Modular structure of brain networks in healthy and diseased brains

Anvar Kurmukov, Yulia Dodonova

NRU HSE
IITP RAS

Nizhny Novgorod, NET2017, 22-24 June

Introduction

- With rapid aging of human population **brain diseases** become an increasingly important problem
- **Magnetic resonance imaging** is a powerful, non-invasive technique that can be used for understanding progression of these diseases
- One possible way to understand how these diseases affect brain structure is to study macroscale brain graphs called **connectomes**
- It is expected that brain pathology primarily affects **modular structure** of brain graphs and changes the way how communicating brain regions join into communities

Goal

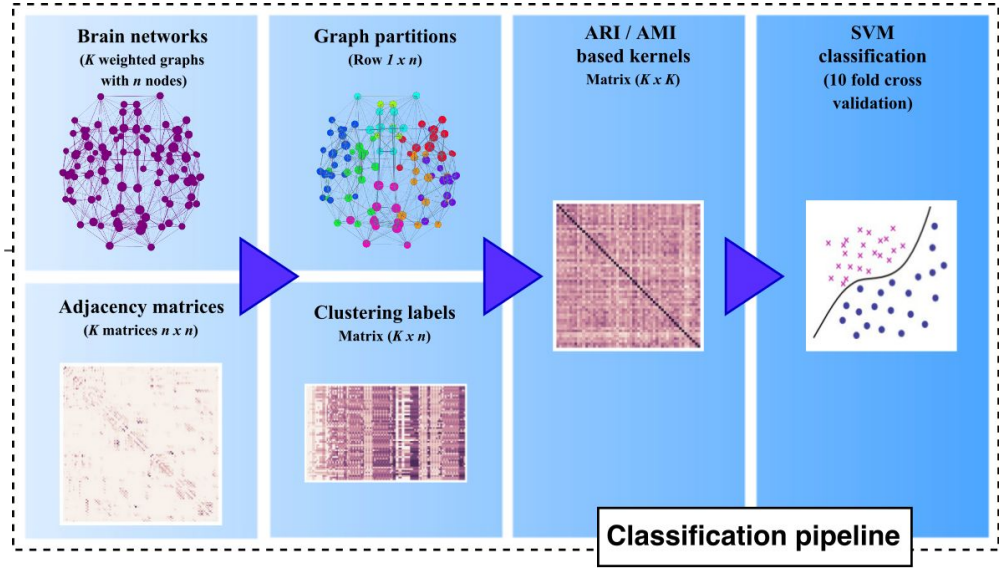
Develop a model for discriminating normal versus pathological phenotypes based on modular community structure of brain graphs.

Approaches

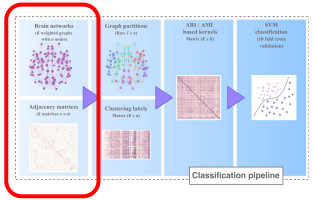
- Develop an algorithm for **evaluating similarity of brain graphs** based on similarity in the community structure of brain regions
- Develop a **classification pipeline** that uses the obtained measure of graph similarity and for each new incoming graph returns a label of its unknown phenotype
- **Validate** the proposed pipeline in different classification tasks based on real-life datasets

