Outline for LLG

May 1, 2016

1 Introduction

- The mean of a collection of graphs can be defined in various ways. One natural definition is the proportion of the existence of an edge between any pair of vertices. Estimating the mean of a population of graphs based on a sample is becoming more and more important both in statistical inference and in various applications like connectomics, social networks, etc.
- Element-wise maximum likelihood estimate is a reasonable estimator if we only consider the independent edge graph model without taking any graph structure into account. However, it does not perform very well especially when we have a few observations, which is likely the case in real world.
- (Challenges) Generally, we don't have any information about the structure of the graphs. So it is hard to take advantage of the unknown graph structure.
- One of the most important structures is the community structure in which vertices are clustered into different communities such that vertices of the same community behave similarly. The stochastic blockmodel (SBM) captures such structural property and is widely used in modeling networks.
- Meanwhile, the latent positions model (LPM), a much more general model compared to SBM, proposes a way to parameterize the graph structure by latent positions associated with each vertex. However, the random dot product graph (RDPG) which is a special case of LPM stays in between and motivates our estimator. In this paper, we analyze our estimator in terms of RDPG specifically.
- Using the estimates of the latent positions in an RDPG setting based on a truncated eigen-decomposition of the adjacency matrix, we consider a new estimator for the mean of the collection of graphs which captures the low-rank structure. And we prove via both theory and simulations/real data analysis that it is better than element-wise MLE.
- (Future work) Robust estimation, dimension selection, diagonal augmentation, etc.

2 Models and Estimators

2.1 Independent Edge Model

- Under the independent edge model (IEM), each edge A_{ij} independently follows the Bernoulli distribution with parameter P_{ij}.

2.2 Estimator $\bar{A}$

- Under the IEM, element-wise mean of the adjacency matrices $\bar{A}$ is the MLE as well as the least squared estimate.
- $\bar{A}$ is unbiased for IEM and has entry-wise variance $\text{Var}(\bar{A}_{ij}) = P_{ij}(1 - P_{ij})/M$. And it is the UMVUE under IEG with no constraints. But, it doesn't exploit any structure.

2.3 Random Dot Product Graph

- Hoff et. al. (2002) proposed a model for random graphs called Latent Positions Graph Model.
- A specific instance of this model that we will examine is the random dot product graph model (RDPG) in which the link function is the dot product, i.e. the probability of an edge being present between two nodes is the dot product of their latent vectors.

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

- In order to take advantage of the underlying low rank structure of the RDPG, we use the adjacency spectral embedding (ASE) studied by Sussman et. al. to enforce a low rank approximation on the entry-wise mean matrix $\bar{A}$, which will decrease the variance without losing much in bias if we embed it into the right dimension.
- Detailed description of the algorithm for our estimator $\hat{P}$.

2.5 Stochastic Block Model as a Random Dot Product Graph

- One of the most important structures is the community structure in which vertices are clustered into different communities such that vertices of the same community behave similarly. Such structural property is captured by the SBM, where each vertex is assigned to a block and the probability that an edge exists between two vertices depends only on their respective block memberships.
- Formally, the SBM is determined by the number of blocks K (generally way less than the number of vertices N), block proportion vector ρ, and block probability matrix B.
- Now if we consider the SBM as a random dot product graph, all vertices in the same block would have identical latent positions.

3 Results

3.1 Theoretical Results

- To compare the performance between $\hat{P}$ and $\bar{A}$, we examine the relative efficiency (RE), in mean squared error (MSE), among the two defined as: $RE_{ij} = \frac{MSE(\hat{P}_{ij})}{MSE(\bar{A}_{ij})}$.

