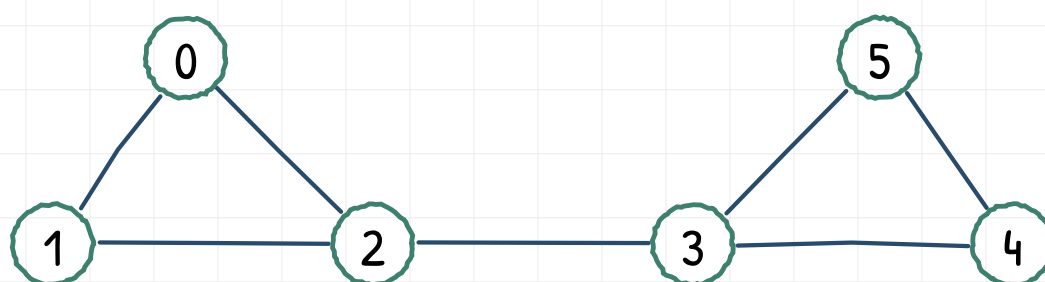


Dấu của MỘT vector chia đúng hai cụm

Hai cụm $\{0,1,2\}$ và $\{3,4,5\}$, nối bởi ĐÚNG một cạnh 2–3:

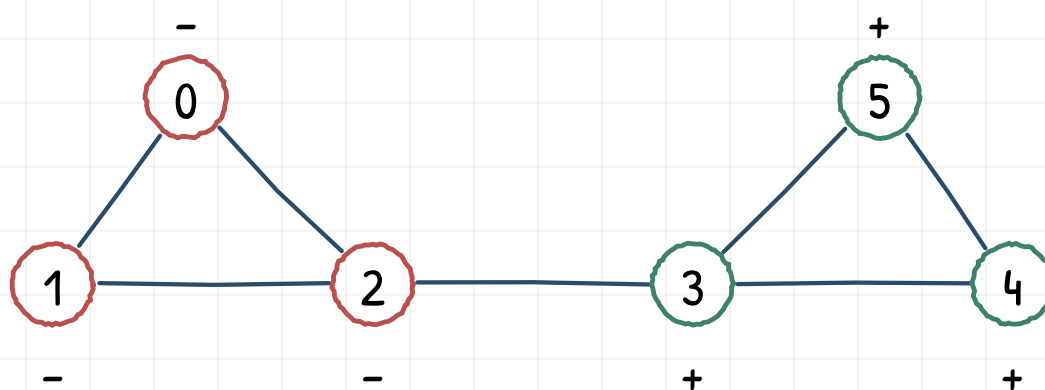


$L = \text{np.diag}(A.\text{sum}(1)) - A \leftarrow$ toàn bộ định nghĩa, một dòng

trị riêng: 0, 0.4384, 3, 3, 3, 4.5616

eigenvector thứ hai $u_2 = (-0.46, -0.46, -0.26, 0.26, 0.46, 0.46)$

dấu ($u_2 > 0$) = [0 0 0 1 1 1]



Không nói số cụm, không hạt giống, không thuật toán phân cụm.

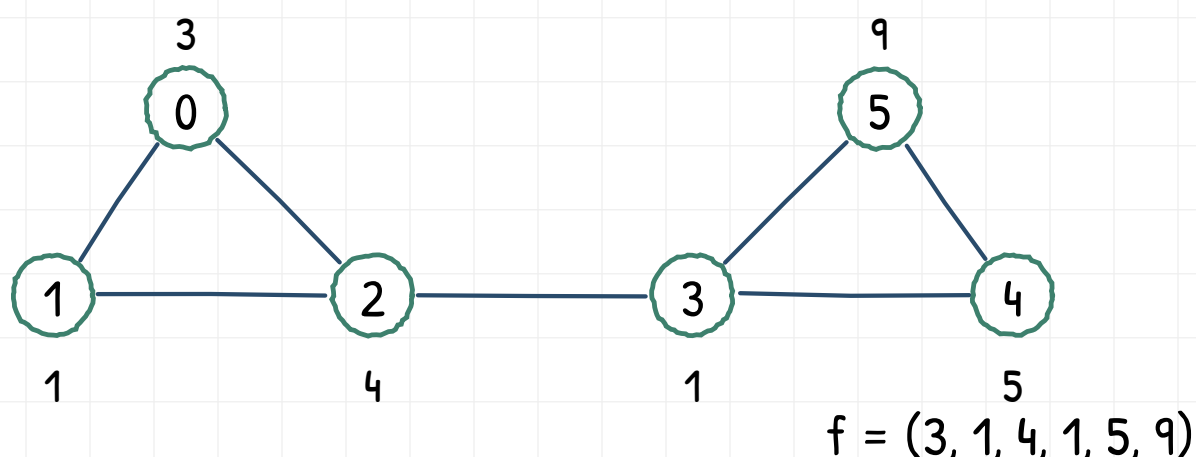
Chỉ lấy eigenvector thứ hai của một ma trận, và DẤU của nó chia đúng hai cụm. Đó là toàn bộ bài này.

