# JointNets: an End-to-end R package for sparse Gaussian graphical model

Zhaoyang Wang, Beilun Wang, Arshdeep Sekhon, Yanjun Qi

2019/5/3

# Motivation

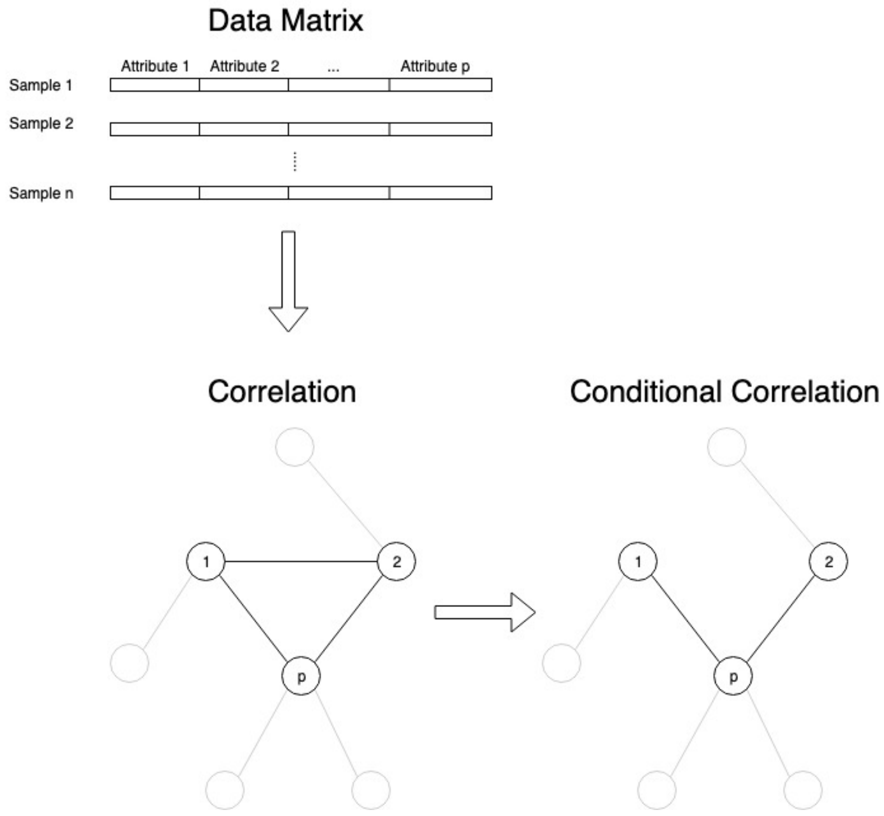
- Advances in sparse Gaussian graphical model learning
  - Joint estimation, difference estimation and etc
- Heterogeneous existing source codes and packages
  - Different forms of inputs, outputs and etc
- Existing tools focus solely on estimation and visualization
  - Lack of comprehensiveness

# Background

- Gaussian graphical model
  - Nodes follow multivariate normal distribution
  - Undirected graphical model
  - Precision matrix -> graph
- Samples to graph
  - Number of observations n << number of dimensions p
  - Estimate sparse precision matrix

