

Spectral Clustering

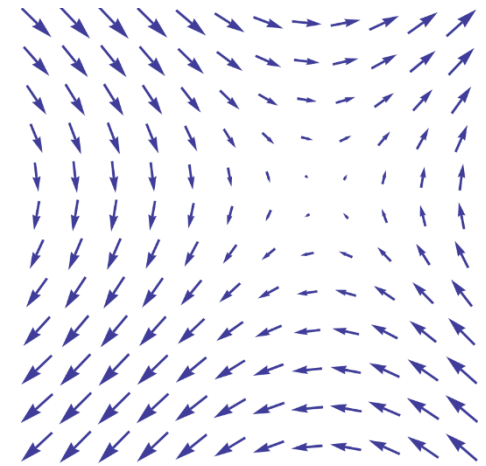
Part 1: The Graph Laplacian

Ng Yen Kaow

Laplacian of a function

□ Given a multivariate function $f: \mathbb{R}^n \rightarrow \mathbb{R}$

□ $\nabla f(\mathbf{x})$, the gradient at $f(\mathbf{x})$, is a vector pointing at the steepest ascent of $f(\mathbf{x})$



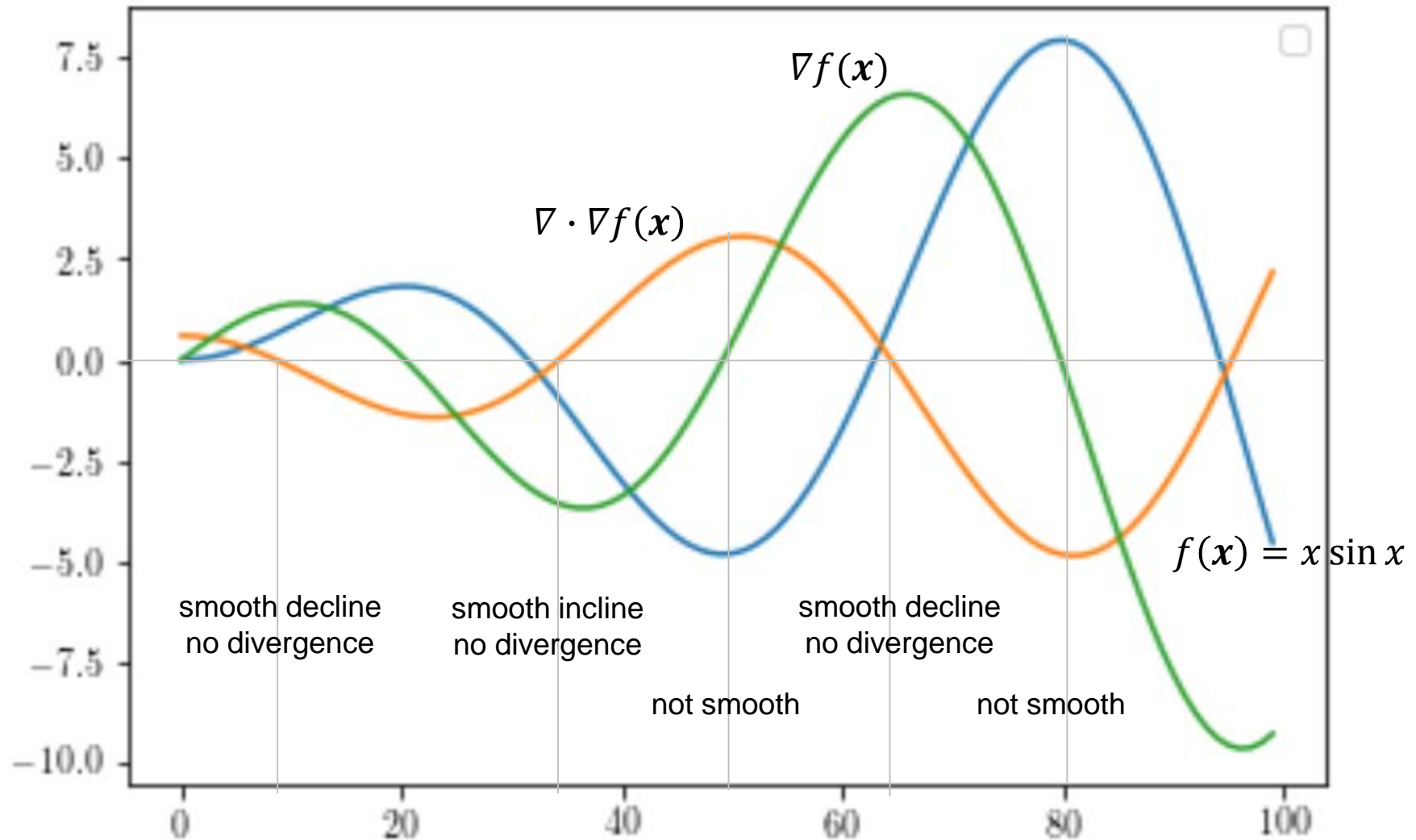
Vector field ∇f

□ Δf , the Laplacian of f , is the divergence of ∇f , that is, $\Delta f(\mathbf{x}) = \nabla \cdot \nabla f(\mathbf{x})$

■ A scalar measurement of the smoothness in $\nabla f(\mathbf{x})$ about point \mathbf{x}