Bài 8 đo "gần" bằng lan truyền ngẫu nhiên; bài 10 đo bằng PHỔ.

$L = D - A$ là MÁY ĐO CHÊNH LỆCH

Đừng nhìn ma trận: cho nó ăn một vector. Gán mỗi nút một số $\rightarrow f$.

$$(Lf)[i] = \sum_{j \text{ kề } i} (f[i] - f[j])$$



$$L \cdot f = (1, -5, 7, -15, 0, 12)$$

Nút 4: $f = 5$, hàng xóm 3 ($f = 1$) và 5 ($f = 9$):

$$(5 - 1) + (5 - 9) = 0: \text{CÂN BẰNG với hàng xóm.}$$

Nhân A: thu gom giá trị hàng xóm. Nhân L: đo mình lệch bao nhiêu với hàng xóm.

Bản rời rạc của toán tử Laplace trong giải tích: lý do của cái tên.

Thêm một bước: $f^T L f$

$$f^T L f = \frac{1}{2} \cdot \sum_{i,j \text{ kề}} (f[i] - f[j])^2 = 119$$

Tổng qua mọi cặp có cạnh; $\frac{1}{2}$ vì mỗi cạnh đếm hai lần trong tổng đôi.

$f^T L f$ nhỏ $\Leftrightarrow$ nút nối nhau có giá trị gần nhau $\Leftrightarrow f$ TRON trên đồ thị.

Cấu trúc đồ thị $\rightarrow$ hàm năng lượng

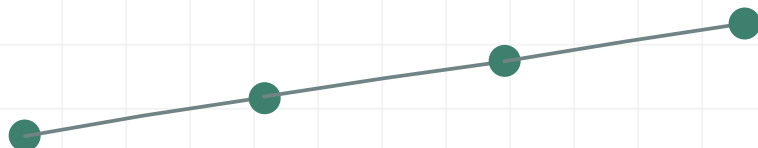
Fiedler, phân cụm phổ, vẽ đồ thị, over-smoothing trong mạng đồ thị: tất cả chỉ là CỰC TIỂU HOÁ hoặc KIỂM SOÁT đúng dạng toàn phương này. Không có ý tưởng thứ hai.

Hệ quả miễn phí

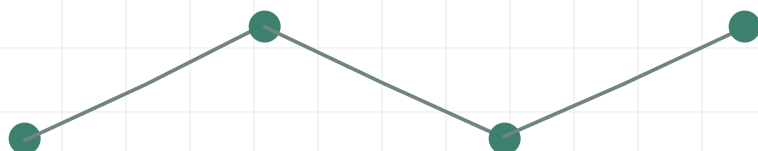
$f^T L f$ là tổng bình phương $\Rightarrow$ luôn $\geq 0 \Rightarrow L$ nửa xác định dương
 $\Rightarrow$ mọi trị riêng ≥ 0 . Chứng minh xong mà không tính trị riêng nào.
 $L \cdot \mathbf{1} = 0$: mỗi nút lệch 0 so với hàng xóm khi mọi giá trị bằng nhau.



trơn: $f^T L f = 0$



dốc đều: $f^T L f = 3$



gồ ghề: $f^T L f = 27$

(đồ thị đường 4 nút: năng lượng = tổng bình phương chênh trên cạnh)

Dạng L từ ma trận liên thuộc

B cỡ $n \times m$: mỗi CỘT là một cạnh, -1 ở nút đầu, $+1$ ở nút cuối.