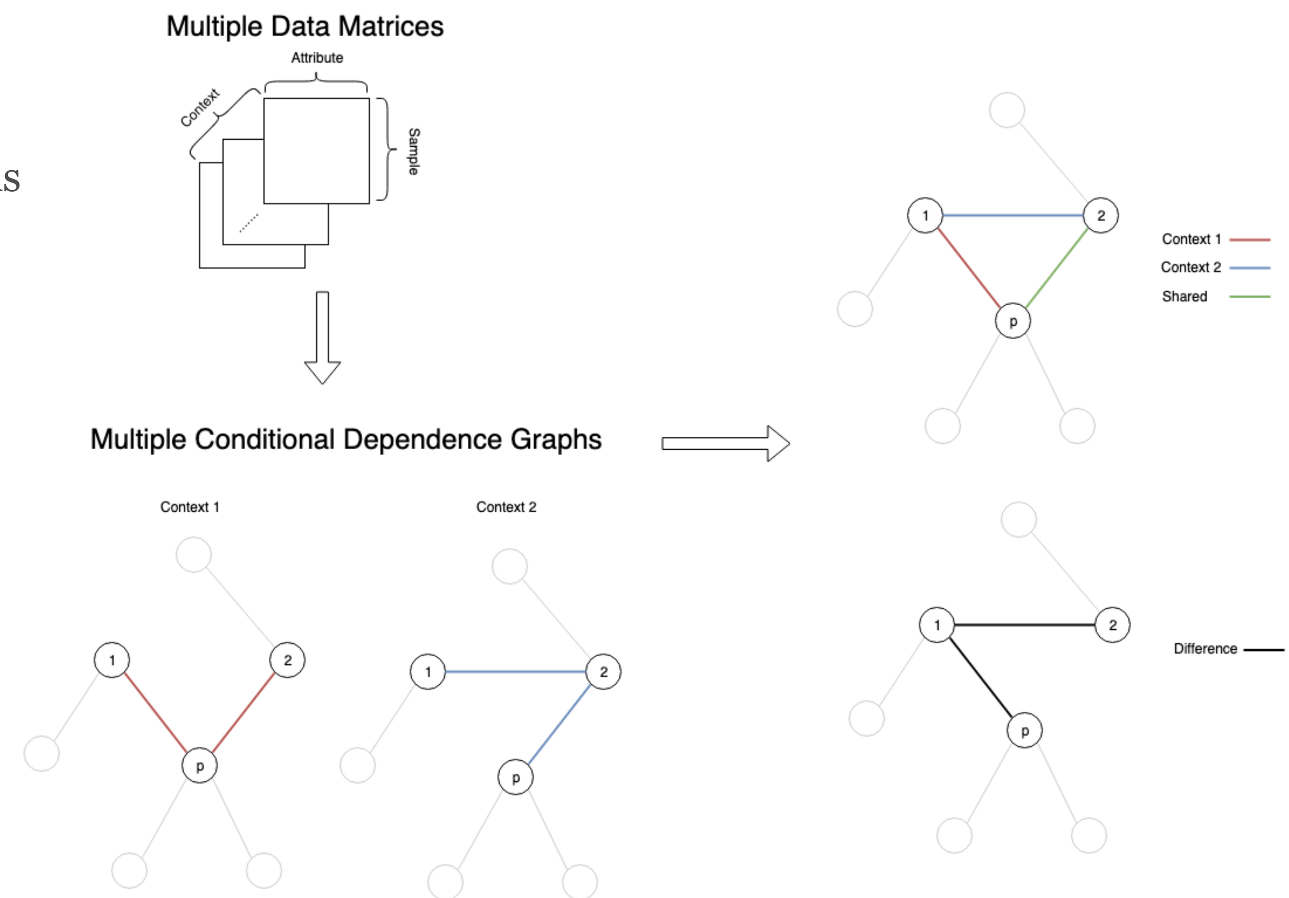
# Tasks

- Learning graph from samples
  - Data matrix => correlation matrix => conditional correlation matrix
  - Gaussian assumption:
    - Conditional correlation => conditional dependence



