Community Detection with Common Structure

Kiyeon Jeon

University of Texas at Austin

kiyeonj@utexas.edu

December 10, 2018

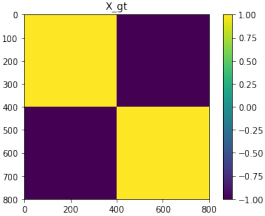
Overview

1. Community Detection
2. SDP relaxation
3. Multiple Networks

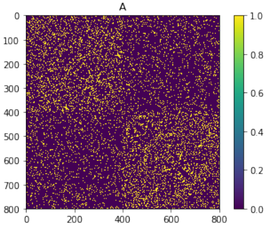
Community Detection

- Recover communities C_1 and C_2 from one instance of a random graph from $G(n, p, q)$

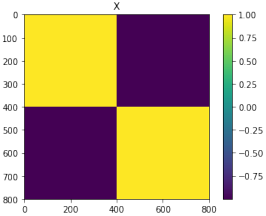
Community Detection

- Recover communities C_1 and C_2 from one instance of a random graph from $G(n, p, q)$
- 

(a) Ground Truth



(b) Adjacency Matrix



(c) Optimal Solution

SDP relaxation - two clusters

- $\bar{x} \in \{-1, +1\}^n$, $\bar{x}_i = \begin{cases} 1, & i \in C_1 \\ -1, & i \in C_2 \end{cases}$

SDP relaxation - two clusters

- $\bar{x} \in \{-1, +1\}^n$, $\bar{x}_i = \begin{cases} 1, & i \in C_1 \\ -1, & i \in C_2 \end{cases}$
- $\hat{x} = \underset{x \in \{-1,+1\}^n}{\arg\max}\ x^T A x \,(= \langle A, xx^T \rangle)$ (NP-hard)

SDP relaxation - two clusters

- $\bar{x} \in \{-1,+1\}^n$, $\bar{x}_i = \begin{cases} 1, & i \in C_1 \\ -1, & i \in C_2 \end{cases}$
- $\hat{x} = \underset{x \in \{-1,+1\}^n}{\arg\max}\ x^T A x \, (= \langle A, xx^T \rangle)$ (NP-hard)
- $\hat{Z} = \underset{Z_{ii}=1, Z \succeq 0}{\arg\max} \langle A - \lambda E_n, Z \rangle$

SDP relaxation - two clusters

- $\bar{x} \in \{-1,+1\}^n$, $\bar{x}_i = \begin{cases} 1, & i \in C_1 \\ -1, & i \in C_2 \end{cases}$
- $\hat{x} = \underset{x \in \{-1,+1\}^n}{\arg\max}\ x^T A x \; (= \langle A, xx^T \rangle)$ (NP-hard)
- $\hat{Z} = \underset{Z_{ii}=1, Z \succeq 0}{\arg\max} \langle A - \lambda E_n, Z \rangle$

Theorem ([GV16])

Under some condition, $\|\hat{Z} - \bar{x}\bar{x}^T\|_F^2 \leq \epsilon \|\bar{x}\bar{x}^T\|_F^2$ *with high probability*

SDP relaxation - multiple clusters(> 2)

- $\bar{X} \in \{0,1\}^{n \times K}, \bar{X}_{ij} = \begin{cases} 1 & \text{if } i \text{ is in } j\text{-th cluster} \\ 0 & \text{otherwise} \end{cases}$

SDP relaxation - multiple clusters(> 2)

- $\bar{X} \in \{0,1\}^{n\times K}, \bar{X}_{ij} = \begin{cases} 1 & \text{if } i \text{ is in } j\text{-th cluster} \\ 0 & \text{otherwise} \end{cases}$
- $\hat{X} = \underset{X\in\{0,1\}^{n\times K}}{\arg\max} < A, XX^T >$

SDP relaxation - multiple clusters(> 2)

- $\bar{X} \in \{0,1\}^{n\times K}, \bar{X}_{ij} = \begin{cases} 1 & \text{if } i \text{ is in } j\text{-th cluster} \\ 0 & \text{otherwise} \end{cases}$
- $\hat{X} = \underset{X\in\{0,1\}^{n\times K}}{\arg\max} < A, XX^T >$
- $\bar{Z} = \bar{X}\bar{X}^T \in \{0,1\}^{n\times n}$ defined as
  $\bar{Z}_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are same community} \\ 0 & \text{otherwise} \end{cases}$

SDP relaxation - multiple clusters(> 2)

