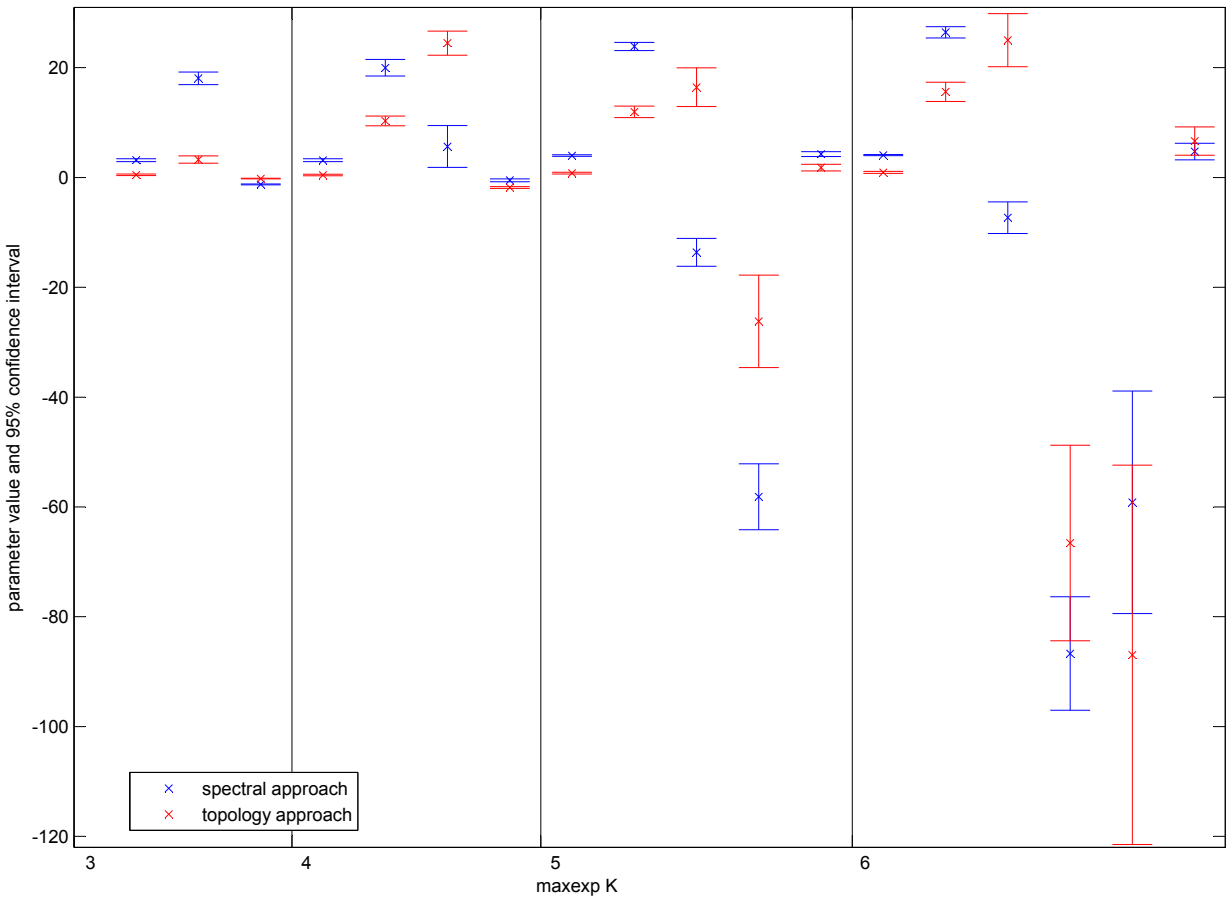


Figure 8: Plot of the estimated coefficient values for different maximal exponents K with their 95% confidence interval as an errorbar for the mapping $f^{-1} : W_{MEG} \to A$ for the group-averaged data set. The spectral and topology approach are marked in blue and red, respectively.

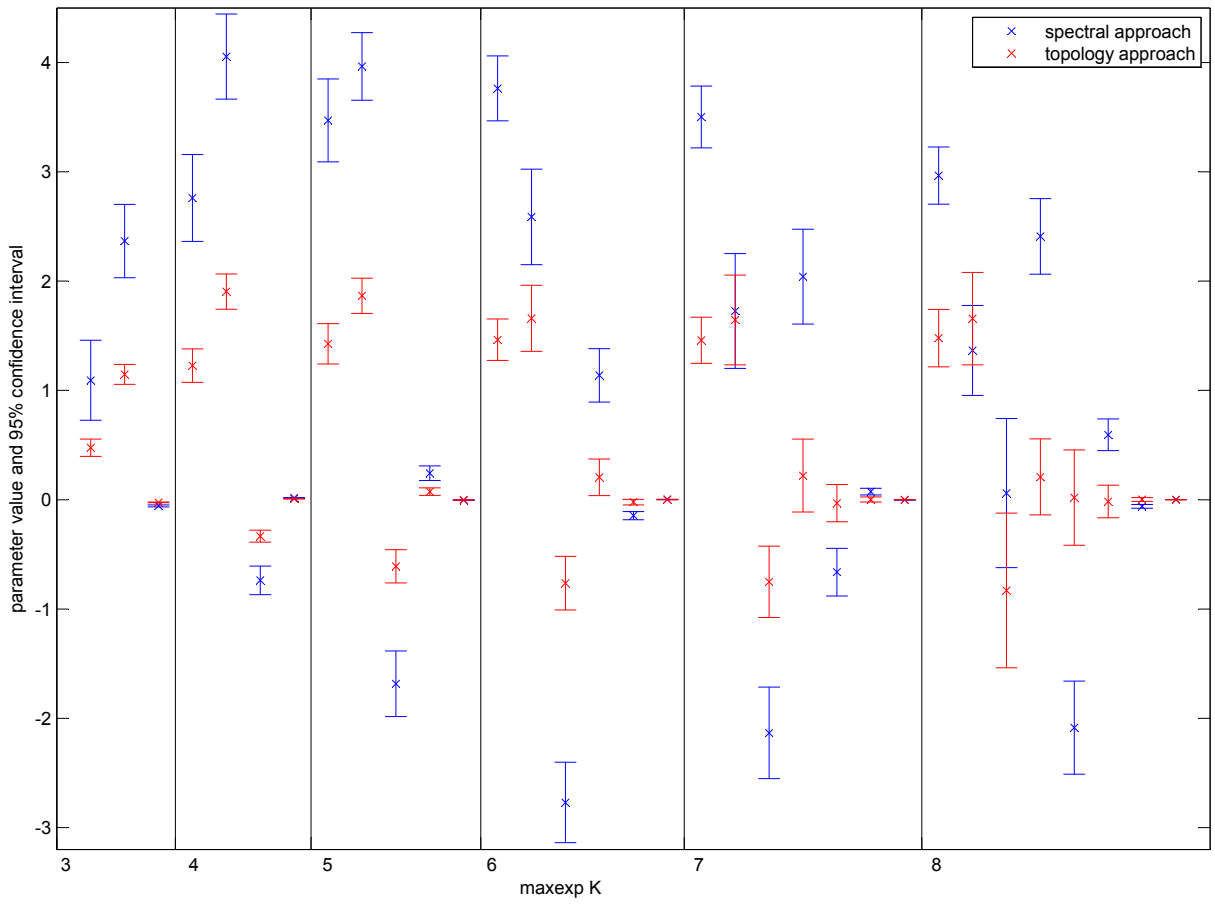


Figure 9: Plot of the estimated coefficient values for different maximal exponents K with their 95% confidence interval as an errorbar for the mapping $f^{-1} : W_{fMRI} \to A$ for the group-averaged data set. The spectral and topology approach are marked in blue and red, respectively.

to structure. In both displayed figures, the 95% confidence intervals for the estimated coefficients of the different modalities do not overlap indicating significantly different coefficients using the 5% significance level. This table and these confidence intervals show clearly the differences between the different analyzed mappings pointing towards a modality-dependent mapping.

Table 1: Estimated coefficient values for a maximal exponent of $K = 6$ and $z_0 = 0$ for the group-averaged data set.

parameter in front of	f: A -> W_{MEG}	f: A -> W_{fMRI}	parameter in front of	f^{-1}: W_{MEG} -> A	f^{-1}: W_{fMRI} -> A
I	-0.186244926	-0.622088682	I	0.931541504	1.472348531
E	0.170142739	0.417817587	E	-0.226229939	-0.098880043
A	-0.002689163	0.073378816	W	15.61376486	1.633796259
A^2	0.000397359	0.027965465	W^2	25.77141943	-0.760008801
A^3	0.00136165	0.010963972	W^3	-64.46444579	0.241736267
A^4	-7.34E-05	-0.000932557	W^4	-92.90312812	-0.03937664
A^5	-5.42E-05	-0.000401534	W^5	-6.670779246	0.002673357
A^6	5.47E-06	3.96E-05	W^6	0.993949647	-4.75E-05

D Interpretation with walks

In Figure 12, we visualized our mapping in terms of the number of walks of the structural brain network (without the error matrix E).

