PROCEEDINGS OF SCIENCE

Investigating Community Detection Algorithms and their Capacity as Markers of Brain Diseases

Eva Výtvarová
Faculty of Informatics, Masaryk University
Central European Institute of Technology, Masaryk University
E-mail: eva.vytvarova@mail.muni.cz

Jan Fousek
Faculty of Informatics, Masaryk University
E-mail: izaak@mail.muni.cz

Michal Mikl
Central European Institute of Technology, Masaryk University
E-mail: michal.mikl@ceitec.muni.cz

Irena Rektorová
Central European Institute of Technology, Masaryk University
E-mail: irena.rektorova@fnusa.cz

Eva Hladká*
Faculty of Informatics, Masaryk University
E-mail: eva@fi.muni.cz

In this paper, we present a workflow for evaluating resting-state brain functional connectivity with different community detection algorithms and their strengths to discriminate between health and Parkinson's disease (PD) and mild cognitive impairment preceding Alzheimer's disease (AD-MCI). We further analyze the complexity of particular pipeline steps aiming to provide guidelines for both execution on computing infrastructure and further optimization efforts.

On a dataset of 50 controls and 70 patients we measured an increased modularity coefficient with 81.8% accuracy of classifying PD versus controls and 76.2% accuracy of classifying AD-MCI versus controls. Significantly higher modularity coefficient values were measured when the random matrix theory decomposition was adapted for network construction. These results were observed on networks of 82 nodes based on AAL atlas and 317 nodes based on multimodal parcellation atlas.

International Symposium on Grids and Clouds 2017 -ISGC 2017-
5-10 March 2017
Academia Sinica, Taipei, Taiwan

*Speaker.

PoS(ISGC2017)018

1. Introduction

The human brain consists of 10^{11} neurons interconnected with approximately 10^{14} connections creating a large, complex structure. This anatomical connectivity is a substrate for brain function which we can both measure and model on various scales. The goal of current research is understanding the brain as a system and its behavior as a whole, i.e. identifying the spatiotemporal structure of brain function. This large-scale functional connectivity of neuronal populations is defined as statistical relations between neural activities of distinct brain cortical regions and is usually measured by fMRI (functional magnetic resonance imaging), EEG or MEG (electro/magnetoencephalography). A very promising tool is to model the system as a network using network science approach, drawing inspiration from previous applications in research areas such as economics, transportation, communication and immunology [1, 2].

Network analysis models a complex system as a set of discrete units, describes the interactions between them, and provides ways to measure the importance of nodes and to detect communities. Using diverse metrics, it describes system topology and whole-system events and can be used to model growth in system size [1]. Network analysis is primarily a data-driven technique and is an excellent tool to study real systems and to understand them. It requires interdisciplinary and big data processing approaches.

As the median age of the population increases and life expectancy shifts upward, neurodegenerative diseases are becoming more prevalent. Early-stage detection of different kinds of dementia and cognitive impairments is an extremely important task considering that we are unable to cure them and are only able to slow down their progression and treat their symptoms. The neuroscience community is devoting great effort into finding potential markers of disease that would be noninvasive, easy to establish, and stable.

In this paper, we evaluate the strengths of community detection methods to differentiate between networks of healthy and diseased subjects using the examples of AD-MCI and PD-MCI/PD-NC (Mild Cognitive Impairment preceding Alzheimer's / Parkinson's Disease, Parkinson's Disease with Normal Cognition) as illnesses known to influence network topology [3, 4, 5, 6]. Mild cognitive impairment refers to an intermediate stage between normal aging and dementia. It includes changes in cognitive performance, mostly memory, language, thinking and judgment [7]. The decline in these functions is more serious in MCI than in normal aging. The goal of studying MCI is early diagnosis and slowing down the onset of dementia. We have chosen community detection because it was shown that community structure is influenced by such diseases as Alzheimer's disease or schizophrenia [8, 9]. A possible biological interpretation is that communities represent groups of nodes that serve different cognitive functions. It has also been shown that these groups can change over time [10].

We apply classification analysis, including parameter sweeps of methods for data preprocessing (since there is no consensual preprocessing pipeline), to quantify the ability of community detection algorithms to distinguish between patient and control and thus evaluate each one of them as a promising or inconclusive marker of a disease. The pipeline consists of several subsequent steps. We also analyze time complexity, which will guide further optimization efforts. Given the substantial amount of data, we used the computing infrastructure of MetaCentrum, CESNET, Czech Republic, to be able to analyze the data.

2. Methods

We divide the pipeline into six steps, as shown in Fig. 1. Time complexity was evaluated for each step. Data acquisition from healthy subjects and patients was followed by preprocessing and construction of a functional connectivity network for each subject. The community structure of these networks was then detected and used as a feature for classification between health and disease. We then computed the strength of community detection as a possible disease marker.