- **Theorem 3.1** *For any i and j, conditioning on $X_i = \nu_{\tau_i}$ and $X_j = \nu_{\tau_j}$, we have*

$$\mathrm{ARE}(\bar{A}_{ij}, \hat{P}_{ij}) = 0.$$

And for N large enough, conditioning on $X_i = \nu_{\tau_i}$ and $X_j = \nu_{\tau_j}$, we have

$$\mathrm{RE}(\bar{A}_{ij}, \hat{P}_{ij}) \approx \frac{1/\rho_{\tau_i} + 1/\rho_{\tau_j}}{N}.$$

- **Lemma 3.2** *In the same setting as above, for any i, j, conditioning on $X_i = \nu_{\tau_i}$ and $X_j = \nu_{\tau_j}$, we have*

$$\lim_{n\to\infty} N \cdot \mathrm{Var}(\hat{P}_{ij}) = \frac{1/\rho_{\tau_i} + 1/\rho_{\tau_j}}{M} P_{ij}(1 - P_{ij}).$$

And for N large enough, conditioning on $X_i = \nu_{\tau_i}$ and $X_j = \nu_{\tau_j}$, we have

$$E[(\hat{P}_{ij} - P_{ij})^2] \approx \frac{1/\rho_{\tau_i} + 1/\rho_{\tau_j}}{MN} P_{ij}(1 - P_{ij}).$$

3.2 Validation with Simulations

- We demonstrate the theoretical results in Section 3.1, the variance of $\hat{P}$ and the relative efficiency, via various Monte Carlo simulation experiments. Specifically, we consider the 2-block SBM parameterized by

$$B = \begin{bmatrix} 0.42 & 0.2 \\ 0.2 & 0.7 \end{bmatrix}, \qquad \rho = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}.$$

- For each Monte Carlo replicate, we generate M random graphs with known vertex correspondence under the SBM described above. And details.

- Given M graphs, we can calculate $\bar{A}$ and $\hat{P}$ assuming that $d = \mathrm{rank}(B) = 2$ is known.

- By checking the averaging MSE and RE of the two estimates $\hat{P}$ and $\bar{A}$ over 1000 Monte Carlo replicates, we demonstrate that the theoretical results in Section 3.1.

- Figure 1 plots the scaled average RE with different N and fixed M of 1000 Monte Carlo replicates. Blue lines denote the simulated scaled RE associated with the edges we are averaging over. Solid line in black represents the theoretical value for scaled RE. From the figure, we see that $N \cdot \mathrm{RE}_{st}(\bar{A}, \hat{P})$ converges to $1/\rho_s + 1/\rho_t$ represented as the black solid line, as suggested in Lemma 3.1. Notice that this means $\mathrm{RE}_{st}(\bar{A}, \hat{P})$ is decreasing at rate $1/N$.

- To verify Theorem 3.1 and Lemma 3.2 hold with different ρ, Figure 2 shows the average MSE and average RE with $N = 500$ and $M = 100$ while changing ρ_1 from 0.1 to 0.9.