E Comparison with a previous study

In Figures 13 and 14, we compare our results with a previous study (Tewarie et al., 2014). Tewarie and coworkers (Tewarie et al., 2014) demonstrated that the degree product between nodes in the structural network together with the Euclidean distance has the ability to predict the functional connections between these nodes. If we merely focus on the degree product, we observe that our approach with the sum of structural matrices A^k (see Eq. (5)) is correlated with the degree product (SI Figure 13). The correlation between those two measures is indeed positive with a Spearman correlation of $R = 0.57$ (p-value < 0.001). There are two clouds in the scatterplot: the upper cloud corresponds to direct connections whereas the lower cloud corresponds to all possible indirect connections, consisting of all walks larger than one. If we investigate the relationship between the previous model (including degree product and Euclidean distance as predictors for functional connectivity) and the mapping approach from this paper, the Spearman correlation R becomes higher ($R = 0.64$, p-value < 0.001, see Figure 14). This result raises the question whether the Euclidean distance as a separate term in a model for explaining functional connections is required (Alexander-Bloch et al., 2013). In our approach, we only incorporated topological distance, which means the distance with respect to intermediary nodes and links in the structural network, and not Euclidean distance; however, these findings might suggest that topological distance and Euclidean distance between nodes are related.

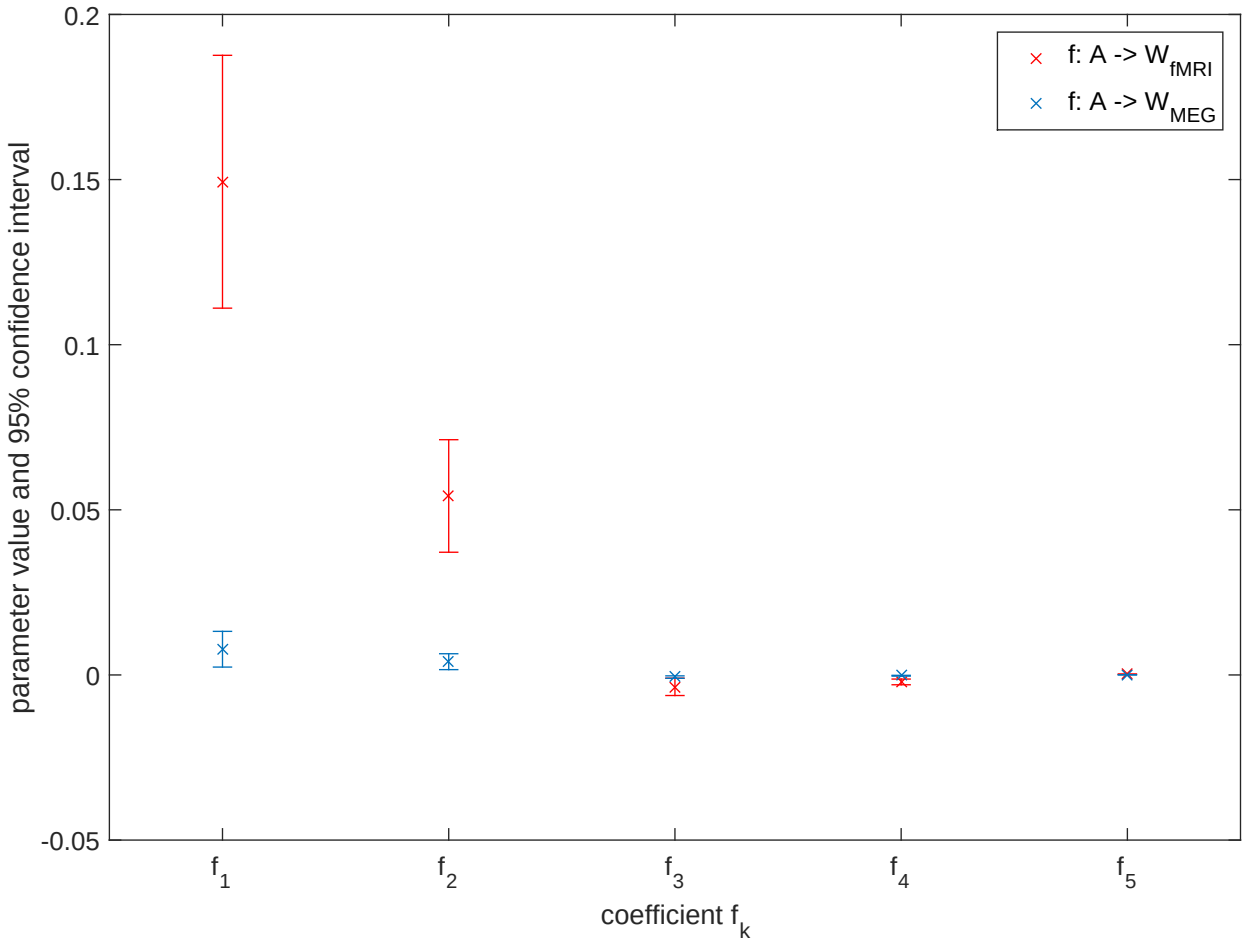


Figure 10: Plot of the estimated coefficient values for maximal fitted exponent $K = 5$ with their 95% confidence interval as an errorbar for the mapping from structural to functional networks for the group-averaged data set. Note that the displayed intervals do not overlap, thus we face here significantly different estimated values.

F Error analysis

The equation (1) assumes that W and A are known exactly. In reality, all types of error mask the true structure so that we actually measure

$$\widetilde{W} = W + \varepsilon_W R_W$$

where R_W is a realization of a random matrix with unit norm and ε_W is the maximum amplitude of the error. Similarly,

$$\widetilde{A} = A + \varepsilon_A R_A$$

and the assumption becomes

$$\widetilde{W} = f\left(\widetilde{A}\right)$$

or

$$W + \varepsilon_W R_W = f\left(A + \varepsilon_A R_A\right)$$

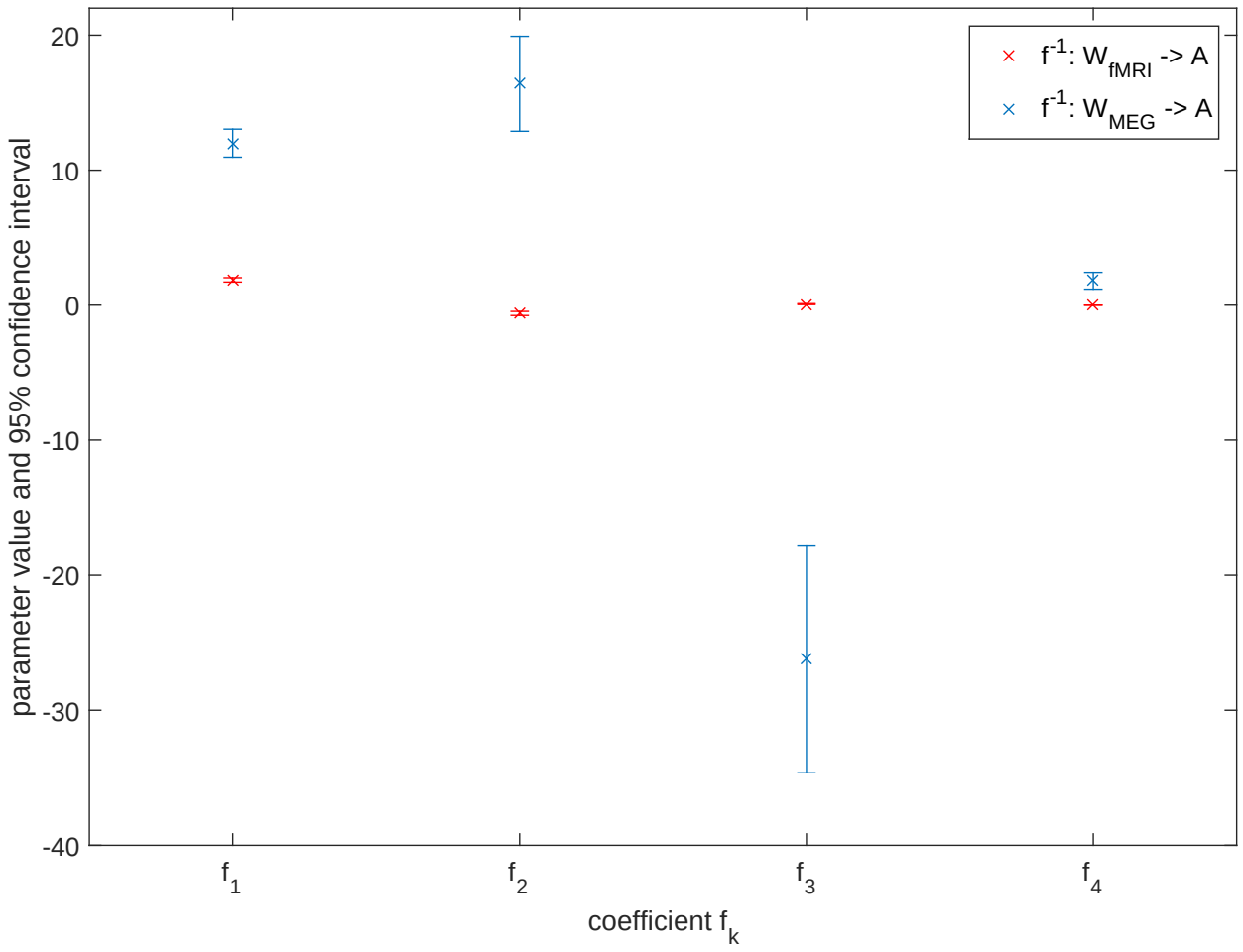


Figure 11: Plot of the estimated coefficient values for maximal exponents $K = 4$ with their 95% confidence interval as an errorbar for the mapping from functional to structural networks for the group-averaged data set. Note that the displayed intervals do not overlap, thus we face here significantly different estimated values.

