Asymptotic Relative Efficiency for Robust Estimation of the Mean of Contaminated Graphs under a Low Rank Model

Runze Tang

Department of Applied Mathematics & Statistics
Johns Hopkins University

July 31st, 2016

Joint work with Minh Tang, Michael Ketcha,
Carey E. Priebe, Joshua T. Vogelstein.

Problem

- **Given:** m networks on n vertices.
  - Each graph is represented as an adjacency matrix $A^{(t)} \in \mathbb{R}_{\geq 0}^{n \times n}$, $1 \leq t \leq m$.
  - Assuming that $A^{(t)}$ is symmetric and non-negative for $1 \leq t \leq m$.
  - We assume that the vertex correspondence is known.
- **Goal:** Estimate the mean of the collection of graphs.
  - Here mean of the graphs is defined as the average weight of an edge between each pair of vertices.
  - We are interested in the mean of the population of graphs.

Entrywise MLE $\hat{P}^{(1)}$

- Under the independent edge model (IEM), each edge A_{ij} independently follows some distribution (e.g. Bernoulli) with parameter P_{ij}.
- In this case, the entrywise MLE $\hat{P}^{(1)}$ is the entrywise mean $\bar{A}$.
    - Unbiased
    - UMVUE
- Problem solved!

Challenge

- What if we observe the graphs with contaminations?
- Even assuming that there is no contaminations, entrywise MLE $\hat{P}^{(1)}$ doesn't exploit any graph structure!
  - How can we take advantage of the unknown graph structure?

Stochastic Blockmodel

- 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.

Stochastic Blockmodel

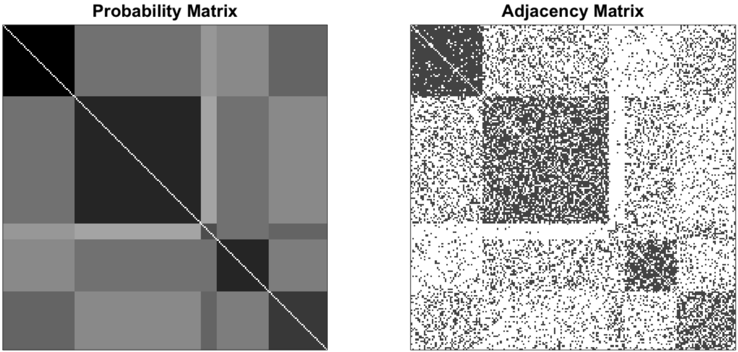


Figure: Example illustrating the SBM. The figure on the left is the probability matrix P that follows a SBM with $K = 5$ blocks; The other figure shows the adjacency matrix A for 200 vertices generated from the SBM with probability matrix P.

Stochastic Blockmodel

- For simplicity, the concepts below are all for Bernoulli distribution.
- Each of n vertices is assigned to exactly one of K blocks.
- The probability of an edge between two vertices depends only on their respective block memberships, and the presence of edges are conditionally independent given block memberships.
- It is parameterized by the block probability matrix $B \in [0,1]^{K \times K}$, and the block proportions $\rho \in (0,1)^K$ with $\sum_{k=1}^{K} \rho_k = 1$.
- Let τ_i denote the block to which vertex i is assigned, then they are independent with $P(\tau_i = k) = \rho_k$.
- The probability of an edge between vertices i and j is B_{τ_i,τ_j}.

Random Dot Product Graph

- Random dot product graph (RDPG)
  - Each vertex is associated with a latent position.
  - The presence or absence of edges are independent Bernoulli random variables conditioning on these latent positions.
  - The probability of an edge between two vertices is given by the dot product of the corresponding latent positions.
- An RDPG can be parameterized as an SBM with K blocks if there are only K distinct latent positions.
- We are going to analyze SBM in the RDPG setting.
- It motivates the estimator $\widetilde{P}^{(1)}$ based on adjacency spectral embedding.