$B^T f$ = độ chênh trên từng cạnh: $(B^T f)_e = f_{\text{cuối}} - f_{\text{đầu}}$.

cạnh	(0,1)	(0,2)	(1,2)	(3,4)	(3,5)	(4,5)
chênh	-2	1	3	4	8	4

(2,3): -3 . Bình phương: $4 + 1 + 9 + 16 + 64 + 16 + 9 = 119 \checkmark$

$$f^T L f = \|B^T f\|^2 = f^T (B B^T) f$$

$(B B^T)_{ii}$ = hàng i với chính nó = BẬC của i .

$(B B^T)_{ij}$ = -1 nếu i, j kề (hai dấu ngược nhau), 0 nếu không. $\Rightarrow B B^T = D - A$.

Hai đầu cầu: hàng 2 $\cdot$ hàng 3 = $-1 = L[2,3]$; hàng 2 $\cdot$ hàng 2 = 3 = bậc nút 2.

Bốn hệ quả

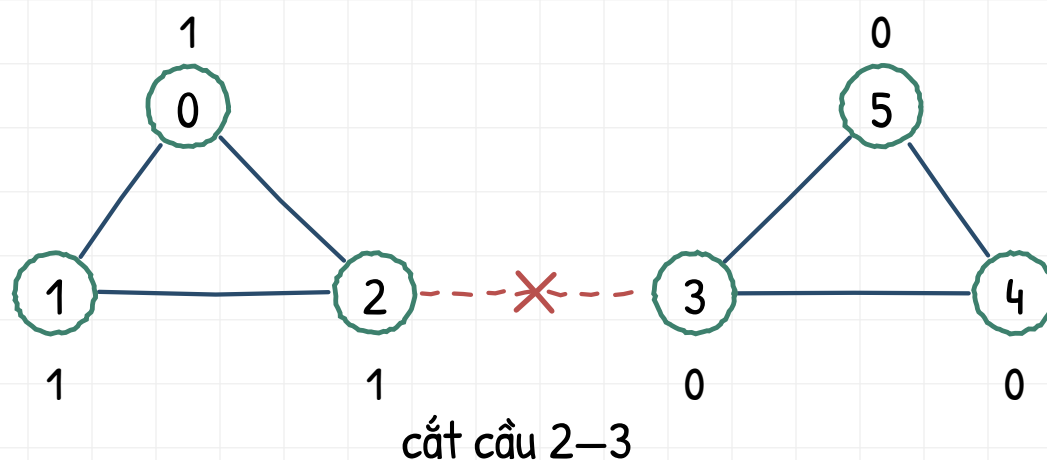
- $\|B^T f\|^2 \geq 0$: nửa xác định dương, không cần hệ số $\frac{1}{2}$.
- Hướng cạnh vô can: lật cạnh = đổi dấu cột b , mà $b b^T = (-b)(-b)^T$.
- Mỗi cột B tổng 0 $\Rightarrow B^T 1 = 0 \Rightarrow L 1 = B(B^T 1) = 0$.
- Đếm mảnh bằng hạng: số trị riêng 0 = $n - \text{rank}(B)$.
Liên thông 6 nút: $\text{rank}(B) = 5 \Rightarrow 1$. Cắt cầu: $\text{rank} = 4 \Rightarrow 2$.

Dirichlet energy (VMLS): tổng bình phương độ chênh trên cạnh.

Trị riêng 0 ĐẾM số mảnh

$f^T L f = 0 \Leftrightarrow$ mọi số hạng $= 0 \Leftrightarrow f[i] = f[j]$ trên MỌI cạnh
 $\Leftrightarrow f$ hằng số trên mỗi mảnh liên thông.

Cơ sở của nhân L : vector chỉ báo từng mảnh (1 trên mảnh, 0 ngoài).



trị riêng của L sau khi cắt: 0, 0, 3, 3, 3, 3

Bộ của trị riêng 0 = số mảnh liên thông.