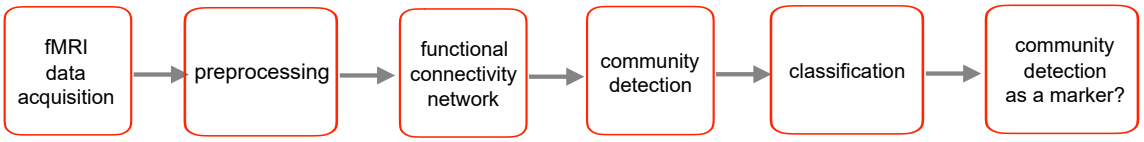

Figure 1: The sequence of steps performed for analyses in this paper.

2.1 Data acquisition

For this study, we obtained data from 70 patients (35 women; 66.7±9.4 years) and 50 healthy controls (34 women; 66.7±7.4 years), all of the same age category. Patients were further divided into AD-MCI and PD subgroups according to their detailed diagnosis. There were 34 AD-MCI patients (25 women; 70.6±7.6 years) and 36 PD patients (10 women; 63.1±9.6 years; patients were scanned in the ON medication state without dyskinesias). The PD group contained PD patients with normal cognition and with mild cognitive impairment [11]. A significant decrease in gray matter volume was detected in the hippocampus bilaterally in the AD-MCI group by means of voxel-based morphometry. No atrophy was measured in the PD group. Basic demographic and clinical variables are shown in Table 1. For further details on the subjects included in this study, please see the supplementary material of [12].

All subjects were scanned using a 3 T Siemens Prisma scanner and 64-channel head-neck receive coil array. Functional scans of resting-state activity (subjects were instructed to lie still with their eyes closed and not to think extensively about anything) were acquired using a gradient-echo echo-planar imaging sequence: TR = 2.08 ms, TE = 30 ms, FOV = 192 mm, FA = 90°, matrix size 64×64, slice thickness = 3 mm, 39 transversal slices, isotropic voxel size 3×3×3 mm, 200 scans (volumes). Structural data were acquired as anatomical T1-weighted images using MPRAGE sequence with 240 sagittal slices, TR = 2300 ms, TE = 2.36 ms, FOV = 256 mm, FA = 8°, matrix size 256×256, slice thickness = 1 mm, isotropic voxel size 1×1×1 mm.

2.2 Preprocessing

Certain data preprocessing is needed between data acquisition and network construction. The essential steps consisted of realigning scans shifted by subject's movement, co-registration with a more detailed anatomical scan, spatial normalization to a Montreal Neurological Institute template, and spatial smoothing with Gaussian kernel of FWHM = 5 mm.

However, there are some steps regarding noise suppression the usage of which is not consensual and depends on the user. We examined 4 steps creating 16 total variants of preprocessing. The steps are shown in Fig. 2 and are as follow:

	HC, n = 50	**MCI-AD**, n = 34	**PD**, n = 36
Age (years)	66.7 (7.4)	70.6 (7.6)	63.1 (9.6)
Education (years)	15.6 (2.6)	14.3 (4.0)	14.8 (4.3)
Sex, % male	32	26	72
MMSE	28.4 (1.3)	25.9 (2.5)	27.5 (2.1)
Average and standard deviation values are displayed. MMSE – mini mental state exam			

Table 1: Demographic and clinical variables.

- High pass filtering with cutoff at 128 s;
- Filtering of movement by 24 movement regressors: 3 translations (in x, y and z axes), 3 rotations, their time-differences, and all of them squared [13];
- Filtering of signals from white matter (WM) and cerebrospinal fluid (CSF): 4 WM and 6 CSF signals; and
- Global signal filtering: function `spm_global` from SPM12[1].

A global signal is an average signal coming from the whole brain and is sometimes filtered out in connectivity analyses. This step, however, is highly controversial [14, 15], and we therefore chose to include it in our analysis in order to show its influence on functional connectivity networks and community structures.

The preprocessing was done using SPM12 by regressing out the nuisance variables by GLM model.

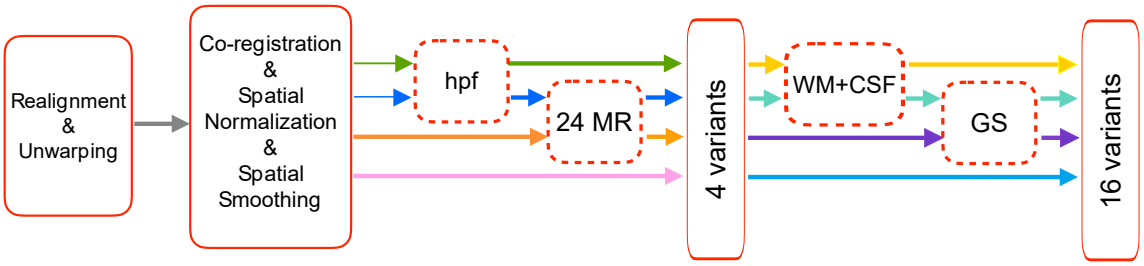

Figure 2: Preprocessing steps adopted in this analysis.