Uncontaminated Model

- Under the stochastic block model with parameters (B, ρ), we have $X_i \overset{iid}{\sim} \sum_{k=1}^{K} \rho_k \delta_{\nu_k}$, where $\nu = [\nu_1, \cdots, \nu_K]^T \in \mathbb{R}^{K \times d}$ satisfies $B = \nu^T \nu$. Define the block assignment τ such that $\tau_i = k$ if and only if $X_i = \nu_k$. Let $P = XX^T$ where $X = [X_1, \cdots, X_n]^T$.
- First sample τ from the multinomial distribution with parameter ρ. Then we are going to sample m conditionally i.i.d. symmetric graphs $G^{(1)}, \cdots, G^{(m)}$ such that conditioning on τ, $G_{ij}^{(k)} \overset{ind}{\sim} \mathrm{Exp}(P_{ij})$ for each $1 \le k \le m$, $1 \le i \le j \le n$.

Adjacency Spectral Embedding

- For an RDPG, the probability matrix $P = XX^T$ is symmetric, positive semidefinite and has rank at most d if $X \in \mathbb{R}^{n \times d}$.
- Let $A = [U_A|\widetilde{U}_A][S_A \oplus \widetilde{S}_A][U_A|\widetilde{U}_A]^T$ be the eigen-decomposition of A, where $S_A \in \mathbb{R}^{d \times d}$ is diagonal matrix with the d largest eigenvalues of A, and $U_A \in \mathbb{R}^{n \times d}$ is the matrix with orthonormal columns of the corresponding eigenvectors. Then the **adjacency spectral embedding (ASE)** of A to dimension d is given by $\widehat{X} = U_A S_A^{1/2}$.

Estimator $\widetilde{P}^{(1)}$

- Obtain $\hat{X}$ by applying ASE to entrywise mean $\hat{P}^{(1)}$ with dimension d.
- Define $\widetilde{P}^{(1)} = \hat{X}\hat{X}^T$ as an estimator of P.

Comparison between $\hat{P}^{(1)}$ and $\widetilde{P}^{(1)}$

To compare the performance between $\hat{P}^{(1)}$ and $\widetilde{P}^{(1)}$, since both of them are asymptotic unbiased, we examine the asymptotic relative efficiency (ARE), defined as: $\mathrm{ARE}(\hat{P}_{ij}^{(1)}, \widetilde{P}_{ij}^{(1)}) = \lim_{n\to\infty} \frac{\mathrm{Var}(\widetilde{P}_{ij}^{(1)})}{\mathrm{Var}(\hat{P}_{ij}^{(1)})}$.

Theorem

Fix m, for any i and j, conditioning on $X_i = \nu_{\tau_i}$ and $X_j = \nu_{\tau_j}$, we have

$$\mathrm{ARE}(\hat{P}_{ij}^{(1)}, \widetilde{P}_{ij}^{(1)}) = 0.$$

And for n large enough, we have the approximation

$$\mathrm{RE}(\hat{P}_{ij}^{(1)}, \widetilde{P}_{ij}^{(1)}) \approx \frac{1/\rho_{\tau_i} + 1/\rho_{\tau_j}}{n}.$$

Contaminated Model

- Under the stochastic block model with parameters (B, ρ), we have $X_i \overset{iid}{\sim} \sum_{k=1}^{K} \rho_k \delta_{\nu_k}$, where $\nu = [\nu_1, \cdots, \nu_K]^T \in \mathbb{R}^{K \times d}$ satisfies $B = \nu^T \nu$. Define the block assignment τ such that $\tau_i = k$ if and only if $X_i = \nu_k$. Let $P = XX^T$ where $X = [X_1, \cdots, X_n]^T$.
- Now we assume the observed edges are contaminated with probability ϵ.
- First sample τ from the multinomial distribution with parameter ρ. Then we are going to sample m conditionally i.i.d. symmetric graphs $A^{(1)}, \cdots, A^{(m)}$ such that conditioning on τ, $A_{ij}^{(t)} \overset{ind}{\sim} (1-\epsilon)\mathrm{Exp}(P_{ij}) + \epsilon \mathrm{Exp}(C_{ij})$ for each $1 \leq t \leq m$, $1 \leq i \leq j \leq n$, where the contamination is $C = YY^T$.