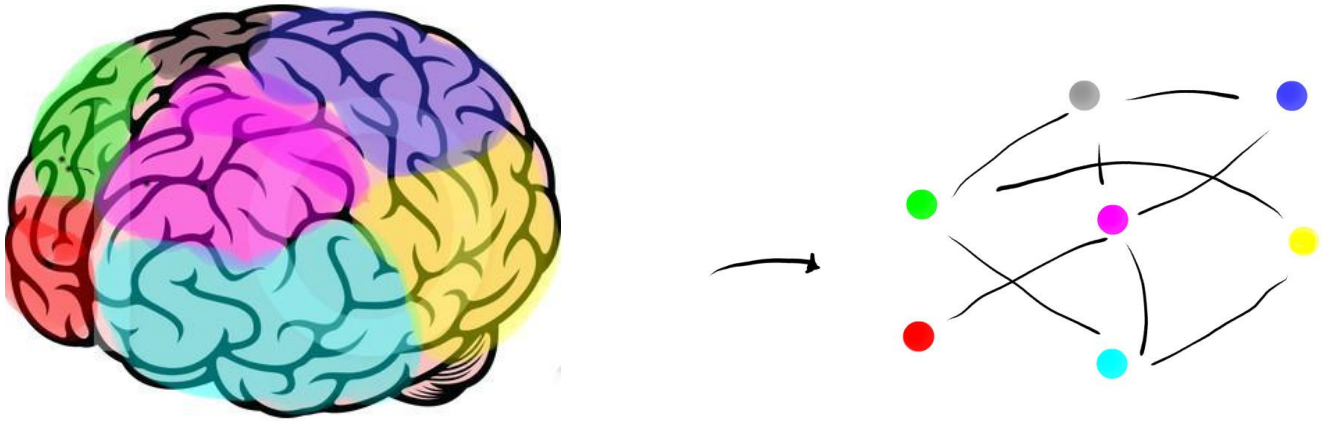
Content

1. What is a connectome?
2. Our proposed approach
3. Experiments
4. Conclusions



What is a connectome?

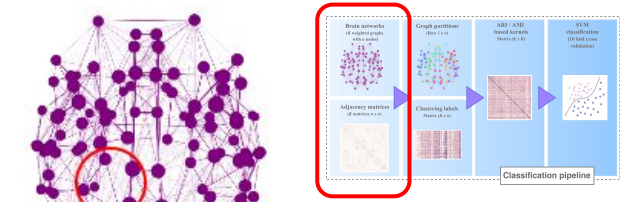


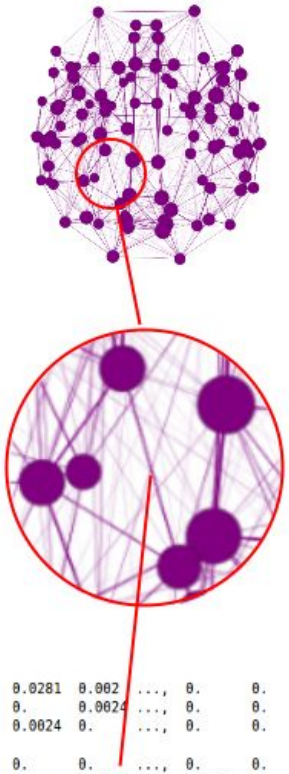


At a macroscale, **connectome** is a graph which nodes correspond to different brain regions, and edges are connections between these regions. We consider only structural connectomes (edges - neural pathways from one region to another)

Connectome: properties

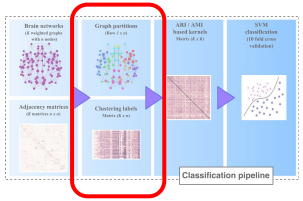
- connectomes are relatively **small** graphs, usually with at most few hundreds of nodes
- the graphs are **undirected**, i.e. the adjacency matrices are symmetric
- edges are **weighted**
- graphs are **connected**
- **each node is uniquely labeled (according to the brain region), and the set of labels is the same across connectomes**
- nodes are positioned in **3D space**