2.3 Network construction

The acquired data were parceled to 82 out of 90 anatomically defined regions of the AAL atlas [16] – regions not fulfilling the criterion of 5% minimal coverage were not included in a network. For each region (containing tens or hundreds of voxels) we computed a representative signal by averaging the signals from voxels in the region. In this way we obtained 82 time series

[1] `http://www.fil.ion.ucl.ac.uk/spm/`

that are 200 time-points long. By computing Pearson's correlation coefficients between every pair of time series, we obtain a correlation matrix that describes the functional connectivity network for which the Pearson's correlation coefficients represent the edge weights. The parcellation enables physiological interpretation of results on a regional level.

We apply community detection methods on the networks that are first thresholded, leaving only 15% of the strongest connections, and then further binarized. This simplification allows us more easily to detect communities than would be possible in a complete network and also prevents the existence of links with negative weights defined by negative correlation coefficients.

We conducted analyses also on the data parceled according to a multimodal parcellation map based on data from the Human Connectome Project (hereinafter referred to as the HCP MMP atlas) [17] into 317 out of 360 regions (respecting the minimal coverage criterion). The same procedure as mentioned above for the AAL atlas was followed also for the case of HCP MMP atlas-based regions.

2.4 Community detection algorithms

Community structure is a property of real complex systems. Communities can also be called modules, clusters, subnetworks, or spatial or temporospatial patterns. Nodes in a network with a community structure tend to be densely connected within communities and have sparse connections between modules [18].

Community detection is an NP-complete problem which uses optimization methods and heuristics. Repetitive computation is required to get representative results. We used iterative community fine-tuning. We started by computing community structure and then repeated the computation using the community structure computed in the previous step to maximize the coefficient value which indicates how well a network is able to form clusters. We repeated the fine-tuning 100 times and compute a representative partition across these repetitions and across subjects using an algorithm from Brain Connectivity Toolbox [19].

We used two algorithms for community detection: Louvain and Potts modularity methods. The Louvain method [20] is a simple greedy agglomerative algorithm that is computationally fast. It starts with the state of each node being in its own community, computing community structure and supplying this structure again and again to the algorithm until the maximum gain in modularity is obtained. The Potts model [21] represents the system as a spin glass where each node has a spin state. The goal of optimization is to find a state corresponding to the maximum value for the modularity.

Since we were dealing with the correlation character of networks, we tested not only these two modularity methods, but also included a preliminary step using random matrix theory. The random matrix theory identifies nonrandom properties of correlation matrices. It is based on eigenvalues computation and decomposes the correlation matrix into three components: a component for random mode, a component for group mode, and a component for market mode. Market mode captures seasonality in the underlying time series, while random mode captures the noise [22, 23]. By using RMT decomposition, we extracted only the group mode and used it as an input in modularity methods.

Combining modularity methods and random matrix theory decomposition, we obtained four approaches to community detection. We evaluated the following features on these community detection methods:

- Modularity coefficient Q, which reflects the ability of a network to form clusters: $Q \in (0;1), Q \to 0 \ldots$ random network without clusters, $Q \to 1 \ldots$ network with clearly defined modules;
- Node classification to a community (spatial localization of modules); and
- Computational demand of methods.

2.5 Statistical evaluations

Before training a classifier to differentiate between health and disease, the statistical significance between modularity coefficients in A) variants with and without RMT decomposition step and B) controls and patient groups was computed using nonparametric Wilcoxon tests, because some of the variables did not have normal distribution.

2.6 Classification

The classification analysis used modularity coefficients computed for each subject, preprocessing variant, and community detection method. The results of basic statistical testing were used to eliminate some of the combinations of preprocessing steps and modularity methods (see Results section for detailed information). For the remaining variables, we randomly sampled subjects to train (75% of samples) and test (25%) samples (leading to 14 healthy controls and 7 AD-MCI in test samples in HC—AD-MCI classification, and 12 HC and 10 PD subjects in test samples in HC—PD classification), used 10-fold cross-validation with 1000 iterations. The support vector machine (SVM) approach with radial basis function kernel was used. Age and gender were taken into consideration as additional variables. Classification training and testing were conducted separately for each patient group against the control group. The classification accuracy was obtained together with values for sensitivity and specificity. The procedure was repeated for diverse combinations of variables (modularity coefficients computed for networks based on different preprocessing pipelines).

2.7 Time complexity

Finally, we also measured the time complexity of individual steps of the processing pipeline. To better examine this, we have further created semi-simulated networks of 634, 984 and 1268 nodes. These artificial networks are based on time series of AAL and HCP MMP regions. A random noise with an amplitude of 50% of the original time series was added in each time point. Given that preprocessing is not influenced by network size, we evaluated only the network construction, community detection, and classification steps for these networks.

