Gaussian weighted stochastic block model

Exact recovery: statistical and algorithmic thresholds

Aaradhya Pandey
(Joint work with Sanjeev Kulkarni)

Operations Research and Financial Engineering,
Princeton University

December 2023

Background: The binary Stochastic block model

- (Unobserved) Community labels $\sigma^* : [n] \to \{\pm 1\}$ such $\langle \sigma^*, 1 \rangle = 0$.
- (Observed) A random graph $G = ([n], E)$ with Adjacency matrix A such that

$$A(i,j) = A(j,i) \sim \begin{cases} Ber(p) & \text{if } \sigma^*(i) = \sigma^*(j) \\ Ber(q) & \text{if } \sigma^*(i) \neq \sigma^*(j) \end{cases}$$

- (Task) Recover σ^* exactly up to a global sign flip.
- (Output) $\hat{\sigma} : [n] \to \{\pm 1\}$ such that $\#\{i \in [n] : \hat{\sigma}(i) = \sigma^*(i)\} = n$.
- (Regime of interest) $p = \frac{a \log n}{n}$, $q = \frac{b \log n}{n}$ with $a > b > 0$ constants.
- (Parameter of interest) SNR $(a,b) = \frac{|\sqrt{a} - \sqrt{b}|}{\sqrt{2}}$.

The Gaussian weighted stochastic block model (our model)

- (Unobserved) Community labelling $\sigma^* : [n] \to \{\pm 1\}$ such $\langle \sigma^*, 1 \rangle = 0$.
- (Observed) A weighted random graph $G = ([n], (w_e)_{e \in \binom{n}{2}})$ such that

$$A(i,j) = A(j,i) \sim \begin{cases} \mathcal{N}(\mu_1, \tau^2) & \text{if } \sigma^*(i) = \sigma^*(j) \\ \mathcal{N}(\mu_2, \tau^2) & \text{if } \sigma^*(i) \neq \sigma^*(j) \end{cases}$$

- (Task) Recover σ^* exactly up to a global sign flip.
- (Output) $\hat{\sigma} : [n] \to \{\pm 1\}$ such that $\#\{i \in [n] : \hat{\sigma}(i) = \sigma^*(i)\} = n$.
- (Regime of interest) $\mu_1 = \alpha \sqrt{\frac{\log n}{n}}$, $\mu_2 = \beta \sqrt{\frac{\log n}{n}}$ with $\alpha > \beta$ and $\tau > 0$ constants.
- (Parameter of interest) SNR $(\alpha, \beta) = \frac{|\alpha - \beta|}{\tau \sqrt{2}}$.

The objects of interest

- $\hat{\sigma}_{MLE} = \text{argmax } \sum_{i,j} A_{ij}\sigma_i\sigma_j = Tr(A\sigma\sigma^T)$ subject to $\sigma \in \{\pm 1\}^n, \ \langle \sigma, 1 \rangle = 0.$
- $\hat{Y}_{SDP} = \text{argmax } Tr(AY)$ subject to $Y \succeq 0, \ Y_{ii} = 1 \ \forall \ i \in [n], \ Tr(JY) = 0.$
- $\hat{\sigma}_{Spec} = \text{sign}(\text{argmax } \sum_{i,j} A_{ij} v_i v_j = Tr(Avv^T)$ subject to $v \in \mathbb{S}^{n-1}, \ \langle v, 1 \rangle = 0).$

Results: Binary and the Gaussian model

(previous results for the binary model)

- If $\text{SNR}(a,b) = \frac{|\sqrt{a}-\sqrt{b}|}{\sqrt{2}} < 1$

$$\mathbb{P}(\hat{\sigma}_{MLE} = \sigma^*) \nrightarrow 1[ABH, 2014]$$

- If $\text{SNR}(a,b) = \frac{|\sqrt{a}-\sqrt{b}|}{\sqrt{2}} > 1$

$$\mathbb{P}(\hat{\sigma}_{MLE} = \sigma^*) \to 1[ABH, 14]$$

$$\mathbb{P}(\hat{Y}_{SDP} = \sigma^* \sigma^{*T}) \to 1[HWX, 16]$$

$$\mathbb{P}(\hat{\sigma}_{Spec} = \sigma^*) \to 1[AFYZ, 19]$$

(our results for the Gaussian model)

- If $\text{SNR}\,(\alpha, \beta) = \frac{|\alpha-\beta|}{\tau\sqrt{2}} < 1$

$$\mathbb{P}(\hat{\sigma}_{MLE} = \sigma^*) \nrightarrow 1$$

- If $\text{SNR}\,(\alpha, \beta) = \frac{|\alpha-\beta|}{\tau\sqrt{2}} > 1$

$$\mathbb{P}(\hat{\sigma}_{MLE} = \sigma^*) \to 1$$

$$\mathbb{P}(\hat{Y}_{SDP} = \sigma^* \sigma^{*T}) \to 1$$

$$\mathbb{P}(\hat{\sigma}_{Spec} = \sigma^*) \to 1$$

Proof ideas

- Statistical possibility and impossibility of the MLE is established through First and Second Moment methods.
- Proof for the Semidefinite Programming (SDP) estimator involves a clever dual certificate argument.
- Spectral estimator's proof requires entrywise analysis of eigenvectors.

Connections

- It resolves the exact recovery problem for the planted spin glass (spiked Wigner).
- It is also shown that exactly recovering two equal-sized community is an easier problem than exactly recovering a densely weighted community of size $n/2$.
- More precisely we need $\mathrm{SNR}(\alpha, \beta) > 2$ to be able to exactly recover (through the SDP procedure) a densely weighted community of size $n/2$.

The Gaussian weighted planted dense subgraph model

- (Unobserved) Community labelling $\sigma^* : [n] \to \{\pm 1\}$ such $\langle \sigma^*, 1 \rangle = 0$.
- (Observed) A weighted random graph $G = ([n], (w_e)_{e \in \binom{n}{2}})$ such that

$$A(i,j) = A(j,i) \sim \begin{cases} \mathcal{N}(\mu_1, \tau^2) & \text{if } \sigma^*(i) = \sigma^*(j) = 1 \\ \mathcal{N}(\mu_2, \tau^2) & \textit{otherwise} \end{cases}$$

- (Task) Recover σ^* exactly up to a global sign flip.
- (Output) $\hat{\sigma} : [n] \to \{\pm 1\}$ such that $\#\{i \in [n] : \hat{\sigma}(i) = \sigma^*(i)\} = n$.
- (Regime of interest) $\mu_1 = \alpha \sqrt{\frac{\log n}{n}}, \mu_2 = \beta \sqrt{\frac{\log n}{n}}$ with $\alpha > \beta$ and $\tau > 0$ constants.
- (Parameter of interest) SNR $(\alpha, \beta) = \frac{|\alpha - \beta|}{\tau\sqrt{2}}$.

[ABH] Emmanuel Abbe, Afonso S. Bandeira, Georgina Hall: Exact Recovery in the Stochastic Block Model ABH.

[AFWZ] Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, Yiqiao Zhong: Entrywise Eigenvector Analysis of Random Matrices with Low Expected Rank AFWZ

[HWX] Bruce Hajek, Yihong Wu, Jiaming Xu : Achieving Exact Cluster Recovery Threshold via Semidefinite Programming HWX

[PK] Aaradhya Pandey, Sanjeev Kulkarni (forthcoming draft): Gaussian weighted Stochastic block model: Statistical and algorithmic thresholds

Thank You!