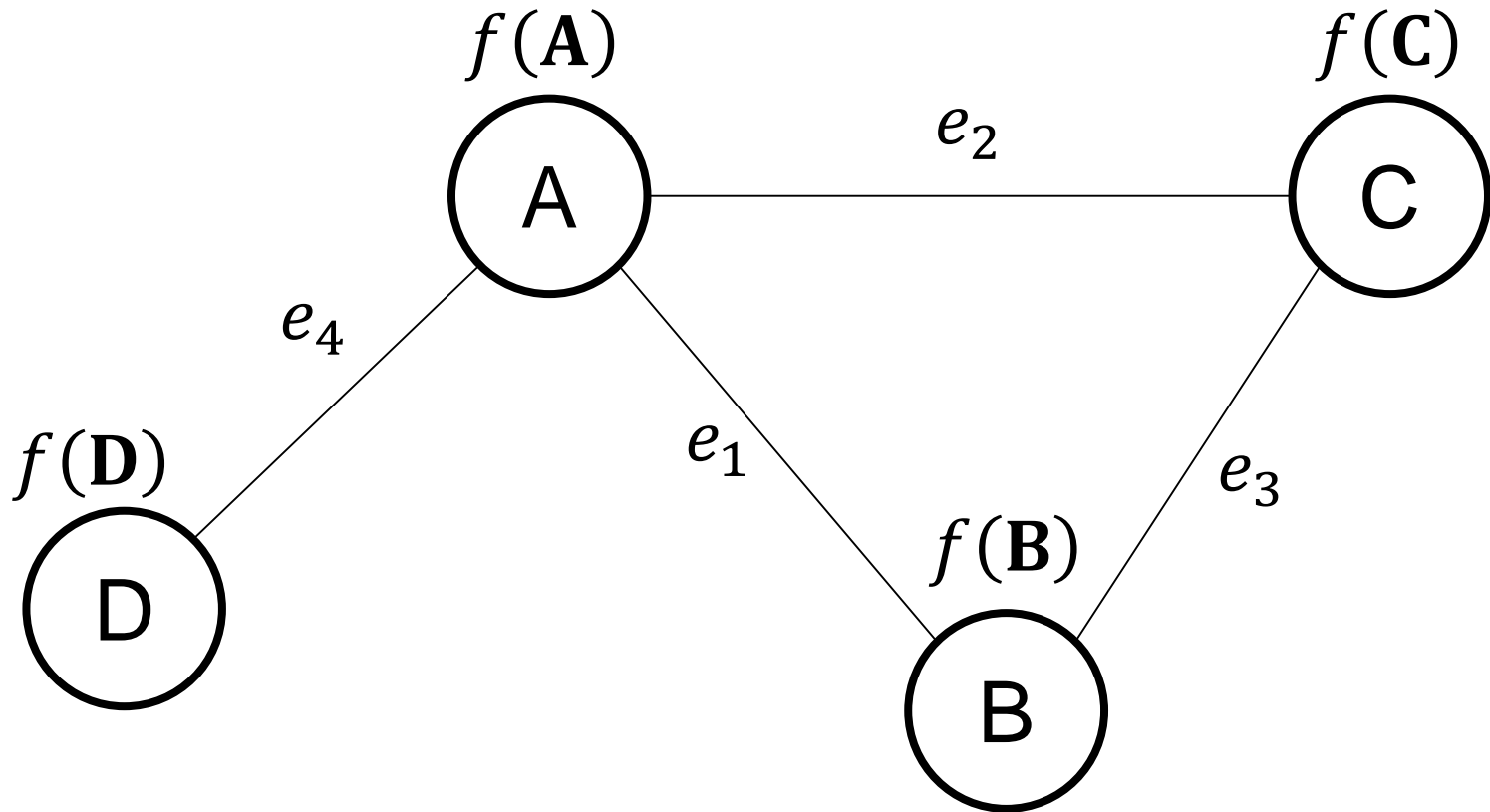
Laplacian of a function

□ 1-D example



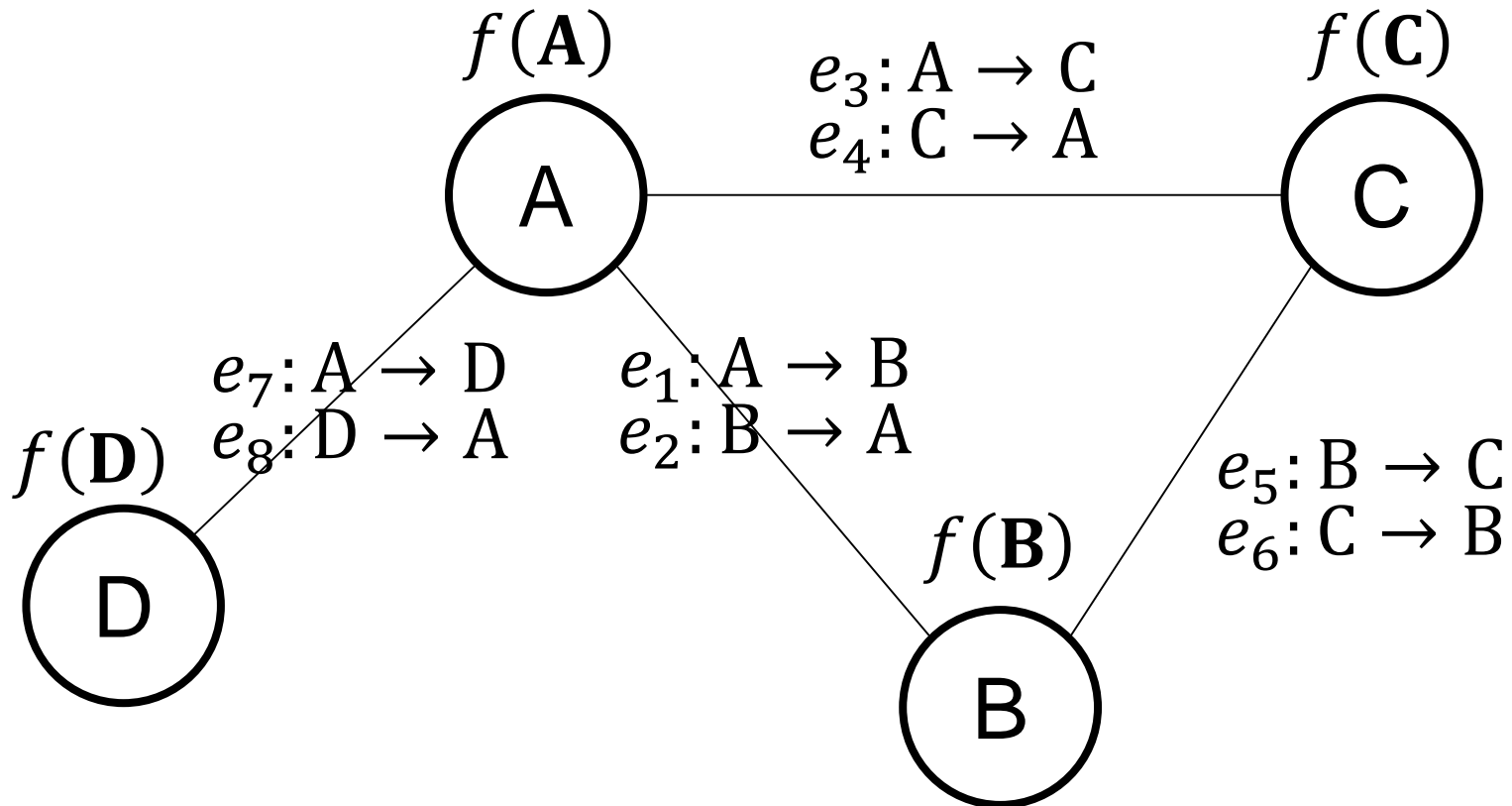
Incidence matrix

- Given a function $f: V \rightarrow \mathbb{R}$ on the undirected graph $G = \{\{A, B, C, D\}, \{e_1, e_2, e_3, e_4\}\}$