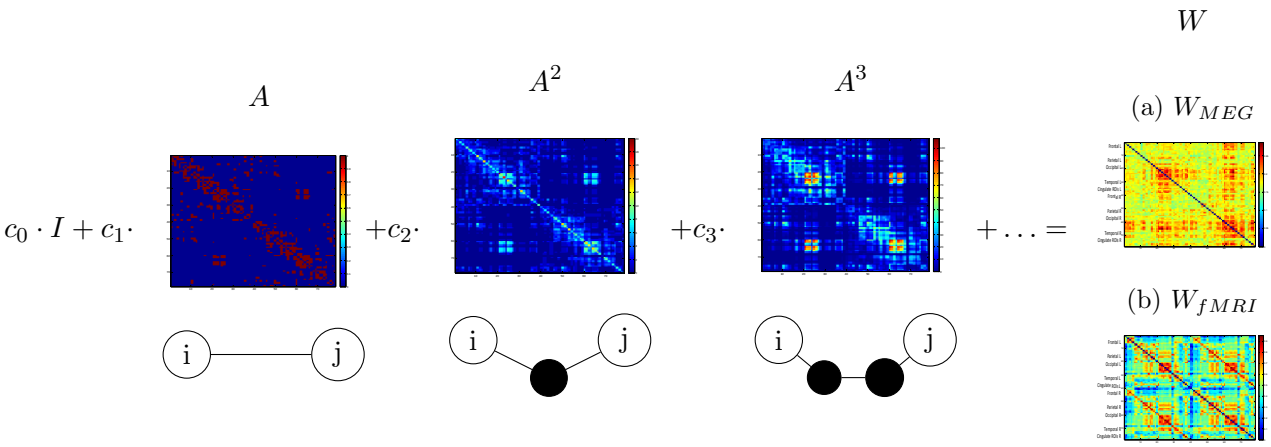


Figure 12: Visualization of a simplified version of our model (from (4) where $c_k = c_k[f]$): Walks of different length between node i and j, first the direct connection, then with one intermediate node, with two and so on, adding up to the functional connectivity matrix.

Using the Taylor expansion (3),

$$f\left(A + \varepsilon_A R_A\right) = \sum_{k=0}^{\infty} f_k\left(A\right) \varepsilon_A^k \left(R_A\right)^k$$

up to first order (assuming that ε_A is sufficiently small!), then

$$W + \varepsilon_W R_W = f\left(A\right) + f_1\left(A\right) \varepsilon_A R_A + O\left(\varepsilon_A^2\right)$$

Invoking the assumption (1) shows a relation between the different types of errors

$$\varepsilon_W R_W = f_1\left(A\right) \varepsilon_A R_A + O\left(\varepsilon_A^2\right)$$

Given that the assumption (1) is correct and that A is known exactly, we could derive a method to improve the measurements $\widetilde{W}$ based on Section SI-A.3, which suggests that all eigenvectors of W are fixed and known (i.e. X is the same as for A), so that $\widetilde{W}$ needs to be modified to incorporate this property. This analysis is a suggestion for future work.

To investigate the influence of false positives in the structural matrix, we randomly added connections (1 % new connections) in the structural matrix of the group-averaged data set and redid the analysis (power series in topology domain with 6 terms). The results can be found in the boxplots of the goodness of fit (SSE_{norm}) in SI Figure 15. If we define the change in SSE_{norm} due to noise as the noise influence NI whereas $NI(SSE_{norm}) := standard\ deviation(SSE_{norm})/mean(SSE_{norm})$, then we can calculated

$$\begin{aligned}
NI(SSE_{norm})(W_{fMRI} \to A) &= 0.0174 \\
NI(SSE_{norm})(W_{MEG} \to A) &= 0.0169 \\
NI(SSE_{norm})(A \to W_{fMRI}) &= 0.0124 \\
NI(SSE_{norm})(A \to W_{MEG}) &= 0.0158.
\end{aligned}$$

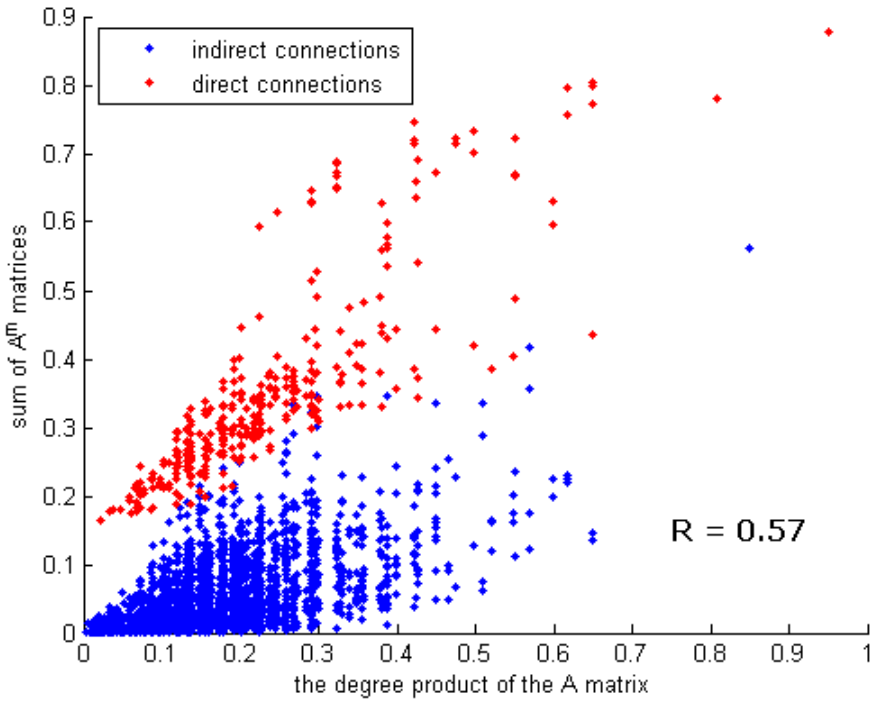


Figure 13: Scatter plot of the structural degree product against the sum of the powers of the structural matrix A (from power $k = 1$ up to $k = 6$).

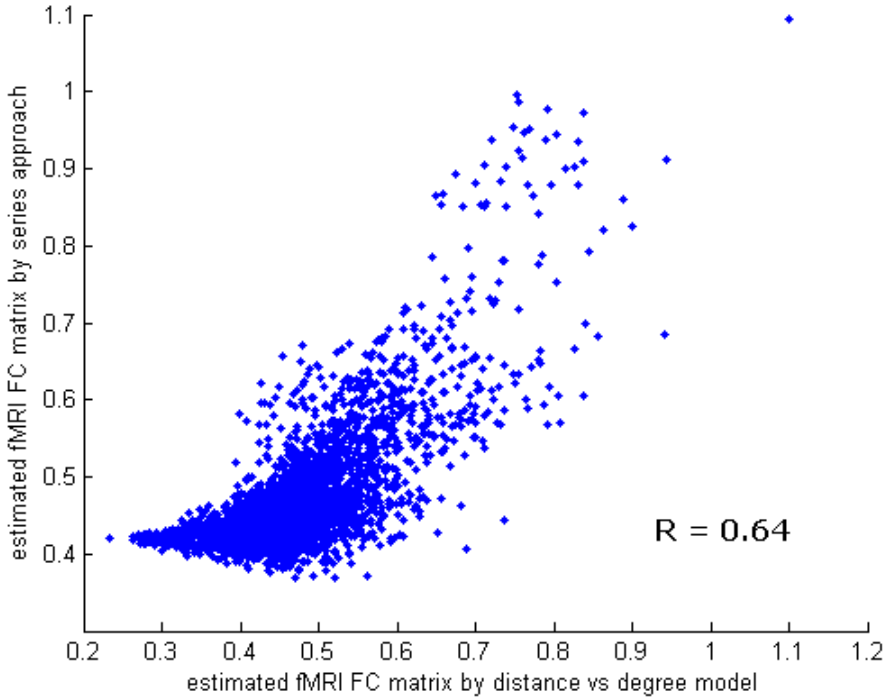
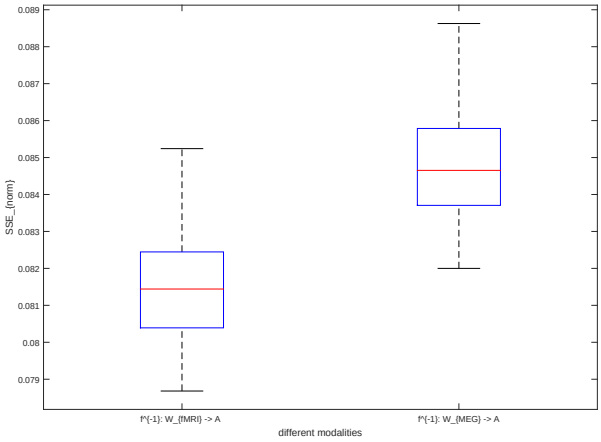
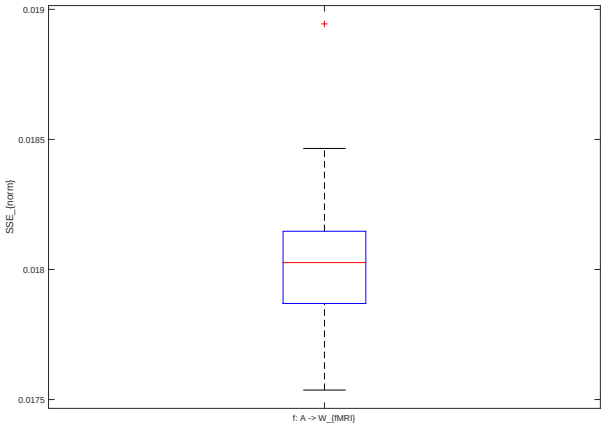


