

Spectral Clustering

Khôi Nguyên, Thăng Long, Vĩnh Toàn, Minh Quân

PiMA 2021



Trình bày: Nhóm 6, SC

8 tháng 8 năm 2021

- 1 Spectral Clustering là gì?
- 2 Các bước tiến hành
 - Dữ liệu đầu vào
 - Mô hình hóa
 - Thuật toán Spectral Clustering
 - Chọn số cụm
- 3 Áp dụng
- 4 Các phiên bản cải tiến

Nội dung

- 1 Spectral Clustering là gì?
- 2 Các bước tiến hành
 - Dữ liệu đầu vào
 - Mô hình hóa
 - Thuật toán Spectral Clustering
 - Chọn số cụm
- 3 Áp dụng
- 4 Các phiên bản cải tiến



Ý nghĩa tên gọi

- Clustering: là một lớp bài toán chia điểm dữ liệu thành các cụm (cluster).



Ý nghĩa tên gọi

- Clustering: là một lớp bài toán chia điểm dữ liệu thành các cụm (cluster).
- Spectral: tính từ của spectrum, tập các giá trị riêng của ma trận.



Ý nghĩa tên gọi

- Clustering: là một lớp bài toán chia điểm dữ liệu thành các cụm (cluster).
 - Spectral: tính từ của spectrum, tập các giá trị riêng của ma trận.
- Spectral graph theory.



Một số thuật ngữ

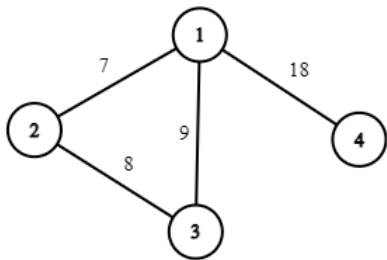
Ma trận kề (adjacency matrix)

Ma trận kề của một đồ thị vô hướng $G = (V, E)$ là một ma trận $W = (w_{ij})_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$ với w_{ij} là trọng số của cạnh nối 2 đỉnh v_i và v_j . Trường hợp $w_{ij} = 0$ nghĩa là không có cạnh nối giữa chúng.



Một số thuật ngữ

Ví dụ: Cho một đồ thị có $n = 4$ đỉnh và ma trận kề tương ứng như sau:



$$W = \begin{bmatrix} 0 & 7 & 9 & 18 \\ 7 & 0 & 8 & 0 \\ 9 & 8 & 0 & 0 \\ 18 & 0 & 0 & 0 \end{bmatrix}$$



Một số thuật ngữ

Ma trận bậc (degree matrix)

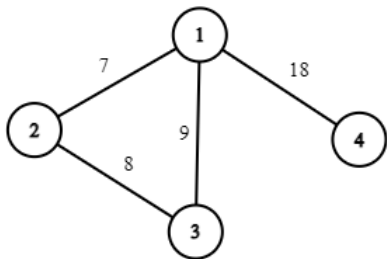
Ma trận bậc là một ma trận đường chéo D sao cho $D_{ii} = d_i$ với d_i là bậc của đỉnh v_i được định nghĩa theo công thức:

$$d_i = \sum_{j=1}^n w_{ij}$$



Một số thuật ngữ

Ví dụ: Cho một đồ thị có $n = 4$ đỉnh như trên và ma trận bậc tương ứng như sau:



$$D = \begin{bmatrix} 34 & 0 & 0 & 0 \\ 0 & 15 & 0 & 0 \\ 0 & 0 & 17 & 0 \\ 0 & 0 & 0 & 18 \end{bmatrix}$$

Nội dung

- 1 Spectral Clustering là gì?
- 2 Các bước tiến hành
 - Dữ liệu đầu vào
 - Mô hình hóa
 - Thuật toán Spectral Clustering
 - Chọn số cụm
- 3 Áp dụng
- 4 Các phiên bản cải tiến



Nội dung

- 1 Spectral Clustering là gì?
- 2 Các bước tiến hành
 - Dữ liệu đầu vào
 - Mô hình hóa
 - Thuật toán Spectral Clustering
 - Chọn số cụm
- 3 Áp dụng
- 4 Các phiên bản cải tiến



Xử lý

Mỗi đỉnh là 1 điểm dữ liệu, giữa 2 đỉnh có trọng số, càng "giống nhau" thì trọng số càng lớn.

→ Similarity graph.