Data were obtained in NIfTI[2] format and processed in MATLAB[3] R2016a with the computational times measured by the MATLAB stopwatch timer. The classification analysis was computed

[2]https://nifti.nimh.nih.gov/

[3]https://www.mathworks.com/products/matlab.html

in STATISTICA[4] 64, version 13. Analysis was carried out on a Lenovo ThinkPad T430s Laptop with Intel i5 CPU processor with 4 physical cores and 8.00 GB RAM running Windows 10 Pro N.

3. Results

3.1 The effect of RMT decomposition on the modularity coefficient

We start the description of results by showing the effect of random matrix theory decomposition on Louvain modularity coefficient. The boxplots in Fig. 3 show different preprocessing pipelines that precede connectivity matrix construction, either including or not including the RMT decomposition step before modularity coefficient computation. In all cases except the one including global signal filtering, we observed a statistically significant increase in modularity when using only the group mode of the correlation matrix, $p<0.05$. This applies both to the healthy control group as well as to the patient groups. The last boxplot in Fig. 3 is for preprocessing with filtering out the global signal. Its usage completely changes the results, with RMT decomposition decreasing the modularity coefficient. Very similar results were observed for networks based on the HCP MMP parcellation atlas. Based on these results we excluded variants with global signal filtering from all follow-up analyses in this paper.

3.2 Modularity alterations in AD-MCI and PD

The statistical difference between modularity coefficients in the health control and AD-MCI and PD groups was computed across preprocessing variants and patient subgroups and is shown in Fig. 4. A statistically significant difference indicated by a p-value less than 0.05 was observed only for Louvain modularity, both with and without RMT decomposition. The statistically significant differences are indicated in the boxplots by asterisks (significant differences are between groups which are at the ends of the lines below asterisks). Modularity was higher in the PD group than in the control group consistently across the majority of preprocessing variants, whereas the AD-MCI group's modularity increased as compared to the control only in one preprocessing setting. Blue boxes are for the control group, orange for the AD-MCI patient group, and green for the PD group. No differences between groups were measured by modularity computed with Potts algorithm no matter the use or no use of RMT decomposition step as is shown also in Fig. 4.

Considering the results for the HCP MMP parcelated networks, the difference between controls and PD was measured in networks created with high pass filtering (hpf), hpf and CSF + WM filtering, and hpf and movement filtering. The increase in the modularity coefficient in AD-MCI as compared to controls was measured only in hpf, WM + CSF filtering variant with random matrix theory decomposition step.

3.3 Classification accuracy

First we established the classification accuracy for the modularity coefficients computed on networks for each of the preprocessing pipelines individually – 8 preprocessing pipelines (excluding those with global signal regression) with and without the random matrix decomposition step, together creating 16 preprocessing pipelines separately evaluated for their effects on modularity

[4]http://software.dell.com/products/statistica/

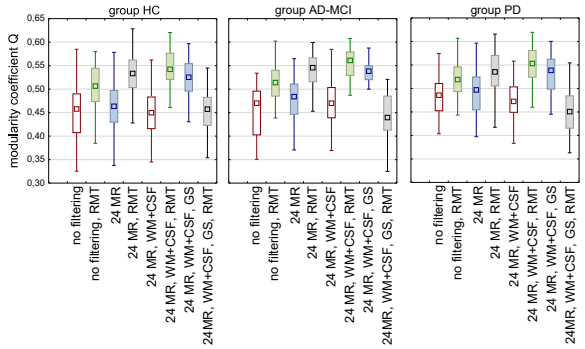

Figure 3: The effect of RMT decomposition, Louvain modularity; shown on networks computed from data with different preprocessing steps and AAL atlas parcellation.

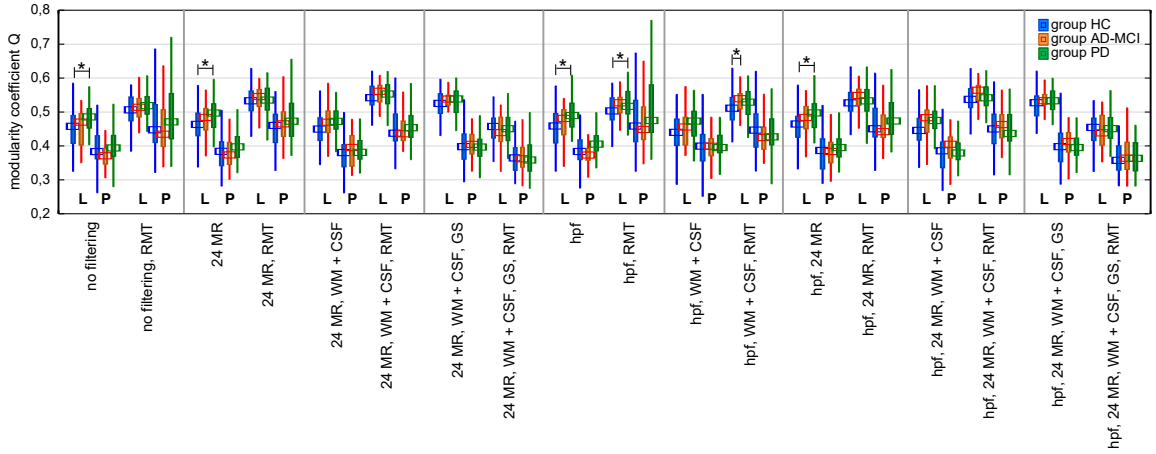

Figure 4: The effect of AD-MCI and PD on modularity coefficient using Louvain (**L**) and Potts (**P**) methods; shown on networks computed from data with different preprocessing steps and AAL atlas parcellation.