- $\bar{X} \in \{0,1\}^{n \times K}, \bar{X}_{ij} = \begin{cases} 1 & \text{if } i \text{ is in } j\text{-th cluster} \\ 0 & \text{otherwise} \end{cases}$
- $\hat{X} = \underset{X \in \{0,1\}^{n \times K}}{\arg\max} < A, XX^T >$
- $\bar{Z} = \bar{X}\bar{X}^T \in \{0,1\}^{n \times n}$ defined as
  $\bar{Z}_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are same community} \\ 0 & \text{otherwise} \end{cases}$
- $\hat{Z} = \underset{Z_{ii}=1, Z \succeq 0, 0 \leq Z \leq 1}{\arg\max} \langle A - \lambda E_n, Z \rangle$

SDP relaxation - multiple clusters(> 2)

- $\bar{X} \in \{0,1\}^{n\times K}, \bar{X}_{ij} = \begin{cases} 1 & \text{if } i \text{ is in } j\text{-th cluster} \\ 0 & \text{otherwise} \end{cases}$
- $\hat{X} = \underset{X\in\{0,1\}^{n\times K}}{\arg\max} < A, XX^T >$
- $\bar{Z} = \bar{X}\bar{X}^T \in \{0,1\}^{n\times n}$ defined as
  $\bar{Z}_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are same community} \\ 0 & \text{otherwise} \end{cases}$
- $\hat{Z} = \underset{Z_{ii}=1\,, Z\succeq 0\,, 0\le Z\le 1}{\arg\max} \langle A - \lambda E_n, Z\rangle$

Theorem ([GV16])

Under some condition, $\|\hat{Z} - \bar{Z}\|_F^2 \le \epsilon n^2$ *with high probability*

Multiple Networks

- Multiple Networks with shared structure

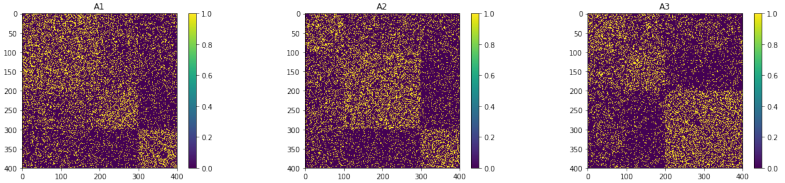
- Multiple Networks with shared structure
- 

Figure: adjacency matrices

Multiple Networks

- Multiple Networks with shared structure

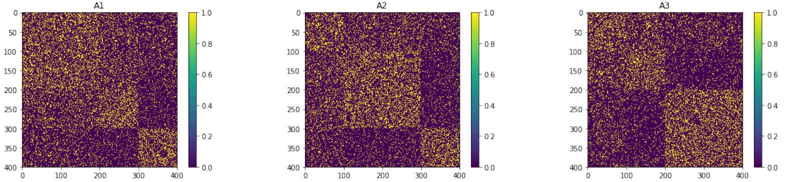
- 

Figure: adjacency matrices

- Motivation: we can use shared structure to estimate cluster matrices

Multiple Networks

Previously,

$$\text{for each } i,\ \hat{X}_i := \arg \max_{\substack{\mathrm{diag}(X_i)=1_n \\ 0 \le X_i \le 1,\, X_i \succeq 0}} \langle A_i, X_i \rangle - \lambda_i \langle E_n, X_i \rangle$$

Multiple Networks

Previously,

$$\text{for each } i,\ \hat{X}_i := \arg \max_{\substack{\operatorname{diag}(X_i)=1_n \\ 0 \le X_i \le 1,\, X_i \succeq 0}} \langle A_i, X_i \rangle - \lambda_i \langle E_n, X_i \rangle$$

1. Let $X_i = C_i + Y_i$ where C_i is common structure

$$\hat{C}_i, \hat{Y}_i := \arg\max \sum_{i=1}^{N} \langle A_i - \lambda_i E_n, C_i + Y_i \rangle - \sum_{i \neq j} \lambda_{ij} \| C_i - C_j \|_F^2$$

Multiple Networks

Previously,

$$\text{for each } i,\ \hat{X}_i := \arg \max_{\substack{\mathrm{diag}(X_i)=1_n \\ 0\le X_i \le 1,\ X_i \succeq 0}} \langle A_i, X_i \rangle - \lambda_i \langle E_n, X_i \rangle$$

1. Let $X_i = C_i + Y_i$ where C_i is common structure

$$\hat{C}_i, \hat{Y}_i := \arg\max \sum_{i=1}^{N} \langle A_i - \lambda_i E_n, C_i + Y_i \rangle - \sum_{i \neq j} \lambda_{ij} \| C_i - C_j \|_F^2$$

