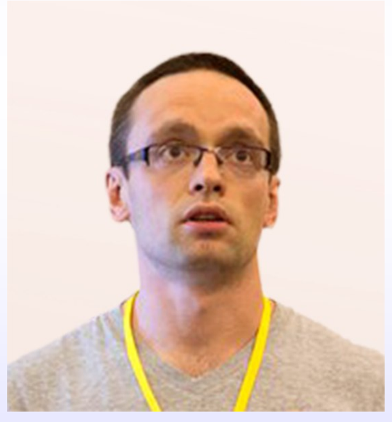

Boosting Connectome Classification via Combination of Geometric and Topological Normalizations

Dmitry Petrov, Yulia Dodonova, Leonid Zhukov and Mikhail Belyaev
to.dmitry.petrov@gmail.com, ya.dodonova@mail.ru, lzhukov@hse.ru and belyaevmichel@gmail.com

RUSSIAN ACADEMY OF SCIENCES
INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS (Kharkevich Institute)

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Abstract

We propose a data prepossessing method that combines geometric and topological connectome normalizations and significantly improves classification results. We validate this approach by performing classification between autism spectrum disorder (ASD) and typical development (TD) connectomes. **We demonstrate a significant enhancement in performance using weighted node degrees of normalized connectomes** over the best model trained on baseline features.

Normalizations

We use three different weighting schemes. First, weights a_{ij} from original connectomes. Second, we obtain binary weights: $a^b_{ij} = 1$ if $a_{ij} > 0$, 0 else. Finally, **we propose weighting by squared distance between nodes (geometric normalization)**:

$$\tilde{a}_{ij} = \frac{a_{ij}}{l^2_{ij}}, \quad (1)$$

where l_{ij} is the Euclidean distance between centers of regions i and j, which are the same for each subject. Then we apply weighted communicability normalization [1] to each of the three weighted sets of connectomes:

$$\hat{w}_{ij} = \frac{w_{ij}}{\sqrt{d_i d_j}}, \quad (2)$$

where w_{ij} is weight of edge between nodes i and j; d_i is weighted degree of node i. For purpose of comparison, we also work with the respective sets of non-normalized weighted connectomes. Hence, a combination of three weighting schemes and two normalization strategies produces six datasets.

Features

For each of these six datasets **we report results for 264 weighted node degrees as feature vectors. To produce an outer baseline, we calculate six global network metrics** used by the authors of the dataset [2]:

1. Weighted clustering coefficient:

$$CC = \frac{1}{n}\sum_{i \in V} \frac{2t_i}{d_i(d_i - 1)}, \quad (3)$$

where t_i is the number of triangles for the node i.

2. Characteristic path length:

$$CPL = \frac{1}{n}\sum_{i \in V} \frac{\sum_{j \in V, j \neq i} g_{ij}}{n-1}, \quad (4)$$

where g_{ij} is the length of the shortest path (geodesic) between the vertices i and j.

3. Normalized CC: $\lambda = \frac{CC}{CC_{rand}}$, where CC_{rand} is the average CC from simulated random networks.

4. Normalized CPL: $\gamma = \frac{CPL}{CPL_{rand}}$, where CPL_{rand} is the average CPL from the same random networks.

5. Small-worldness: $\sigma = \frac{\lambda}{\gamma}$.

6. Modularity: $Q = \frac{1}{2m}\sum_{ij}[A_{ij} - \frac{d_i d_j}{2m}]\delta(c_i c_j)$, where m is the sum of weighted edges in the network, and c is the community.

Data

We use publicly available UCLA autism dataset [3]. It includes DTI-based connectivity matrices of 51 high-functioning ASD subjects (6 females) and 43 TD subjects (7 females). 264 nodes of the connectomes were defined using parcellation scheme proposed by Power et al. [4] based on a large meta-analysis of fMRI studies combined with whole brain functional connectivity mapping.

Results

We perform binary classification task ASD vs TD on each of the seven datasets described above. **We use four classifiers: logistic regression; SVM with linear kernel; random forest, and boosted decision trees (BDT)**. For linear methods we scale features with min-max scaling and apply elastic net regularization. Prediction quality of algorithms is measured by area under the receiver operating characteristic curve (ROC AUC). To avoid overfitting we test models with the best parameters on 100 10-fold splits and report results in terms of ROC AUC distributions.