coefficients and classification ability. These results are presented in Table 2. The table shows low accuracy for all preprocessing pipelines with the best result for HC *vs.* PD classification by modularity for data with high pass filtering, regression of movement artifacts and RMT decomposition with 77.3% accuracy, 0.73 sensitivity and 0.82 specificity. For the HC *vs.* AD-MCI classification, no individual preprocessing pipelines used for classification was able to correctly sort any patient into a patient group (number of true positives equals zero) leading to zero or not-a-number results in sensitivity.

Second, working with the assumption that different preprocessing procedures filter out different information from the data, we included into the classification various combinations of preprocessing pipelines. We believed this would ensure that the information filtered out in one pipeline but not the other would not be lost and would be included in the classification.

For the HC *vs.* AD-MCI, Table 3 shows that the best classification (76.2% accuracy with specificity at 0.80 and sensitivity at 0.67) was obtained when modularity coefficients computed for networks based on AAL atlas and three different preprocessing pipelines entered the classification:

AAL, HC *vs.* AD-MCI						
Individual preprocessing pipelines for classification	Accuracy [%]		Sensitivity		Specificity	
	-	RMT	-	RMT	-	RMT
No preprocessing	47.1	66.7	0	NaN	0.63	0.67
WM + CSF	66.7	66.7	NaN	NaN	0.67	0.67
24 MR	66.7	61.9	NaN	0	0.67	0.65
24 MR + WM + CSF	66.7	66.7	NaN	NaN	0.67	0.67
hpf	66.7	66.7	NaN	NaN	0.67	0.67
hpf, WM + CSF	66.7	57.1	NaN	0.40	0.67	0.73
hpf, 24 MR	66.7	57.1	NaN	0	0.67	0.63
hpf, 24 MR + WM + CSF	66.7	66.7	NaN	NaN	0.67	0.67
AAL, HC *vs.* PD						
No preprocessing	68.2	68.2	0.62	0.62	0.78	0.78
WM + CSF	68.2	68.2	0.62	0.62	0.78	0.78
24 MR	68.2	68.2	0.62	0.62	0.78	0.78
24 MR + WM + CSF	68.2	68.2	0.62	0.62	0.78	0.78
hpf	68.2	68.2	0.62	0.62	0.78	0.78
hpf, WM + CSF	68.2	68.2	0.62	0.62	0.78	0.78
hpf, 24 MR	68.2	77.3	0.62	0.73	0.78	0.82
hpf, 24 MR + WM + CSF	68.2	68.2	0.62	0.62	0.78	0.78

Table 2: HC *vs.* patients classification by modularity; individual preprocessing variants. NaN . . . not-a-number; division by zero (true positives + false negatives = 0).

preprocessing with 1. high-pass filtering, movement regression and RMT decomposition, 2. high-pass filtering, movement regression and WM + CSF filtering, 3. high-pass filtering, movement regression, WM + CSF filtering and RMT decomposition. Each of the preprocessing pipeline filters out different information and this combination contains information that best differentiates between HC and AD-MCI. All other combinations for networks constructed on the AAL atlas and all combinations for networks constructed on the HCP MMP atlas were not able to correctly sort any patient into the AD-MCI patient group.

For the HC *vs.* PD groups, the best combination yielded 81.8% classification accuracy with specificity at 0.79 and sensitivity at 0.88 when the modularity was measured by Louvain modularity and the classifier included modularity coefficients computed on networks with a higher level of filtering in preprocessing. See Table 3 for detailed information about the combination of preprocessing pipelines. Very similar accuracy was measured for modularity computed on networks based on the HCP MMP atlas with good classification results across more combinations of preprocessing variants.

3.4 Analysis of complexity

Fig. 5a) shows the proportions of processing steps on the time of the whole computation for 120 subjects, 16 preprocessing variants and 4 community detection methods. The data acquisition

AAL, HC *vs.* AD-MCI			
Combination of preprocessing variants used for classification	Accuracy [%]	Sensitivity	Specificity
hpf, 24 MR, RMT hpf, 24 MR + WM + CSF hpf, 24 MR + WM + CSF, RMT	76.2	0.67	0.80
AAL, HC *vs.* PD			
24 MR, RMT 24 MR + WM + CSF 24 MR + WM + CSF, RMT hpf, 24 MR hpf, 24 MR, RMT hpf, 24 MR + WM + CSF hpf, 24 MR + WM + CSF, RMT	81.8	0.88	0.79

Table 3: HC *vs.* patients classification by modularity; a combination of preprocessing variants with the best classification results.

time depends on the MR measuring sequence and is usually around 10 minutes of measuring time and tens of minutes for subject preparation. It was excluded from this figure.

