Classification of Structural Brain Networks Based on Information Divergence of Graph Spectra

Yulia Dodonova, Sergey Korolev, **Anna Tkachev**, Dmitry Petrov, Leonid Zhukov and Mikhail Belyaev
ya.dodonova@mail.ru, annatkachev42@gmail.com

RUSSIAN ACADEMY OF SCIENCES
INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS (Kharkevich Institute)
HIGHER SCHOOL OF ECONOMICS NATIONAL RESEARCH UNIVERSITY

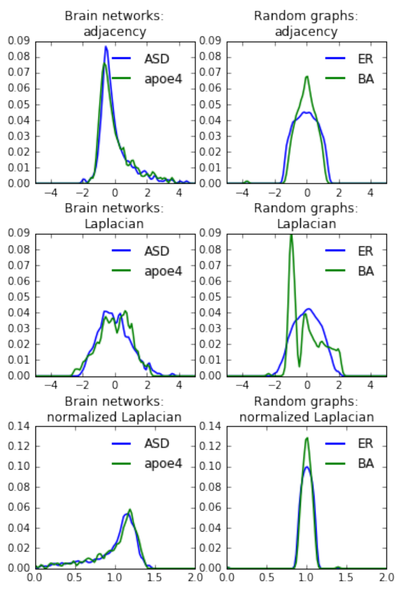
Abstract

We tackle the problem of brain network classification with machine learning algorithms using spectra of networks' matrices. First, linear and tree-based models are trained on the vectors of sorted eigenvalues of the adjacency matrix, the Laplacian matrix and the normalized Laplacian. Next, SVM classifier is trained with kernels based on information divergence between the eigenvalue distributions. The latter approach gives promising results in the classification of autism spectrum disorder versus typical development and of the carriers versus noncarriers of an allele associated with the high risk of Alzheimer disease.

Graph Spectra

One way to encode essential graph properties in a compact manner is to compute spectrum of its matrices:

- **Adjacency matrix**, the $n \times n$ matrix with entries a_{ij}, where a_{ij} is the weight between the respective nodes, n is the number of nodes.
- The **graph Laplacian**: $L = D - A$, where D is a diagonal matrix of weighted node degrees: $d_i = \sum_j a_{ij}$.
- The **normalized Laplacian**: $\mathcal{L} = D^{-1/2} L D^{-1/2}$



We plot spectra distributions of these matrices for the group average brain networks (see data description below) and the spectra of two random graphs: Erdös-Rényi (ER) graph and the Barabási-Albert (BA) preferential attachment network. The distributions for human brain connectivity matrices are very close; both differ from those obtained with the random graphs.

Data

UCLA Autism dataset includes DTI-based connectivity matrices of 51 high-functioning autism spectrum disorder (ASD) and 43 typically developing (TD) subjects. Nodes for connectomes are defined based on a large meta-analysis of fMRI studies; this approach produces 264 equal-size brain regions. Network edges result from brain deterministic tractography.

UCLA APOE-4 dataset includes DTI-based connectivity matrices of 25 carriers and 30 noncarriers of the APOE-4 allele associated with the higher risk of Alzheimer's disease. To define network nodes, each brain is partitioned into 110 regions using the Harvard-Oxford subcortical and cortical probabilistic atlases. Network edges are obtained using the FACT algorithm; raw fiber counts are adjusted for the unequal region volumes.

Network construction

For each dataset, we apply three weighting schemes:

- Original matrices scaled by: $a_{ij}^{scaled} = \frac{a_{ij}}{\sum_{i,j} a_{ij}}$.
- Binarized weights: $a_{ij}^{binarized} = 1 if a_{ij} > 0,\ 0\ else.$
- Original edge weights scaled by the Euclidean distance l_{ij} between centers of the regions i and j: $a_{ij}^{weighted} = \frac{a_{ij}}{l_{ij}^2}$, next scaled by the sum of elements.

