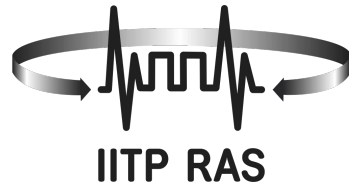


# Structural Connectome Validation using Pairwise Classification

Dmitry Petrov <sup>1 2</sup>, Boris Gutman <sup>1</sup>, Alexander Ivanov <sup>2</sup>, Joshua Faskowitz <sup>3 1</sup>,  
Neda Jahanshad <sup>1</sup>, Mikhail Belyaev <sup>2,4</sup> and Paul Thompson <sup>1</sup>



April 20  
2017

# Motivation

- Predictive modeling using DWI-based features (in particular, structural connectomes) has become a popular subgenre of neuroimaging (Arbabshirani et al., 2016)
- The great variety of possible pre-processing steps (e.g. non-linear registration, parcellation, or tractography) leads to potential challenges in downstream application of the connectomes, for example, in a classification task.
- The performance of a particular case-control classifier may not suffice as a means of data verification due to small samples and high dimensionality
- More objective validation may be needed in addition to frequently used Intra-class Correlation Coefficient (ICC) on test-retest data

# Pairwise classification

- Let's assume we have set of connectomes  $C_j^i$ , where where i-indices correspond to images and j-indices correspond to subjects and feature mapping  $f : C \rightarrow \mathbb{R}^d$
- For each pair of connectome feature vectors  $(f(C_{j_1}^{i_1}), f(C_{j_2}^{i_2}))$  we assign target variable 1 if they are from the same subject, 0 — else; then we construct three pairwise differences of these vectors according to  $l_1, l_2$  and  $l_\infty$  norms
- We run linear classification on these three features and report ROC AUC score
- Two-step validation procedure. First, grid search based on a 10-fold cross-validation with a fixed random state for reproducibility. Second, evaluation of the best parameters on 100 train/test splits (test size was set to 20% of data).

# Datasets: ADNI

- 227 individuals from the Alzheimers Disease Neuroimaging Initiative (ADNI2).
- Mean age at baseline visit  $73.1 \pm 7.4$ , 99 females.
- Each individual has at least 1 brain scan and at most 6 scans.
- The data include 46 people with AD (111 AD scans), 80 individuals with EMCI (247 MCI scans), 40 people with LMCI (120 LMCI scans) and 61 healthy participants (160 scans).
- 227 475 possible pairs of subjects (764 of which were labeled as 0).

More details on data: [adni.loni.usc.edu](http://adni.loni.usc.edu)

# Datasets: PPMI

- 226 individuals from Parkinson's Progression Markers Initiative (PPMI).
- 159 subjects from PD cohort and 67 healthy controls.
- Mean age at the baseline visit  $61.0 \pm 9.8$  years, 79 were females.
- Each individual has at least 1 brain scan and at most 4 scans.
- 152031 pairs from PPMI data (301 of which were labeled as 1)

More details on data: [ppmi-info.org](http://ppmi-info.org)

# Network construction: ADNI and PPMI

## T1w Processing

Cortical reconstruction generated via  
FreeSurfer



Skull-stripped T1w, FreeSurfer parcellations,  
aligned to MNI152 space

## DWI Processing

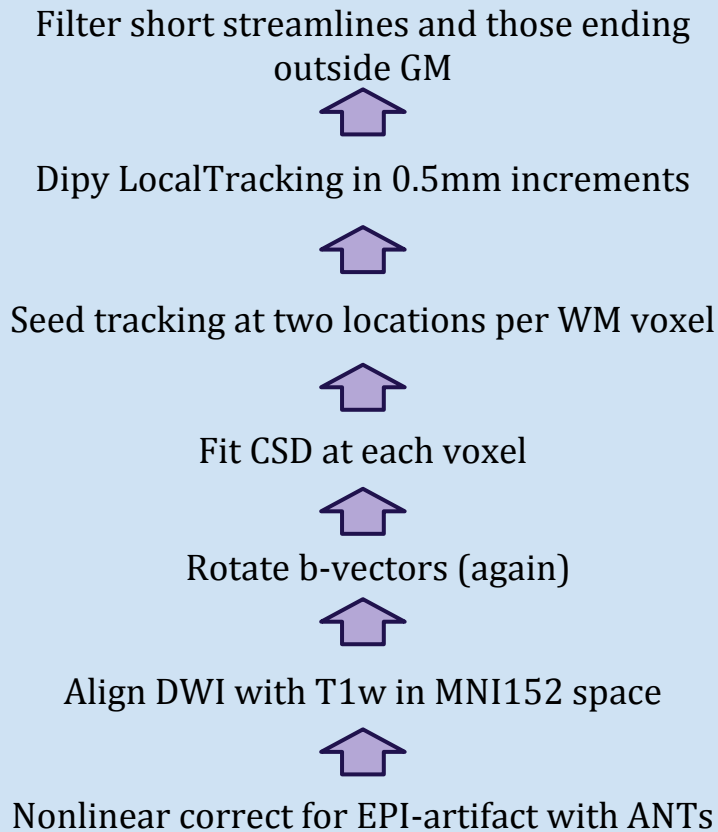
Denoise with NLSAM (PPMI-only)