Figure 14: Scatter plot of the estimated fMRI correlation matrix from the distance and degree model described in a previous study (Tewarie et al., 2014) against the estimated values using the mapping approach on the structural matrix A.

(a) $f^{-1} : W_{fMRI} \to A$ and $f^{-1} : W_{MEG} \to A$



(b) $f : A \to W_{fMRI}$



(c) $f : A \to W_{MEG}$

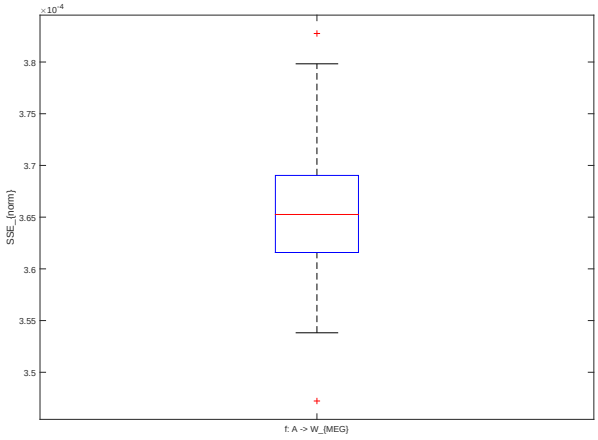


Figure 15: Boxplot of the variations in SSE_{norm} in the topology domain for $K = 6$ with group-averaged data for $z_0 = 0$. To investigate the influence of false positives in the structural connectivity matrix, we randomly added connections (1% new connections) in the structural matrix and redid the analysis.

It can be observed from the above that the change in goodness of fit is small in the presence of little noise (in the order of 1%), thus the mapping does not seem to be sensitive to small noise fluctuations.

If we analyze the region-to-region variability, we find that the inter-hemispheric connections were often quite different between the empirical matrix and its fitted version (see the secondary diagonal in Figures 16 - 19 where we displayed the absolute error). This result confirms our interpretation about the more homologous inter-hemispheric connections in the fitted networks than in the empirically observed networks. Furthermore, we were also interested in which regions benefitted more from an increasing number of coefficients (darker regions in SI Figures 20-23). Those regions that benefitted from an increasing number of fitted coefficients were possibly most influenced by longer walks in the underlying structural network. For the estimated structural and functional networks by our mapping we observe as a result quite a diverse homogenous spreading of the benefitting regions over the entire group of regions except for the diagonal and secondary diagonal. Thus, we can again conclude that the inter-hemispheric connections are probably benefitting most from our mapping approach.

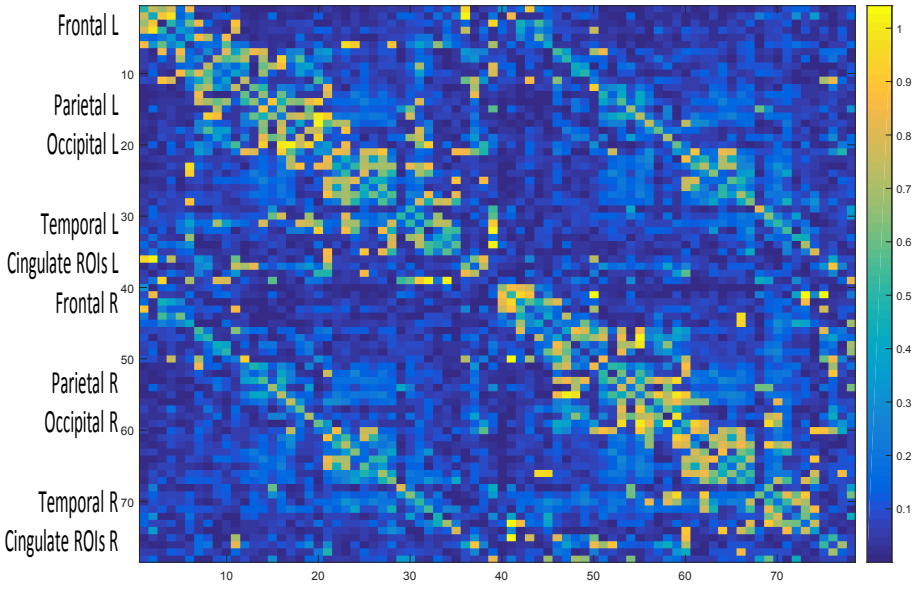


Figure 16: Absolute distance matrix between the best fit for $\widetilde{A} = f(W_{fMRI})$ ($K = 6$, $z_0 = 0$) and the empirical matrix A for the group-averaged data set.

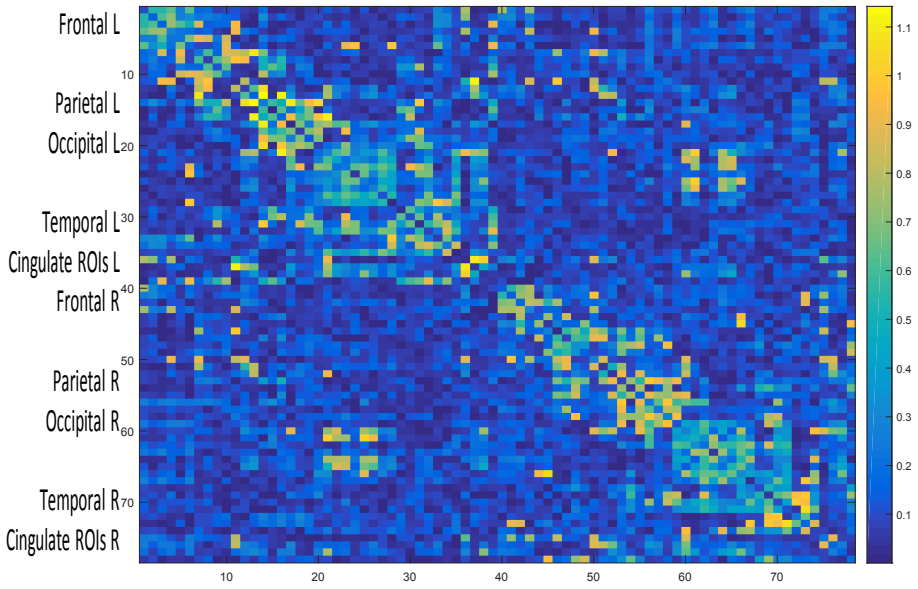


Figure 17: Absolute distance matrix between the best fit for $\widetilde{A} = f(W_{MEG})$ ($K = 6$, $z_0 = 0$) and the empirical matrix A for the group-averaged data set.