Comparison between $\hat{P}^{(1)}$ and $\widetilde{P}^{(1)}$

$\hat{P}^{(1)}$

For any fixed m, or $m = O(n^b)$ for $b > 0$, any $1 \le i, j \le n$, $\lim_{n\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(1)}) = \lim_{n\to\infty} \mathrm{Bias}^2(\widetilde{P}_{ij}^{(1)})$ $\mathrm{Var}(\widetilde{P}_{ij}^{(1)})/\mathrm{Var}(\hat{P}_{ij}^{(1)}) = O(n^{-2}(\log n)^6)$

$\widetilde{P}^{(1)}$

Maximum L-q Likelihood Estimator

- With contaminations, the performance of MLE degrades.
- We should use entrywise robust estimator instead of $\hat{P}^{(1)}$.
- Maximum L-q Likelihood Estimator (MLqE) is an M-estimator with robust property.
- Denote MLqE as $\hat{P}^{(q)}$.
- MLE is a special case of MLqE with $q = 1$.

Comparison between $\hat{P}^{(1)}$ and $\hat{P}^{(q)}$

$$\begin{tikzcd} \hat{P}^{(1)} \arrow[rr, leftrightarrow] \arrow[d, leftrightarrow] & & \hat{P}^{(q)} \\ \widetilde{P}^{(1)} & & \end{tikzcd}$$

For any fixed n, C large enough, any $1 \leq i, j \leq n$,
$$\lim_{m\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(1)}) > \lim_{m\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(q)})$$

$$\lim_{m\to\infty} \mathrm{Var}(\hat{P}_{ij}^{(1)}) = \lim_{m\to\infty} \mathrm{Var}(\hat{P}_{ij}^{(q)}) = 0$$

For any fixed m, or $m = O(n^b)$ for $b > 0$, any $1 \leq i, j \leq n$,
$$\lim_{n\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(1)}) = \lim_{n\to\infty} \mathrm{Bias}^2(\widetilde{P}_{ij}^{(1)})$$
$$\mathrm{Var}(\widetilde{P}_{ij}^{(1)})/\mathrm{Var}(\hat{P}_{ij}^{(1)}) = O(n^{-2}(\log n)^6)$$

Estimator $\widetilde{P}^{(q)}$

- $\hat{P}^{(q)}$ has the robust property while $\widetilde{P}^{(1)}$ has low asymptotic variance.
- Can we have both simultaneously?
- Define $\widetilde{P}^{(q)}$ as applying ASE to $\hat{P}^{(q)}$!

Comparison between $\widetilde{P}^{(q)}$ and $\hat{P}^{(q)}$

For any fixed n, C large enough, any $1 \le i,j \le n$,
$\lim_{m\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(1)}) > \lim_{m\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(q)})$

$\lim_{m\to\infty} \mathrm{Var}(\hat{P}_{ij}^{(1)}) = \lim_{m\to\infty} \mathrm{Var}(\hat{P}_{ij}^{(q)}) = 0$

$$\begin{tikzcd}
\boxed{\hat{P}^{(1)}} \arrow[rrr, leftrightarrow] \arrow[d, leftrightarrow] & & & \boxed{\hat{P}^{(q)}} \arrow[d, leftrightarrow] \\
\boxed{\widetilde{P}^{(1)}} & & & \boxed{\widetilde{P}^{(q)}}
\end{tikzcd}$$

For any fixed m, or $m = O(n^b)$ for $b > 0$, any $1 \le i,j \le n$,
$\lim_{n\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(1)}) = \lim_{n\to\infty} \mathrm{Bias}^2(\widetilde{P}_{ij}^{(1)})$
$\mathrm{Var}(\widetilde{P}_{ij}^{(1)})/\mathrm{Var}(\hat{P}_{ij}^{(1)}) = O(n^{-2}(\log n)^6)$