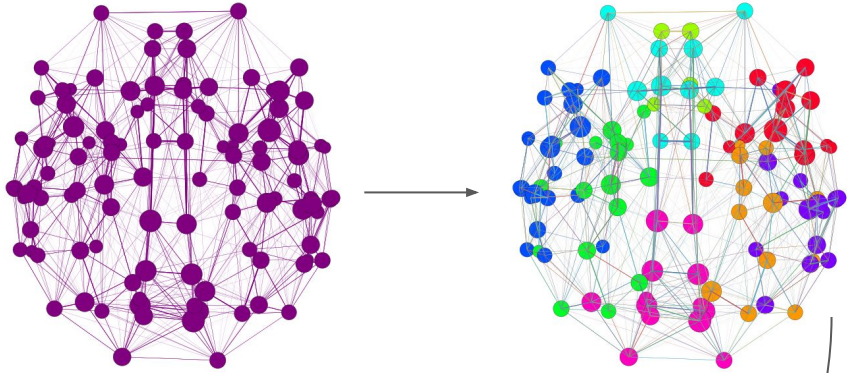
```
[ 0.     0.0281  0.002  ...,  0.     0.     0.    ]
[ 0.0281 0.      0.0024 ...,  0.     0.     0.    ]
[ 0.002  0.0024  0.     ...,  0.     0.     0.    ]
...,
[ 0.     0.      0.     ...,  0.     0.     0.    ]
[ 0.     0.      0.0059 ...,  0.003  0.     0.0004]
[ 0.     0.      0.     ...,  0.     0.0005 0.    ]
[ 0.     0.      0.     ...,  0.     0.     0.    ]
[ 0.     0.      0.     ...,  0.     0.     0.    ]
[ 0.     0.      0.     ...,  0.     0.0059 0.    ]
...,
[ 0.     0.      0.     ...,  0.     0.     0.002 ]
[ 0.     0.0017  0.     ...,  0.     0.     0.001 ]
[ 0.     0.      0.     ...,  0.002  0.001  0.    ]
```

Graph partitioning



- Approximate
  1. **Newman eigenvector**
  2. **Louvain**
  3. **Greedy modularity optimization**

  - *Very fast*
  - *Suboptimal*



[0 0 1 1 2 2 0 ... 4 6 4 5 5 5 5]

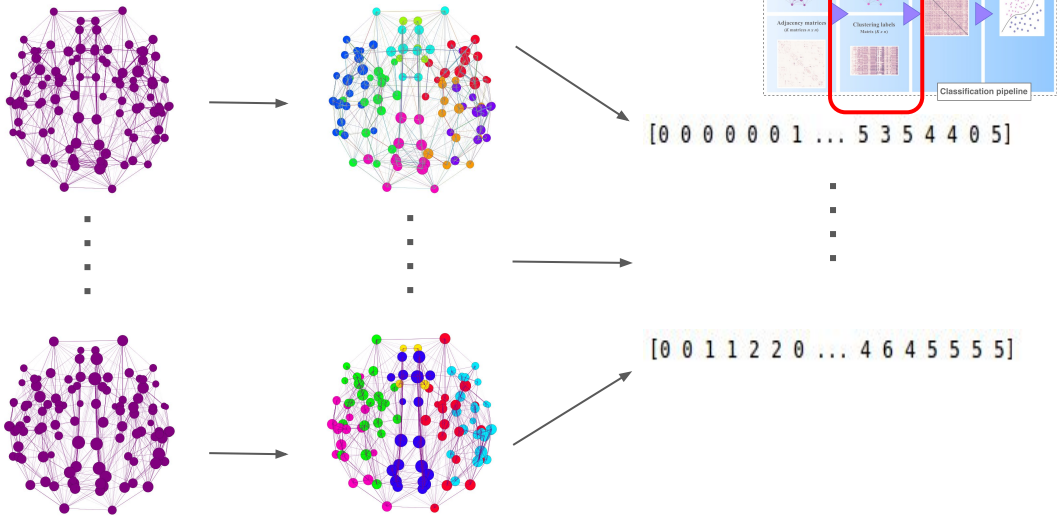
All algorithms optimize modularity Q which is given by the formula:

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(i,j)$$

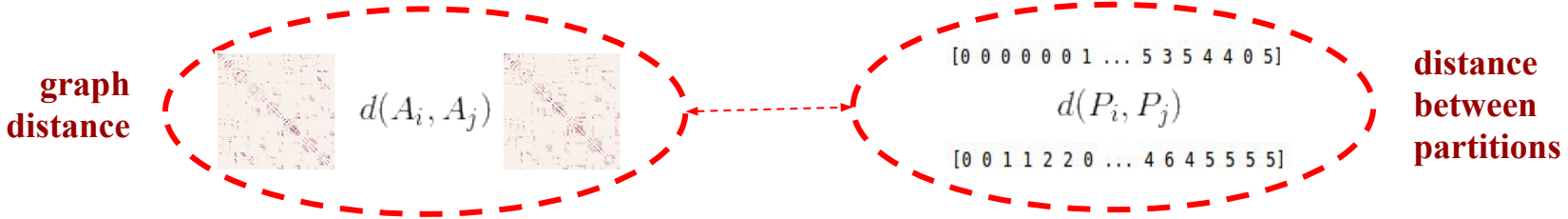
1. M.E.J. Newman, 2006
2. Vincent D Blondel et al., 2008
3. Clauset A. et al, 2004

Similarity of graph partitions

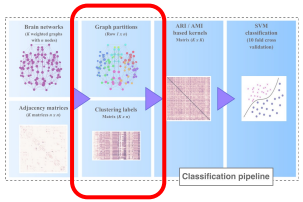
For each graph, we obtain its best partition P which is a vector of length n, where n is the number of nodes.
i-th value in P represents community label of an i-th node.



Given a set of graphs $X = \{G_1, \ldots G_M\}$, we obtain partitions $\{P_1, \ldots P_M\}$.
Now we will compare graphs based on similarity in their partitions into communities.



Similarity between partitions



ARI and AMI are indifferent to cluster relabeling

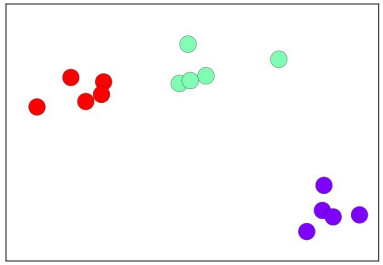
- **Adjusted Rand Index**

  ARI (P_1, P_1) = 1.0
  ARI (P_1, P_2) = 1.0
  ARI (P_1, P_3) = 0.479
  ARI (P_1, P_4) = 0.042

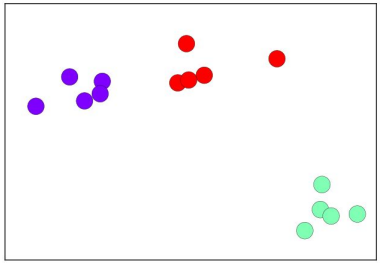
- **Adjusted Mutual Information**

  AMI (P_1, P_1) = 1.0
  AMI (P_1, P_2) = 1.0
  AMI (P_1, P_3) = 0.529
  AMI (P_1, P_4) = 0.049

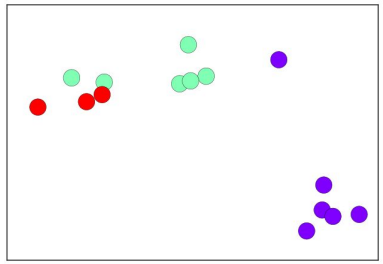
Partition 1 : [0 0 0 0 0 1 1 1 1 1 2 2 2 2 2]



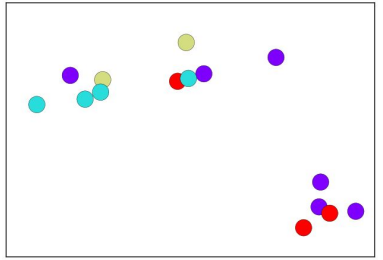
Partition 2 : [1 1 1 1 1 2 2 2 2 2 0 0 0 0 0]



Partition 3 : [0 0 0 0 0 0 1 1 1 1 1 1 2 2 2]



Partition 4 : [0 0 0 3 3 0 2 0 3 1 2 0 1 1 1]

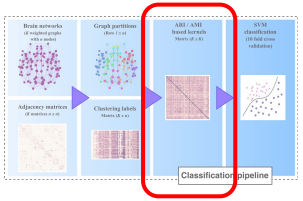


Both ARI and AMI take the value 1 when two partitions are identical and values close to 0 for random labeling

Take (1-ARI) and (1-AMI) to obtain distances

1. W. M. Rand, 1971 2. Vinh N. X., et al., 2010

How to classify graphs?