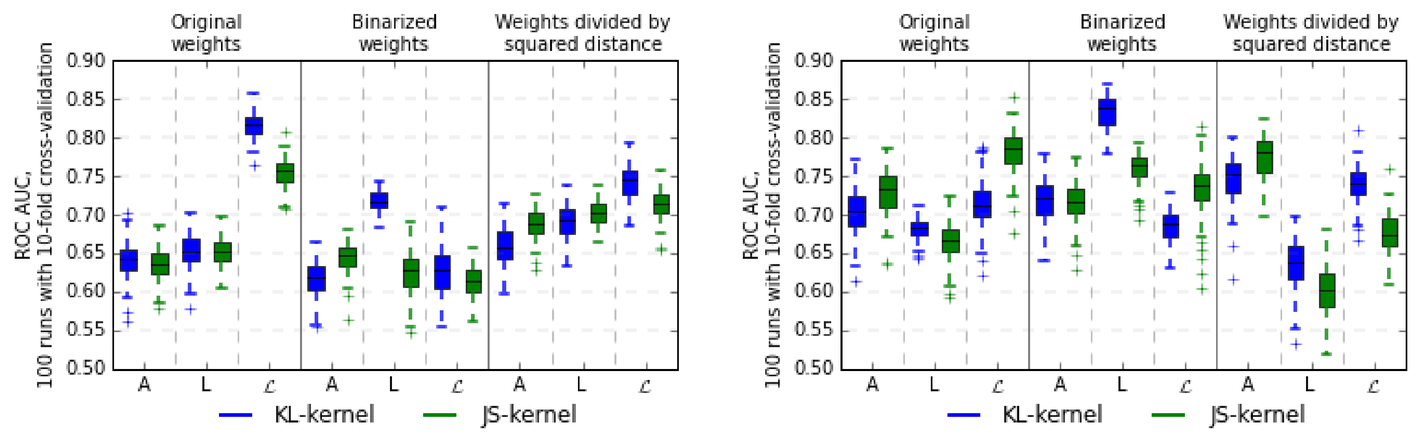
Results: kernel-based classification

The basic idea of our proposed approach is to measure the pairwise differences between the distributions of graph spectra, construct kernels based on these differences and use them in an SVM classifier.

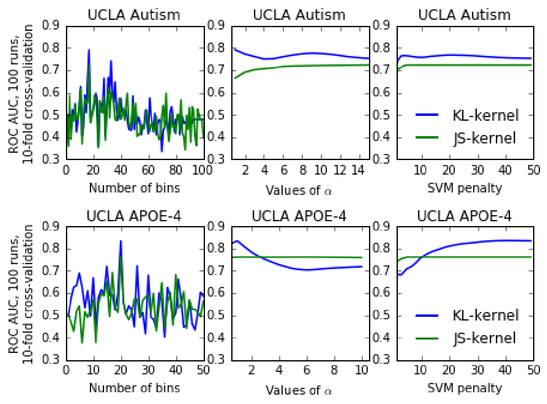
For two probability distributions with densities $p(x)$ and $q(x)$ the **Kullback-Leibler (KL) divergence** is: $KL(p||q) = \int_{-\infty}^{\infty} p(x) log \frac{p(x)}{q(x)} dx$. The **KL kernel** is obtained by exponentiating the symmetric KL divergence: $K_{KL}(p,q) = e^{-\alpha(KL(p||q)+KL(q||p))}$.

The **Jensen-Shannon (JS) divergence** is: $JS(p||q) = \frac{1}{2}(KL(p||r) + KL(q||r))$, where $r(x) = \frac{1}{2}(p(x) + q(x))$. We compute **JS kernel** by: $K_{JS}(p,q) = e^{-\alpha\sqrt{JS(p||q)}}$.

These kernels work with the probability density functions restored from the samples; we use sample frequencies as a proxy for the probabilities and vary the number of bins. We also vary the parameter α used to compute a kernel and a penalty parameter of the SVM classifier. A figure shows the result for UCLA Autism (left) and UCLA APOE-4 (right).



The best classification result on the **UCLA Autism** dataset is obtained with the **KL-based kernel on the normalized Laplacian spectra** of the matrices with the original weights (**ROC AUC 0.817±0.017**). For the **UCLA APOE-4** dataset, the highest classification quality is achieved with the **KL-based kernel on the scaled Laplacian spectra** of the binarized matrices (**ROC AUC 0.834±0.022**).



For these best models, we explore how the results change depending on each of the three parameters. The ROC AUC values remain smooth and almost constant with respect to values of α and the SVM penalty. However, the graph that plots the results as a function of the number of bins shows that **the algorithm is highly sensitive to the procedure of density reconstruction**.

Baseline: classification based on feature vectors

We produce the following sets of features: **bag of edges** (the vectorized upper triangle of the adjacency matrix), weighted **node degrees**, and sorted **eigenvalues of A, L and $\mathcal{L}$**. We run linear (**linear SVM and logistic regression** with elastic-net regularization) models and **boosted decision trees**. A table below shows the mean **ROC AUC** values obtained over 100 runs of the algorithm with different 10-fold cross-validation splits (top and bottom rows in each cell refer to best linear and nonlinear models, respectively).

Features	UCLA Autism			UCLA APOE-4		
	Original weights	Binarized weights	Weights by l^2	Original weights	Binarized weights	Weights by l^2
Edges	0.525 0.595	0.515 0.510	0.531 0.523	0.552 0.570	0.558 0.563	0.550 0.520
Degrees	0.539 0.534	0.515 0.515	0.545 **0.754**	0.550 0.548	0.550 0.545	0.631 0.704
A spectra	0.515 0.687	0.509 0.506	0.516 0.647	0.550 0.717	0.550 **0.780**	0.550 0.533
L spectra	0.585 0.501	0.515 0.508	0.534 0.581	0.550 0.525	0.550 0.638	0.604 0.512
$\mathcal{L}$ spectra	0.514 0.556	0.507 0.504	0.515 0.638	0.550 0.584	0.689 0.647	0.592 0.506

For the **UCLA Autism** dataset, the best result is obtained with the **boosted decision trees trained on the weighted node degrees**. For the **UCLA APOE-4** dataset, the best result is achieved with the **boosted decision trees on the adjacency spectra**.

Acknowledgements

The study was supported by the Russian Science Foundation (project 14-50-00150). We gratefully acknowledge UCLA Multimodal Connectivity Database for making the datasets available to the community.