Preprocessing as the first step of fMRI data analysis and the length of its computation is influenced by data quantity (MR sequence parameters), subject specifics (e.g. movement) and level of preprocessing. For the 16 variants used, the computation took 3 to 5 minutes per subject.

Subsequent steps are strongly influenced by network size. For all subjects and all preprocessing variants we measured 11.6 mins spent on network construction for a network of 82 nodes (AAL atlas), 16.4 mins for 317 nodes (HCP MMP atlas), 37.5 mins for 634 nodes (HCP MMP based), 65.9 mins for 984 nodes (AAL based), and 134.3 mins for 1268 nodes (HCP MMP based). These values indicate an exponential relationship between network size and time spent on network construction.

An exponential trend was also observed in the community detection step. Fig. 5b) captures the dependence of the time needed for the community detection step computation on network size. As shown in the pie charts in Fig. 5a) for networks larger than 300 nodes, the community detection step became the most time-consuming component of the whole pipeline. We based the SVM classification on modularity coefficients. Thus the computation did not depend on network size because the whole network was represented by one value. We measured approximately 20 mins for this step.

4. Discussion

Our results show that decomposition using random matrix theory, which captures the group mode and discards the random and seasonal modes of the correlation matrix, increases the modularity coefficient (ability of network to form clusters). A higher level of filtering (we used 24 move-

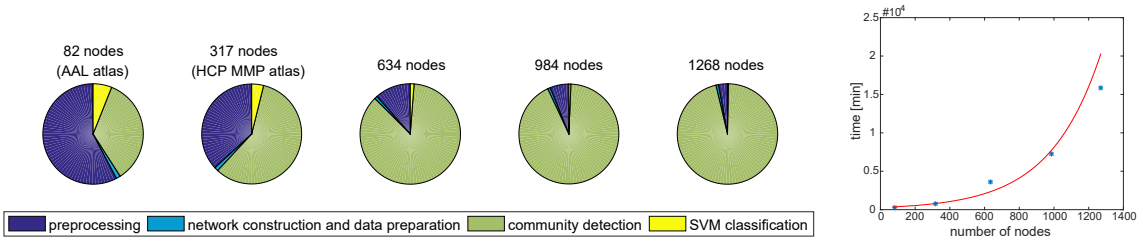

Figure 5: a) Relative time complexity of processing pipeline steps for networks of different sizes. b) Influence of network size on community detection computation length.

ment regressors and filtered out the signals from white matter and cerebrospinal fluid) relates to a higher modularity coefficient in community detection variants using RMT decomposition. However, RMT decomposition filtering out the global mode can be viewed as another kind of global signal filtering. Including GS filtering in the preprocessing pipeline produces worse results with regard to the modularity coefficient when RMT decomposition is applied, which could be caused by this twice-applied similar filtering. Since it has been shown in other studies [14, 15] that global signal filtering has an extreme influence on functional connectivity, we do not recommend using it because this influence transfers also to community structure detection. Our future work will target the appropriate use of RMT decomposition in functional brain time series.

For community detection we used the Louvain modularity method and the Potts spin-glass model. Only with the Louvain algorithm, however, did we measure a significantly altered – increased – modularity coefficient in Parkinson's disease as compared to healthy controls. MCI preceding Alzheimer's disease showed increased modularity versus controls only in one preprocessing variant – including high pass filtering, white matter and cerebrospinal fluid filtering, and random matrix theory decomposition step, albeit consistently in networks created on AAL as well as HCP MMP parcellation. High pass filtering (filtering out low frequencies and leaving frequencies higher than the cutoff threshold at 128s) enhances the difference from healthy controls. Our results differ from [8] who observed reduced modularity with increasing CDR (Clinical Dementia Rating) in preclinical AD. This discrepancy could be caused by different subject characteristics or the preprocessing or modularity method used.

The more prominent difference in PD from the controls can be explained by the characteristics of the group, which included patients already being treated for Parkinson's disease in its early stages. Although the patients were measured in the OFF medication state, the classification could have been even higher. On the other hand, the AD-MCI group was heterogeneous, the patients were not treated with medication, and they only displayed deficiencies in cognitive and blood markers known to be influenced by AD. Furthermore, the compensatory mechanisms and deficits also play a role. A more thorough study will be conducted to map the differences between groups also in terms of atrophy and cognitive decline. A longitudinal study (data has been obtained from patients one year after the first measurement) will help to differentiate between patients who progress to dementia related to AD and stable patients. This could further improve classification accuracy.