Kernel classifiers

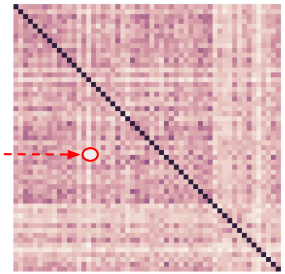
Define a positive semi-definite function (**kernel**) on graphs and feed the resulting Gram matrix to the SVM (support vector machines)

If we introduce a distance $d(G_i, G_j)$ between the two graphs, a kernel can be produced by:

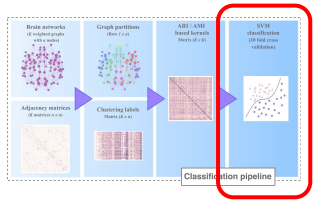
$$K(G_i, G_j) = e^{-\alpha d(G_i, G_j)}$$

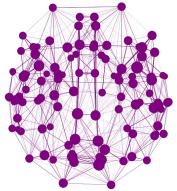
Compute distance between two connectomes as a distance between their partitions

1-AMI (P_i, P_j)
or
1-ARI (P_i, P_j)



Classification pipeline



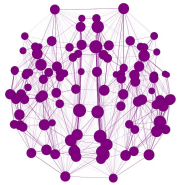


[0 0 0 0 0 0 1 ... 5 3 5 4 4 0 5]

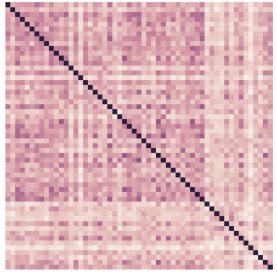
Step 0
Choose clustering algorithm

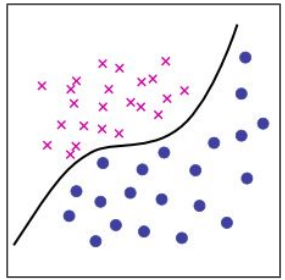
Step 1
obtain best partition for each graph

Step 2a
compute pairwise distances between partitions



[0 0 1 1 2 2 0 ... 4 6 4 5 5 5 5]





Step 3
produce a kernel by exponentiating distances and run kernel SVM

Produce baseline classification quality:

Step 2b
compute pairwise *L1* (Manhattan) and *L2* (Euclidean) distances between graph adjacency matrices

Data description

UCLA Autism

DTI-based connectivity matrices of **51 (ASD)** high-functioning autism spectrum disorder subjects and **43 (TD)** typically developing subjects. Total of 94 graphs with **264 nodes** each.

UCLA APOE

Carriers versus non-carriers of the APOE-4 allele associated with the higher risk of Alzheimer's disease, total of 55 graphs. **30 APOE-4** non carriers and **25 APOE-4** carriers, with **110 nodes** each.

ADNI

Alzheimer‘s Disease Neuroimaging Initiative (ADNI2) database which comprises a total of **228** individuals (756 scans). The data include 47 people with AD (136 AD scans), 40 individuals with LMCI (147 LMCI scans), 80 individuals with EMCI (283 EMCI scans), and 61 healthy participants (190 scans), with **68 nodes** each.

Comparison results

Task / Kernel	Autism vs Normal		APOE-4 vs APOE-3		Alzheimer's vs Normal	
Edge type	binarized	weighted	binarized	weighted	binarized	weighted
Louvian	.60 / .57	.59 / .52	.57 / .57	.69 / .58	.80 / **.83**	.73 / .73
Newman	.43 / .45	.56 / .58	.53 / .56	**.70** / .63	.76 / .76	.66 / .68
Greedy optimization	.63 / .61	.50 / .49	.53 / .67	.59 / .56	.72 / .72	.69 / .72
Euclidean distance	.54	**.64**	.47	.52	.80	.72
Manhattan distance	.48	.48	.45	.58	.54	.49