Correct for motion and eddy-current distortion



Rotate b-vectors

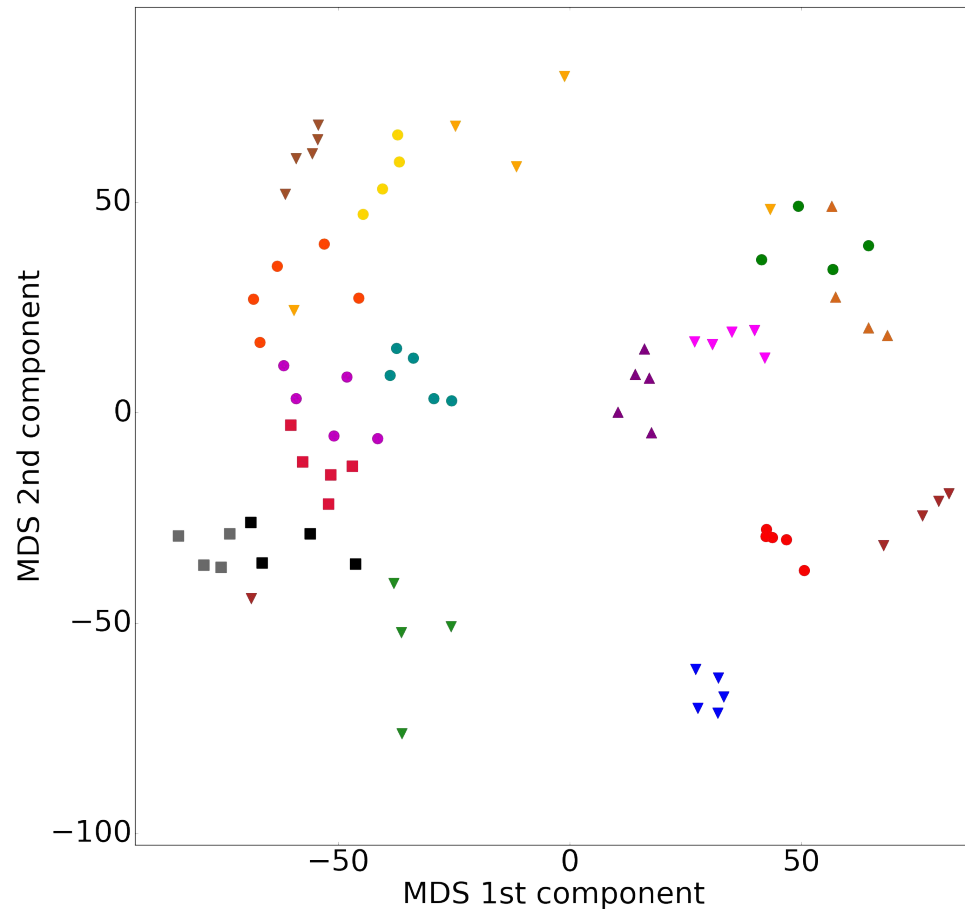


# Connectome normalizations and features

- **Four normalizations:** by mean, by max, binary and no normalization
- **Weighted nodes' degrees** or strength.
- **Closeness centrality.** The higher it is, the more close node to others.
- **Betweenness centrality.** Represents the degree of which nodes stand between each other.
- **Eigenvector centrality.** Measure of influence of a node in a network.
- **Local efficiency.** Measure of how efficiently node exchanges information.
- **Clustering coefficient.** Measures degree to which nodes in a network tend to cluster together
- **Weighted number of triangles** around node.
- **PageRank** (Brin and Page, 1998). Another estimate of node's importance related to random walks on a network.