# Tasks

- Joint Learning
  - Multiple graphs
  - Difference



# Related Work

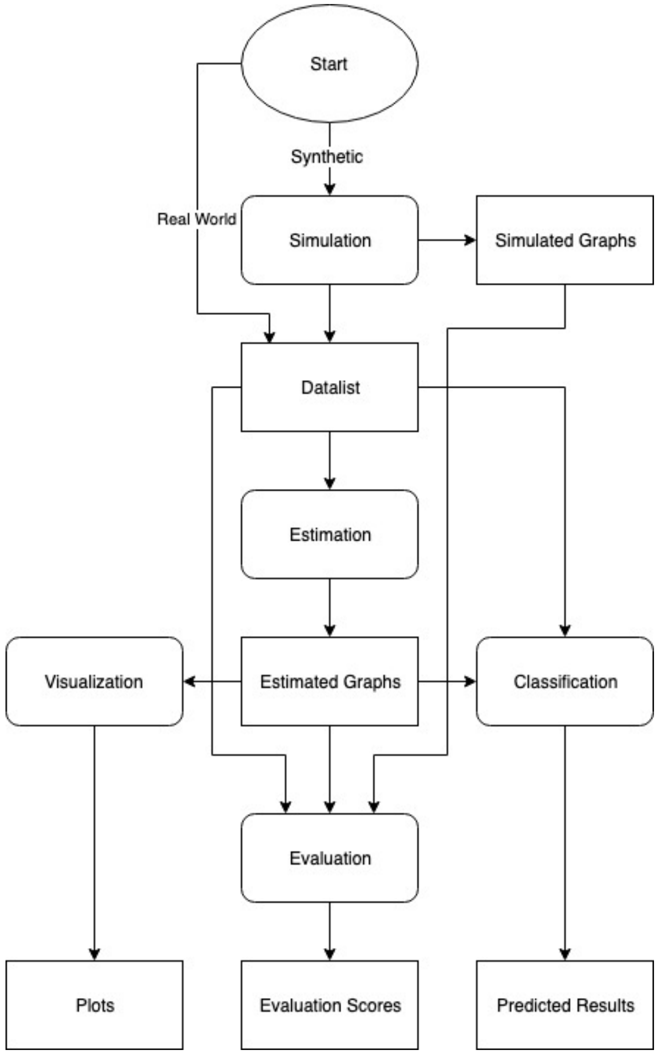
- Baselines
  - Joint Graphical Lasso (JGL)
  - Constrained l1 minimization (CLIME)
  - Elementary estimator (EE)
- Our algorithms
  - FASJEM
  - JEEK
  - (W)SIMULE
  - DIFFEE(K)

# Goal

- An end-to-end and unified framework for sparse Gaussian graphical model learning

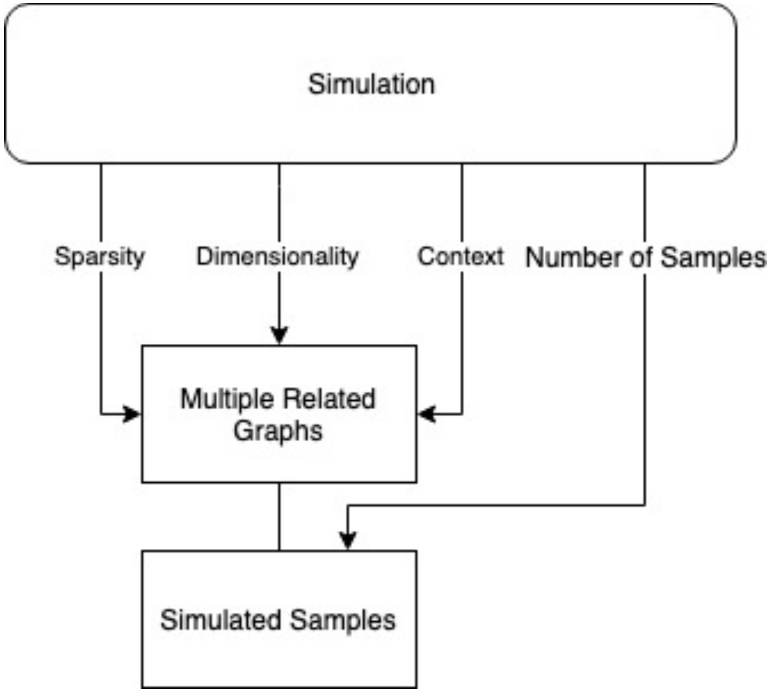
# Proposed Solution

- Connected Modules
  - Simulation
  - Estimation
  - Visualization
  - Evaluation
  - Classification