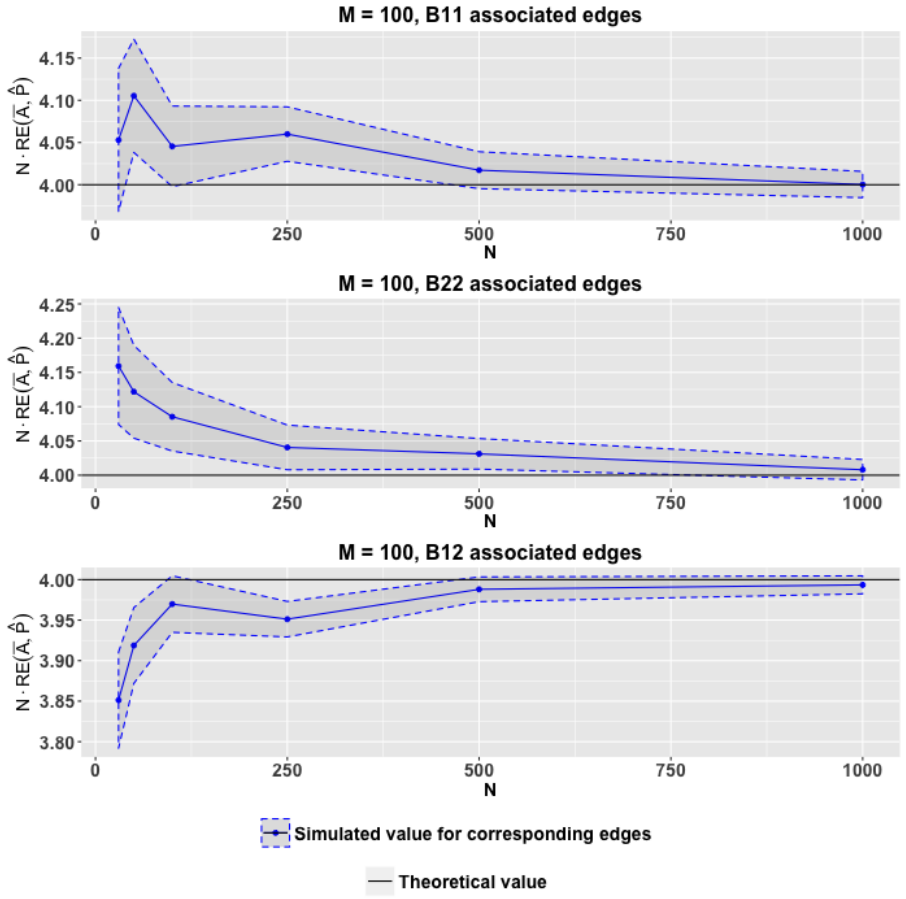

Figure 1: Scaled average RE with different N and fixed M of 1000 Monte Carlo replicates. Blue lines denote the simulated scaled RE associated with the edges we are averaging over. Dashed lines represent the 95% confidence interval. Solid line in black represents the theoretical value for scaled RE. Observe that $N \cdot \mathrm{RE}_{st}(\bar{A}, \hat{P})$ converges to $1/\rho_s + 1/\rho_t$ as expected.

3.3 CoRR Brain Graphs: Cross-Validation

- In practice, the graphs may not perfectly follow an RDPG, or even not IEM. But we are still interested in the mean graph. To demonstrate that the $\hat{P}$ estimate is still valid in such cases, we examine three datasets, JHU, desikan and CPAC200, which are sets of 454 brain connectomes with different number of nodes generated from fMRI scans available at the Consortium for Reliability and Reproducibility (CoRR).