Câu hỏi TỔ HỢP rời rạc ("mấy mảnh?") có câu trả lời LIÊN TỤC ("số chiều của một không gian con"). Chuyển pha nhận thức của cả bài: vector chỉ báo cụm CHÍNH LÀ một eigenvector.

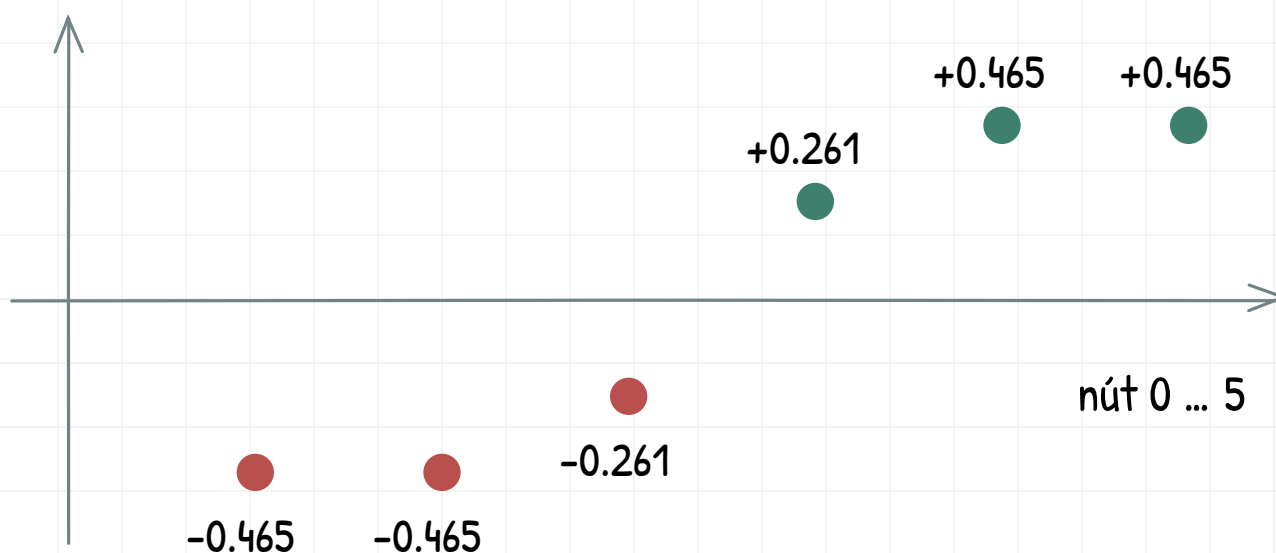
networkx: `number_connected_components = 2 = số trị riêng  $< 1e-10$ .`

Fiedler λ_2 : vì sao trị riêng THỨ HAI cắt

1. Trị riêng = cực trị của thương Rayleigh $R(f) = f^T L f / f^T f$ (Rayleigh–Ritz).
2. Với L , tử số $= \sum (f[i] - f[j])^2$: cạnh nào bị "cắt qua" thì bị tính tiền.
Tử số Rayleigh là CHI PHÍ CẮT mềm.
3. Cực tiểu tuyệt đối: f hằng, $R = 0$: "không cắt gì". Vô dụng.
Ép $f \perp 1$ (tổng = 0) $\Rightarrow$ BẮT BUỘC có nút dương, nút âm $\Rightarrow$ phải cắt thật.
4. Cực tiểu dưới ràng buộc đó $= \lambda_2$; vector đạt nó = Fiedler vector u_2 .

λ_2 nhỏ $\Leftrightarrow$ có cách chia đôi rất rẻ $\Leftrightarrow$ đồ thị gần đứt.

