

Britta Pester*, Christoph Schmidt, Karl-Jürgen Bär and Lutz Leistritz

Methodological aspects of analyzing high resolved brain connectivity for multiple subjects

Abstract: Analyzing directed interactions within brain networks of high spatial resolution is always associated with a limited interpretability due to the high amount of possible connections. Here, module detection algorithms have proven to helpfully subsume the information of the resulting networks for each proband. However, the between-subject comparison of clusters is not straightforward since identified modules are not matched to each other across different subjects. Tensor decomposition has successfully been applied for the detection of group-wide connectivity patterns. Yet, a thorough investigation of the effect of the involved analysis parameters and data properties on decomposition results has still been missing. In this study we filled this gap and found that - given appropriate parameter choices - tensor decomposition of functional connectivity data reveals meaningful, group-specific insights into the brain's information processing.

Keywords: Large scale Granger causality, parallel factor analysis, network analysis, module detection, fMRI.

<https://doi.org/10.1515/cdbme-2017-0088>

1 Introduction

The analysis of spatially highly resolved functional brain connectivity suffers from the problem that the number of possible interactions between remote regions increases as the

square of the considered network nodes. As a solution, insights into functional connectivity can be obtained by identifying clusters of strongly connected brain functional network nodes, called modules [1]. However, this approach leads to the problem that the integration of multiple subjects is not straightforward. Tensor decomposition has been proposed as a tool for the rearrangement of the subject-specific networks into subject-independent networks together with subject-specific loadings and thus enables a group-wide segmentation of the resulting networks. Even though a first proof-of-principle study has outlined the promising application of the new strategy [2], the influence of basic estimator and data parameters has not been studied yet. The objective of this work is to fill that gap and to systematically investigate the impact of the following key factors:

- *Data properties:* signal-to-noise ratio (SNR), number of temporal samples and network nodes;
- *Model parameters:* order of time series model, explained variance after dimension reduction, number of factors for the tensor decomposition.

Networks of different sizes were used to simulate data [3]; this allows the study of the impact of all parameters by comparing the results with the known ground truth. Also, resting state fMRI (rs-fMRI) data were used to evaluate the influence of model parameters in the context of real data.

The study provides much needed background information on an appropriate use of the novel combination of spatially highly resolved connectivity measures and tensor decomposition, and helps users to achieve more reliable interpretations of the results.

2 Material

2.1 Synthetic data

Simulated data offer the possibility to examine the impact of data properties systematically. For this purpose, synthetic

*Corresponding author: **Britta Pester:** Institute of Medical Statistics, Computer Sciences and Documentation, Jena University Hospital, Friedrich Schiller University, Jena, Germany, Bachstr. 18, 07743 Jena, e-mail: Britta.Pester@med.uni-jena.de
Christoph Schmidt, Lutz Leistritz: Institute of Medical Statistics, Computer Sciences and Documentation, Jena University Hospital, Friedrich Schiller University, Jena, Germany
Karl-Jürgen Bär: Department of Psychiatry, Jena University Hospital, Friedrich Schiller University, Jena, Germany

ground truth networks (GTN) with differing numbers of network nodes ($D = 100, \dots, 800$) were generated; they were realized on the basis of multivariate autoregressive models, where the number of temporal samples was chosen between $N = 100, \dots, 1000$. The SNR was varied by systematically adding white noise with increasing standard deviation. To account for group effects, two different GTNs, A and B , were generated for each parameter constellation and were then used to realize $S = 10$ ('subjects') data sets, where five 'subjects' followed GTN A and the others followed GTN B .

2.2 Clinical data

Supplementary to simulated data, rs-fMRI data were used to investigate the influence of varying model parameters. For it, fMRI Images of $S = 154$ subjects were used. During the scan, subjects were instructed to stay still and keep eyes closed. 240 T2*-weighted images were obtained using a gradient-echo EPI sequence each consisting of 45 transversal slices (in-plane resolution = $2.5 \times 2.5 \text{ mm}^2$, slice thickness = 2.5 mm) covering the whole brain including the lower brain stem. $D = 171$ non-overlapping ROIs were extracted, where 160 ROIs were derived directly from established networks, while another 11 were drawn manually.