Xử lý

Mỗi đỉnh là 1 điểm dữ liệu, giữa 2 đỉnh có trọng số, càng "giống nhau" thì trọng số càng lớn.

→ Similarity graph.

Công thức để đo độ "giống nhau":

$$s(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$



Các loại similarity graph

The ε -neighborhood graph

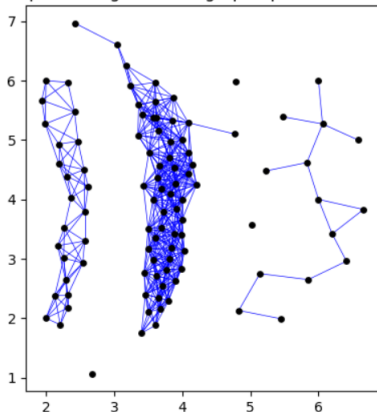
Trong ε -neighborhood graph ta nối cạnh những đỉnh có khoảng cách bé hơn ε bằng những cạnh không trọng số.



Các loại similarity graph

Ví dụ: ε -neighborhood graph xây trên tập dữ liệu tự sinh:

epsilon-neighborhood graph(epsilon = 0.75)



- Khó chọn tham số ε .
- Trong dataset với các vùng không cùng "mật độ", vùng có mật độ thưa hơn sẽ khó kết nối với nhau.



Các loại similarity graph

k-nearest neighbor graphs

Trong k-nearest neighbor graphs, với mỗi đỉnh ta nối nó với k đỉnh gần đỉnh đang xét nhất. Có 2 loại k-nearest neighbor graphs:

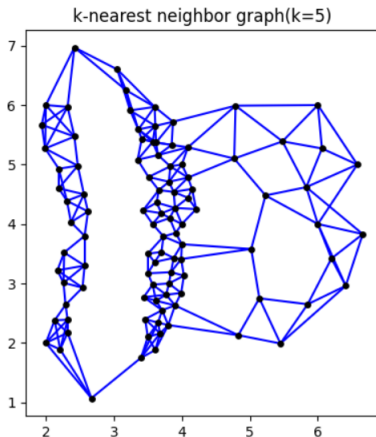
1. *k-nearest neighbor graphs*: Mỗi đỉnh nối cạnh vô hướng tới k đỉnh gần nó nhất.
2. *Mutual k-nearest neighbor graphs*: Chỉ nối cạnh v_i với v_j khi v_j nằm trong k đỉnh gần nhất với v_i và ngược lại.

Cả hai trường hợp đều là đồ thị có trọng số bằng độ giống nhau giữa 2 đỉnh.



Các loại similarity graph

Ví dụ: k-nearest neighbor graph trên tập dữ liệu tự sinh:

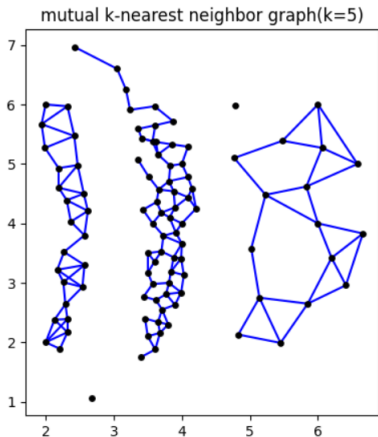


- Có thể nối các vùng có mật độ khác nhau lại với nhau.
- Có thể tách các vùng có mật độ thưa nếu chúng đủ xa.



Các loại similarity graph

Ví dụ: Mutual k-nearest neighbor graph trên tập dữ liệu tự sinh:



- Thường nối các vùng có chung mật độ nhưng không nối các vùng có mật độ khác nhau lại với nhau.



Các loại similarity graph

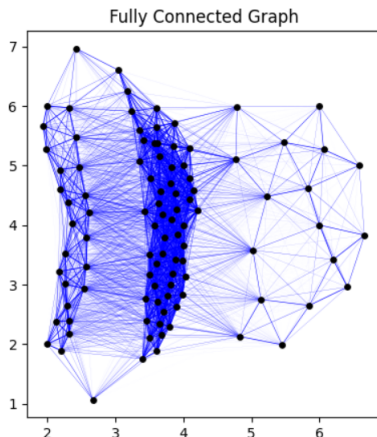
The fully connected graph

Trong fully connected graph, tất cả các cặp cạnh đều được nối với nhau bởi các cặp cạnh vô hướng với trọng số là độ giống nhau giữa 2 đỉnh.



