
A brief description:

This is a graduate course. In the previous years, the enrollees are PhD and master students from Statistics, Biostatistics, Economics, and Data Science. There will be no homework and no exams. The only requirement is a group project. Please see Slides 39-41 of last year's Introduction Slide ([Lec1-Intro.pdf](#)). The course project requirement from last year can be found here ([Projects.pdf](#)).

The list of topics covered last year:

- Chapter 1: Introduction
- Chapter 2: Multiple Testing (normal means models, Rare/Weak signals and phase transitions, global detection, signal recovery, the Higher Criticism tests)
- Chapter 3: Community Detection (common network data sets and models, SBM, spectral clustering, modularity optimization, DCBM and degree heterogeneity, the SCORE method, semi-supervised community detection).
- Chapter 4: Topic Modeling (anchor word condition, connection to nonnegative matrix factorization, simplex geometry, latent Dirichlet allocation, anchor-word recovery algorithms, the Topic-SCORE method)
- Chapter 5: Variable Selection (sparse linear models, penalization methods, estimation errors and model selection consistency of Lasso, non-convex penalties, Screen and Clean methods, phase transitions for Hamming errors, forward-backward selection, LARS).
- Chapter 6: Other Topics in Network Analysis (link prediction, networks in epidemiology, hypergraph networks and tensor-decompositions).
- Chapter 7: More Text Models (hierarchical/dynamic/seeded/correlated topic models, multinomial testing for text data, topic ranking models, bi-gram models, pre-trained large language models such as transformers).
- Chapter 8: Clustering and PCA (pitfall of classical PCA in high dimension, sparse PCA, spectral clustering and feature selection)
- Chapter 9: Selective Topics in Hypothesis Testing.