3 Methods

Analysis of data was carried out in three consecutive steps. First, spatially highly resolved functional connectivity of each individual was quantified by means of lsGCI [3] that enables the quantification of a node-by-node (e.g. voxel-by-voxel) connectivity even for high dimensional time-series. The lsGCI incorporates a dimension reduction and thus introduces the model parameter 'variance explanation'.

Second, tensor decomposition was used for the identification of group-wide networks. The basic idea is to rearrange the individual networks into a conjoint tensor and to subsequently decompose this tensor into a sum of rank-1 outer products. More precisely, let $\Gamma_{d,s} \in \mathbb{R}^{D(D-1) \times S}$ be the tensor containing the lsGCI edge weights of S subjects with D network nodes each. Then, this multi-subject lsGCI tensor is decomposed into the modes 'location of directed interaction' and 'subjects': $\Gamma_{d,s} = \sum_{m=1}^M a_{d,m} \cdot b_{s,m} + \eta_{d,m}$, where the coefficients $a_{d,m}$ represent spatial loadings (where are strong connections?), $b_{s,m}$ denotes subject-wise loadings (who possesses those strong connections?) and $\eta_{d,m}$ contains the model residuals. Each of the M spatial loading vectors

$a_{d,m}$ can then be rearranged into a matrix representing a

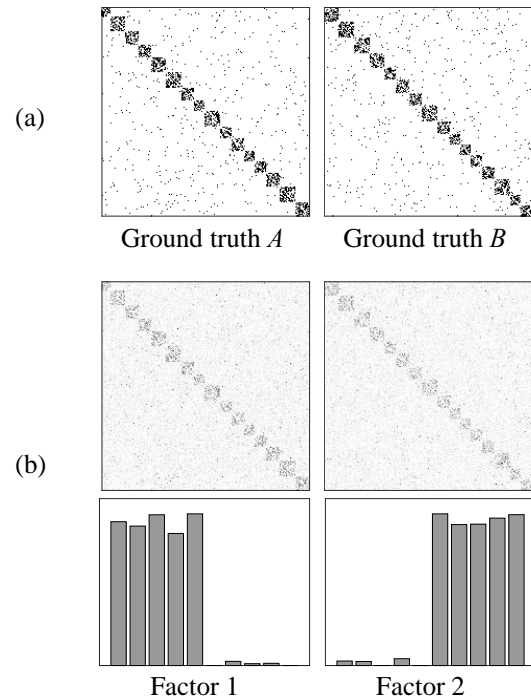


Figure 1: Example of the tensor decomposition results for $N = 1000$, $D = 200$, $M = 2$, explained variance = 85 %.

group-wide weighted network.

The resulting weighted networks were directly used without any edge weight rescaling or edge weight thresholding. Module structure was detected using the algorithm of Blondel et al. [4] in combination with consensus clustering [5]. For the fMRI-based functional networks, in addition module structure of the binarized networks was identified using the same technique as for the weighted networks. Module structure quality was assessed using the modularity, coverage and performance measures [1]. The recoverability of network partitions, i.e. the distance of identified partitions to ground truth partitions, was assessed using normalized variation of information (VI) [6]. The quality of the networks identified by means of tensor decomposition was quantified by means of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rates.

4 Results

4.1 Synthetic data

For a broad range of parameters we found that the proposed approach yields robust results. Fig. 1a shows two GTNs for $D = 200$; Fig. 1b represents the corresponding results for

$N = 1000$ realized samples, 85% explained variance and $M = 2$ factors. Obviously, the spatial mode of one factor reflects GTN A and the other factor reflects GTN B (Fig. 1b, first row); considering the loadings of the ‘subject’ mode, all ‘subjects’ are correctly segregated into the both GTN groups (Fig. 1b, second row).

4.1.1 Data properties

Signal-to-noise ratio: As expected, network detection accuracy is reduced with decreasing SNR. This is true for TP and TN rate, although TP being more affected than TN.

With decreasing SNR this distance between identified and ground truth module structure (VI) increases in line with the decrease of network detection accuracy. Interestingly, modularity attains its highest values for intermediate ranges of reduced SNR.