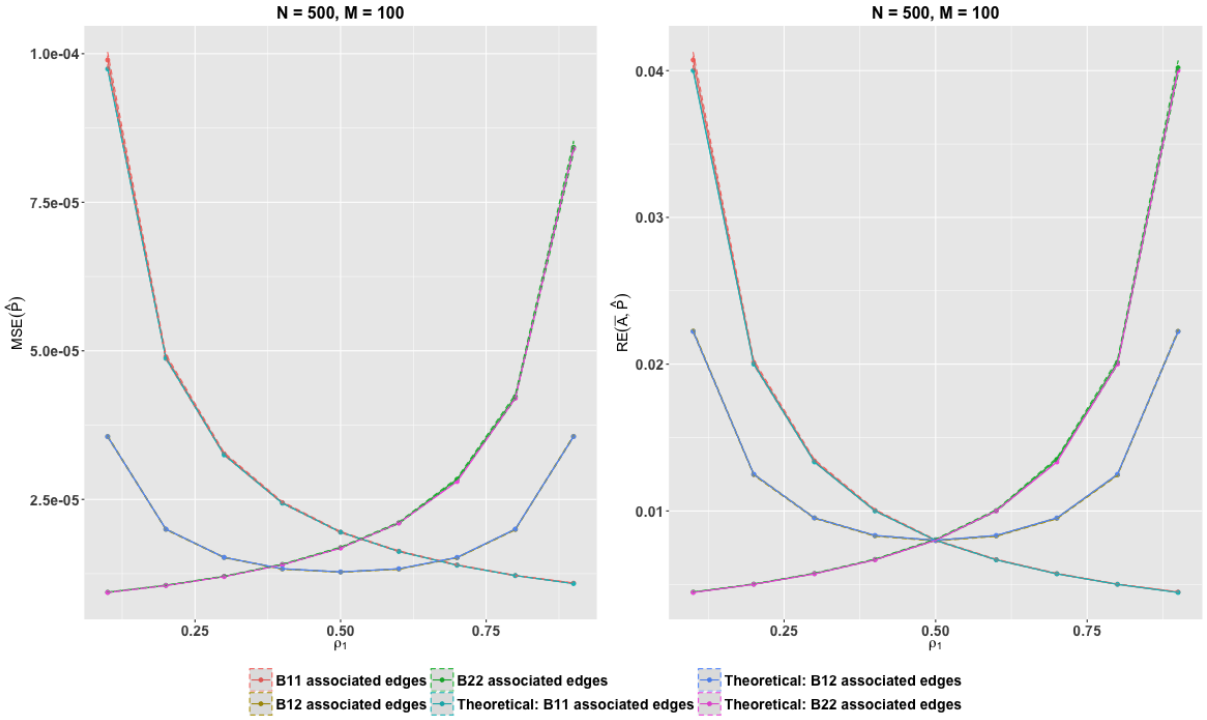

Figure 2: Simulated results for (a) $\text{MSE}_{st}(\hat{P})$ and (b) $\text{RE}_{st}(\bar{A}, \hat{P})$ with $N = 500$ and $M = 100$ of 1000 Monte Carlo replicates while changing ρ_1 from 0.1 to 0.9. Dashed lines represent the 95% confidence interval. The simulated values for the average MSE and average RE measurements match perfectly with the theoretical values.

- To compare $\bar{A}$ and $\hat{P}$ we perform a cross-validation study to examine the impact of the number of available graphs M.
- Figure 3, Figure 4 and Figure 5 demonstrate that our algorithm gives a better estimate $\hat{P}$ according to all three datasets.
- When M is small, $\bar{A}$ has large variance which leads to large MSE. Meanwhile, $\hat{P}$ reduces the variance by taking advantages of the graph structure and outperforms $\bar{A}$ dramatically.
- Moreover, Zhu and Ghodsi's algorithm and USVT algorithm both do a good job for selecting the dimension to embed.
- 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.

