Outline for LLG

April 19, 2016

1 Introduction

- Estimating the mean of a collection of graphs is becoming important but the general element-wise MLE doesn't take advantage of the graph structure.
- SBM is a good model which captures the graph structure in reality.
- SBM as a RDPG is what we are going to consider.
- Under the model, we propose a new estimator based on ASE and prove theoretically that it is better than element-wise MLE.

2 Model

- Goal is to estimate the mean of a collection of unweighted simple graphs by observing their adjacency matrices with known vertex correspondence.
- Introduce problem setting, i.e. how M graphs are generated in the model.

2.1 Entry-Wise Least Squares Estimate

- MLE is the right thing to do under Independent Edge Graph Model, i.e. without taking graph structure into account.

2.2 Random Dot Product Graph

- Introduce Latent Positions Graph Model.
- RDPG is a special case of Latent Positions Graph Model.

2.3 Stochastic Block Model as a Random Dot Product Graph

- Intuition behind SBM.
- Formal definition of SBM.
- Consider SBM as RDPG in this paper.

2.4 Estimator $\hat{P}$ Based on Adjacency Spectral Embedding

- Introduce the new estimator based on ASE.
- Intuition why this should be better.

2.5 Performance Evaluation: Relative Efficiency

- Definition of RE.

3 Results

3.1 Theoretical Results

- $\mathrm{RE}_{ij} = \frac{\mathrm{MSE}(\hat{P}_{ij})}{\mathrm{MSE}(\bar{A}_{ij})} \approx \frac{1/\rho_i + 1/\rho_j}{N}$
- $\mathrm{MSE}(\hat{P}_{ij}) \approx \frac{(1/\rho_i + 1/\rho_j) P_{ij}(1-P_{ij})}{NM}$
-

3.2 Validation with Simulations

- Consider a specific SBM.
- Figure shows that simulated RE agrees with theoretical values when changing N.
- Figure shows that simulated MSE agrees with theoretical values when changing ρ.

3.3 CoRR Brain Graphs: Cross-Validation

- Consider three datasets, JHU, desikan and CPAC200.
- Explain how we sample from the whole dataset.
- Figure shows the performance of $\bar{A}$ and $\hat{P}$ when embedding into different dimension.
- $\hat{P}$ outperforms $\bar{A}$ when M is small.
- ZG and USVT are both doing good jobs in dimension selection, while the result is not very sensitive to the dimension selection.
- Simulation with P being the mean graph of the real data shows that $\hat{P}$ still does a good job when the low rank assumption is violated.

4 Discussion

- Scheinerman's method is better than row sum divided by $N-1$ in diagonal augmentation.

5 Methods

5.1 Algorithm

- Detailed description of the algorithm for our estimator $\hat{P}$.

5.2 Choosing Dimension

- ZG.
- USVT.

5.3 Graph Diagonal Augmentation

- Reason for diagonal augmentation.
- Introduce the iterative method developed by Scheinerman and Tucker.

5.4 Source Code

5.5 Dataset Description

Give detailed description of the data we are using.

5.6 Outline for the Proof of the Theorems