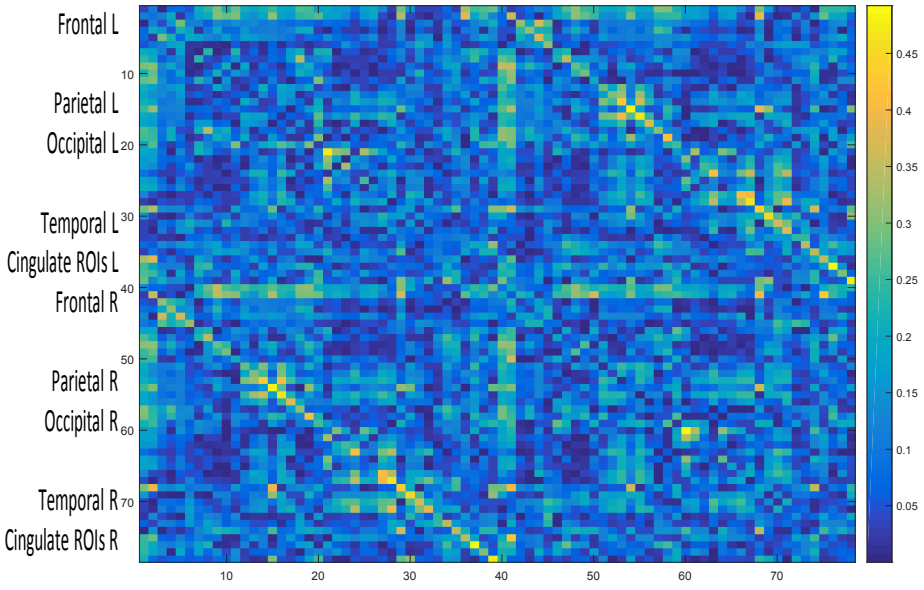


Figure 18: Absolute distance matrix between the best fit for $\widetilde{W}_{fMRI} = f(A)$ ($K = 6$, $z_0 = 0$) and the empirical matrix W_{fMRI} for the group-averaged data set.

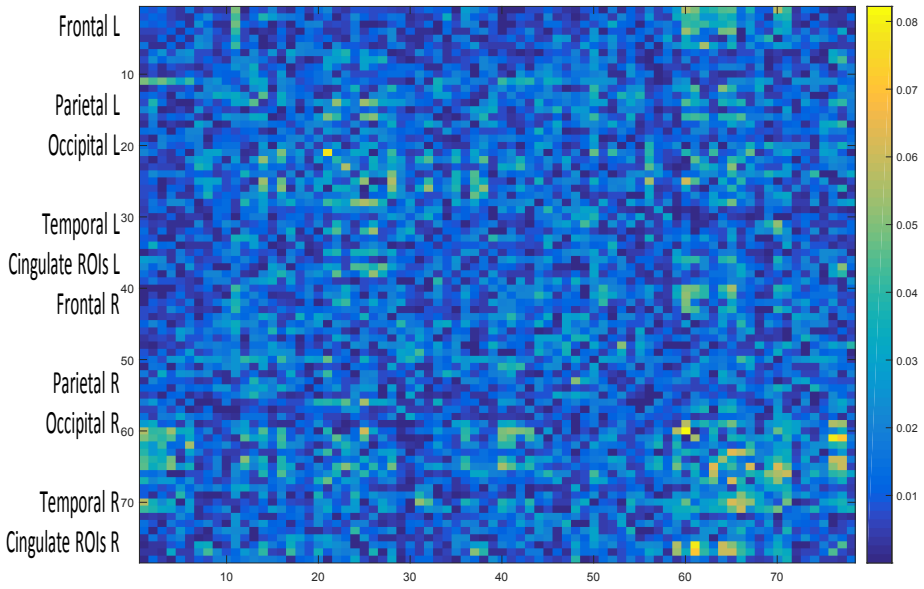


Figure 19: Absolute distance matrix between the best fit for $\widetilde{W}_{MEG} = f(A)$ ($K = 6$, $z_0 = 0$) and the empirical matrix W_{MEG} for the group-averaged data set.

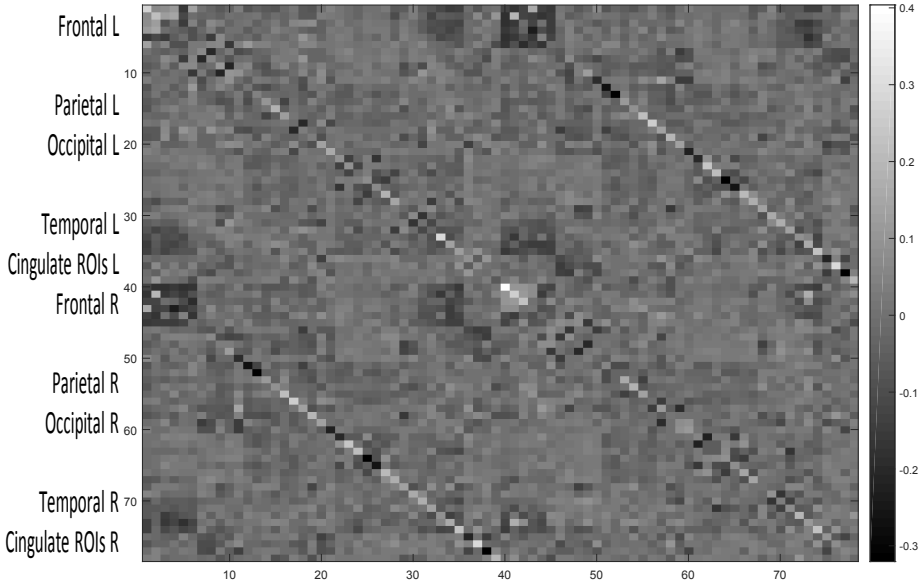


Figure 20: Absolute differences in the error for 2 in comparison with the error for 6 fitted coefficients ($K = 6$, $z_0 = 0$) for the mapping $f^{-1} : W_{fMRI} \to A$ for the group-averaged data set. Note that darker areas correspond here to regions that benefitted more from a higher number of coefficients.

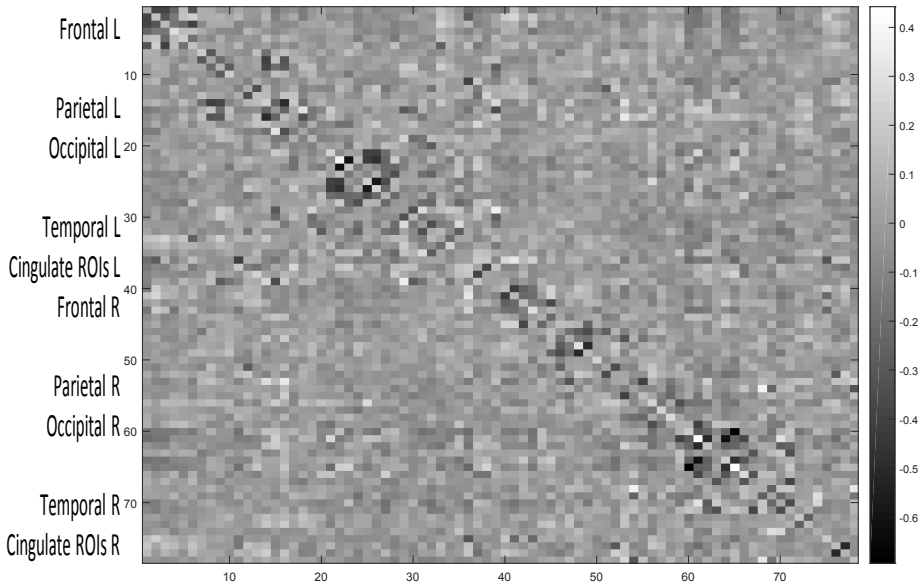


Figure 21: Absolute differences in the error for 2 in comparison with the error for 6 fitted coefficients ($K = 6$, $z_0 = 0$) for the mapping $f^{-1} : W_{MEG} \to A$ for the group-averaged data set. Note that darker areas correspond here to regions that benefitted more from a higher number of coefficients.

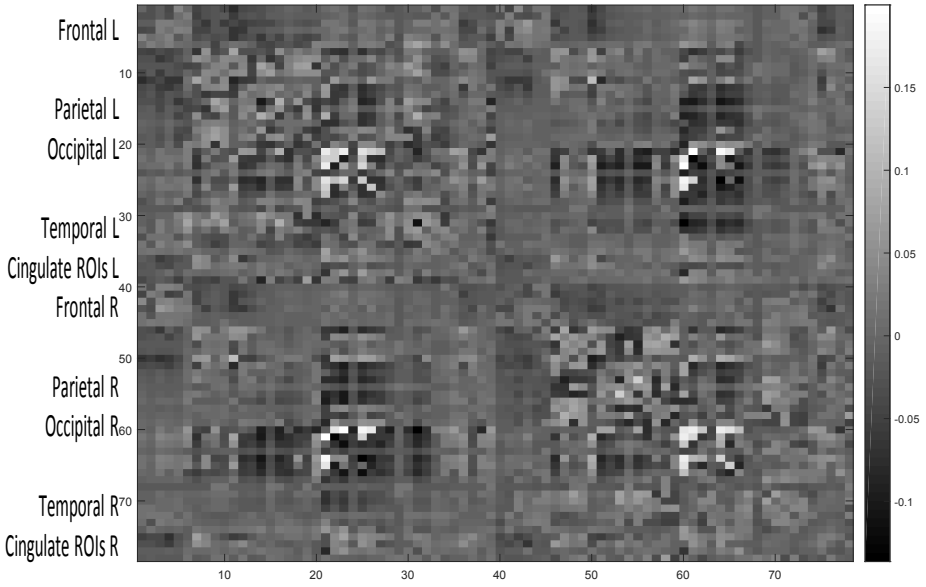


Figure 22: Absolute differences in the error for 2 in comparison with the error for 6 fitted coefficients ($K = 6$, $z_0 = 0$) for the mapping $f : A \to W_{fMRI}$ for the group-averaged data set. Note that darker areas correspond here to regions that benefitted more from a higher number of coefficients.

