Introduction to spectral methods

Cong Ma

University of Chicago, Autumn 2021

Outline

- A motivating application: community detection
- A general recipe for spectral methods (with more applications)

A motivating application: community detection

Community detection / graph clustering

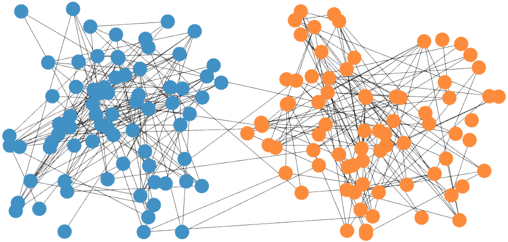
Community structures are common in many social networks

figure credit: The Future Buzz

figure credit: S. Papadopoulos

Goal: partition users into several clusters based on their friendships / similarities

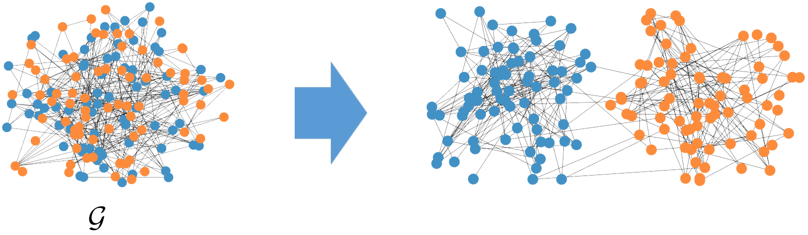
A simple model: stochastic block model (SBM)



$x_i^\star = 1$: 1st community $\qquad$ $x_i^\star = -1$: 2nd community

- n nodes $\{1, \ldots, n\}$
- 2 communities
- n unknown variables: $x_1^\star, \ldots, x_n^\star \in \{1, -1\}$
  - encode community memberships

A simple model: stochastic block model (SBM)



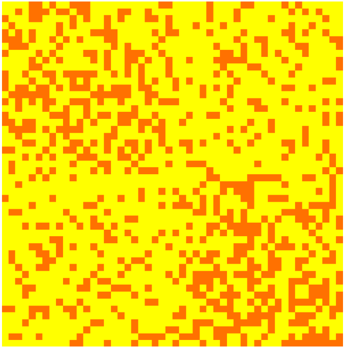
- observe a graph $\mathcal{G}$

$$(i,j) \in \mathcal{G} \text{ with prob. } \begin{cases} p, & \text{if } i \text{ and } j \text{ are from same community} \\ q, & \text{else} \end{cases}$$

  Here, $p > q$

- **Goal:** recover community memberships of all nodes, i.e., $\{x_i^\star\}$

Adjacency matrix

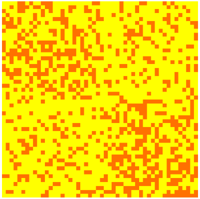


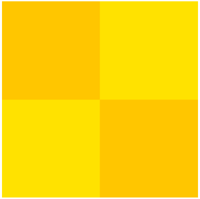
Consider the adjacency matrix $\boldsymbol{A} \in \{0,1\}^{n \times n}$ of $\mathcal{G}$:

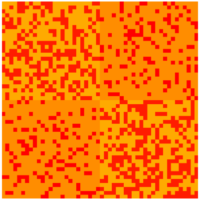
$$A_{i,j} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{G} \\ 0, & \text{else} \end{cases}$$

- WLOG, suppose $x_1^\star = \cdots = x_{n/2}^\star = 1$; $x_{n/2+1}^\star = \cdots = x_n^\star = -1$