(Overview of all metrics except PageRank can be found in Rubinov and Sporns, 2009)

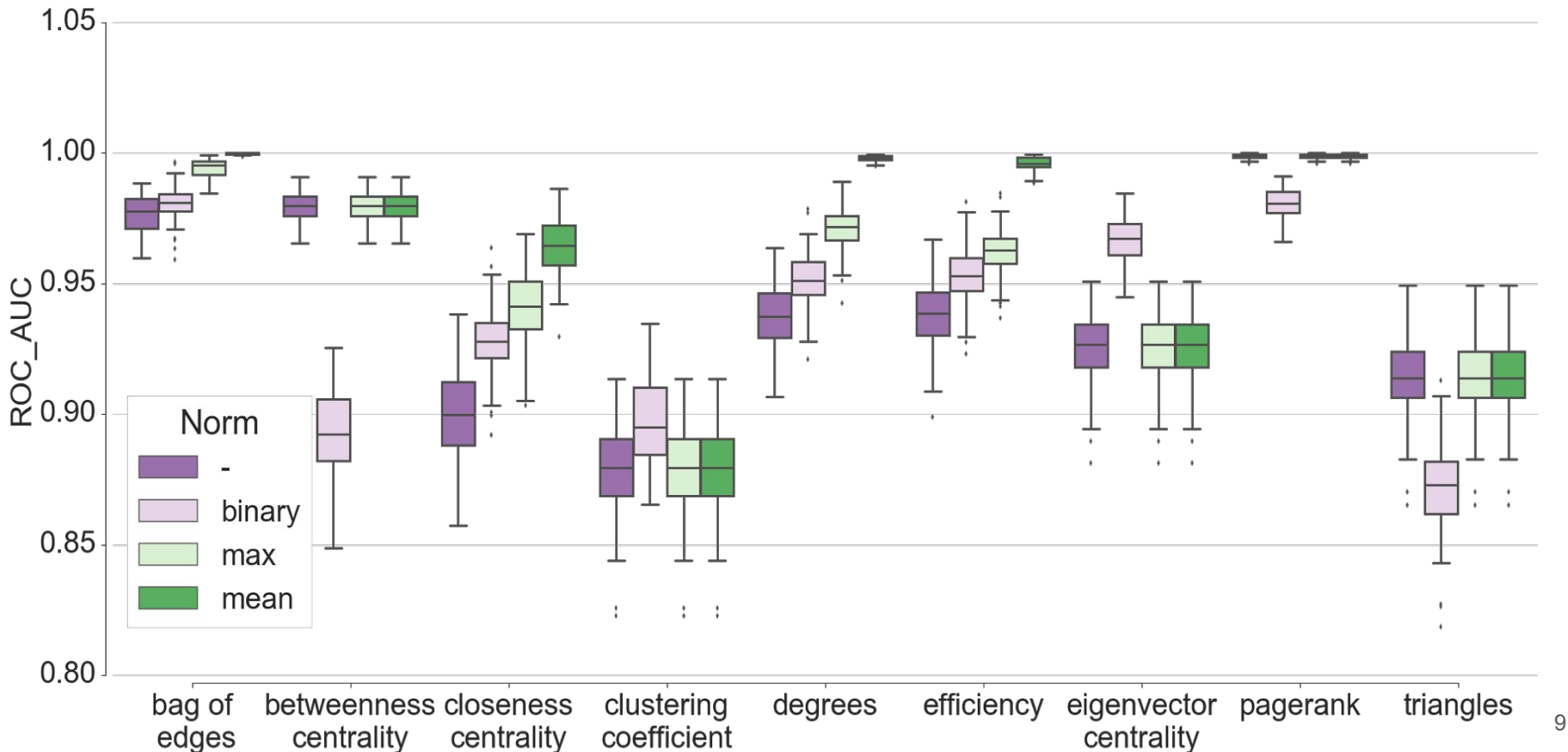
# Results: ADNI MDS on bag of edges



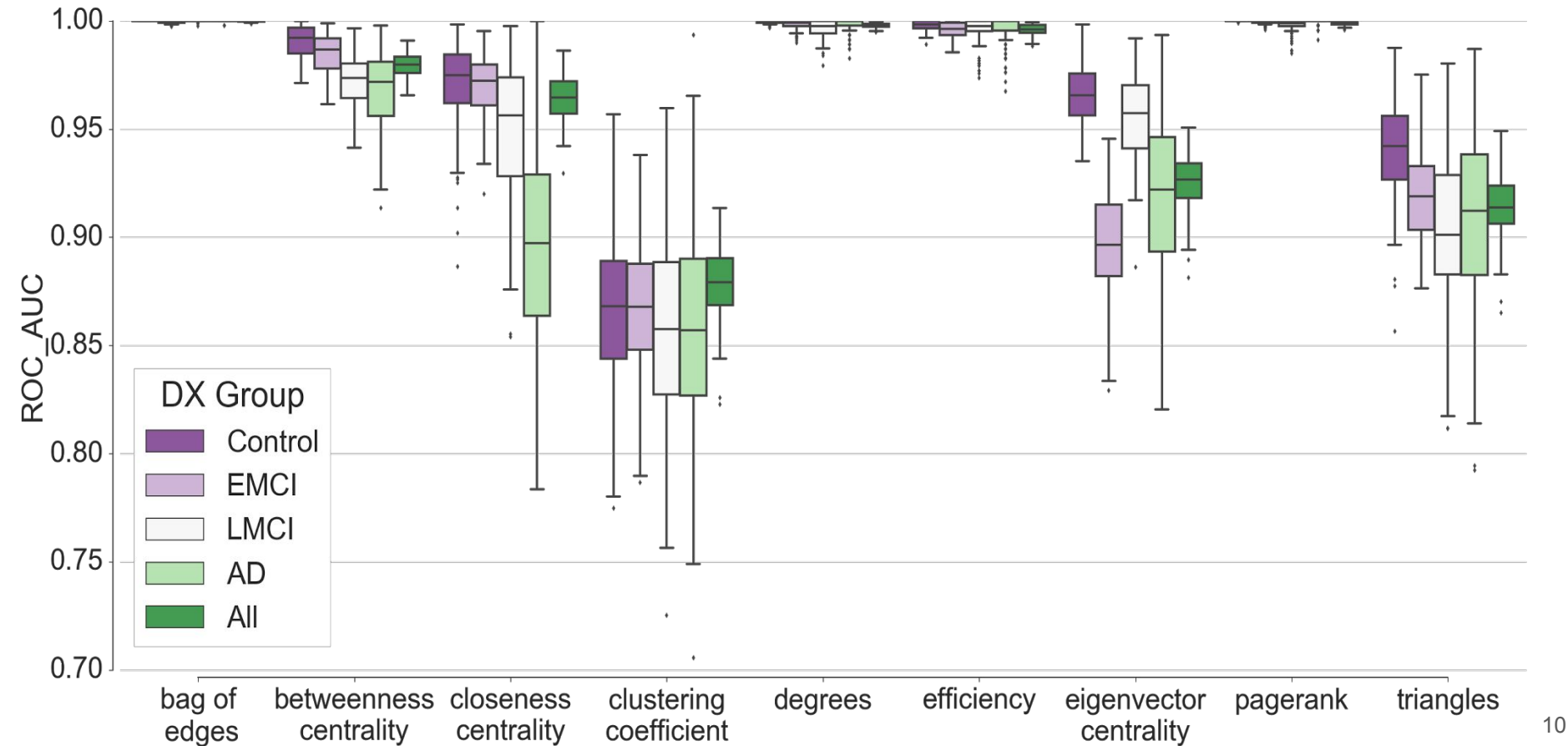
- Multidimensional scaling based on  $l_2$ -distance between bag of edges for ADNI subjects
- Only 17 subjects shown for purpose of visualisation
- Different colors represent different subjects
- Shapes of markers represent diagnostic group:  $\square$  — AD,  $\blacktriangle$  — MCI,  $\blacktriangledown$  — EMCI,  $\square$  — HC