The classification analysis distinguishes between health and disease with 81.8% accuracy for PD and with 76.2% accuracy for AD-MCI. The classification accuracy measured in this study corresponds with that found in the literature. The study of [24] using metrics derived from diffusion

magnetic resonance imaging shows 78.2% accuracy in classifying AD from HC, 59.2% accuracy in classifying controls versus early AD-MCI, and 62.8% accuracy for controls versus late AD-MCI. A higher percentage was achieved in classifying patients with amnestic AD-MCI from healthy subjects by mean node strength of an NBS-network (network based statistics) computed on resting state fMRI [3]. They achieved an accuracy of 85.7%. An accuracy of 89.6% was measured when classifying Alzheimer's disease patients by network measures on networks of cortical thickness [25].

Since this work is mostly preliminary, we will further implement feature selection for SVM classifier based on t-values of inter-group statistics, which could increase the accuracy of our results even more.

Preprocessing is the most computationally demanding step for networks of small size (approximately up to 100 nodes). The subsequent steps—network construction and community structure detection—strongly depend on the number of network nodes and became the most time-consuming components for larger networks. Except for removing variants with global signal filtering and an a priori specification of preprocessing variants, we did not apply any sophisticated feature selection for classification analysis. Therefore, this step is more time demanding than it might otherwise be but still minimal when compared to other steps. It is not dependent on network size.

There are several possibilities and tools for faster and more efficient computation of individual pipeline steps. Thanks to the Parallel Computing Toolbox, a MATLAB user can run parallel computations (maximum for 12 processors). If still not sufficient, MATLAB also offers the Distributed Computing Server. Both require a paid license. A parallel version of SPM (the tool we used for preprocessing) is not yet available. The NeuroImaging Analysis Kit (NIAK)[5] includes parallel computing for preprocessing in MATLAB and OCTAVE. The fmriprep: A Robust Preprocessing Pipeline for fMRI Data[6] is a tool that includes parallelized steps that are significantly faster than typical preprocessing steps. The efficiency of metrics for computing relationships between nodes (such as mutual information and phase synchronization) is studied by [26], who show efficient computation of network metrics. Their FastFC tool[7] achieved 2 to 3 orders of magnitude faster computation of path length and betweenness centrality than the standard BCT toolbox [19] but slower performance for clustering coefficient and not so much difference in strength computation. Efficient SVM classification has been implemented, for example, by [27, 28] and many others.

5. Conclusions

In this work we measured the differences in community structure of resting-state fMRI functional connectivity between healthy controls and patients with mild cognitive impairment preceding Alzheimer's disease and patients with Parkinson's disease with mild cognitive impairment or normal cognition. These differences were measured on data with diverse preprocessing and community method variants. Increases in modularity coefficients were found in PD patients across preprocessing variants. The SVM classifier was able to distinguish PD patients from controls with 81.8% accuracy and 87.5% sensitivity. Differences between AD-MCI and controls in modularity

[5] `https://www.nitrc.org/projects/niak`
[6] `https://fmriprep.readthedocs.io/`
[7] `http://juangpc.github.io/FastFC/`

were found only for one preprocessing variant, and the classification accuracy was not so high (76.2% with 66.7% sensitivity). From these results we conclude that modularity is a promising marker for AD and PD although more extensive study of preprocessing and community detection variants is needed. Random matrix theory decomposition influences the strength of the modularity coefficient and its effect on localization of modules will be further studied.

Considering the complexity of the computation, the approach we used (and which is mostly used also by other neuroscientists) becomes time consuming (especially the community detection step of the pipeline) for networks with more than 300 nodes. Thanks to advances in acquisition parameters and more detailed parcellation atlases, the study of larger networks is becoming frequent and meaningful. This provides opportunities for implementation of more efficient algorithms for all pipeline steps.

Acknowledgments

This work was supported by the project Central European Institute of Technology 2020 (LQ1601) by the Ministry of Education, Youth and Sports of the Czech Republic; the EU Joint Programme – Neurodegenerative Disease Research (JPND) grant *Pre-clinical genotype-phenotype predictors of Alzheimer's disease and other dementias (APGeM)*; and the 15-33854A grant from the Czech Ministry of Health. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures".

References

[1] A.-L. Barabási, *Network science*. Cambridge University Press, 2016.

[2] A. Fornito, A. Zalesky and E. Bullmore, *Fundamentals of brain network analysis*. Academic Press, 2016.

[3] J. Wang, X. Zuo, Z. Dai, M. Xia, Z. Zhao, X. Zhao et al., *Disrupted functional brain connectome in individuals at risk for Alzheimer's disease*, *Biological Psychiatry* **73** (2013) 472–481.

[4] Z. Yao, Y. Zhang, L. Lin, Y. Zhou, C. Xu and T. Jiang, *Abnormal cortical networks in mild cognitive impairment and alzheimer's disease*, *PLoS Computational Biology* **6** (2010) .

[5] T. M. Nir, N. Jahanshad, A. W. Toga, M. A. Bernstein, C. R. Jack, M. W. Weiner et al., *Connectivity network measures predict volumetric atrophy in mild cognitive impairment*, *Neurobiology of Aging* **36** (jan, 2015) S113–S120.