$\lambda_2 = 0 \Leftrightarrow$ đứt thật (tờ 5). u_2 = bản LÀM MỀM của vector chỉ báo cụm.



quan sát trong u_2

3 nút âm, 3 nút dương

nút 2, 3 có giá trị nhỏ nhất

tổng = 0

nghĩa

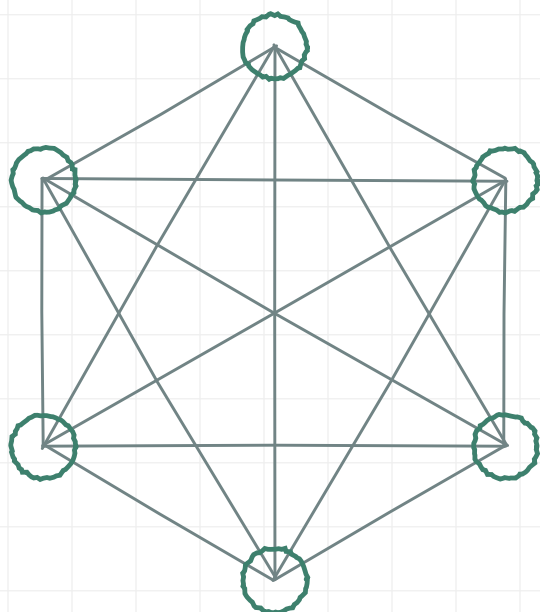
phép cắt $\{0,1,2\} \mid \{3,4,5\}$

hai đầu cầu: sát ranh giới

chính là ràng buộc $u_2 \perp 1$

Bầy: λ_2 lớn KHÔNG phải "cụm tốt"

Đồ thị đầy đủ K_6 : trị riêng của $L = 0, 6, 6, 6, 6, 6$.



$\lambda_2 = 6$: lớn gấp ~ 14 lần
đồ thị hai cụm (0.4384).

Nhưng K_6 KHÔNG có cụm nào.

λ_2 lớn = khó cắt = không có cụm. λ_2 nhỏ = có cụm rõ.

u_2 LUÔN trả về một phép cắt, kể cả khi không có cụm nào: nó buộc phải.
Phải nhìn λ_2 để biết phép cắt đó có ý nghĩa không. Chỉ nhìn u_2 là bạn sẽ
tìm ra "cụm" trong nhiều.

Chạy phân cụm thì bao giờ cũng ra cụm; câu hỏi thật là cụm có THẬT không (môn 6).

Chế độ hỏng thứ hai

Đồ thị gần đầy đủ: mọi trị riêng phụ xấp xỉ bằng nhau, không có KHE HỞ phổ
 $\Rightarrow$ phương pháp phổ không phát hiện cụm nào, kể cả cụm yếu có thật.
 K_6 : năm trị riêng phụ bằng nhau chẵn chẵn.

Bài 8 và bài 10: hai câu hỏi khác nhau

	bài 8: Markov	bài 10: Laplacian
ma trận	$P = D^{-1}A$	$L = D - A$
đối xứng?	KHÔNG	CÓ $\Rightarrow$ định lý phổ
eigenvector	trội ($\lambda = 1$)	thứ hai từ đáy (λ_2)
trả lời	nút nào QUAN TRỌNG?	cắt ở ĐÂU?
kết quả	một thứ hạng	một phép chia

$L_{rw} = D^{-1}L = I - D^{-1}A = I - P$

P và L_{rw} có CHUNG eigenvector, trị riêng đảo: λ của P $\leftrightarrow$ $1 - \lambda$ của L_{rw} .
Trội của P ($\lambda = 1$) $\leftrightarrow$ đáy của L_{rw} (0). Kiểm bằng số: khớp.

Phân phối dừng của bước ngẫu nhiên và eigenvector đáy: MỘT vật, hai phía.

Bẫy: P và L_{rw} KHÔNG đối xứng

Phải dùng eigvals, không dùng eigh. Không đối xứng $\Rightarrow$ không có định lý phổ: không đảm bảo cơ sở trực chuẩn, không đảm bảo trị riêng thực.

$L_{sym} = D^{(-1/2)} \cdot L \cdot D^{(-1/2)}$

Đồng dạng với L_{rw} (cùng phổ) nhưng ĐỐI XỨNG (có định lý phổ).
Lý do thật không phải "ổn định số": cần định lý phổ thì phải có ma trận đối xứng.