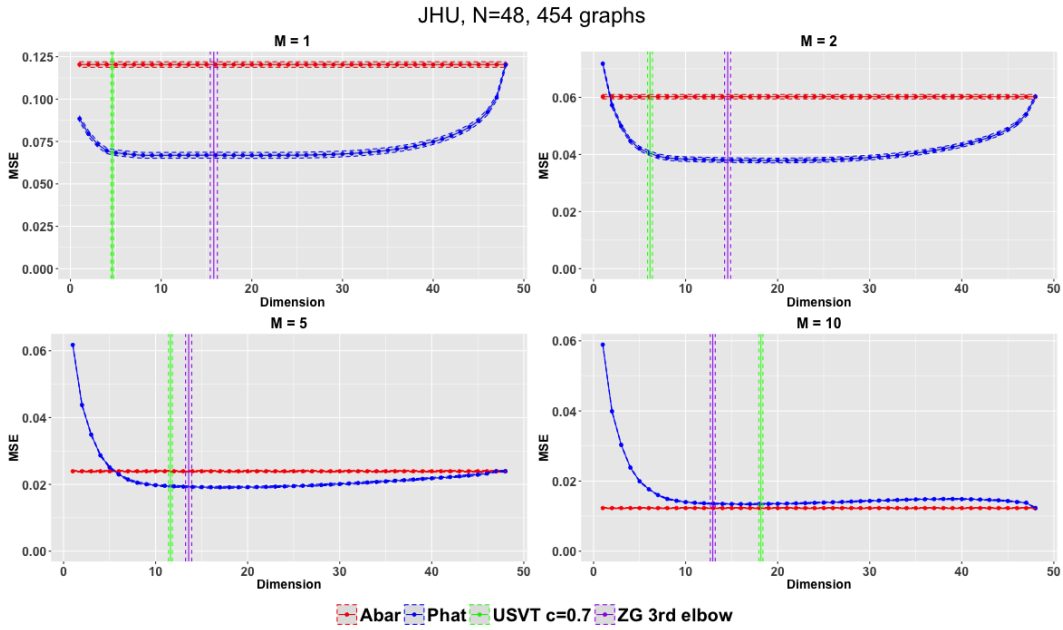

Figure 3: Comparison of MSE between $\bar{A}$ (red) and $\hat{P}$ (blue) for JHU dataset while embedding the graphs into different dimensions with different size M of the subsamples. The dimension chosen by the 3rd elbow of Zhu and Ghodsi is denoted in purple, and chosen by USVT with threshold equals 0.7 is denoted in green. Dashed lines represent the 95% confidence interval. When M is small, $\hat{P}$ outperforms $\bar{A}$ with a flexible range of the embedding dimension including what Zhu and Ghodsi selects.

4 Discussion

4.1 Summary

- In this paper, we propose a better way to estimate the mean of a collection of graphs by taking advantage of the low rank structure of the graphs.

4.2 Future Work

- Generally the observations we have are always contaminated in practice. In this case, improved robust estimator based on the low rank structure of the graphs is preferred.
- Estimating the rank of the graph structure accurately will certainly help improve the results.
- In this paper, we are using Scheinerman's method with 1 iteration for diagonal augmentation.

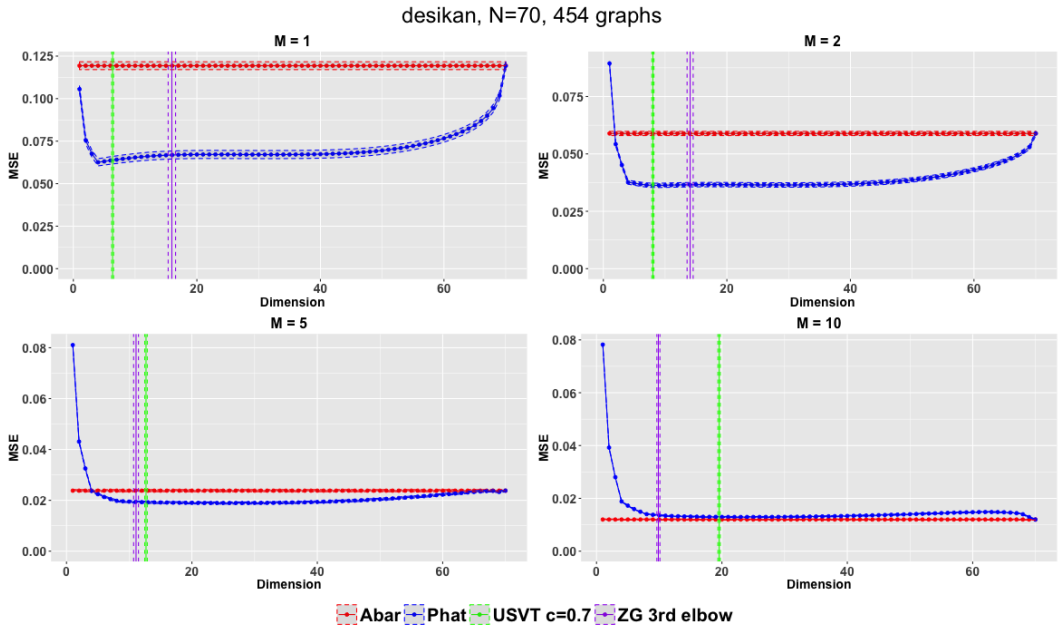

Figure 4: Comparison of MSE between $\bar{A}$ (red) and $\hat{P}$ (blue) for desikan dataset while embedding the graphs into different dimensions with different size M of the subsamples. The dimension chosen by the 3rd elbow of Zhu and Ghodsi is denoted in purple, and chosen by USVT with threshold equals 0.7 is denoted in green. Dashed lines represent the 95% confidence interval. When M is small, $\hat{P}$ outperforms $\bar{A}$ with a flexible range of the embedding dimension including what Zhu and Ghodsi selects.