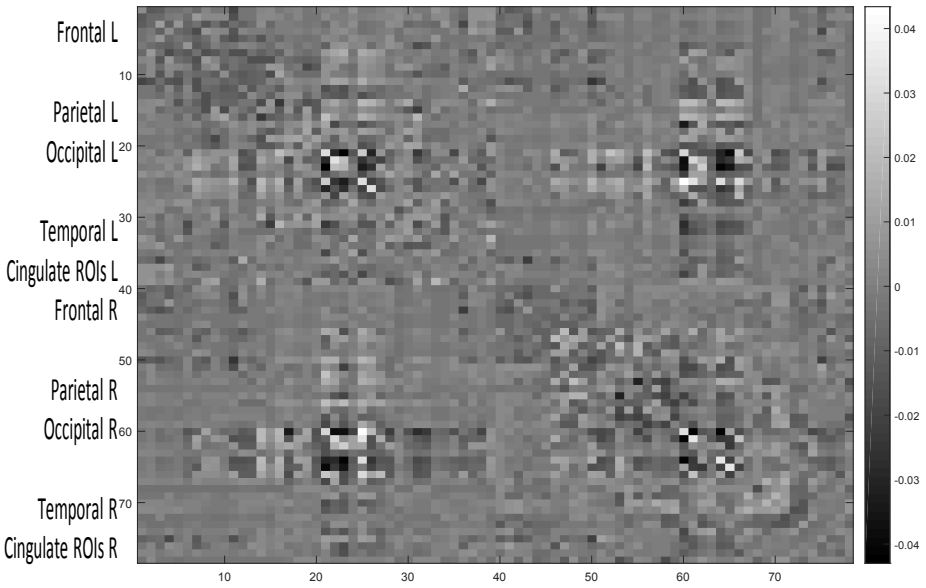


Figure 23: Absolute differences in the error for 2 in comparison with the error for 6 fitted coefficients ($K = 6$, $z_0 = 0$) for the mapping $f : A \to W_{MEG}$ for the group-averaged data set. Note that darker areas correspond here to regions that benefitted more from a higher number of coefficients.

G Reshuffled matrices

We used one reshuffling technique on all matrices: we selected two matrix entries at random and then exchanged their entries in the matrix (their link weights in the network). We repeated this step 5000 times to obtain a reshuffled version of every matrix, which is again symmetric. Thereby we preserved the distribution of the weights. After generating 100 reshuffled matrices, we compared the goodness of fit distribution of the mappings using reshuffled matrices to our empirical results (see SI Figures 24 and 25). More precisely, for each mapping from any matrix M to N, we first replaced the underlying matrix M with a reshuffled version of this adjacency matrix and ran the same mapping analysis on it. Then we replaced the image matrix N with a randomly reshuffled version of itself (keeping the underlying matrix M as the original empirical matrix) and ran the fitting algorithm on that combination (thus always one empirical matrix with one reshuffled version of the other matrix together).

For the group-averaged data set, the fit errors of 100 reshuffled matrices were larger compared to those obtained using the experimental data for all analyzed mappings (see SI Figures 24 and 25). Therefore, we can conclude that the mapping from structure to function (and vice versa) fails when reshuffled connections are used. Therefore, the empirical matrices seem to possess a special structure making the relationship between the structural and functional brain network closer than when reshuffled version of those matrices are used.

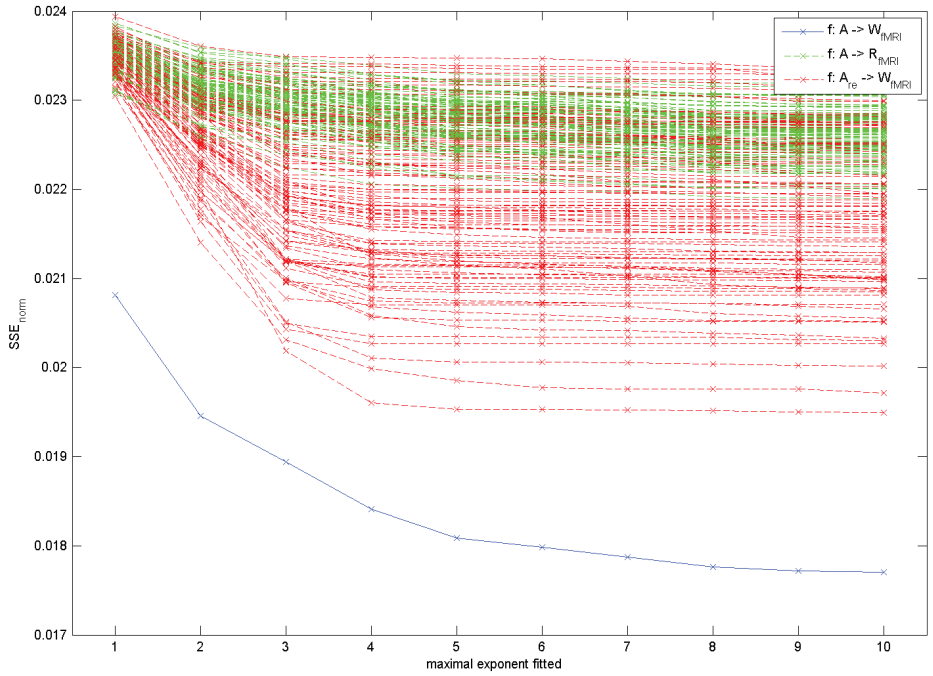
For the data set of individual healthy controls, we also used 100 reshuffled matrices of their structural and functional networks, respectively, and displayed the percentage of those matrices that achieved better goodness of fit than our empirical data (see SI Figures 26 and 27).

When using 5 coefficients or more, we obtained good results for the mapping $f : V \to W_{fMRI}$ for nearly all subjects. Only subject number 5 seems to be an exception which could be due to some measurement errors or specific individual attributes of that subject. We also identified two individuals as outliers whose goodness of fit level did not outperform the random reshuffles for the mapping $f^{-1} : W_{fMRI} \to V$. The function f^{-1} starting from MEG networks obtained good results for $K > 2$. The only mapping that was not able to outperform the reshuffled matrices as a benchmark for most of the subjects was $f : V \to W_{MEG}$ in the individual healthy control data. To sum up, for the individual structure-function relationships, in most of the cases the mapping performs worse when we apply it to random reshuffles indicating a high goodness of fit level for the mapping between the original empirical matrices.

H Details of the fitting procedure

In order to use the non-linear regression algorithm in MATLAB (using the routine *nlinfit.m* version *R2015a*) to determine the coefficients in (5) by iterative least squares estimation, we need to adapt our data first. Because all involved matrices are symmetric, we only need to fit the lower triangular matrices and the diagonal to get our fitting results. Thus, we first write all matrices in a vectorized form only containing their lower triangular and diagonal entries. For any matrix M of dimension $N \times N$ this vector will be denoted by $ltd(M)$. To be able to use the standard equation $Y = X \cdot \beta$ (with X design matrix, β parameter vector and Y image matrix) for a linear model, we need to define the