[6] F. Vecchio, F. Miraglia, C. Marra, D. Quaranta, M. G. Vita, P. Bramanti et al., *Human Brain Networks in Cognitive Decline: A Graph Theoretical Analysis of Cortical Connectivity from EEG Data.*, *Journal of Alzheimer's disease : JAD* **41** (2014) 113–127.

[7] R. C. Petersen, R. Doody, A. Kurz, R. C. Mohs, J. C. Morris, P. V. Rabins et al., *Current Concepts in Mild Cognitive Impairment*, *Archives of Neurology* **58** (dec, 2001) 1985.

[8] M. R. Brier, J. B. Thomas, A. M. Fagan, J. Hassenstab, D. M. Holtzman, T. L. Benzinger et al., *Functional connectivity and graph theory in preclinical Alzheimer's disease*, *Neurobiology of Aging* **35** (2014) 757–768.

[9] A. F. Alexander-Bloch, P. E. Vértes, R. Stidd, F. Lalonde, L. Clasen, J. Rapoport et al., *The anatomical distance of functional connections predicts brain network topology in health and schizophrenia*, *Cerebral cortex* (2012) bhr388.

[10] D. S. Bassett, M. a. Porter, N. F. Wymbs, S. T. Grafton, J. M. Carlson and P. J. Mucha, *Robust detection of dynamic community structure in networks*, *Chaos* **23** (2013) , [1206.4358].

[11] I. Litvan, J. G. Goldman, A. I. Tröster, B. A. Schmand, D. Weintraub, R. C. Petersen et al., *Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines*, *Movement Disorders* **27** (2012) 349–356.

[12] L. Anderkova, M. Barton and I. Rektorova, *Striato-cortical connections in Parkinson's and Alzheimer's diseases: Relation to cognition*, *Movement Disorders* **00** (2017) 1–6.

[13] K. J. Friston, S. Williams, R. Howard, R. S. J. Frackowiak and R. Turner, *Movement-Related effects in fMRI time-series*, *Magnetic Resonance in Medicine* **35** (mar, 1996) 346–355.

[14] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen and M. E. Raichle, *From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks*, *Proceedings of the National Academy of Sciences* **102** (2005) 9673–9678.

[15] A. Weissenbacher, C. Kasess, F. Gerstl, R. Lanzenberger, E. Moser and C. Windischberger, *Correlations and anticorrelations in resting-state functional connectivity MRI: A quantitative comparison of preprocessing strategies*, *NeuroImage* **47** (oct, 2009) 1408–1416.

[16] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix et al., *Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.*, *NeuroImage* **15** (jan, 2002) 273–89.

[17] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub et al., *A multi-modal parcellation of human cerebral cortex.*, *Nature* **536** (2016) 171–8, [NIHMS150003].

[18] E. Bullmore and O. Sporns, *Complex brain networks: graph theoretical analysis of structural and functional systems.*, *Nature reviews. Neuroscience* **10** (2009) 186–198.

[19] M. Rubinov and O. Sporns, *Complex network measures of brain connectivity: Uses and interpretations*, *NeuroImage* **52** (2010) 1059–1069.

[20] V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, *Fast unfolding of communities in large networks*, *Journal of Statistical Mechanics: Theory and Experiment* **10008** (2008) 6, [0803.0476].

[21] M. Blatt, S. Wiseman and E. Domany, *Superparamagnetic Clustering of Data*, *Physical Review Letters* **76** (apr, 1996) 3251–3254.

[22] M. L. Mehta, *Random matrices*, vol. 142. Elsevier, 2004.

[23] M. MacMahon and D. Garlaschelli, *Community detection for correlation matrices*, *Physical Review X* **021006** (2015) 1–34.

[24] G. Prasad, S. H. Joshi, T. M. Nir, A. W. Toga and P. M. Thompson, *Brain connectivity and novel network measures for Alzheimer's disease classification*, *Neurobiology of Aging* **36** (jan, 2015) S121–S131.

[25] Y. Li, Y. Wang, G. Wu, F. Shi, L. Zhou, W. Lin et al., *Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dyna\mic and network features*, *Neurobiology of Aging* **33** (2012) 427—-e15.

[26] J. García-Prieto, R. Bajo and E. Pereda, *Efficient Computation of Functional Brain Networks: toward Real-Time Functional Connectivity*, *Frontiers in Neuroinformatics* **11** (2017) 1–18.

[27] S. Maji, A. C. Berg and J. Malik, *Classification using intersection kernel support vector machines is efficient*, in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, IEEE, IEEE, jun, 2008. DOI.

[28] B. Catanzaro, N. Sundaram and K. Keutzer, *Fast support vector machine training and classification on graphics processors*, in *Proceedings of the 25th international conference on Machine learning - ICML '08*, (New York, New York, USA), pp. 104–111, ACM, ACM Press, 2008. DOI.