For any fixed m, or $m = O(n^b)$ for $b > 0$, any $1 \le i,j \le n$,
$\lim_{n\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(q)}) = \lim_{n\to\infty} \mathrm{Bias}^2(\widetilde{P}_{ij}^{(q)})$
$\mathrm{Var}(\widetilde{P}_{ij}^{(q)})/\mathrm{Var}(\hat{P}_{ij}^{(q)}) = O(mn^{-2}(\log n)^6)$

Summary

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
\hat{P}^{(1)} \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \hat{P}^{(q)} \arrow[d, leftrightarrow] \\
\widetilde{P}^{(1)} \arrow[r, leftrightarrow] & \widetilde{P}^{(q)}
\end{tikzcd}
$$

For any fixed n, C large enough, any $1 \le i, j \le n$,
$$\lim_{m\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(1)}) > \lim_{m\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(q)})$$
$$\lim_{m\to\infty} \mathrm{Var}(\hat{P}_{ij}^{(1)}) = \lim_{m\to\infty} \mathrm{Var}(\hat{P}_{ij}^{(q)}) = 0$$

For any fixed m, or $m = O(n^b)$ for $b > 0$, any $1 \le i, j \le n$,
$$\lim_{n\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(1)}) = \lim_{n\to\infty} \mathrm{Bias}^2(\widetilde{P}_{ij}^{(1)})$$
$$\mathrm{Var}(\widetilde{P}_{ij}^{(1)})/\mathrm{Var}(\hat{P}_{ij}^{(1)}) = O(n^{-2}(\log n)^6)$$

For any fixed m, or $m = O(n^b)$ for $b > 0$, any $1 \le i, j \le n$,
$$\lim_{n\to\infty} \mathrm{Bias}^2(\hat{P}_{ij}^{(q)}) = \lim_{n\to\infty} \mathrm{Bias}^2(\widetilde{P}_{ij}^{(q)})$$
$$\mathrm{Var}(\widetilde{P}_{ij}^{(q)})/\mathrm{Var}(\hat{P}_{ij}^{(q)}) = O(mn^{-2}(\log n)^6)$$

For sufficiently large m, n and C, any $1 \le i, j \le n$,
$$\lim_{m\to\infty} \mathrm{Bias}^2(\widetilde{P}_{ij}^{(1)}) > \lim_{m\to\infty} \mathrm{Bias}^2(\widetilde{P}_{ij}^{(q)})$$

For any fixed m, or $m = o(n^2(\log n)^{-6})$, any $1 \le i, j \le n$,
$$\lim_{n\to\infty} \mathrm{Var}(\widetilde{P}_{ij}^{(1)}) = \lim_{n\to\infty} \mathrm{Var}(\widetilde{P}_{ij}^{(q)}) = 0$$

Simulation Settings

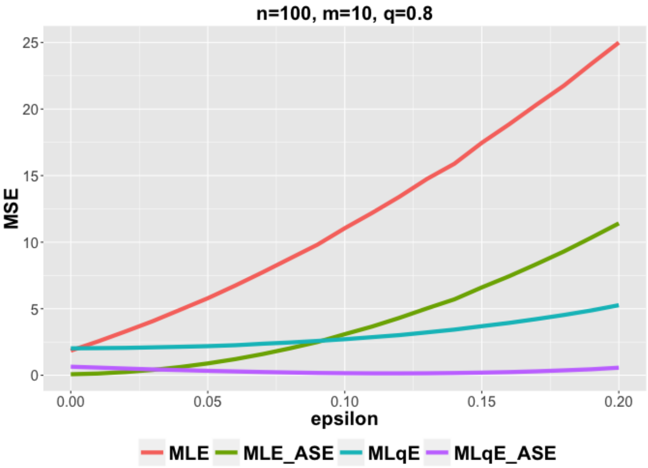
- Consider the stochastic blockmodel parameterized by

$$B = \begin{pmatrix} 4.2 & 2 \\ 2 & 7 \end{pmatrix} \text{ and } \rho = (0.5, 0.5).$$