Adjacency matrix



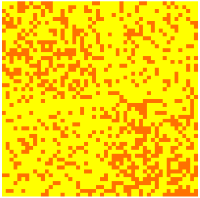


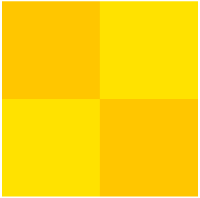


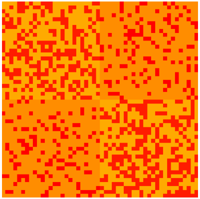
$$\boldsymbol{A} \quad = \quad \underbrace{\mathbb{E}[\boldsymbol{A}]}_{\text{rank 2}} \quad + \quad \boldsymbol{A} - \mathbb{E}[\boldsymbol{A}]$$

$$\mathbb{E}[\boldsymbol{A}] = \begin{bmatrix} p\mathbf{1}\mathbf{1}^\top & q\mathbf{1}\mathbf{1}^\top \\ q\mathbf{1}\mathbf{1}^\top & p\mathbf{1}\mathbf{1}^\top \end{bmatrix} = \underbrace{\frac{p+q}{2}\mathbf{1}\mathbf{1}^\top}_{\text{uninformative bias}} + \frac{p-q}{2} \underbrace{\begin{bmatrix} \mathbf{1} \\ -\mathbf{1} \end{bmatrix}}_{=\boldsymbol{x}^\star = [x_i]_{1\le i\le n}} [\mathbf{1}^\top, -\mathbf{1}^\top]$$

Spectral clustering







$$\boldsymbol{A} = \underbrace{\mathbb{E}[\boldsymbol{A}]}_{\text{rank 2}} + \boldsymbol{A} - \mathbb{E}[\boldsymbol{A}]$$

1. computing the leading eigenvector $\boldsymbol{u} = [u_i]_{1 \leq i \leq n}$ of $\boldsymbol{A} - \frac{p+q}{2}\mathbf{1}\mathbf{1}^\top$
2. rounding: output $x_i = \begin{cases} 1, & \text{if } u_i > 0 \\ -1, & \text{if } u_i < 0 \end{cases}$

Rationale behind spectral clustering

Recovery is reliable if $\underbrace{\boldsymbol{A} - \mathbb{E}[\boldsymbol{A}]}_{\text{perturbation}}$ is sufficiently small

- if $\boldsymbol{A} - \mathbb{E}[\boldsymbol{A}] = \boldsymbol{0}$, then

$$\boldsymbol{u} \ \propto \ \pm \begin{bmatrix} \boldsymbol{1} \\ -\boldsymbol{1} \end{bmatrix} \quad \Longrightarrow \quad \text{perfect clustering}$$

A general recipe for spectral methods

Three key steps:

- identify a key matrix $\boldsymbol{M}^{\star}$, whose eigenvectors disclose crucial information
- construct a surrogate matrix $\boldsymbol{M}$ of $\boldsymbol{M}^{\star}$ using data
- compute corresponding eigenvectors of $\boldsymbol{M}$

Low-rank matrix completion

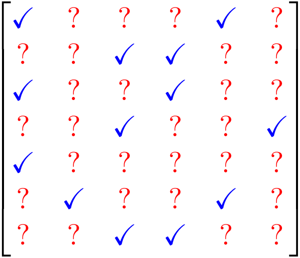


figure credit: Candès

- consider a low-rank matrix $\boldsymbol{M}^{\star} = \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{V}^{\star\top}$
- each entry $M^{\star}_{i,j}$ is observed independently with prob. p
- **intermediate goal:** estimate $\boldsymbol{U}^{\star}, \boldsymbol{V}^{\star}$

Spectral method for matrix completion

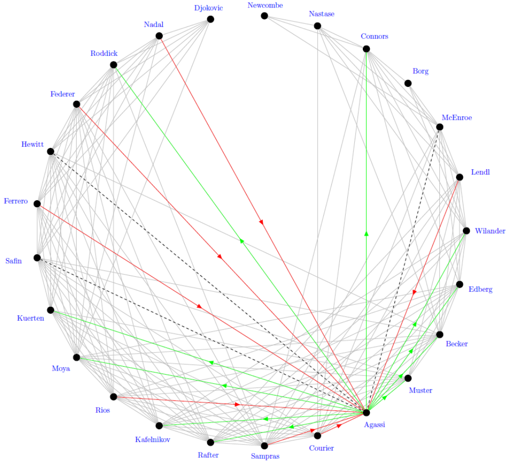
1. identify the key matrix $\boldsymbol{M}^\star$
2. construct surrogate matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ as

$$M_{i,j} = \begin{cases} \frac{1}{p} M^\star_{i,j}, & \text{if } M^\star_{i,j} \text{ is observed} \\ 0, & \text{else} \end{cases}$$