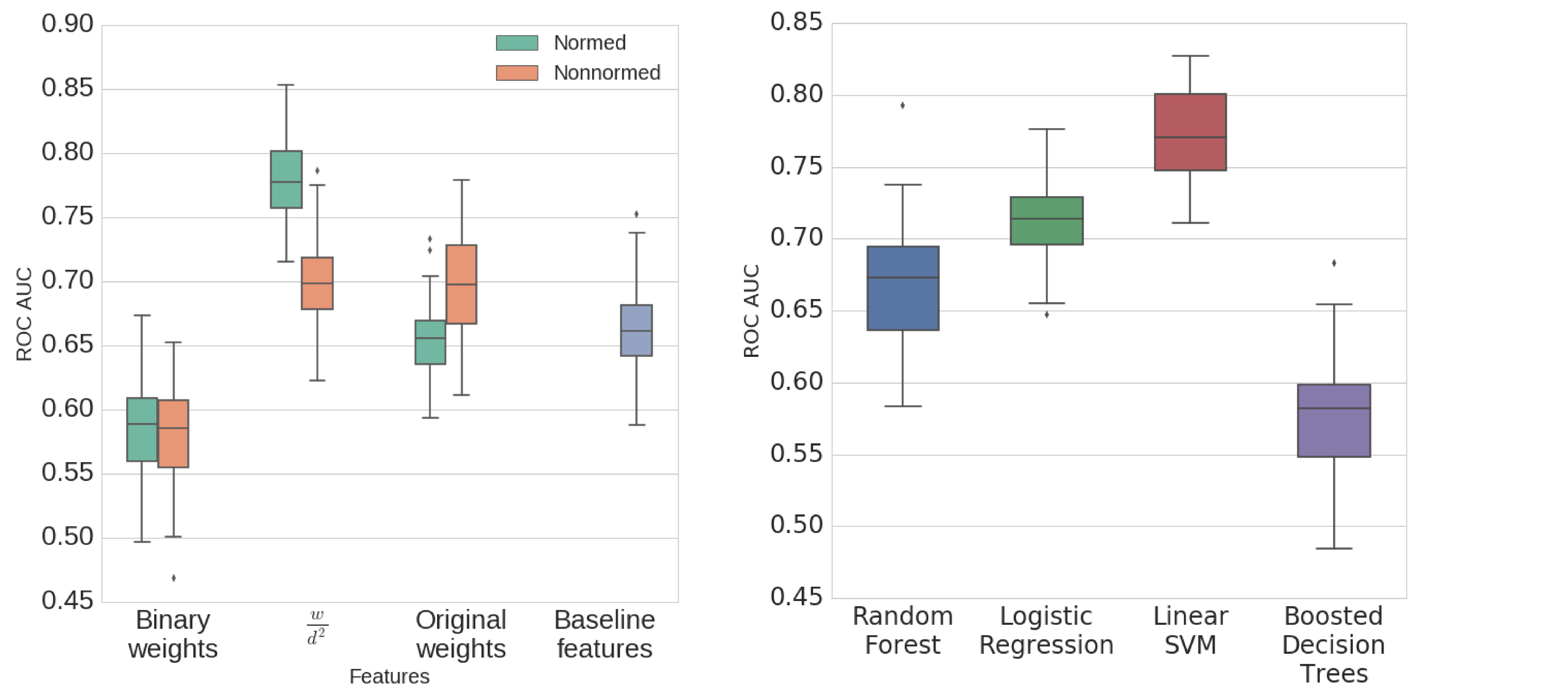

Figure 1: Left: best models' results on each of seven sets of features. Right: performance of classifiers on the best set of features – weighted node degrees of the connectomes normalized by proposed combination.

As seen in Figure 1 **combination of normalization by distance and topological normalization gives best classification performance on weghted node degrees.** Best model (linear SVM) on this set of features performs (0.77 ROC AUC) significantly better ($p = 7.8 \times 10^{-18}$, Wilcoxon test with Bonferroni post-hoc correction) than the best model (BDT) on the baseline feature set with 0.66 mean ROC AUC. All pairwise differences between the results on our datasets and the baseline are significant (Wilcoxon test on ROC AUC values with Bonferroni post-hoc correction has p-values less than 10^{-8}, except the difference between the baseline and the degrees on normalized original weights with $p = 0.028$).