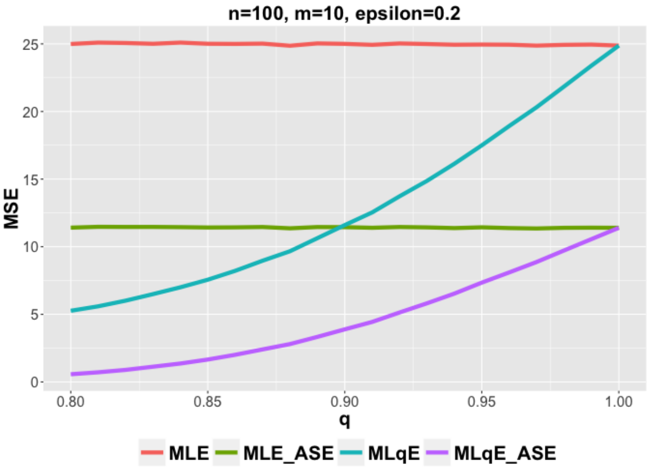
- The contamination is also a SBM parameterized by

$$C = \begin{pmatrix} 20 & 18 \\ 18 & 25 \end{pmatrix} \text{ and } \rho = (0.5, 0.5).$$

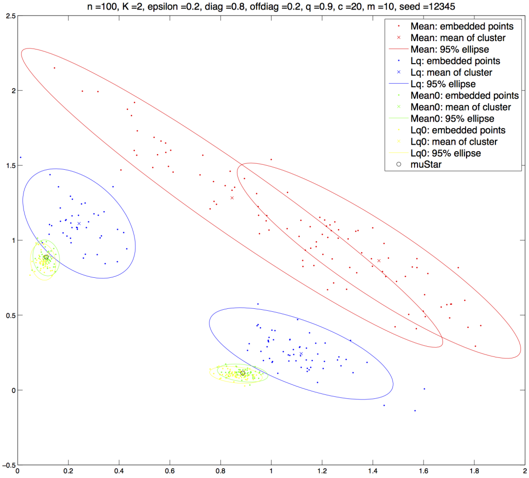
Simulation Results



Simulation Results



Scatter Plot of the Estimated Latent Positions $\widehat{X}_i$



Conclusion

- We propose an estimator for the mean of a collection of weighted graphs under a low rank random graph model.
- It not only inherits robustness from element-wise robust estimators but also has small variance due to application of a rank-reduction procedure.
- Under appropriate conditions, we prove that our estimator outperforms standard estimators via asymptotic relative efficiency.
- We illustrate our theory and methods by Monte Carlo simulation studies and experimental results, more is coming!

References

Athreya, A., V. Lyzinski, D. J. Marchette, C. E. Priebe, D. L. Sussman, and M. Tang (2013).
A limit theorem for scaled eigenvectors of random dot product graphs.
Sankhya A 1–18.

Scheinerman, E. R. and K. Tucker (2010).
Modeling graphs using dot product representations.
Computational Statistics 25(1), 1–16.

Sussman, D. L., M. Tang, D. E. Fishkind, and C. E. Priebe (2012).
A consistent adjacency spectral embedding for stochastic blockmodel graphs.
Journal of the American Statistical Association 107(499), 1119–1128.

Young, S. and E. Scheinerman (2007).
Random dot product graph models for social networks.
Algorithms and models for the web-graph, 138–149.

Zhu, M. and A. Ghodsi (2006).
Automatic dimensionality selection from the scree plot via the use of profile likelihood.
Computational Statistics & Data Analysis 51(2), 918–930.

Oliveira, RI (2006).
Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges.
arXiv preprint arXiv:0911.0600.

Ferrari, D. and Yang, Y. (2006).
Maximum Lq-likelihood estimation.
The Annals of Statistics 753–783.