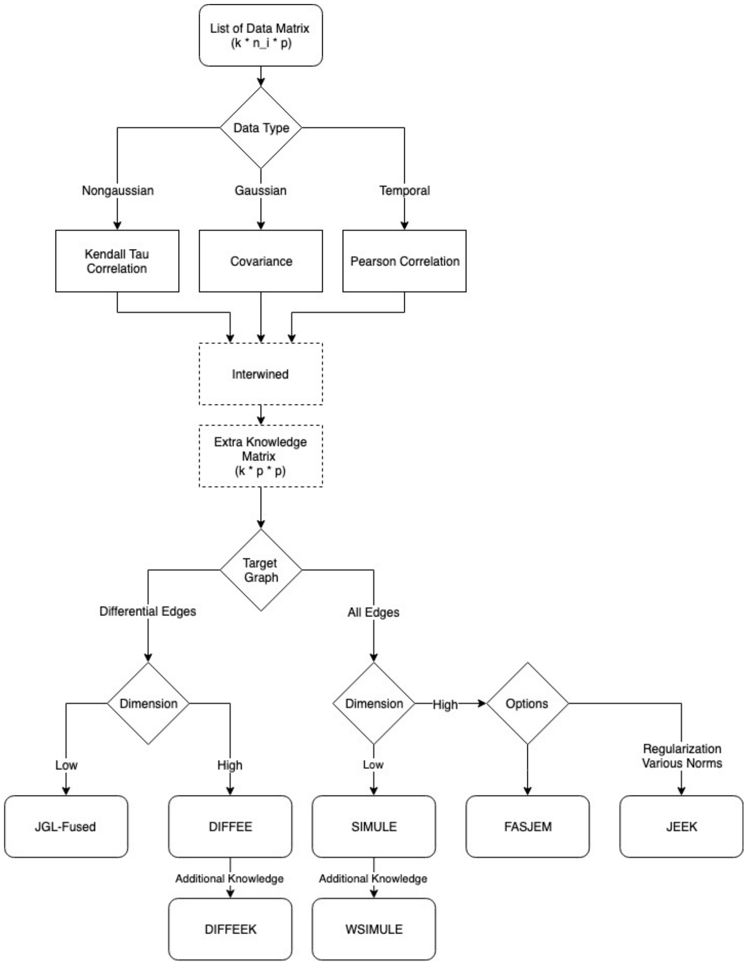
# Simulation

- simulation()
  - Sparsity
    - Shared
    - Individual
  - Dimensions
  - Contexts
  - Samples



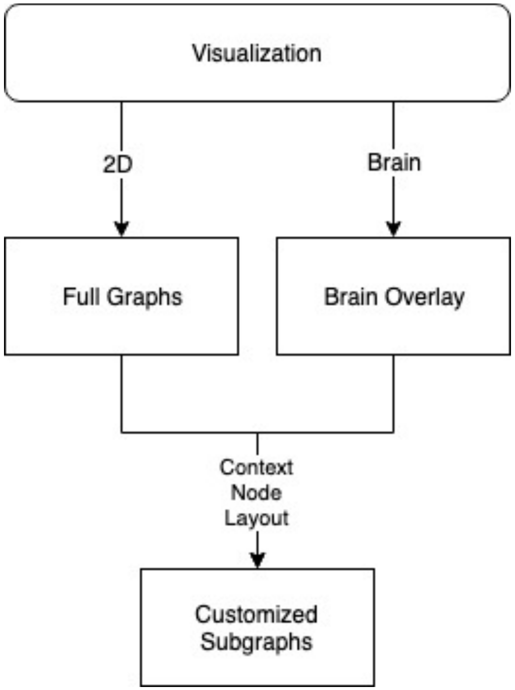
# Estimation

- method_name()
  - Correlation options
  - Intertwined
  - Prior knowledge
  - Joint Estimation
  - Difference Estimation