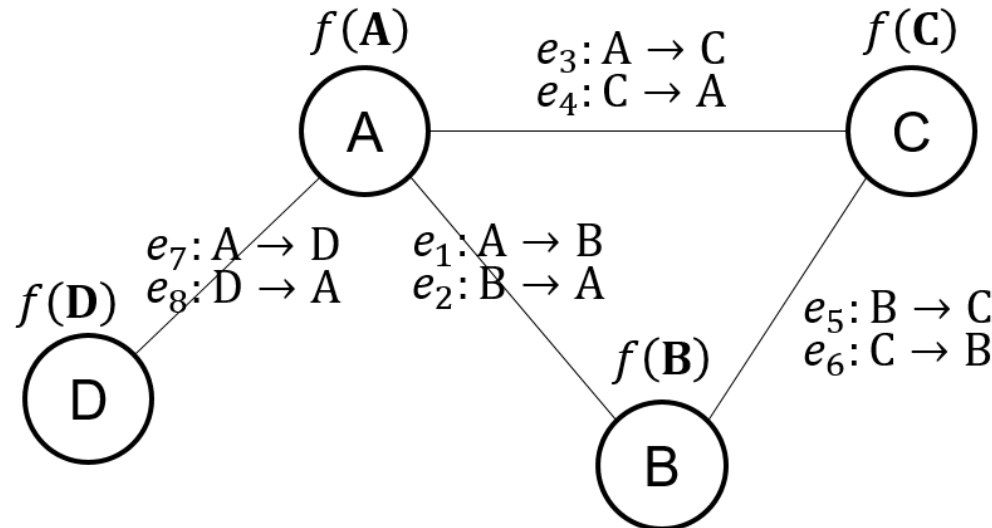


Incidence matrix

- Given a function $f: V \rightarrow \mathbb{R}$ on the undirected $G = \{\{A, B, C, D\}, \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}\}$



Incidence matrix



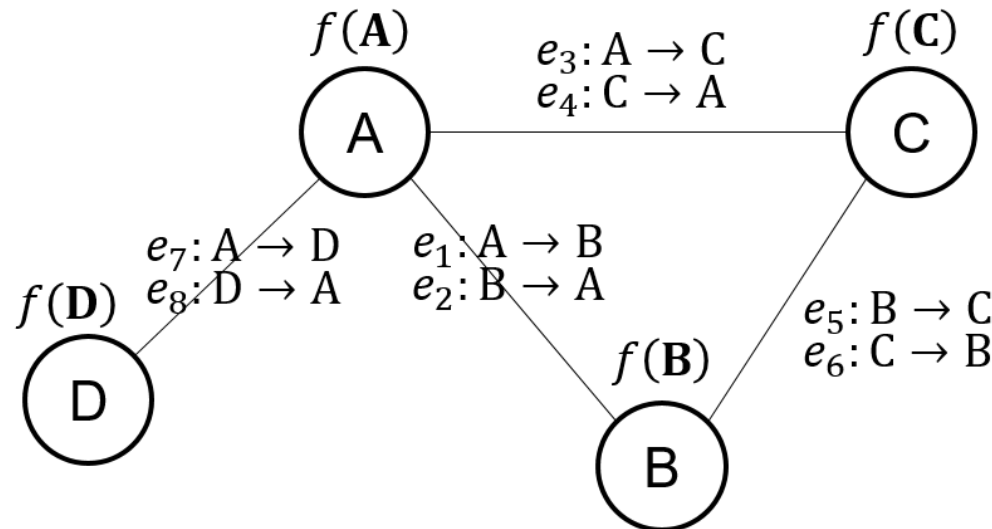
- (Oriented) incidence matrix

$$\begin{matrix} & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 1 & -1 & 1 & -1 & 0 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \end{matrix}$$