(a) $f : A \to W_{fMRI}$



(b) $f : A \to W_{MEG}$

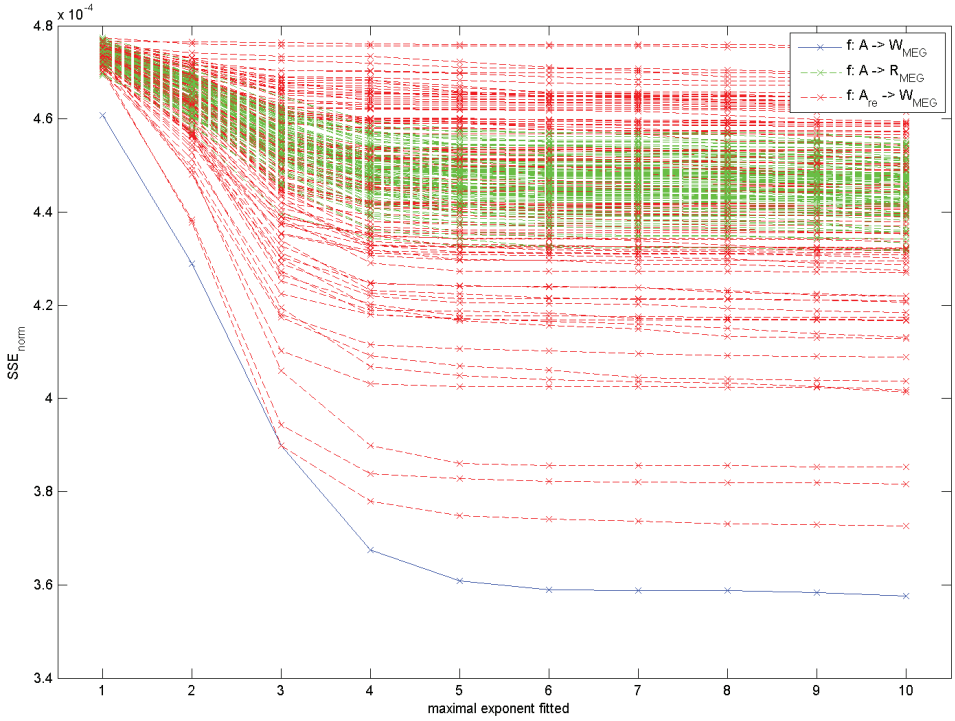
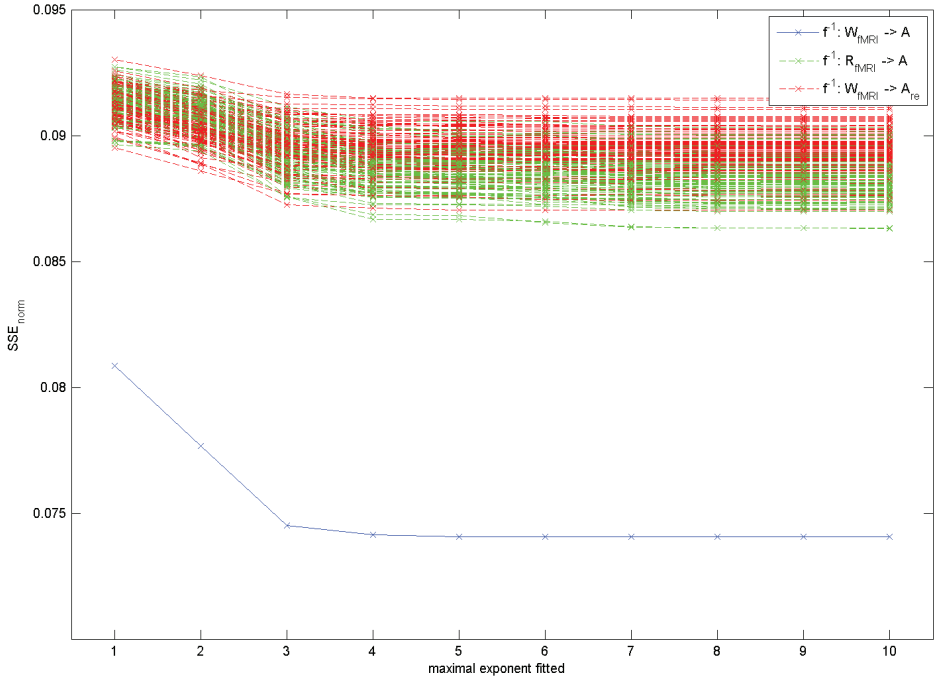


Figure 24: Plot of the normalized sum of squared errors (SSE_{norm}) of the function f in the topology domain for different maximal fitted exponents K and always in combination with a randomly reshuffled matrix R (R_{MEG}, R_{fMRI} and A_{re} denoting the reshuffled versions of W_{MEG}, W_{fMRI} and A, respectively) averaged over a range of z_0 values from -3 till 3 (always including an error matrix E) for the group-averaged data set.

(a) $f^{-1} : W_{fMRI} \rightarrow A$



(b) $f^{-1} : W_{MEG} \rightarrow A$

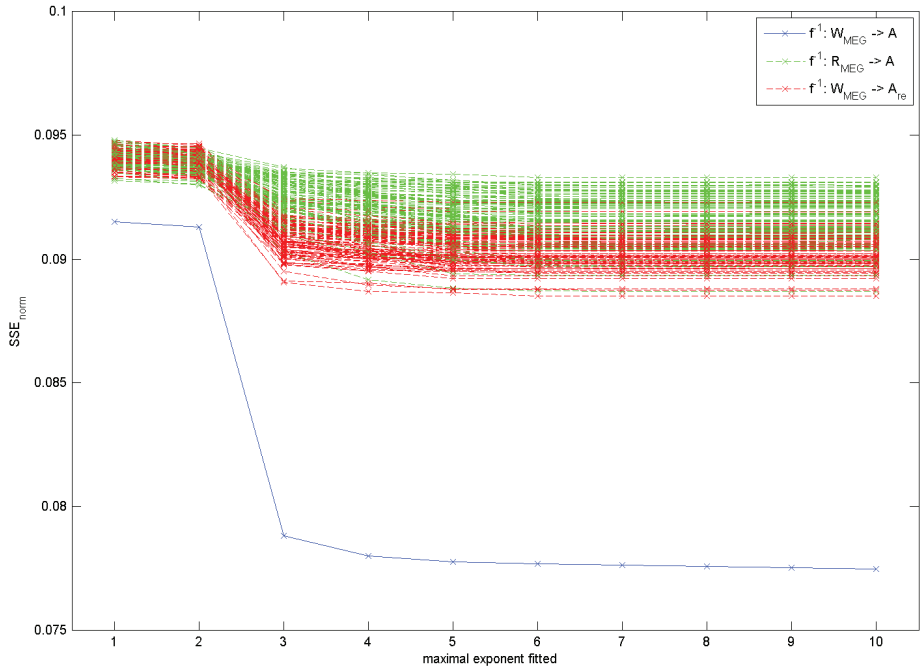
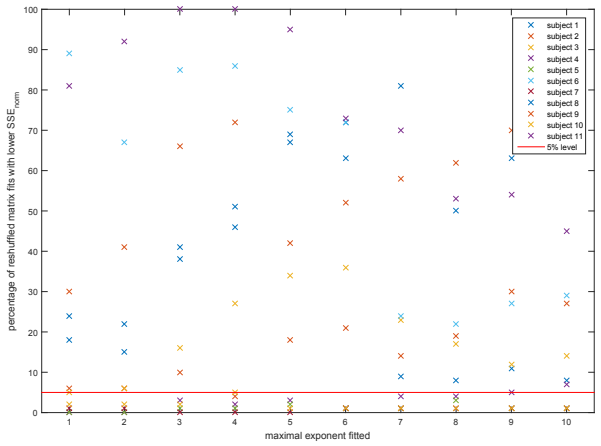
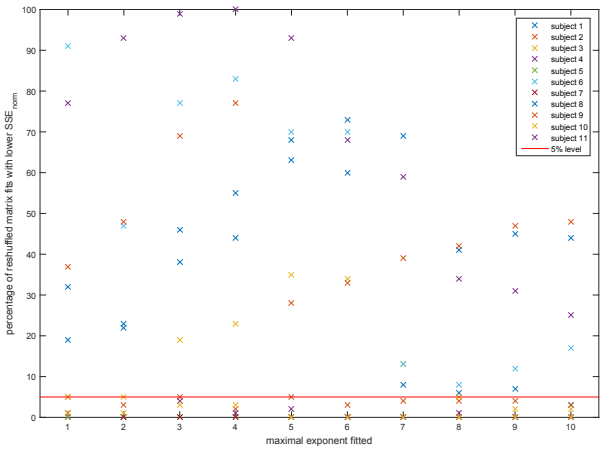


Figure 25: Plot of the normalized sum of squared errors (SSE_{norm}) of the function f^{-1} in the topology domain for different maximal fitted exponents K and always in combination with a randomly reshuffled matrix R (R_{MEG}, R_{fMRI} and A_{re} denoting the reshuffled versions of W_{MEG}, W_{fMRI} and A, respectively) averaged over a range of z_0 values from -3 till 3 (always including an error matrix E) for the group-averaged data set.