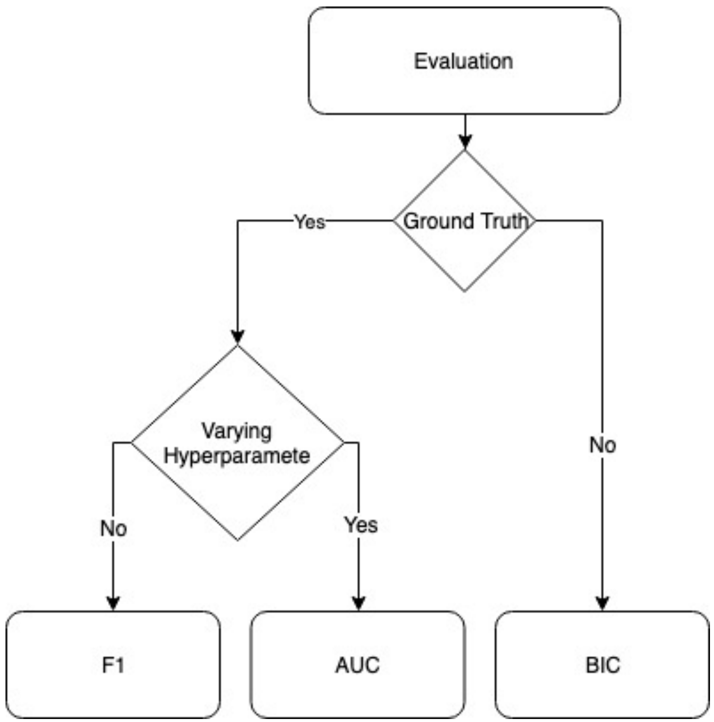
# Visualization

- plot()
  - Subgraph
  - Title and legend
  - Layout
- plot_brain()
  - 3D brain overlay



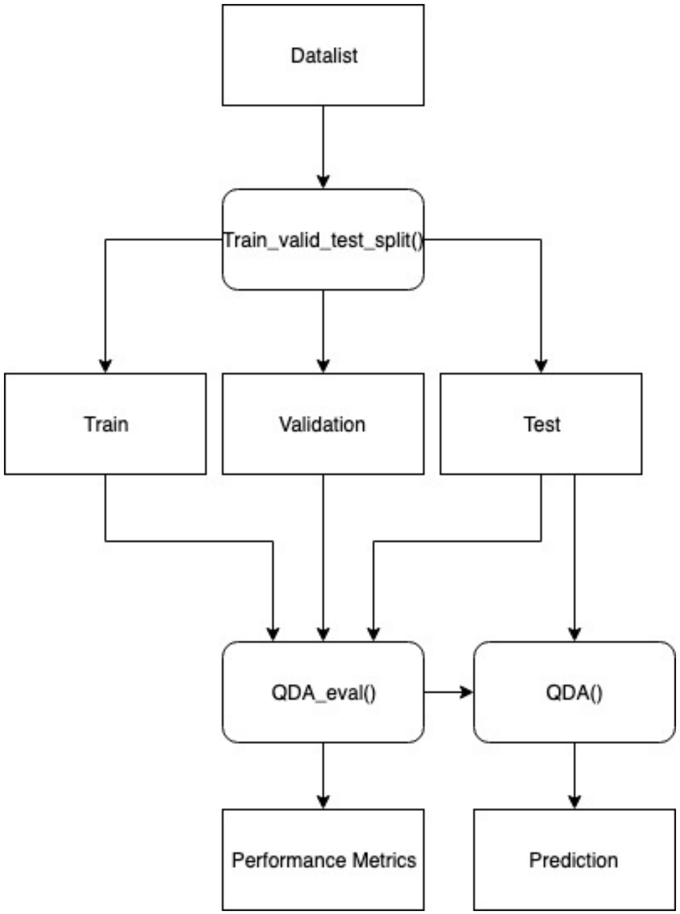
# Evaluation

- F1()
- AUC()
- BIC()



# Classification

- QDA_eval()
- QDA()