Number of samples in time N : TP and TN rates of identified networks continuously rise with the number of temporal samples. Most likely, the reason is that lsGCI itself achieves better results for high N , which is obviously passed on to the subsequent processing steps. This is also reflected by network analyses, since the recoverability of module structure improves with increasing N . In addition, module structure quality improves, as seen in the increases of modularity, although this effect is less pronounced.

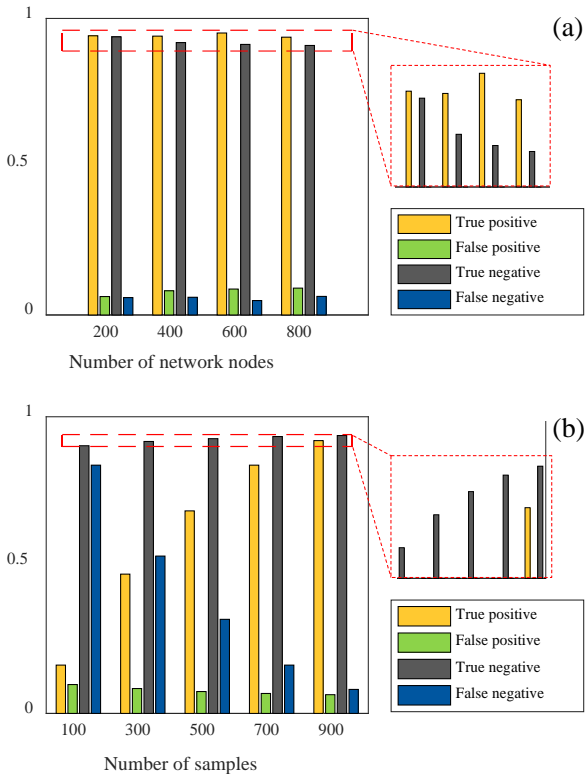


Figure 2: Quality criteria for binary classification of simulated data in dependence on (a) the spatial dimension of the network and (b) the number of simulated samples in time.

Number of network nodes D : Surprisingly, the network size had no influence on the correctly identified connections (TP) by tensor decomposition, whereas the number of correctly discarded connection (TN) is negatively affected by increasing D (Fig 2a). Due to the resulting increase in spurious connections (FP), the underlying ground truth module structure becomes increasingly hard to correctly identify with increasing D . A further consequence of increased FP is that modularity also decreases with D .

4.1.2 Model parameters

MVAR model order p : Fig. 2b asserts that the best performance is achieved when the correct model order $p = 2$ is chosen. Obviously, $p < 2$ prevents a sufficient model fit, while $p > 2$ involves too much MVAR parameters leading to an overfitting. The findings from connectivity analysis were confirmed from the perspective of module identification, where VI was minimal for $p = 2$ and substantially increased for $p > 3$. Modularity is maximal for $p = 1$ and $p = 2$, while differences to the values obtained for all other investigated model orders are rather small.

Explained variance: It became clear that the percentage of explained variance after dimension reduction by principal component analysis in the course of lsGCI computation had only minor impact on the quality of detected networks. That means, although dimension reduction of the lsGCI leads to a loss of information, the decomposition still extracts the fundamental properties of the resulting networks. The recoverability of ground truth module structure is also largely unaffected by varying levels of explained variance.

Number of factors M : Our results show that if the number of factors M is less than the number of underlying GTNs (i.e. $M = 1$), both GTNs are conflated into one single factor. The correct choice $M = 2$ yields a perfect separation of the GTNs. If M is chosen too high ($M > 2$), ‘subjects’ of the same group are split into different factors; the corresponding spatial weights still reflect the correct GTNs. VI is minimal for $M = 2$, whereas modularity is largely unaffected by the choice of M .

4.2 Clinical data

In real data, the properties N , D , and SNR are naturally predetermined by the experimental setup. Consequently, the evaluation of influencing factors is limited to model parameters. In general, it became apparent that irrespective of the choice of parameters, only one, or in some cases two, factors applied for the whole group (further referred to as

‘group factor’); the other factors mainly represented subject-specific connectivity patterns. The observed effects of model parameters can be summarized as follows.