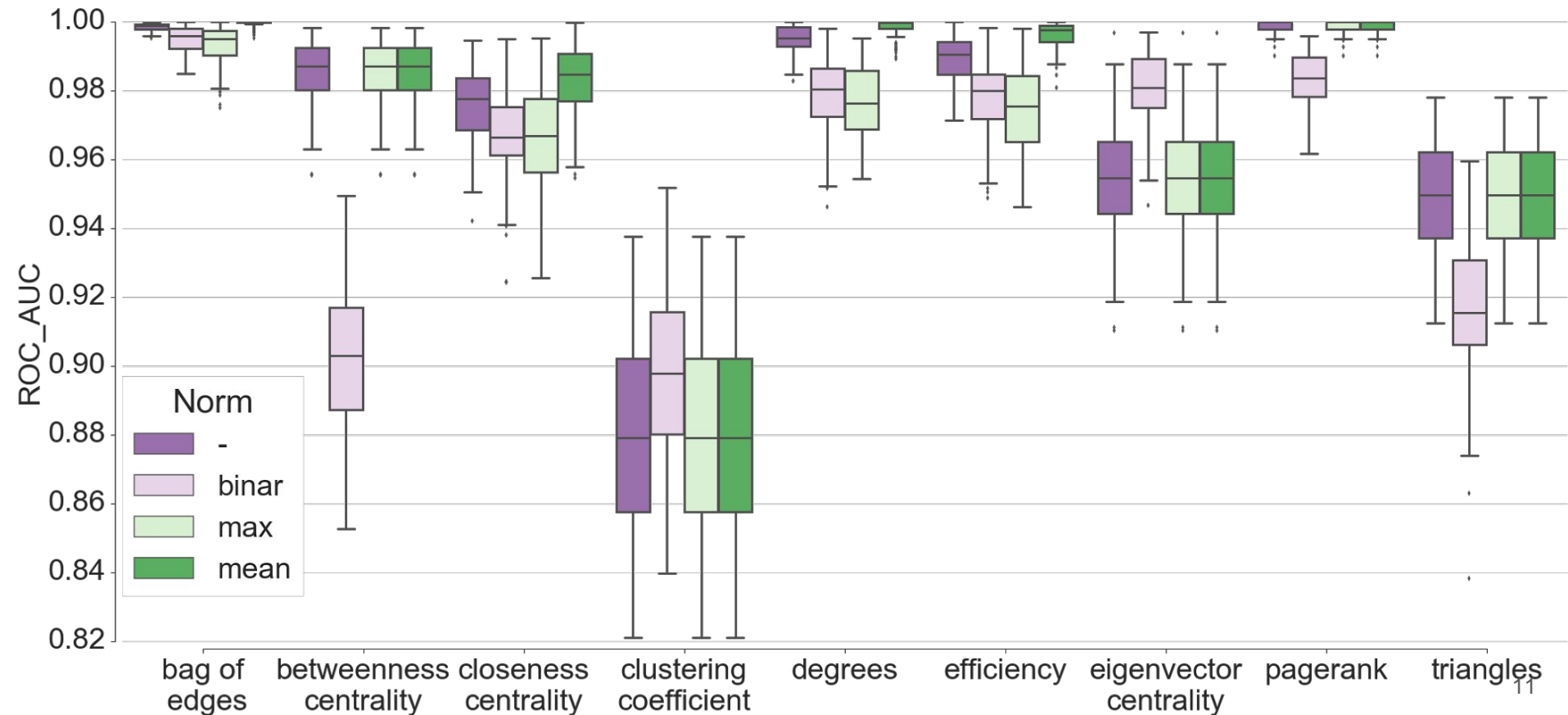
# Results: ADNI by norms and features



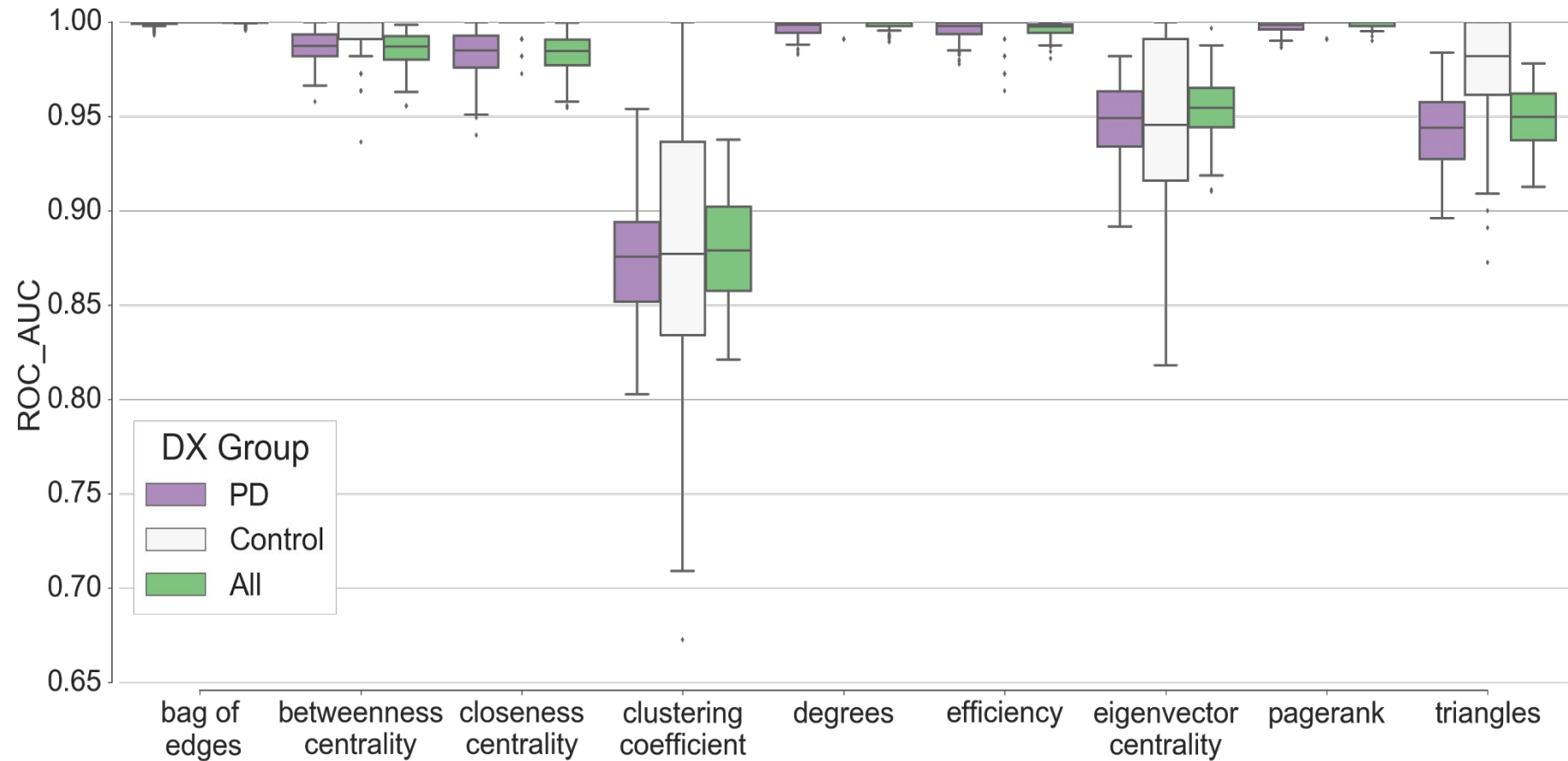
# Results: ADNI by DX groups and features



# Results: PPMI by norms and features



# Results: PPMI by DX group and features



# Conclusion

- Pairwise classification may be useful for comparison of preprocessing pipelines and particular features
- Results consistent across two datasets and inside diagnostic groups
- Images for both ADNI and PPMI data were obtained with time difference at least a year. We can't be sure that subject connectomes didn't change due to some reasons (i.e. disease progression)
- Pairwise classification is not a feature selection technique for classification tasks in a common sense.

# Related work

— Finn, Emily S., et al. "Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity." *Nature neuroscience* (2015). Similar work for fMRI data. Authors successfully identified subjects based on fMRI connectivity. Results were consistent across scan sessions and even between task and rest conditions.

— Yeh, Fang-Cheng, et al. "Quantifying Differences and Similarities in Whole-brain White Matter Architecture Using Local Connectome Fingerprints." *PLOS Computational Biology* 12.11 (2016). Authors used a local structural connectome, different features, datasets and connectome construction pipelines and arrived at similar conclusions.

# Possible extensions

- Evaluation of different connectome building pipelines using both pairwise classification and Intraclass Correlation Coefficient on test-retest data (done, under review).
- Investigation of how pairwise classification changes with each pre-processing and connectome building step.
- Time difference regression on connectome pairs.
- Structural/functional connectome pairs classification.

# Links and acknowledgements

Source code:

[github.com/neuro-ml/structural-connectome-validation-pairwise](https://github.com/neuro-ml/structural-connectome-validation-pairwise)

Library for creating and curating reproducible pipelines:

[github.com/neuro-ml/reskit](https://github.com/neuro-ml/reskit)

Funding: NIH U54 EB020403 (ENIGMA CENTER FOR WORLDWIDE MEDICINE, IMAGING & GENOMICS), PI Paul M. Thompson; Russian Science Foundation (project 14-50-00150)



# Thank you!

[to.dmitry.petrov@gmail.com](mailto:to.dmitry.petrov@gmail.com)