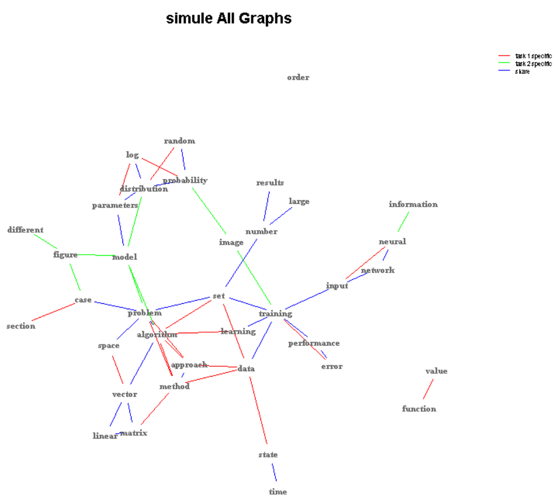


# Demos and Data Summary

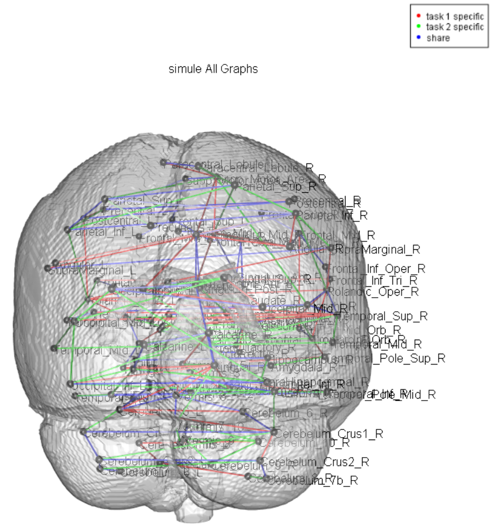
- Simulated data p = 20, 100
- cancer genomics p = 50
    - High dimensional microarray p >1000
- Natural language text data p = 37
- Brain imaging data (ABIDE) p = 116