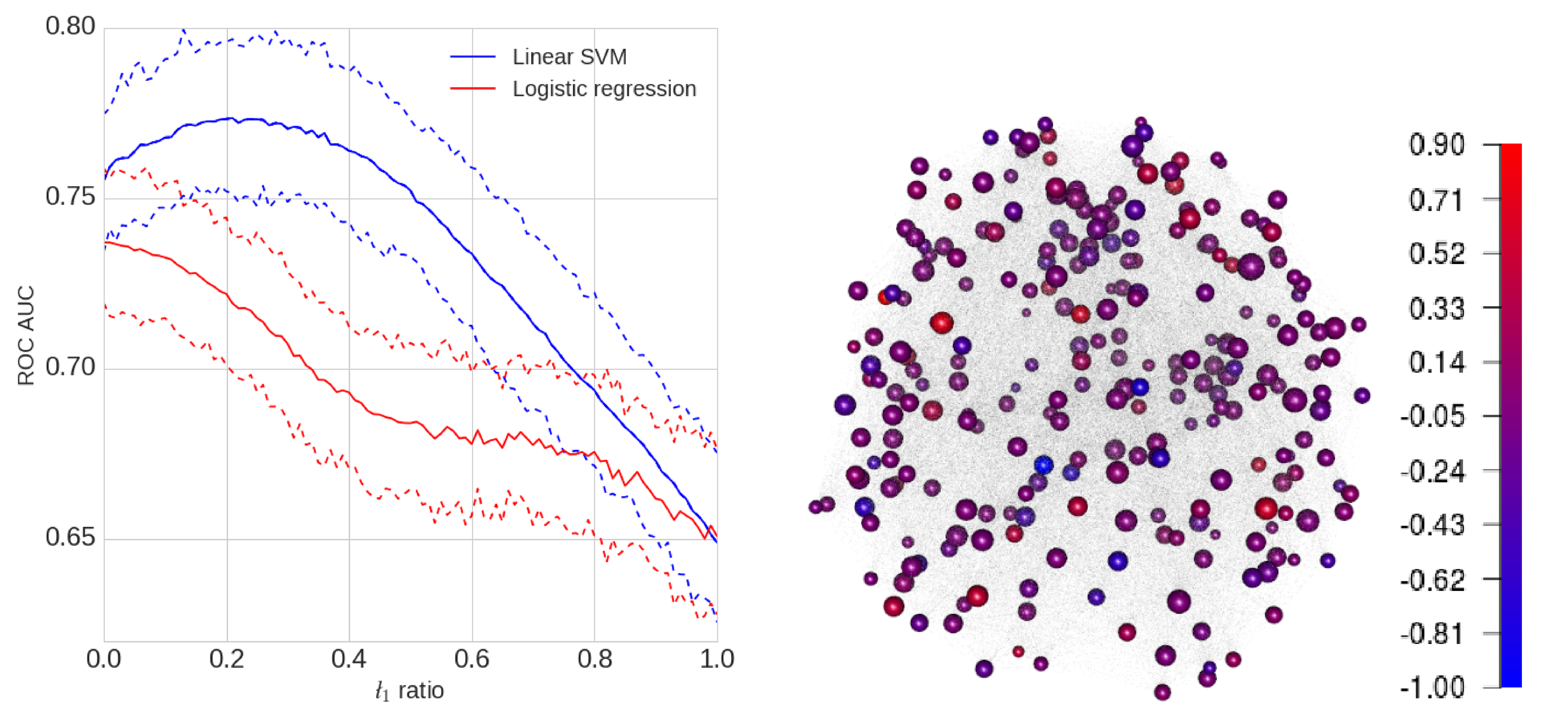

Figure 2: Left: Effects of regularization in terms of ROC AUC values for linear SVM and LR for 101 l_1-ratios from 0 to 1 in 0.01 increments with α's fixed at 0.01 for SVM and 0.0008 for LR. Right: Zone centers in their physical coordinates (axial view). Node color represents mean absolute SVM weight of the respective node. Node size is proportional to group average node degrees on the best set of features.

It is possible that we found a specific pattern for this particular classification task. For example, short connections may be more important for ASD patients. **If you have data to test our pipeline or suggestions for collaboration, please contact me at the conference or by email to.dmitry.petrov@gmail.com**.

References

[1] Crofts, J.J., Higham, D.J. A weighted communicability measure applied to complex brain networks. Journal of The Royal Society Interface 6, 411–414 (2009).

[2] Rudie, J.D., Brown, J.A., Beck-Pancer, D., Hernandez, L.M., Dennis, E.L., Thompson, P.M., et al. Altered functional and structural brain network organization in autism. Neuroimage Clin 2, 79–94 (2013)

[3] Available online at: http://umcd.humanconnectomeproject.org

[4] Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E. Functional net- work organization of the human brain. Neuron 72, 665–678 (2011)