Các loại similarity graph

Ví dụ: Fully connected graph trên tập dữ liệu tự sinh:



- Ma trận đầu ra không phải là một ma trận thưa.



Nội dung

- 1 Spectral Clustering là gì?
- 2 Các bước tiến hành
 - Dữ liệu đầu vào
 - Mô hình hóa
 - Thuật toán Spectral Clustering
 - Chọn số cụm
- 3 Áp dụng
- 4 Các phiên bản cải tiến



Bài toán đồ thị

- **Trực giác:** Phân hoạch similarity graph thành k phần A_1, \dots, A_k sao cho 2 cụm A_i, A_j khác nhau.



Bài toán đồ thị

- **Trực giác:** Phân hoạch similarity graph thành k phần A_1, \dots, A_k sao cho 2 cụm A_i, A_j khác nhau.

→ Đo "độ khác nhau":

$$W(A, B) = \sum_{\substack{v_i \in A \\ v_j \in B}} w_{i,j}$$



Bài toán đồ thị

- **Trực giác:** Phân hoạch similarity graph thành k phần A_1, \dots, A_k sao cho 2 cụm A_i, A_j khác nhau.

→ Đo "độ khác nhau":

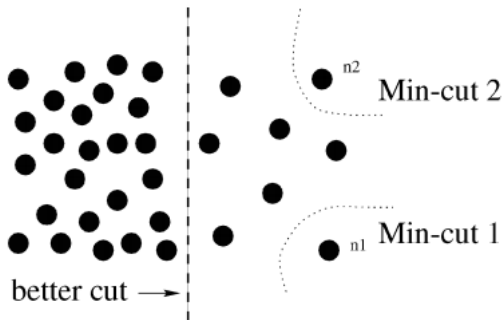
$$W(A, B) = \sum_{\substack{v_i \in A \\ v_j \in B}} w_{i,j}$$

→ Cần $\sum_{i=1}^k W(A_i, \overline{A_i})$ đạt GTNN.



Vấn đề

- Cách làm này thường chia thành từng điểm riêng lẻ [WL93].



Hình: Ví dụ về lát cắt không tối ưu [SM00]



Bài toán đồ thị

Bài toán đồ thị

Cho đồ thị $G = (V, E)$ và k nguyên dương. Tìm phân hoạch A_1, \dots, A_k của V sao cho:

$$\mathcal{P} = \sum_{i=1}^k \frac{W(A_i, \overline{A_i})}{|A_i|} \text{ đạt GTNN}$$



Một vài kết quả lý thuyết

Định nghĩa

Cho đồ thị $G = (V, E)$ với D, W lần lượt là ma trận bậc và ma trận kề. Khi đó (unnormalized) Laplacian matrix L được định nghĩa là:

$$L = D - W$$



Một vài kết quả lý thuyết

Mệnh đề

Với mọi vector $v = (v_1, \dots, v_n)^T \in \mathbb{R}^n$ ta có:

$$v^T L v = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (v_i - v_j)^2$$



Một vài kết quả lý thuyết

$$\begin{aligned} \mathbf{v}^T D \mathbf{v} - \mathbf{v}^T W \mathbf{v} &= \sum_{i=1}^n v_i^2 d_{i,i} - \sum_{i,j=1}^n v_i v_j w_{i,j} \\ &= \frac{1}{2} \left(\sum_{i=1}^n v_i^2 d_{i,i} - \sum_{i,j=1}^n 2v_i v_j w_{i,j} + \sum_{j=1}^n v_j^2 d_{j,j} \right) \\ &= \frac{1}{2} \left(\sum_{i=1}^n v_i^2 \sum_{j=1}^n w_{i,j} - \sum_{i,j=1}^n 2v_i v_j w_{i,j} + \sum_{j=1}^n v_j^2 \sum_{i=1}^n w_{j,i} \right) \\ &= \frac{1}{2} \sum_{i,j=1}^n w_{ij} (v_i - v_j)^2. \end{aligned}$$



Một vài kết quả lý thuyết

Hệ quả

L có n giá trị riêng thực không âm $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.



Bài toán tối ưu

Mệnh đề

Ta có đẳng thức:

$$\frac{W(A_i, \overline{A_i})}{|A_i|} = h_i^T L h_i,$$

trong đó $h_i = (h_{1,i}, \dots, h_{n,i})^T$ và $h_{j,i} = \begin{cases} \frac{1}{\sqrt{|A_i|}}, & v_j \in A_i \\ 0, & \text{TH khác} \end{cases}$.

