Investigating community detection algorithms and their capacity as markers of brain diseases

Eva Hladka

Eva Vytvarova, Jan Fousek, Michal Mikl, Irena Rektorova

SITOLA
The Laboratory of Advanced Network Technologies

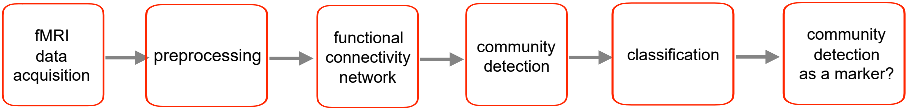
CEITEC

FACULTAS ARTIS INFORMATICAE · UNIVERSITAS MASARYKIANA ·

Schema

Aim: workflow for evaluation of brain functional connectivity with different community detection algorithms, and their strengths to discriminate between health and brain disease.

- fMRI measures brain function
- no consensual preprocessing pipeline



- community structure influenced by disease, e.g. Alzheimer's disease or schizophrenia (Brier, 2014; Alexander-Bloch, 2012)

Possible biological interpretation: communities represent groups of nodes that have different cognitive function (sight, memory, etc.). These groups can change in time (Bassett, 2013).

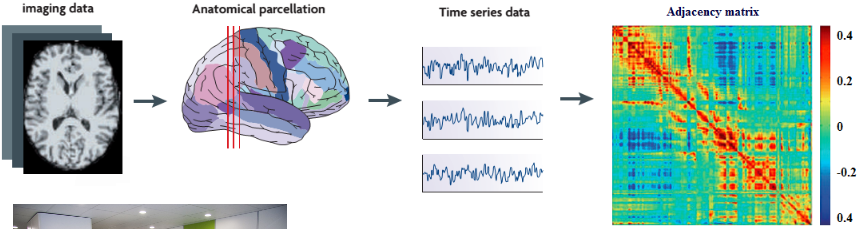
Data

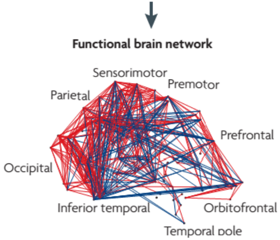
70 patients with mild cognitive impairment (MCI): 35 women; 66.71±9.44 years
50 healthy controls (HC): 34 women; 66.74±7.35 years

MCI

- intermediate stage between the expected cognitive decline of normal aging and the more-serious decline of dementia
- problems with memory, language, thinking and judgment greater than normal age-related changes
- goal of studying MCI: early diagnosis and slowing down the onset of dementia

Functional connectivity

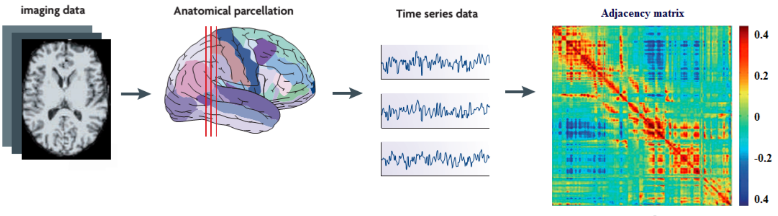


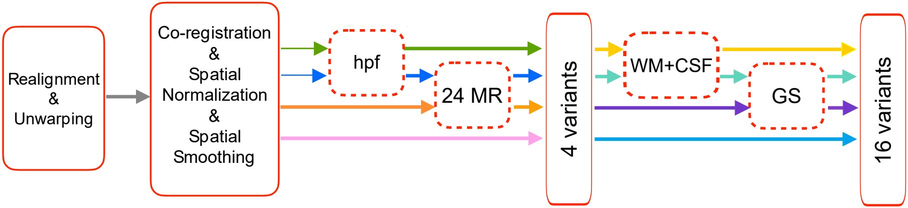


Sporns, Bullmore 2009

- 3 T resting-state fMRI, 7min, 200 scans
- 3 x 3 x 3 mm voxels, TR = 2.08 s
- nodes: 82 regions of AAL atlas (Tzourio-Mazoyer, 2002)
- edges: Pearson's correlation coefficients

Preprocessing





hpf . . . high pass filtering, cutoff at 128 s
MR . . . movement regressors
WM . . . white matter