2. Using shared sparsity,

$$\hat{C}_i, \hat{Y}_i := \arg \max_{\substack{\mathrm{diag}(C_i+Y_i)=1_n \\ C_i+Y_i \le 1,\ C_i+Y_i \succeq 0 \\ C_i, Y_i \ge 0,\ \|C\|_{1,\infty} \le R}} \sum_{i=1}^{N} \langle A_i - \lambda_i E_n, C_i + Y_i \rangle$$

where $C = [vec(C_1)| \cdots |vec(C_n)] \in \mathbb{R}^{n^2 \times n}$ and $\|C\|_{1,\infty} = \sum_{i=1}^{n^2} \|C^i\|_\infty$

Multiple Networks

$$\begin{cases} \tilde{C}_i := & \arg\max_{\substack{\mathrm{diag}(C_i) \le 1_n \\ 0 \le C_i \le 1, C_i \succeq 0}} \langle A_i - \lambda_i E_n, C_i \rangle \quad \text{for each } i \in [n] \\ \hat{C} := & \mathcal{P}_{\|\cdot\|_{1,\infty} \le R}(\tilde{C}) \\ \hat{Y}_i := & \arg\max_{\substack{\mathrm{diag}(\hat{C}_i + Y_i) = 1_n \\ 0 \le \hat{C}_i + Y_i \le 1, \hat{C}_i + Y_i \succeq 0}} \langle A_i - \lambda_i E_n, Y_i \rangle \end{cases}$$

Multiple Networks

$$\begin{cases} \tilde{C}_i := & \arg\max_{\substack{\mathrm{diag}(C_i)\le 1_n \\ 0\le C_i\le 1, C_i\succeq 0}} \langle A_i - \lambda_i E_n, C_i\rangle \quad \text{for each } i\in[n] \\ \hat{C} := & \mathcal{P}_{\|\cdot\|_{1,\infty}\le R}(\tilde{C}) \\ \hat{Y}_i := & \arg\max_{\substack{\mathrm{diag}(\hat{C}_i+Y_i)=1_n \\ 0\le \hat{C}_i+Y_i\le 1, \hat{C}_i+Y_i\succeq 0}} \langle A_i - \lambda_i E_n, Y_i\rangle \end{cases}$$

[Sra12]

$$\mathcal{P}_{\|X\|_{1,\infty}}(C) = \arg\min_{\|X\|_{1,\infty}\le R} \|X - C\|_F^2$$

$$\begin{aligned} prox_{\lambda\|X\|_{1,\infty}}(C) &= \arg\min_X \frac{1}{2}\|X - C\|_F^2 + \lambda\|X\|_{1,\infty} \\ &= \{\arg\min_{X^i} \frac{1}{2}\|X^i - C^i\|_2^2 + \lambda\|X^i\|_\infty\}_{i=1,\dots,n} \\ &= \{prox_{\lambda\|X^i\|_\infty}(C^i)\} = \{C^i - prox_{\|X^i\|_1/\lambda}(C^i/\lambda)\} \end{aligned}$$

Simulation

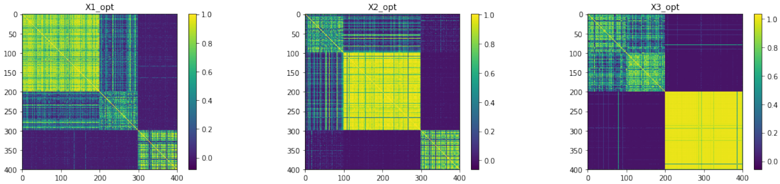


Figure: optimal solution via ADMM

Simulation

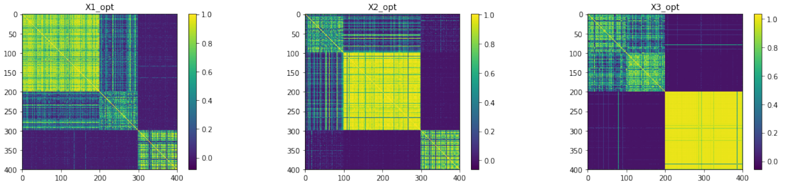


Figure: optimal solution via ADMM

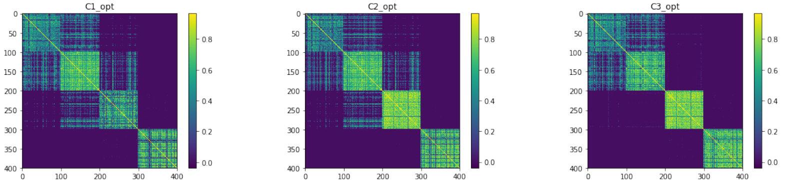


Figure: common structure via $\|\cdot\|_{1,\infty}$ norm projection

Reference

Olivier Guédon and Roman Vershynin.
Community detection in sparse networks via grothendiecks inequality.
Probability Theory and Related Fields, 165(3-4):1025–1049, 2016.

Suvrit Sra.
Fast projections onto mixed-norm balls with applications.
Data Mining and Knowledge Discovery, 25(2):358–377, 2012.

The End