Hơn nữa ma trận H nhận h_i làm cột thỏa $H^T H = I_k$.



Bài toán tối ưu

- Ta có biến đổi:

$$\begin{aligned}
 h_i^T L h_i &= \frac{1}{2} \sum_{j,k=1}^n w_{j,k} (h_{j,i} - h_{k,i})^2 \\
 &= \frac{1}{2} \sum_{\substack{j \in A_i \\ k \in \bar{A}_i}} w_{j,k} \left(\frac{1}{\sqrt{|A_i|}} \right)^2 + \frac{1}{2} \sum_{\substack{j \in A_i \\ k \in \bar{A}_i}} w_{j,k} \left(-\frac{1}{\sqrt{|A_i|}} \right)^2 \\
 &= \frac{W(A_i, \bar{A}_i)}{|A_i|}
 \end{aligned}$$

- $h_i \cdot h_j = 0$ do các A_i đôi một không giao và
 $h_i \cdot h_i = \sum_{j \in A_i} \frac{1}{|A_i|} = 1$



Bài toán tối ưu

Mệnh đề

Ta có đẳng thức

$$\frac{W(A_i, \bar{A}_i)}{|A_i|} = h_i^T L h_i = [H^T L H]_{i,i}$$

Từ đó suy ra $\mathcal{P} = \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{|A_i|} = \text{trace}(H^T L H)$.



Bài toán tối ưu

Mệnh đề

Ta có đẳng thức

$$\frac{W(A_i, \bar{A}_i)}{|A_i|} = h_i^T L h_i = [H^T L H]_{i,i}$$

Từ đó suy ra $\mathcal{P} = \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{|A_i|} = \text{trace}(H^T L H)$.

$$h_i^T L h_i = \sum_{j,k=1}^n h_{j,i} h_{k,i} [L]_{j,k} = \sum_{j,k=1}^n [H^T]_{i,j} [L]_{j,k} [H]_{k,i} = [H^T L H]_{i,i}.$$



Bài toán tối ưu

Bài toán tối ưu 1

Tìm $\min_{H \in \mathbb{R}^{n \times k}} \text{trace}(H^T L H)$ với $H^T H = I_k$.

Định lý (Hệ quả 4.3.39, [HJ13])

Cho $L \in \mathbb{R}^{n \times n}$ có n giá trị riêng $0 \leq \lambda_1 \leq \dots \leq \lambda_n$ và $1 \leq k \leq n$.
Khi đó $\min_{H^T H = I_k} \text{trace}(H^T L H) = \lambda_1 + \dots + \lambda_k$.



Nội dung

- 1 Spectral Clustering là gì?
- 2 Các bước tiến hành
 - Dữ liệu đầu vào
 - Mô hình hóa
 - Thuật toán Spectral Clustering
 - Chọn số cụm
- 3 Áp dụng
- 4 Các phiên bản cải tiến



Thuật toán

Thuật toán Spectral clustering

Cho dữ liệu đầu vào là similarity matrix $S \in \mathbb{R}^{n \times n}$, $k \in \mathbb{N}$ là số cụm.

- 1 Tính (unnormalized) Laplacian matrix L .
- 2 Tính k vector riêng u_1, \dots, u_k ứng với k giá trị riêng nhỏ nhất của L . Gọi $U \in \mathbb{R}^{n \times k}$ là ma trận gồm các cột là các u_i .
- 3 Dùng k -means để phân các y_i thành C_1, \dots, C_k , với y_i là vector dòng của U .

Dữ liệu đầu ra là A_1, \dots, A_k với $A_i = \{v_j | y_j \in C_i\}$.



Nội dung

- 1 Spectral Clustering là gì?
- 2 Các bước tiến hành
 - Dữ liệu đầu vào
 - Mô hình hóa
 - Thuật toán Spectral Clustering
 - Chọn số cụm
- 3 Áp dụng
- 4 Các phiên bản cải tiến



Chọn số cụm

- Là vấn đề chung của bài toán clustering.
- Nhiều phương pháp được phát triển cho nhiều loại mô hình khác nhau.



Chọn số cụm

- Là vấn đề chung của bài toán clustering.
 - Nhiều phương pháp được phát triển cho nhiều loại mô hình khác nhau.
- Eigengap heuristic: Chọn k sao cho $\lambda_{k+1} - \lambda_k$ tương đối lớn.