5 Methods

5.1 Choosing Dimension

- Often in dimensionality reduction techniques, the choice for dimension, d, relies on visually analyzing a plot of the ordered eigenvalues, looking for a "gap" or "elbow" in the scree-plot.
- USVT is a simple estimation procedure that works for any matrix that has "a little bit of structure".

5.2 Graph Diagonal Augmentation

- The graphs examined in this work are hollow, in that there are no self-loops and thus the diagonal entries of the adjacency matrix are 0. This leads to a bias in the calculation of the eigenvectors.
- We minimize this bias by using an iterative method developed by Scheinerman and Tucker.

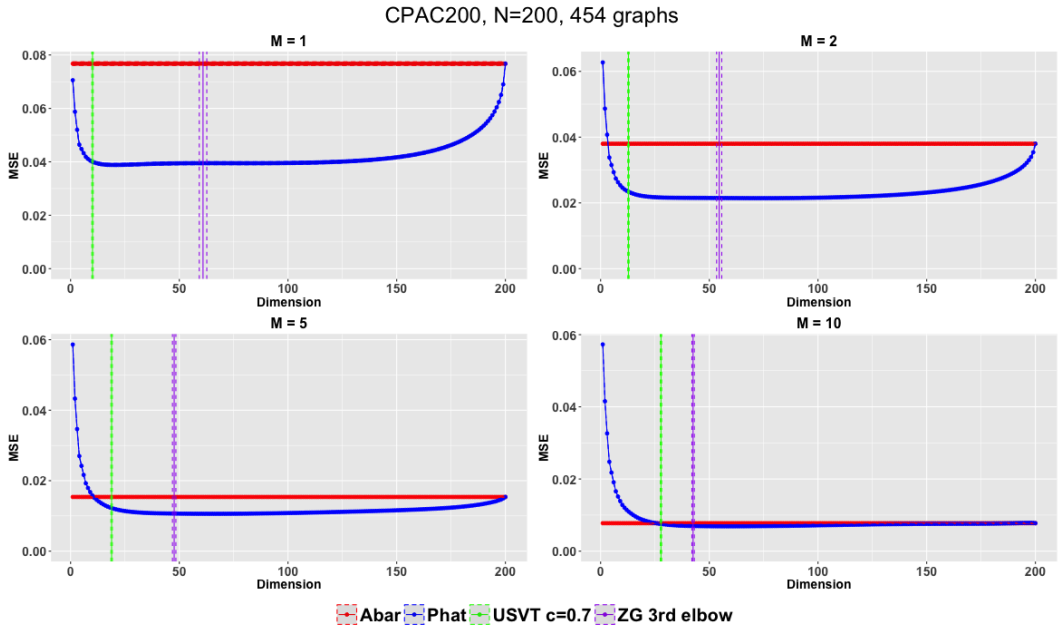

Figure 5: Comparison of MSE between $\bar{A}$ (red) and $\hat{P}$ (blue) for CPAC200 dataset while embedding the graphs into different dimensions with different size M of the subsamples. The dimension chosen by the 3rd elbow of Zhu and Ghodsi is denoted in purple, and chosen by USVT with threshold equals 0.7 is denoted in green. Dashed lines represent the 95% confidence interval. When M is small, $\hat{P}$ outperforms $\bar{A}$ with a flexible range of the embedding dimension including what Zhu and Ghodsi selects.

5.3 Source Code

5.4 Dataset Description

- Detailed description of the data we are using.

5.5 Outline for the Proof of the Theorems