# CS630 Homework 5

Qilin Ye

April 6, 2025

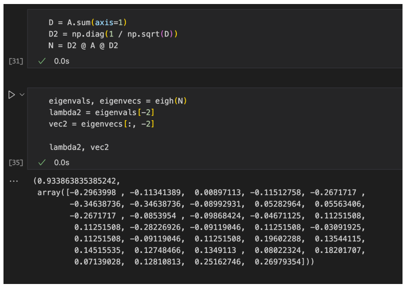
```
import networkx as nx
import numpy as np
from scipy.linalg import eigh

G = nx.read_gml("karate.gml", label="id")
mapping = {node: node - 1 for node in G.nodes()}
G = nx.relabel_nodes(G, mapping)
n = G.number_of_nodes()



A = np.zeros((n, n))
for u, v in G.edges():
    A[u, v] = 1
    A[v, u] = 1
for i in range(n):
    A[i, i] = A[i, :].sum()


D = A.sum(axis=1)
D2 = np.diag(1 / np.sqrt(D))
N = D2 @ A @ D2


eigenvals, eigenvecs = eigh(N)
lambda2 = eigenvals[-2]
vec2 = eigenvecs[:, -2]

lambda2, vec2


def conductance(cut, vol, A):
    Sc = ~cut
    cut_weight = 0
    for i in range(len(cut)):
        if cut[i]:
            for j in range(len(cut)):
                if Sc[j]:
                    cut_weight += A[i, j]
    vol_S = vol[cut].sum()
    vol_Sc = vol[Sc].sum()
    if vol_S == 0 or vol_Sc == 0:
```

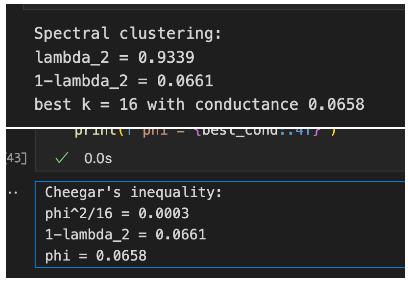
```
        return np.inf
    return cut_weight / min(vol_S, vol_Sc)

order = np.argsort(vec2)
vol = A.sum(axis=1)
best_cond = np.inf
best_cut = None
best_k = None
for k in range(1, n):
    S = np.zeros(n, dtype=bool)
    S[order[:k]] = True
    c = conductance(S, vol, A)
    if c < best_cond:
        best_cond = c
        best_cut = S.copy()
        best_k = k
print("Spectral clustering:")
print(f"lambda_2 = {lambda2:.4f}")
print(f"1-lambda_2 = {1 - lambda2:.4f}")
print(f"best k = {best_k} with conductance {best_cond:.4f}")




# cheegar's inequality
# phi^2/16 vs 1-lambda 2 vs phi
print("Cheegar's inequality:")
print(f"phi^2/16 = {best_cond**2 / 16:.4f}")
print(f"1-lambda_2 = {1 - lambda2:.4f}")
print(f"phi = {best_cond:.4f}")


random_conds = []
for trial in range(10):
    while True:
        rand_assign = np.random.randint(0, 2, size=n).astype(bool)
        if rand_assign.sum() > 0 and (n - rand_assign.sum()) > 0:
            break
    random_conds.append(conductance_custom(rand_assign, A, m))

print("Baseline random cut conductance:")
for i, rc in enumerate(random_conds):
    print(f"\ttrial {i+1}: {rc:.4f}")

avg_random = np.mean(random_conds)
print(f"\taverage = {avg_random:.4f}")
```

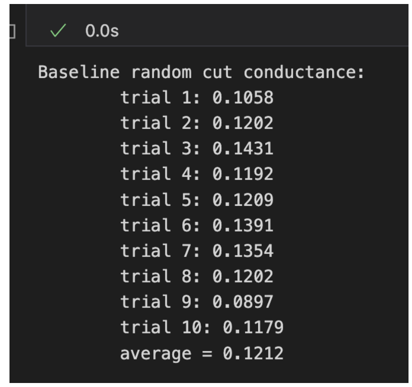
*Solution to problem 2.* First suppose $G$ has $k$ connected components. For a $n = |V|$-dimensional vector $v$, consider the quadratic term

$$v^T Lv = \frac{1}{2} \sum_{(i,j)\in E} \left( \frac{v(i)}{\sqrt{d(i)}} - \frac{v(j)}{\sqrt{d(j)}} \right)^2 .$$

Therefore, if $v$ is an eigenvector of $L$ with eigenvalue $0$ we have $v^T Lv = 0$, which forces each edge $(i,j) \in E$




to have $v(i)/\sqrt{d(i)} = v(j)/\sqrt{d(j)}$. This shows that $v(i)/\sqrt{d(i)}$ is constant on each connected component. This implies that every connected component contributes one independent eigenvector with eigenvalue $0$. Thus the multiplicity of eigenvalue $0$ of $L$ is precisely $k$.