Motivation

Định lý

Cho G là đồ thị vô hướng và với trọng số không âm. Khi đó bội k của giá trị riêng 0 của L bằng số thành phần liên thông của G .

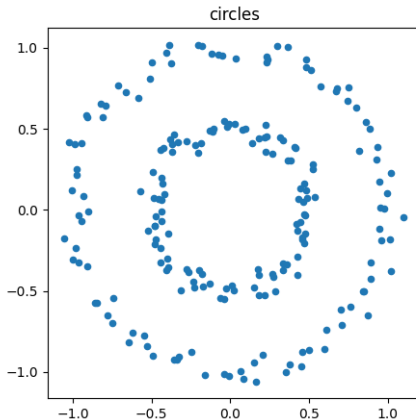


Nội dung

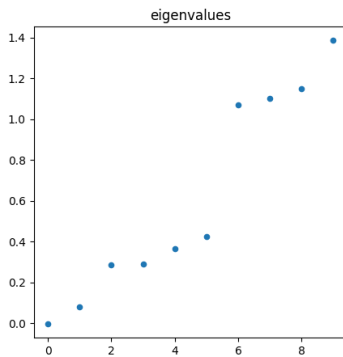
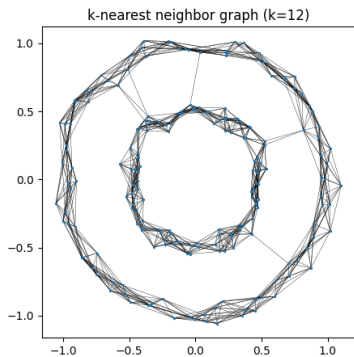
- 1 Spectral Clustering là gì?
- 2 Các bước tiến hành
 - Dữ liệu đầu vào
 - Mô hình hóa
 - Thuật toán Spectral Clustering
 - Chọn số cụm
- 3 **Áp dụng**
- 4 Các phiên bản cải tiến



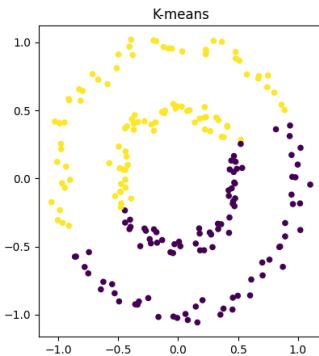
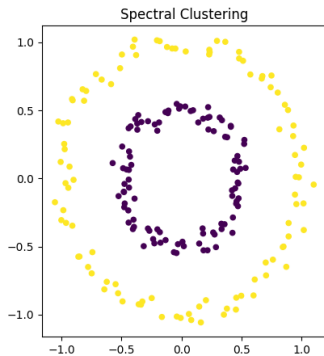
Áp dụng



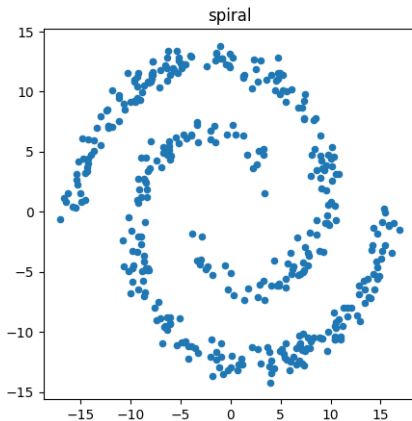
Áp dụng



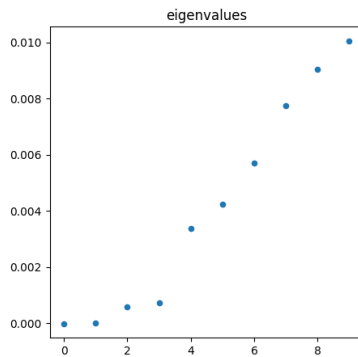
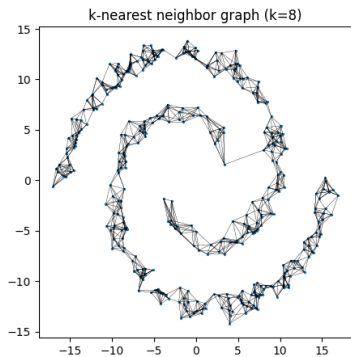
Áp dụng