Order of MVAR model p : The model order was chosen variably between 1 and 10. It appeared that from $p \geq 2$ the order had only marginal influence on the connectivity patterns of the group factor. Every other factor is assigned to a single subject, whose brain functional network shows seemingly random, irregular edge patterns. Network analysis revealed that in general module structure quality was highest for model orders $p = 1$ and $p = 5$.

Explained variance: The proportion of explained variance had the most pronounced impact on rs-fMRI networks. A variance explanation less than 70 % led to diffusively distributed networks of all factors. With increasing explained variance levels, module structure quality was generally decreased.

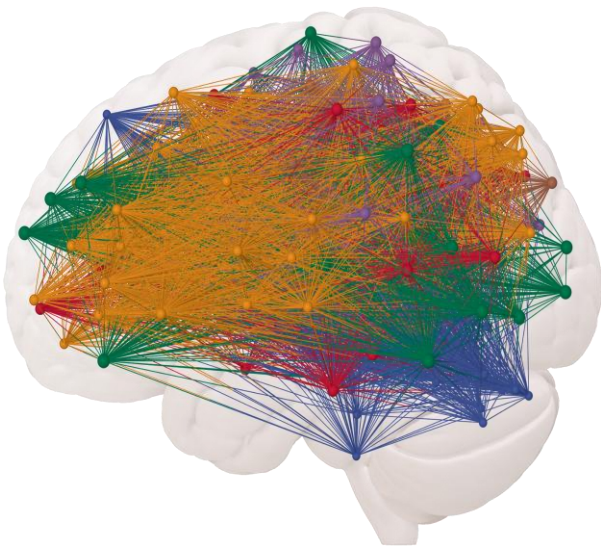


Figure 3: Module structure identified in the rs-fMRI network obtained by tensor decomposition.

Number of factors M : It did not severely change the basic properties of the detected group network. However, a clear pronounced group network only emerged for $M \geq 3$. By tendency, the number of decomposition factors had no distinct effect on the module structure quality.

As an example, Fig. 3 depicts the identified modules in the network of one spatial factor ($M = 5$, $p = 5$, 50 % explained variance). The figure illustrates the spatial arrangement of modules was not unsystematic, e.g. one module is situated at the cerebellum, whereas another is located at frontoparietal brain regions.

5 Discussion

We proposed tensor decomposition of spatially highly resolved functional networks at the group level as an approach to gain better insights into functional brain networks. For this novel strategy we evaluated the effect of various data properties and analysis parameters on obtained results. The analysis was twofold: from the perspective of connectivity analysis and from the analysis of corresponding brain functional networks.

Simulations have shown that decreasing SNRs and a reduced number of temporal samples result in an impaired accuracy of detected networks. Most likely this is primarily caused by impairments on the level of lsGCI network computation, rather than the subsequent tensor decomposition. By contrast, the number of network nodes D did not have a decisive influence on the resulting network patterns given by correctly identified connections. The module structure analysis of the resulting networks is largely in line with the findings from connectivity analysis. Expected detrimental effects were confirmed. Surprisingly, the explained variance after dimension reduction had almost no effect on the detection of simulated networks while fMRI data suggest a negative influence of too large explained variance levels.

In summary, we found that our proposed method is effective for spatially high resolved connectivity analysis given an appropriate choice of the involved parameters.

Author's Statement

Research funding: This study was supported by the German Research Foundation (DFG Pe 2546/1-1). **Conflict of interest:** Authors state no conflict of interest. **Informed consent:** Informed consent has been obtained from all individuals included in this study. **Ethical approval:** The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

- [1] Fortunato, S., Community detection in graphs. *Physics reports*, 2010. 486(3): p. 75-174.
- [2] Pester, B., et al. Detecting spatially highly resolved network modules: A multi subject approach. in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. 2016. IEEE.
- [3] Schmidt, C., et al., A Multivariate Granger Causality Concept towards Full Brain Functional Connectivity. *PLoS one*, 2016. 11(4): p. e0153105.

- [4] Blondel, V.D., et al., Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008. 2008(10): p. P10008.
- [5] Lancichinetti, A. and S. Fortunato, Consensus clustering in complex networks. *Scientific reports*, 2012. 2.
- [6] Meilă, M., Comparing clusterings—an information based distance. *Journal of multivariate analysis*, 2007. 98(5): p. 873-895.