CSF . . . cerebro-spinal fluid
GS . . . global signal

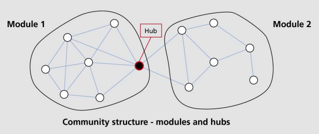
16 variants of preprocessing

Community structure and its detection

modules / communities / clusters / subnetworks / (temporo-)spatial patterns

- property of real complex networks
- dense connectivity within communities
- sparse connections between modules

- NP-complete problem
- optimization methods
- heuristics, some non-deterministic requiring repetitive computations



(Bullmore & Sporns, 2009)

Community detection

iterative community finetuning repeated 100x
representative/consensual partition across repetitions and subjects

Used methodology

- Louvain modularity method (Blondel, 2008)
- Potts spin-glass model (Blatt, 1996)
- random matrix theory – RMT (Mehta, 2004; MacMahon, 2015): identification of non-random properties of correlation matrices $\mathbf{C}$. It is based on eigenvalues computation. $\mathbf{C} = \mathbf{C}^{(r)} + \mathbf{C}^{(g)} + \mathbf{C}^{(m)}$, $\mathbf{C}^{(r)}$ ... random mode, $\mathbf{C}^{(g)}$ ... group mode, $\mathbf{C}^{(m)}$ ... market mode

Community detection – used approaches

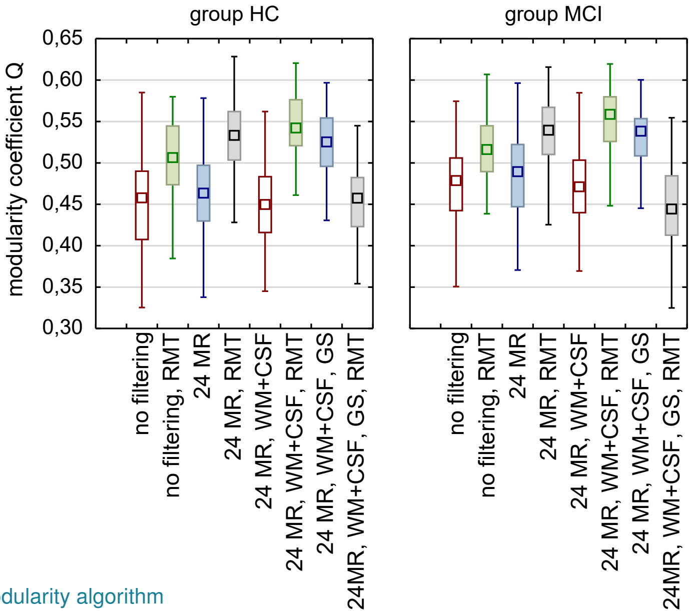
binary network, 15% sparsity threshold

- Louvain modularity method
- RMT + Louvain modularity method
- Potts modularity model
- RMT + Potts modularity model

Evaluated features

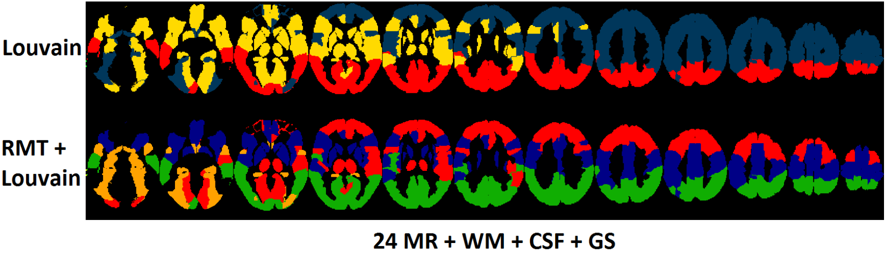
- modularity coefficient: ability of network to form clusters
- node classification to a community
- computational demand

Results: RMT step increases modularity



Louvain modularity algorithm

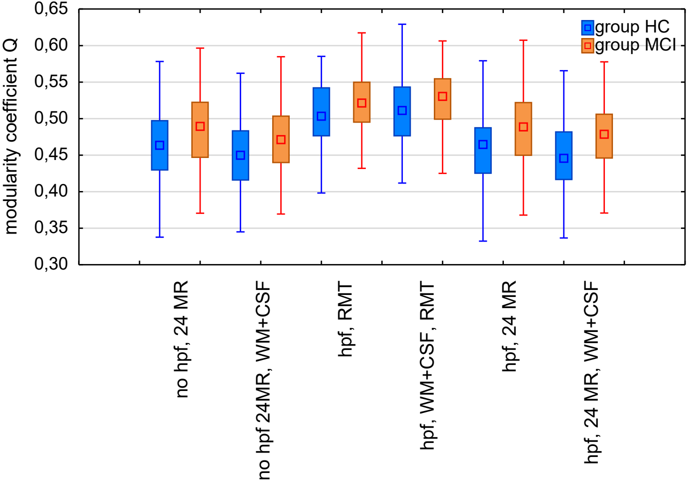
Results: global signal influences community structure



Extreme effect of global signal filtering on functional connectivity (Murphy, 2009).