Nối vào AI: ba chỗ gặp Laplacian

(a) Phân cụm phổ

1. dựng L (hoặc L_{sym})
 2. lấy k eigenvector ứng k trị riêng NHỎ nhất
 3. nút $i \rightarrow$ hàng i của ma trận eigenvector (điểm trong $\mathbb{R}^k$)
 4. k -means.
- Bước 3 = nhúng đồ thị vào không gian Euclid: graph embedding thuần khiết nhất.
 Phải BỎ u_1 (vector hằng): như PCA trừ trung bình, trục mọi mẫu giống nhau vô ích.

(b) Vẽ đồ thị đẹp

Mỗi cạnh là lò xo; năng lượng $\sum w \|p(v_i) - p(v_j)\|^2 =$ đúng dạng toàn phương tờ 2.
 Lấy u_2 LIÊN TỤC để vẽ, lấy DẤU của u_2 để cắt: cùng một thuật toán.

(c) Mạng nơ-ron đồ thị

Bộ lọc phổ tổng quát: cần phân rã trị riêng (đắt) và mỗi nút phụ thuộc cả đồ thị.
 Ép bộ lọc = đa thức bậc K của L : $\sum \theta_k L^k$, hết phân rã trị riêng.
 $L^k[i, j] = 0$ khi khoảng cách $> k \Rightarrow$ đa thức bậc $K \equiv$ chỉ nhìn hàng xóm trong K bước.
 Spectral và spatial hoá ra là MỘT. Tính định xứ rơi ra miễn phí từ ràng buộc.

Bài 8 cần lại: over-smoothing.

Bỏ phi tuyến, chồng L lớp = lũy thừa một ma trận lan truyền = power iteration:
 hội tụ về eigenvector trội (chỉ chứa bậc của nút), mất sạch đặc trưng đầu vào.
 PageRank MUỐN quên điều kiện ban đầu; mạng đồ thị thì điều kiện ban đầu là DỮ LIỆU.
 Điều kiện định lý: liên thông và KHÔNG hai phía. Đồ thị user-item thì hai phía.

Một kết quả toán không tự nó tốt hay xấu: tùy bạn muốn gì.

Đối chứng: cầu thật và nhiễu

(a) Hai cụm 10 nút ĐẦY ĐỦ, nối bởi k cạnh cầu:

k	1	2	5	10
λ_2	0.1690	0.3431	0.9010	2.0000
dấu u_2 đúng?	✓	✓	✓	✓

λ_2 TĂNG theo k : dính càng chặt càng khó cắt. Nhưng phép cắt vẫn đúng ở cả bốn k . Muốn làm hỏng: tăng k nữa, hay làm hai cụm THỪA đi? Thử cả hai.

(b) Erdős-Rényi 20 nút, CÙNG mật độ với (a) $k = 5$ ($p = 0.5$):

	seed 0	1	2	3
λ_2 (ER)	3.4266	4.3562	4.4954	4.5645

Hai cụm thật $k = 5$: $\lambda_2 = 0.9010$. ER lớn gấp $\sim 4-5$ lần.

u_2 VẪN trả một phép chia ở cả bốn seed: nó buộc phải.

λ_2 lớn gấp nhiều lần là dấu hiệu DUY NHẤT cho biết phép chia vô nghĩa. Phương pháp luôn trả đáp án; cần một con số KHÁC để biết đáp án có nghĩa. Không có đối chứng: "phát hiện cấu trúc" trong nhiễu, bảng trông y như thật.

(c) $k = 2$: k-means trên 3 eigenvector đáy và trên hàng ma trận kẻ đều 100 %.

Bài toán dễ thì phương pháp phức tạp không hơn baseline tầm thường.

Việc thật: tìm cấu hình (cụm thừa, lồng nhau) mà phổ thắng.