- Every column in the incidence matrix describes an edge

$$\begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} f(A) \\ f(B) \\ f(C) \\ f(D) \end{bmatrix} = f(A) - f(B)$$

Incidence matrix



□ (Oriented) incidence matrix

$$\begin{matrix} & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 1 & -1 & 1 & -1 & 0 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \end{matrix}$$

□ Every row in the incidence matrix describes a vertex

$$\begin{bmatrix} 1 & -1 & 1 & -1 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} w(e_1) \\ w(e_2) \\ w(e_3) \\ w(e_4) \\ w(e_5) \\ w(e_6) \\ w(e_7) \\ w(e_8) \end{bmatrix} = w(e_1) - w(e_2) + w(e_3) - w(e_4) + w(e_7) - w(e_6)$$

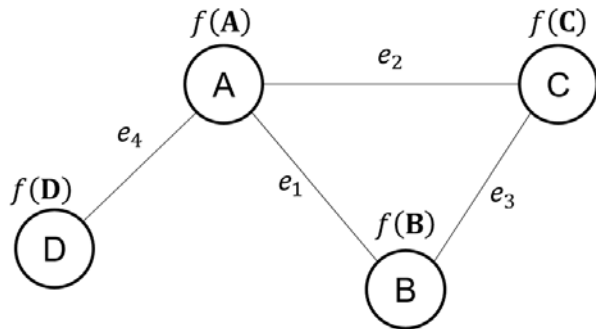
$w(e)$ is the weight of edge e

Incidence matrix

- Incidence matrix encodes the graph structure
- What constitutes an incidence matrix is not strictly defined
 - Open to re-definition
 - Results may differ
 - **Let's look at some examples**

(Unoriented) incidence matrix

https://en.wikipedia.org/wiki/Incidence_matrix



$$\begin{array}{c} \begin{matrix} & e_1 & e_2 & e_3 & e_4 \\ A & \begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix} \\ B & \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \\ C & \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix} \\ D & \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} \end{array}$$

$$\begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} f(A) \\ f(B) \\ f(C) \\ f(D) \end{bmatrix} = f(A) + f(B)$$

$$\begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} w(e_1) \\ w(e_2) \\ w(e_3) \\ w(e_4) \end{bmatrix} = w(e_1) + w(e_2) + w(e_4)$$

(Oriented) incidence matrix

As shown earlier, this is

https://en.wikipedia.org/wiki/Incidence_matrix

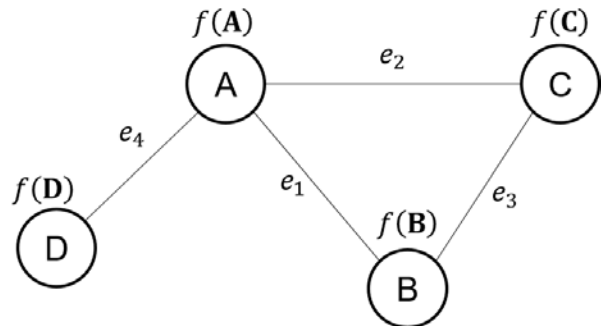
$$\begin{array}{c} A \\ B \\ C \\ D \end{array} \begin{array}{cccccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 \\ \left[\begin{array}{cccccccc} 1 & -1 & 1 & -1 & 0 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{array} \right] \end{array}$$

$$\begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} f(A) \\ f(B) \\ f(C) \\ f(D) \end{bmatrix} = f(A) - f(B)$$

$$\begin{bmatrix} 1 & -1 & 1 & -1 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} w(e_1) \\ w(e_2) \\ w(e_3) \\ w(e_4) \\ w(e_5) \\ w(e_6) \\ w(e_7) \\ w(e_8) \end{bmatrix} = w(e_1) - w(e_2) + w(e_3) - w(e_4) + w(e_7) - w(e_6)$$

(Oriented) incidence matrix++

Ordering $A < B < C < D$



$$\begin{matrix} & e_1 & e_2 & e_3 & e_4 \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \end{matrix}$$

https://en.wikipedia.org/wiki/Laplacian_matrix#Incidence_matrix

Define a fixed ordering over the vertices, then define the (oriented) incidence matrix++ M with each element M_{ve} for vertex v and edge $e = (v, u)$,

$$M_{ve} = \begin{cases} 1 & \text{if } v < u \\ -1 & \text{if } v > u \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} f(A) \\ f(B) \\ f(C) \\ f(D) \end{bmatrix} = f(A) - f(B)$$

$$\begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} w(e_1) \\ w(e_2) \\ w(e_3) \\ w(e_4) \end{bmatrix} = w(e_1) + w(e_2) + w(e_4)$$