Conversely, suppose $N$ has $k$ eigenvalues equal to $0$. The same equality $v(i)/\sqrt{d(i)} = v(j)/\sqrt{d(j)}$ must hold. Thus, any such $v$ must be constant on every connected component. Since we have $k$ linearly independent eigenvectors of $L$ with eigenvalues $0$, we must have precisely $k$ connected components.

*Solution to problem 3.* If $|S| \geqslant k+1$ then clearly nothing in $C$ can recover all of $S$ by intersection. Hence the VC dimension is bounded by $k$.

On the other hand, a symmetry argument implies that the VC dimension is also bounded by $n-k$.

To see this dimension can b attained, WLOG assume $k \leqslant n/2$. Let $S$ be any set of size $k$. For any subset $A \subset S$, one can extend $A$ to a $k$-element subset by adding $k - |A|$ elements from $X \backslash S$.

*Solution to problem 4.* (1) First consider a diamond, e.g., $\{(0,\pm1),(\pm1,0)\}$. It is clear that such $4$ points can be shattered by axis-aligned rectangles. On the other hand, for any set $S$ of $5$ points, assuming we can break ties arbitrarily, there exists one point with rightmost $x$-coordinate, one with leftmost $x$-coordinate, and two with largest/smallest $y$-coordinate. At least one out of the $5$ points is not covered by the previous $4$ extremes. Call it $p$. Then $S\backslash\{p\}$ cannot be obtained by intersecting $S$ with some axis-aligned rectangle, since the intersection must include a continuum of $x$-values, making it impossible to include points with largest and smallest $x$-values but not $p$.

(2) Immediately we see the $3$ vertices of an equilateral triangle can be easily shattered by the set of axis aligned squares.

To see that a $4$-point set cannot, consider a four-point set $S$. Let $p_1$ be the rightmost point (w.r.t. $x$-coordinate), $p_2$ leftmost, $p_3$ topmost (w.r.t. $y$-coordinate), and $p_4$ bottom-most. WLOG assume $p_{1,x} - p_{2,x} \geqslant p_{3,y} - p_{4,y}$, i.e., the $x$-span is larger, making the set more horizontal. Then no square $C$ can satisfy $C \cap S = \{p_1, p_2\}$, for they have to also include either $p_3, p_4$, or both.

(3) Intuitively, to construct a $5$-point set, consider one point $s$ very far away, while the remaining $4$ points $p_1, p_2, p_3, p_4$ are aligned along an arc, close to each other, and form a convex cone with $c$. For example, given $\epsilon > 0$, consider on the complex plane $c = (0,0)$ and $p_k = e^{2k\epsilon\pi i}$ for $k \in \{0,1,2,3\}$. It's trivial to shatter the $\{p_i\}$. On the other hand, for sufficiently small $\epsilon$, it's also easy to shatter anything consisting of both $c$ and some $p_i$'s. (Sorry, too lazy to draw all diagrams now...) No idea how to prove the 6-point case, except

I assume it will involve some results in convex analysis

(4) The $d$-dimensional simplex $S$ can clearly be shattered. Suppose $A$ is a subset of these $d+1$ points. The binary vector corresponding to the complement of $A$ (i.e. entry has $0$ if correpsonding dimension is not chosen) defines a half space that separates $A$ against $S \backslash A$.

To show that a set of size $d+2$ cannot be shattered we use Radon's theorem, which claims that any such set can be partitioned into two sets $A, B$ whose convex hulls intersect.Therefore, any halfspace containing $A$ must contain the convex hull of $A$ and therefore something from $B$. This implies $A, B$ cannot be shattered, completing the proof that the set of $d$-dimensional half spaces has VC-dim $d+1$.

*Solution to problem 5.* If the VC dimension of $S$ is $\infty$ then the result trivially holds. Otherwise, if a set $A$ is shattered by $R'$ then it certainly is shattered by $R$: $R' \subset R$ implies $\{A \cap C' \mid C' \in R'\} \subset \{A \cap C \mid C \in R\}$. Therefore, $\mathrm{VCdim}(R') \leqslant \mathrm{VCdim}(R)$.