Table 1. Comparison results for all tasks (UCLA Autism dataset, UCLA APOE dataset, ADNI2 dataset). First value for partition based kernels is ARI based kernel, second is AMI based kernel. Best kernels for each classification task are bolded. Perfomance is measured in terms of area under ROC

Comparison results

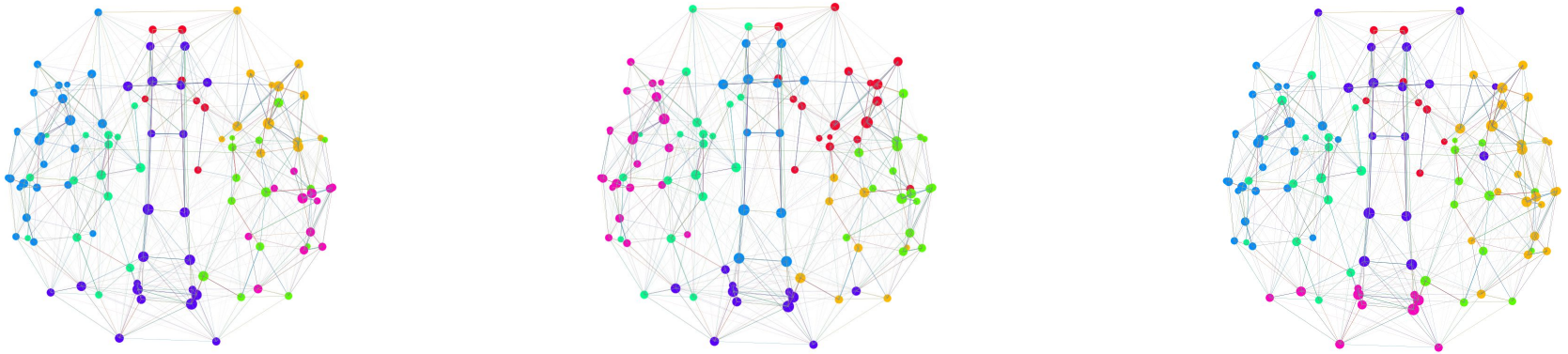


Figure 1. From left to right: Louvain, Newman, Greedy optimization. All partitions obtained for a single graph from UCLA APOE data set

- **Different** in terms of partitioning
- **Very simular** in terms of modularity score
- Using **soft partitioning** instead of strict ones might improve the situation

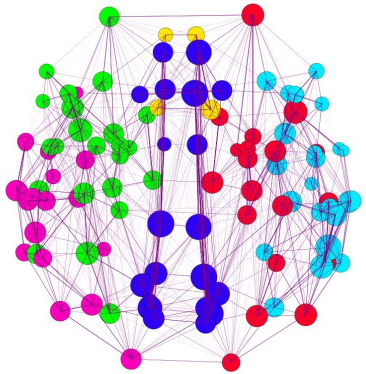
	Louvain	Newman	Greedy optimization
Louvain	1	0.57	0.66
Newman	0.57	1	0.60
Greedy optimization	0.66	.60	1

Table 2. Comparison of partitions in terms Adjusted Rand Index

Conclusions

- **We proposed a pipeline for classifying normal and pathological phenotypes based on the community structure of brain graphs**
- We proposed to evaluate **distance between brain graphs** based on whether or not brain regions in these graphs similarly cluster into communities
- We developed a **classification model** that uses information about these distances in deciding whether a newly incoming graph represents normal or diseased brain
- Based on **real-life datasets** on both neurodegenerative and psychiatric disorders, we demonstrated that the proposed model outperforms the baselines thus supporting the idea that brain pathology crucially affects the entire structure of brain networks

Thank you for your attention!



Modular structure of brain networks in healthy and diseased brains

Anvar Kurmukov, Yulia Dodonova

National Research University Higher School of Economics
The Institute for Information Transmission Problems, Russian Academy of Science

kurmukovai@gmail.com