Graph Laplacian

- Extend the Laplacian $\Delta f(\mathbf{x}) = \nabla \cdot \nabla f$ on $f: \mathbb{R}^n \rightarrow \mathbb{R}$ to one for $f: V \rightarrow \mathbb{R}$
- We have
 - $(M_{ev})f(x)$ (i.e. $(M_{ve})^T f(x)$) gives the edges from each node $f(x)$
 - $(M_{ve})(M_{ev})f(x)$ gives the divergence of the edges
- Our Laplacian is $L = (M_{ve})(M_{ev})$, or
$$L = (M_{ve})(M_{ve})^T$$

Normalized Graph Laplacian

- The **unnormalized** graph Laplacian L of an undirected graph is defined as $L = (M_{ve})(M_{ve})^T$ or $(M_{ev})^T(M_{ev})$
 - The oriented incidence matrix++ is typically implied
- The **normalized** graph Laplacian of an undirected graph is $D^{-1/2}LD^{-1/2}$, where D is the diagonal matrix indicating the degree of each vertex

Laplacian L for earlier matrices

- Unoriented incidence matrix ($L = D + A$)

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Adjacency
matrix

- Oriented incidence matrix ($L = D - 2A$)

$$\begin{bmatrix} 1 & -1 & 1 & -1 & 0 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & -2 & -2 & -2 \\ -2 & 4 & -2 & 0 \\ -2 & -2 & 4 & 0 \\ -2 & 0 & 0 & 2 \end{bmatrix}$$

- Oriented incidence matrix++ ($L = D - A$)

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

$D^{-1/2}LD^{-1/2}$ for earlier matrices

□ Unoriented incidence matrix

$$\begin{bmatrix} 0.58 & 0 & 0 & 0 \\ 0 & 0.71 & 0 & 0 \\ 0 & 0 & 0.71 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.58 & 0 & 0 & 0 \\ 0 & 0.71 & 0 & 0 \\ 0 & 0 & 0.71 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.41 & 0.41 & 0.58 \\ 0.41 & 1 & 0.5 & 0 \\ 0.41 & 0.5 & 1 & 0 \\ 0.58 & 0 & 0 & 1 \end{bmatrix}$$

□ Oriented incidence matrix

$$\begin{bmatrix} 0.41 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.71 \end{bmatrix} \begin{bmatrix} 6 & -2 & -2 & -2 \\ -2 & 4 & -2 & 0 \\ -2 & -2 & 4 & 0 \\ -2 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0.41 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.71 \end{bmatrix} = \begin{bmatrix} 1 & -0.41 & -0.41 & -0.58 \\ -0.41 & 1 & -0.5 & 0 \\ -0.41 & -0.5 & 1 & 0 \\ -0.58 & 0 & 0 & 1 \end{bmatrix}$$

□ Oriented incidence matrix++

$$\begin{bmatrix} 0.58 & 0 & 0 & 0 \\ 0 & 0.71 & 0 & 0 \\ 0 & 0 & 0.71 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.58 & 0 & 0 & 0 \\ 0 & 0.71 & 0 & 0 \\ 0 & 0 & 0.71 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -0.41 & -0.41 & -0.58 \\ -0.41 & 1 & -0.5 & 0 \\ -0.41 & -0.5 & 1 & 0 \\ -0.58 & 0 & 0 & 1 \end{bmatrix}$$

□ Normalization unifies the oriented incidence matrices

Significance of the graph Laplacian

- Each row in L describes the dependency of a vertex with respect to the others
- Let the adjacency matrix $A = (a_{ij})$, then

$$x^T L x = \frac{1}{2} \sum_{i,j=1}^m a_{ij} (x_i - x_j)^2$$

$$\begin{aligned} x^T L x &= x^T D x - x^T A x = \sum_{i=1}^m d_i x_i^2 - \sum_{i,j=1}^m a_{ij} x_i x_j \\ &= \frac{1}{2} \left(\sum_{i=1}^m d_i x_i^2 - 2 \sum_{i,j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m d_i x_i^2 \right) \\ &= \frac{1}{2} \sum_{i,j=1}^m a_{ij} (x_i - x_j)^2 \end{aligned}$$