We show this influence on communities' localization.

Results: HC *vs.* MCI changes in modularity coefficient



– t-tests between groups

– statistically significant differences ($p<0.05$) observed only for Louvain modularity with or without RMT decomposition

Classification: modularity as a marker of MCI

- variants without global signal filtering
- random sampling to train (75%) and test (25%) samples
- 10-fold cross-validation, 1000 iterations
- support vector machine (SVM) kernel: radial basis function
- age and gender taken into consideration

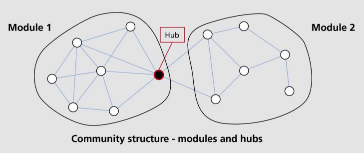
SVM using all preprocessing and filtering variants

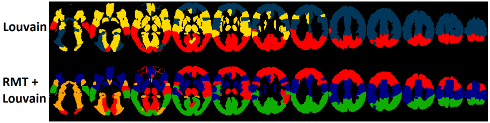
- classification accuracy = 78.9% (Train), 50.0% (Test)
- 75 support vectors (63 bounded)

SVM using preprocessing with hpf and RMT+Louvain modularity

- classification accuracy = 75.6% (Train), 63.3% (Test)
- 72 support vectors (60 bounded)

Physiological conclusions I.





24 MR + WM + CSF + GS

- Communities represent functionally specific clusters / modules.
- Decomposition by random matrix theory increases modularity coefficient.
- Higher level of filtering (24 MR, CSF+WM) relates to higher value of modularity coefficient in RMT variants.
- We do not recommend global signal filtering.

Physiological conclusions II.

- Louvain modularity is significantly increased in mild cognitive impairment.
- High pass filtering enhances the difference from healthy controls.
- However, the increase is not enough for classification analyses.
- Classification accuracy similar to literature:
  - 62.8% using diffusion MRI (Prasad, 2015)
  - 84% using fMRI when classifying Alzheimer's disease patients (Zhang, 2015)
  - 89.6% using cortical thickness when classifying Alzheimer's disease patients (Li, 2012)

Computational and time complexity of pipeline steps

per subject (120 subjects):

- data acquisition: subject preparation (tens of mins) + scanning ($\sim$10min). MR provider and sequence dependent.
- preprocessing: compulsory steps ($\sim$3min) + additional filtering ($\sim$1min); $\sim$160 thousand voxels per scan (200 scans). MATLAB.
- network construction: representative signal + Pearson's correlation ($<$ 1s); 82 nodes of 200 time-points signals. MATLAB.
- community detection algorithms: data loading and preparation ($\sim$7s) + community detection ($<$ 2s in average, in extreme up to 11s); RMT decomposition, 100 repetitions of finetuning, null models generating. MATLAB.

Computational and time complexity of pipeline steps

group level:

- community detection algorithms: $\sim$5 hours of computing community structure for all preprocessing variants and all subjects + representative partition computation ($\sim$7s per preprocessing variant, $\sim$2min in total); 120 subjects with 100 repetitions, 16 preprocessing variants. MATLAB.
- classification analysis: $\sim$5s for each combination of parameters, $\sim$20min in total; repetitions of train/test divisions, 1000 iterations of cross-validation. STATISTICA.

Computational and time complexity conclusions

- Preprocessing is the most computationally demanding step.
- Strongly depends on number of network nodes.
- Classification analysis time demanding because of missing feature selection.

Future work

- More sophisticated feature selection for classification is needed.
- More sophisticated parameter of community structure could better reveal the differences between groups.
- Community structure algorithms considering temporal evolution of connectivity may show more prominent difference between health and MCI (we're working on it).
- Better (parallel) implementation is needed for easier use.

Thank you.