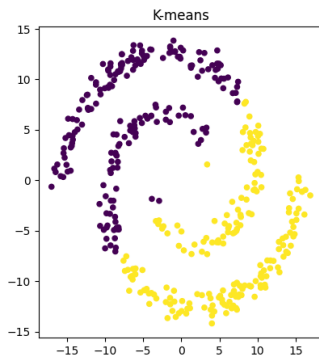
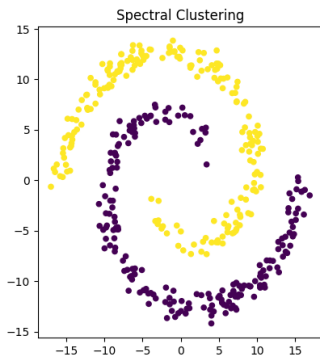
Áp dụng



Áp dụng



Áp dụng



Ưu điểm

- K-means đưa ra một giả sử mạnh, đó là các cụm là các khối tròn. Một số trường hợp thì spectral clustering sẽ phân hiệu quả hơn.
- Không cần chạy nhiều lần như k-means.



Một số vấn đề

- Chọn các siêu tham số.
- Nếu ma trận Laplacian dense thì tính các vector riêng sẽ tốn thời gian.



Nội dung

- 1 Spectral Clustering là gì?
- 2 Các bước tiến hành
 - Dữ liệu đầu vào
 - Mô hình hóa
 - Thuật toán Spectral Clustering
 - Chọn số cụm
- 3 Áp dụng
- 4 Các phiên bản cải tiến



Normalized symmetric Laplacian

Bằng cách tiếp cận hoàn toàn tương tự như trong phiên bản unnormalized Laplacian matrix nhưng thay $|A_i|$ bằng

$$\text{vol}(A) = \sum_{i: v_i \in A} d_i$$

Thì ta thu được bài toán tối ưu

Bài toán tối ưu 2

Tìm $\min_{K \in \mathbb{R}^{n \times k}} \text{trace}(K^T D^{-\frac{1}{2}} L D^{\frac{1}{2}} K)$ với $K^T K = I_k$.

Ma trận $L_{\text{sym}} = D^{-\frac{1}{2}} L D^{\frac{1}{2}}$ được gọi là normalized Laplacian matrix.



Random walk Laplacian

- Một quá trình đi từ 1 đỉnh v_i tới đỉnh v_j một cách ngẫu nhiên, với xác suất là p_{ij} tỉ lệ với w_{ij} , tức $p_{ij} = \frac{w_{ij}}{d_i}$.
- Ta muốn tìm một phân hoạch sao cho quá trình này ở càng lâu trong mỗi cụm càng tốt.



Random walk Laplacian

- Một quá trình đi từ 1 đỉnh v_i tới đỉnh v_j một cách ngẫu nhiên, với xác suất là p_{ij} tỉ lệ với w_{ij} , tức $p_{ij} = \frac{w_{ij}}{d_i}$.
- Ta muốn tìm một phân hoạch sao cho quá trình này ở càng lâu trong mỗi cụm càng tốt.

→ $P = D^{-1}W$ là ma trận lưu xác suất và $L_{rw} = I - P$ cũng được gọi là normalized Laplacian matrix.



Random walk Laplacian

Định lý

Ta có đẳng thức

$$\frac{W(A, \bar{A})}{\text{vol}(A)} = P(A|\bar{A}) + P(\bar{A}|A).$$

Trong đó $P(A|B)$ là xác suất để bước qua đỉnh thuộc A biết đang đứng ở B , với A, B không giao nhau.



Nhận xét

Khi áp dụng thuật toán, ta muốn điểm giữa 2 cụm khác nhau và điểm trong cùng cụm giống nhau.



Nhận xét

Khi áp dụng thuật toán, ta muốn điểm giữa 2 cụm khác nhau và điểm trong cùng cụm giống nhau.

- Unnormalized Laplacian: chỉ thực hiện mục tiêu 1, do $|A|$ không đánh giá được mục tiêu 2.
- Normalized Laplacian: thực hiện cả 2 mục tiêu do

$$W(A, A) = W(A, V) - W(A, \bar{A}) = \text{vol}(A) - W(A, \bar{A}).$$



References

- [HJ13] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2013.
- [SM00] Jianbo Shi and Jitendra Malik. “Normalized cut and image segmentation”. In: *IEEE Transactions On Pattern Analysis And Machine Intelligence* 22.8 (2000).
- [WL93] Zhenyu Wu and Richard Leahy. “An optimal graph theoretic approach to data clustering: Theory and its Application to image segmentation”. In: *IEEE Transactions On Pattern Analysis And Machine Intelligence* 15.11 (1993).