# Visualization Results

- 2D

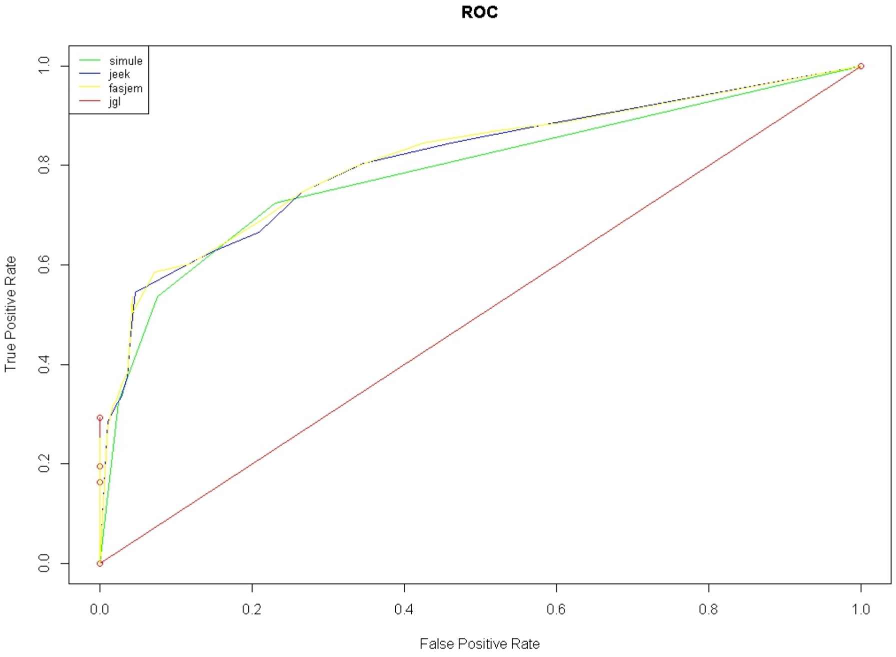


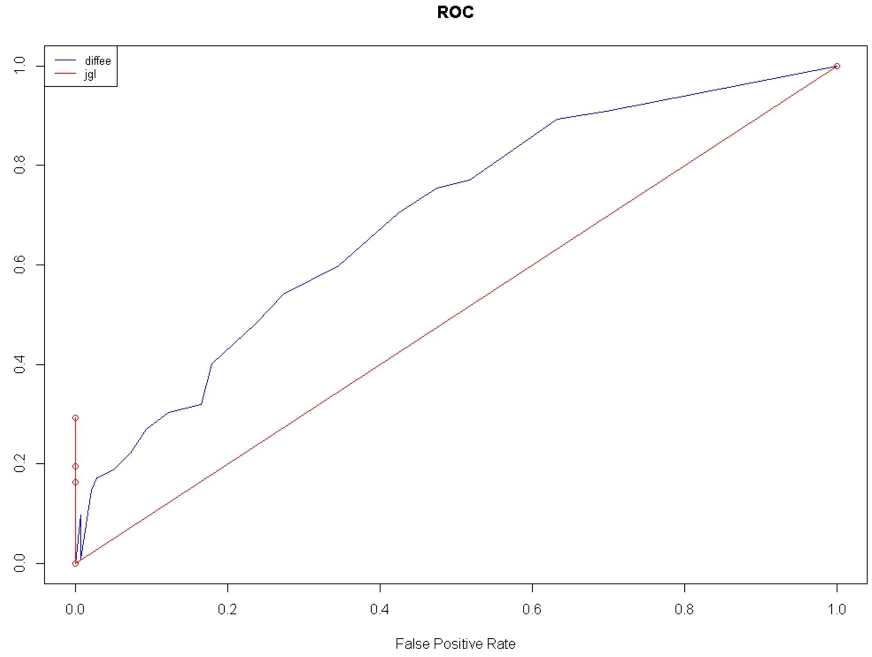

- 3D(brain)



# Experimental Results

- ROC
  - Synthetic dataset, p = 20, k = 2



# Experimental Results

- Classification Accuracy
  - ABIDE
  - NIPS conference paper word count

ABIDE prediction accuracy (from publications, p = 160, binary)

| Method | DIFFEE | WSIMULE | JGL |
|---|---|---|---|
| Accuracy | 57.58% | 58.62% | 56.90% |

Text data prediction accuracy (p = 37, binary)

| Method | DIFFEE | SIMULE | JEEK | FASJEM | JGL |
|---|---|---|---|---|---|
| Accuracy | 74.50% | 76.02% | 66.8% | 52.00% | 50.7% |

# Demo

- R markdown

# Conclusion and Future Work

- JointNets
  - End-to-end package
  - Closing gaps between fields by showcasing various applications
  - Standard and scalable interface / API
- Future works
  - Sparse QDA
  - Full GUI

# References


- Beilun Wang, Arshdeep Sekhon: "A Fast and Scalable Joint Estimator for Integrating Additional Knowledge in Learning Multiple Related Sparse Gaussian Graphical Models", 2018, International Conference on Machine Learning. 2018; [http://arxiv.org/abs/1806.00548 arXiv:1806.00548].
- Beilun Wang, Arshdeep Sekhon: "Fast and Scalable Learning of Sparse Changes in High-Dimensional Gaussian Graphical Model Structure", 2017; [http://arxiv.org/abs/1710.11223 arXiv:1710.11223].
- Beilun Wang, Ji Gao: "A Fast and Scalable Joint Estimator for Learning Multiple Related Sparse Gaussian Graphical Models", 2017; [http://arxiv.org/abs/1702.02715 arXiv:1702.02715].
- Beilun Wang, Ritambhara Singh: "A constrained L1 minimization approach for estimating multiple Sparse Gaussian or Nonparanormal Graphical Models", 2016; [http://arxiv.org/abs/1605.03468 arXiv:1605.03468].
- Chandan Singh, Beilun Wang: "A Constrained, Weighted-L1 Minimization Approach for Joint Discovery of Heterogeneous Neural Connectivity Graphs", 2017; [http://arxiv.org/abs/1709.04090 arXiv:1709.04090].
- Patrick Danaher, Pei Wang: "The joint graphical lasso for inverse covariance estimation across multiple classes", 2011; [http://arxiv.org/abs/1111.0324 arXiv:1111.0324].