   - **rationale for rescaling:** ensures $\mathbb{E}[\boldsymbol{M}] = \boldsymbol{M}^\star$

3. compute the rank-r SVD $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ of $\boldsymbol{M}$, and return $(\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V})$

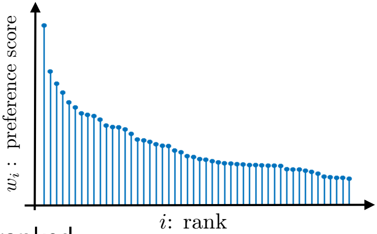
Ranking from pairwise comparisons



pairwise comparisons for ranking tennis players

figure credit: Bozóki, Csató, Temesi

Bradley-Terry-Luce (logistic) model



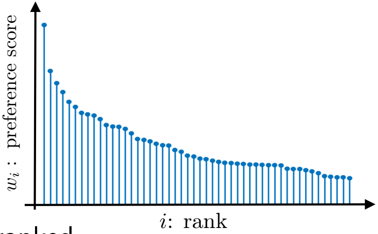
- n items to be ranked
- assign a latent score $\{w_i^\star\}_{1 \le i \le n}$ to each item, so that

$$\text{item } i \succ \text{item } j \quad \text{if} \quad w_i^\star > w_j^\star$$

- each pair of items (i, j) is compared independently

$$\mathbb{P}\left\{\text{item } j \text{ beats item } i\right\} = \frac{w_j^\star}{w_i^\star + w_j^\star}$$

Bradley-Terry-Luce (logistic) model



- n items to be ranked
- assign a latent score $\{w_i^\star\}_{1\leq i\leq n}$ to each item, so that

$$\text{item } i \succ \text{item } j \quad \text{if} \quad w_i^\star > w_j^\star$$

- each pair of items (i,j) is compared independently

$$y_{i,j} \overset{\text{ind.}}{=} \begin{cases} 1, & \text{with prob. } \frac{w_j^\star}{w_i^\star + w_j^\star} \\ 0, & \text{else} \end{cases}$$

- **intermediate goal:** estimate score vector $\boldsymbol{w}^\star$ (up to scaling)

Spectral ranking

1. identify key matrix $\boldsymbol{P}^{\star}$—probability transition matrix

$$P^{\star}_{i,j} = \begin{cases} \frac{1}{n} \cdot \frac{w_j^{\star}}{w_i^{\star} + w_j^{\star}}, & \text{if } i \neq j \\ 1 - \sum_{l: l \neq i} P^{\star}_{i,l}, & \text{if } i = j \end{cases}$$

Rationale:

- $\boldsymbol{P}^{\star}$ obeys

$$w_i^{\star} P^{\star}_{i,j} = w_j^{\star} P^{\star}_{j,i} \qquad \text{(detailed balance)}$$

- Thus, the stationary distribution $\boldsymbol{\pi}^{\star}$ of $\boldsymbol{P}^{\star}$ obeys

$$\boldsymbol{\pi}^{\star} \; = \; \frac{1}{\sum_l w_l^{\star}} \boldsymbol{w}^{\star} \qquad \text{(reveals true scores)}$$

Spectral ranking

2. construct a surrogate matrix $\boldsymbol{P}$ obeying

$$P_{i,j} = \begin{cases} \frac{1}{n} y_{i,j}, & \text{if } i \neq j \\ 1 - \sum_{l: l \neq i} P_{i,l}, & \text{if } i = j \end{cases}$$

3. return leading left eigenvector $\boldsymbol{\pi}$ of $\boldsymbol{P}$ as score estimate

— closely related to PageRank

Spectral ranking

2. construct a surrogate matrix $\boldsymbol{P}$ obeying

$$P_{i,j} = \begin{cases} \frac{1}{n} y_{i,j}, & \text{if } i \neq j \\ 1 - \sum_{l:l \neq i} P_{i,l}, & \text{if } i = j \end{cases}$$

3. return leading left eigenvector $\boldsymbol{\pi}$ of $\boldsymbol{P}$ as score estimate

— closely related to PageRank

Key: stability of eigenspace against perturbation $\boldsymbol{M} - \boldsymbol{M}^{\star}$?