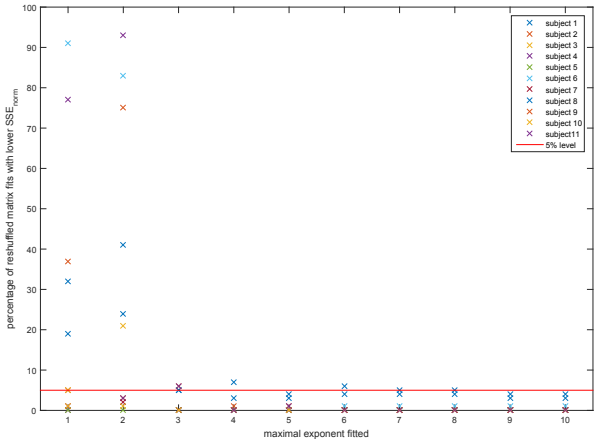
(a) $f : V_{re} \to W_{MEG}$



(b) $f : V \to R_{MEG}$



(c) $f^{-1} : R_{MEG} \to V$



(d) $f^{-1} : W_{MEG} \to V_{re}$

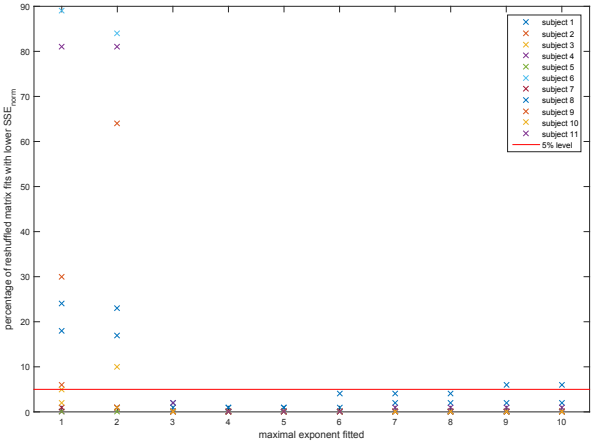
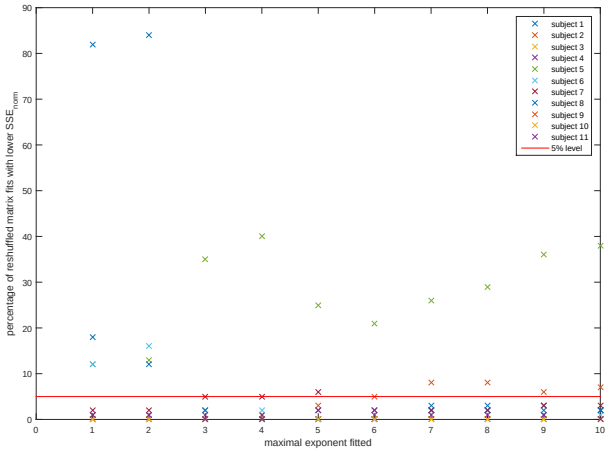
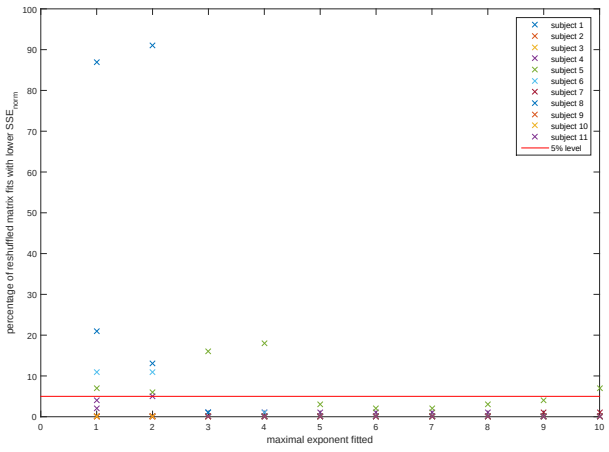


Figure 26: Plot of the percentages of reshuffled matrices that resulted in a lower normalized sum of squared errors (SSE_{norm}) in the topology domain for different maximal fitted exponents K with individual healthy controls results for all mapping including MEG (V denoting the weighted structural matrix and V_{re} denoting its randomly reshuffled version) for $z_0 = 0$ (with an error matrix E).

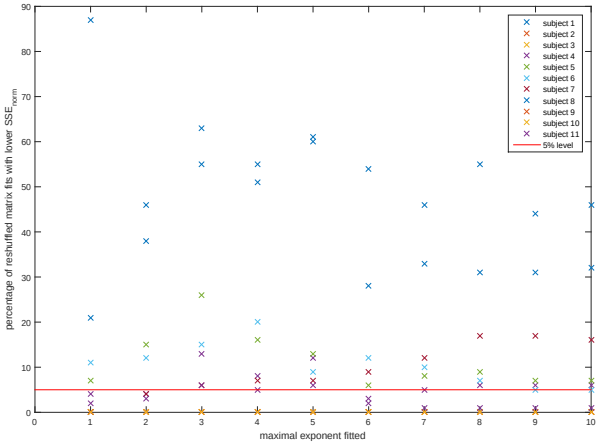
(a) $f : V_{re} \to W_{fMRI}$



(b) $f : V \to R_{fMRI}$



(c) $f^{-1} : R_{fMRI} \to V$



(d) $f^{-1} : W_{fMRI} \to V_{re}$

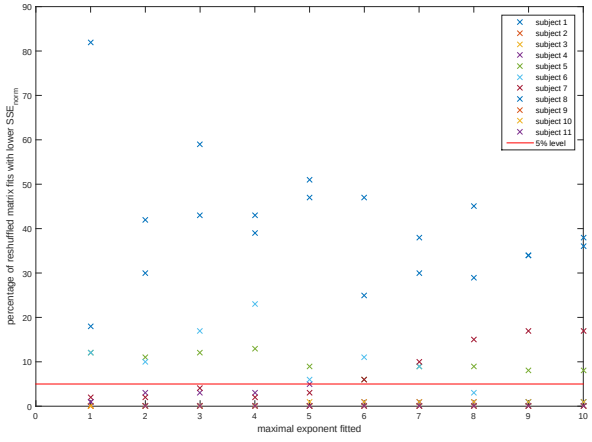


Figure 27: Plot of the percentages of reshuffled matrices that resulted in a lower normalized sum of squared errors (SSE_{norm}) in the topology domain for different maximal fitted exponents K with individual healthy controls results for all mappings including fMRI (V denoting the weighted structural matrix and V_{re} denoting its randomly reshuffled version) for $z_0 = 0$ (with an error matrix E).

variables for our case. In the case of $f(A) = W$, the response Y is just the image matrix W written as a vector containing the lower triangular and diagonal entries, $Y = ltd(W)$. The design matrix X is in the case of K as the maximal exponent

$$X = \Big(ltd(J) \quad ltd(I) \quad ltd(A) \quad ltd(A^2) \quad ... \quad ltd(A^K) \Big)$$

Therefore, the parameter vector β has the length $(K+2)$, where the first entry will be the coefficient c in front of the all-one matrix J and the second one is the coefficient for the identity matrix I and the others are in front of the matrix powers of A. Because the matrix powers of A are exploding in magnitude quickly, we normalize all the matrices beforehand dividing every entry by the absolute maximum entry of each matrix, which has the consequence that all matrices now have values between 0 and 1. Then, the *nlinfit.m* algorithm can be applied to our data using the underlying function *myfun.m* displayed here

```
function F = myfun(beta,xdata)

        F = xdata*beta;
end
```

where **xdata** refers to our design matrix X and **beta** is the parameter vector β. Because our model resembles a GLM, we could also use the pseudo-inverse of our design matrix **xdata** (*pinv(**xdata**)*) and multiply it with the vectorized matrix Y in order to obtain the same estimated coefficients. In order to obtain the coefficient values for the original powers of the A matrix, we have to denormalize the estimated values by dividing the estimated coefficients each by the absolute normalization value from before.

I Dimension differences

In practice, the $m \times m$ measured matrix W^* may be of a different dimension than the $N \times N$ matrix W. If $m \geq N$, then we can transform the measured matrix W^* to W as follows. Since W^* is symmetric, the spectral decomposition is

$$W^* = Y^* \Upsilon^* Y^{*T}$$

where the diagonal matrix $\Upsilon^* = \mathrm{diag}\big(\mu_1^*, \ldots, \mu_N^*, \mu_{N+1}^*, \ldots, \mu_m^*\big)$ with the real eigenvalues ordered as $|\mu_1^*| \geq |\mu_2^*| \geq \ldots \geq |\mu_m^*|$. The ordering here is different than the usual ordering in Section A.3, because eigenvalues of W may be negative (in principle; although those of a correlation matrix are non-negative). Next, we let $\mu_k = \mu_k^*$ for $1 \leq k \leq N$ and $\mu_k = 0$ for $k > N$ and

$$Y^* = \begin{bmatrix} (Y_{11})_{N \times N} & (Y_{12})_{N \times (m-N)} \\ (Y_{21})_{(m-N) \times N} & (Y_{22})_{(m-N) \times (m-N)} \end{bmatrix}$$

so that

$$\widetilde{W^*} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} \Upsilon & O \\ O & O \end{bmatrix} \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}^T = \begin{bmatrix} Y_{11}\Upsilon Y_{11}^T & Y_{11}\Upsilon Y_{21}^T \\ Y_{21}\Upsilon Y_{11}^T & Y_{21}\Upsilon Y_{21}^T \end{bmatrix}$$

from which we choose $W = Y_{11}\Upsilon Y_{11}^T$. This method is well-known in the theory of singular value decompositions (see e.g. (Golub and Loan, 1996)) and provides the best $N \times N$ (in the mean-square sense) approximation of an $m \times m$ matrix.

References

Alexander-Bloch, A. F., Vértes, P. E., Stidd, R., Lalonde, F., Clasen, L., Rapoport, J., Giedd, J., Bullmore, E. T., and Gogtay, N. (2013). The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. *Cerebral cortex*, 23(1):127–138.

Golub, G. H. and Loan, C. F. V. (1996). *Matrix Computations*. The John Hopkins University Press, Baltimore, third edition.

Tewarie, P., Hillebrand, A., van Dellen, E., Schoonheim, M., Barkhof, F., Polman, C., Beaulieu, C., Gong, G., van Dijk, B., and Stam, C. (2014). Structural degree predicts functional network connectivity: A multimodal resting-state fMRI and MEG study. *NeuroImage*, 97:296–307.

Van Mieghem, P. (2011). *Graph Spectra for Complex Networks*. Cambridge University Press.