

## **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*

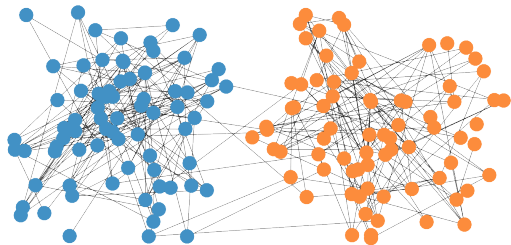


*figure credit: S. Papadopoulos*

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

# A simple model: stochastic block model (SBM)

---

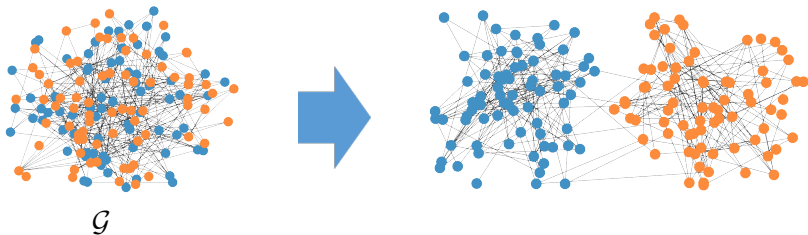


$x_i^* = 1$ : 1<sup>st</sup> community

$x_i^* = -1$ : 2<sup>nd</sup> community

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

# A simple model: stochastic block model (SBM)



- observe a graph  $\mathcal{G}$

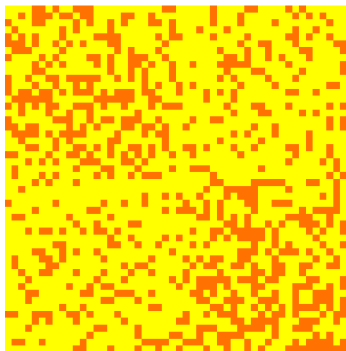
$$(i, j) \in \mathcal{G} \text{ with prob. } \begin{cases} p, & \text{if } i \text{ and } j \text{ are from same community} \\ q, & \text{else} \end{cases}$$

Here,  $p > q$

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

# Adjacency matrix

---

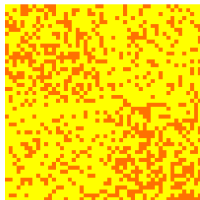


Consider the adjacency matrix  $\mathbf{A} \in \{0, 1\}^{n \times n}$  of  $\mathcal{G}$ :

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

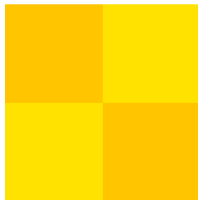
- WLOG, suppose  $x_1^* = \dots = x_{n/2}^* = 1$ ;  $x_{n/2+1}^* = \dots = x_n^* = -1$

# Adjacency matrix



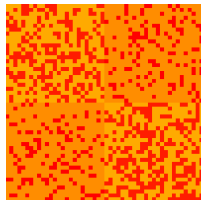
$\mathbf{A}$

=



$\mathbb{E}[\mathbf{A}]$   
rank 2

+



$\mathbf{A} - \mathbb{E}[\mathbf{A}]$

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



# Spectral clustering

---

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

1. computing the leading eigenvector  $\mathbf{u} = [u_i]_{1 \leq i \leq n}$  of  $\mathbf{A} - \frac{p+q}{2} \mathbf{1}\mathbf{1}^\top$
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  $\underbrace{\mathbf{A} - \mathbb{E}[\mathbf{A}]}_{\text{perturbation}}$  is sufficiently small

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

$$\mathbf{u} \propto \pm \begin{bmatrix} 1 \\ -1 \end{bmatrix} \implies \text{perfect clustering}$$

# A general recipe for spectral methods

---

Three key steps:

- identify a key matrix  $M^*$ , whose eigenvectors disclose crucial information
- construct a surrogate matrix  $M$  of  $M^*$  using data
- compute corresponding eigenvectors of  $M$

# Low-rank matrix completion

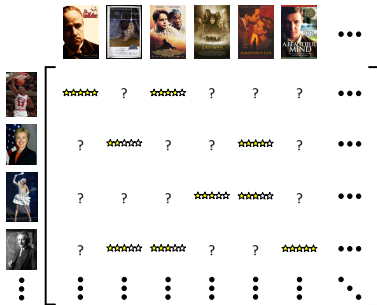


figure credit: Candès

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

# Spectral method for matrix completion

---

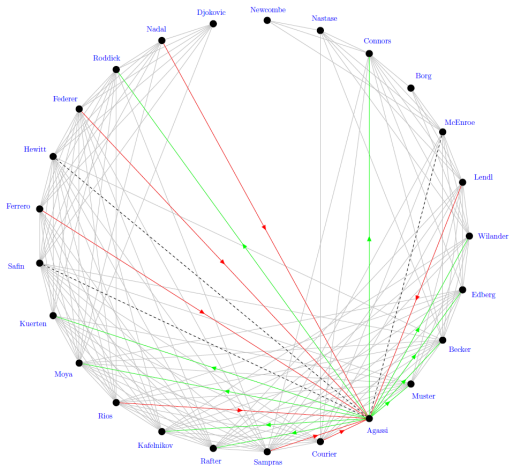
1. identify the key matrix  $M^\star$
2. construct surrogate matrix  $M \in \mathbb{R}^{n \times n}$  as

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

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

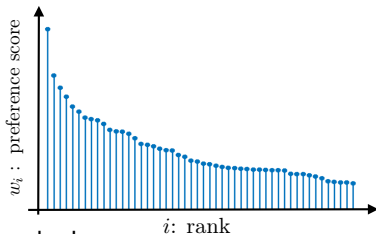
3. compute the rank- $r$  SVD  $U\Sigma V^\top$  of  $M$ , and return  $(U, \Sigma, V)$

## Ranking from pairwise comparisons



# Bradley-Terry-Luce (logistic) model

---



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

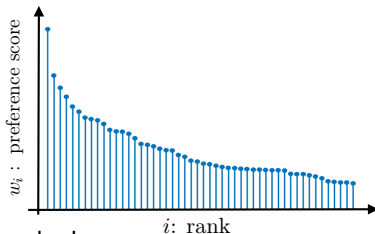
$$\text{item } i \succ \text{item } j \quad \text{if} \quad w_i^* > w_j^*$$

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

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

# Bradley-Terry-Luce (logistic) model

---



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

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

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

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

- **intermediate goal:** estimate score vector  $w^*$  (up to scaling)



# Spectral ranking

---

1. identify key matrix  $P^*$ —probability transition matrix

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

Rationale:

- $P^*$  obeys

$$w_i^* P_{i,j}^* = w_j^* P_{j,i}^* \quad (\text{detailed balance})$$

- Thus, the stationary distribution  $\pi^*$  of  $P^*$  obeys

$$\pi^* = \frac{1}{\sum_l w_l^*} w^* \quad (\text{reveals true scores})$$

# Spectral ranking

---

2. construct a surrogate matrix  $\mathbf{P}$  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  $\mathbf{P}$  as score estimate

— closely related to PageRank

# Spectral ranking

---

2. construct a surrogate matrix  $\mathbf{P}$  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  $\mathbf{P}$  as score estimate

— closely related to PageRank

**Key:** stability of eigenspace against perturbation  $\mathbf{M} - \mathbf{M}^*$