Tự kiểm "tận cùng"

1. Chứng minh $f^T L f = \frac{1}{2} \sum (f[i] - f[j])^2$ bằng khai triển trực tiếp.
2. Từ câu 1, L nửa xác định dương mà không tính trị riêng nào.
3. Vì sao $L \cdot 1 = 0$ với mọi đồ thị? Hai cách: $(Lf)[i]$ và cấu trúc $D - A$.
4. Đồ thị 5 mảnh: bội trị riêng 0? Một cơ sở của nhân L ?
5. Vì sao ép $f \perp 1$ khi tìm λ_2 ? Bỏ đi thì cực tiểu là gì?
6. $\lambda_2 = 6$ trên K_6 , 0.4384 trên hai cụm: đồ thị nào dễ cắt? Dùng u_2 trên K_6 ?
7. Vì sao nút 2, 3 có $|u_2|$ nhỏ nhất? Thông tin đó dùng vào việc gì?
8. $L_{rw} = I - P$ chung eigenvector với P . Vì sao vẫn cần L_{sym} ? (định lý phổ)
9. PageRank MUỐN hội tụ về eigenvector trội, mạng đồ thị SỢ điều đó.
Vì sao cùng một kết quả vừa là tính năng vừa là lỗi?

Câu 1 và 9 là hai ranh giới.

Cheeger: biện minh cho tờ 6 (đồ thị d-đều)

$\kappa/2 \leq h(G) \leq \sqrt{2d\kappa}$, $\kappa = \lambda_2(L)$, $h = \min(\text{cạnh rời } S) / |S|$ (Zhao Thm 3.2.13).
 Q_4 ($d = 4$): $\kappa = 2$, $h = 1 = \kappa/2$: cận dưới CHẠM. Chu trình 20 đỉnh ($d = 2$): $\kappa = 0.0979$,
 $h = 0.2 \gg \kappa/2 = 0.049$, cận trên $\sqrt{2d\kappa} = 0.63$: h cỡ $\sqrt{\kappa}$, cần đủ căn bậc hai.
 λ_2 lớn $\Rightarrow$ không cổ chai (tuyến tính). λ_2 nhỏ $\Rightarrow$ có cắt tốt, nhưng chỉ tới căn bậc hai.
 Đồ thị bậc lệch (như đồ thị memory): cần Laplacian chuẩn hoá (Chung, Spielman).

Đáp nhanh

- $f^T L f = \sum_i d_i f_i^2 - \sum_{ij} A_{ij} f_i f_j = \frac{1}{2} \sum_{ij} A_{ij} (f_i^2 + f_j^2 - 2f_i f_j) = \frac{1}{2} \sum A_{ij} (f_i - f_j)^2$.
- Tổng bình phương ≥ 0 với mọi $f \Rightarrow$ mọi trị riêng ≥ 0 (lấy $f = \text{eigenvector}$).
- $(L1)[i] = \sum (1 - 1) = 0$; hoặc hàng i của $D - A$: bậc - (số 1 trên hàng) = 0.
- Bộ 5; cơ sở: 5 vector chỉ báo (1 trên mảnh k , 0 nơi khác).
- Bỏ ràng buộc $\Rightarrow f$ hằng, $R = 0$: "không cắt" (nghiệm tầm thường).
 $f \perp 1$ bắt buộc có hai dấu $\Rightarrow$ phải chia đôi thật.
- Hai cụm dễ cắt hơn nhiều. Trên K_6 , u_2 vẫn cho một phép cắt nhưng vô nghĩa.
- Chúng là hai đầu cầu, sát đường cắt. Cho biết nút nào ở RANH GIỚI.
- P , L_{rw} không đối xứng: không có định lý phổ (trục chuẩn, trị riêng thực).
 L_{sym} đối xứng, đồng dạng $L_{rw} \Rightarrow$ cùng phổ mà có đủ định lý.
- Cùng định lý hội tụ về eigenvector trội. PageRank cần quên điểm xuất phát;
GNN thì điểm xuất phát là dữ liệu: quên là mất đặc trưng (over-smoothing).

Nguồn

Ma & Tang, Deep Learning on Graphs §2.4-2.5 (tr.26-32), §5.3, §14.2.
 Gallier & Quaintance ch.20 (tr.701-721). VMLS §7.3 (ma trận liên thuộc).
 Zhao, Graph Theory and Additive Combinatorics, Thm 3.2.13 tr.105 (Cheeger).