Significance of the graph Laplacian

- Each row in L describes the dependency of a vertex with respect to the others
- Let the adjacency matrix $A = (a_{ij})$, then

$$x^T L x = \frac{1}{2} \sum_{i,j=1}^m a_{ij} (x_i - x_j)^2$$

- Suppose x is a vector of only the values +1 and -1, indicating the membership of the vertices in a set S

$$x_i = \begin{cases} 1 & \text{if } v_i \in S \\ -1 & \text{if } v_i \in \bar{S} \end{cases}$$

- That is, we want to use x to indicate the result of a 2-partition, S and \bar{S}

Significance of the graph Laplacian

- Each row in L describes the dependency of a vertex with respect to the others
- Let the adjacency matrix $A = (a_{ij})$, then

$$x^T L x = \frac{1}{2} \sum_{i,j=1}^m a_{ij} (x_i - x_j)^2$$

- Suppose x is a vector of only $\{+1, -1\}$, then $x^T L x$ has special significance

$$\begin{aligned} \frac{1}{2} \sum_{i,j=1}^m a_{ij} (x_i - x_j)^2 &= \sum_{i,j=1, i < j}^m a_{ij} (x_i - x_j)^2 \\ &= 4 \sum_{1 \leq i < j \leq m, x_i \neq x_j} a_{ij} \end{aligned}$$

- That is, $x^T L x$ is 4 times the number of edges between adjacent vertices of each from S and \bar{S}

Intuitions of the graph Laplacian

- Compute $x^T L x$ for all x
 - e.g. $x = [1, -1, -1, -1]$ gives $x^T L x = 12$
- This gives us the 2-partition that results in the least number of removed edges
 - $x = \mathbf{1} = [1 \ 1 \ 1 \ 1]$ or $x = -\mathbf{1} = [-1 \ -1 \ -1 \ -1]$ which has $x^T L x = 0$ are trivial solutions
 - Best x is $[1 \ 1 \ 1 \ -1]$, that is, A, B, C in one group and D in another

Group 1	Group 2	$x^T L x$
A	B C D	12
B	A C D	8
C	A B D	8
D	A B C	4
A B	C D	12
A C	B D	12
A D	B C	8
A B C D	∅	0

- The optimal x can be approximately found

Rayleigh Quotient

- Minimize $x^T L x$
 - Consider instead problem of minimizing $\frac{x^T L x}{x^T x}$
 - x is of only +1 and -1 $\Rightarrow x^T x = |x| = \text{const}$

Group 1	Group 2	$x^T L x$	$\frac{x^T L x}{x^T x}$
A	B C D	12	3
B	A C D	8	2
C	A B D	8	2
D	A B C	4	1
A B	C D	12	3
A C	B D	12	3
A D	B C	8	2

Rayleigh Quotient

□ $\frac{x^T L x}{x^T x}$ is known as the Rayleigh quotient

■ By the min-max theorem of Rayleigh quotient,

$$\min_x \frac{x^T L x}{x^T x} = \lambda_k$$

■ where λ_k is the smallest eigenvalue in the decomposition of $Lx = \lambda x$, and

■ $\mu_k = \operatorname{argmin}_x \frac{x^T L x}{x^T x}$

□ However, μ_k is the trivial solution

■ Compromise and use the second best solution μ_{k-1} (which corresponds to the second smallest eigenvalue λ_{k-1})

Eigendecomposition example

□ Eigenvalues

λ_1	λ_2	λ_3	λ_4
4.0000	3.0000	1.0000	0.0000

□ Eigenvectors

μ_1	μ_2	μ_3	μ_4
0.8660	0.0000	0.0000	-0.5000
-0.2887	0.7071	-0.4082	-0.5000
-0.2887	-0.7071	-0.4082	-0.5000
-0.2887	0.0000	0.8165	-0.5000

More precisely, -9.51E-17

- $\lambda_3 = 1 = \text{optimal value for } \frac{1}{2} \sum_{1 \leq i, j \leq m} a_{ij} (x_i - x_j)^2$
- If group by the (\pm) sign, μ_3 correctly places A, B, C in one group ($-$) and D in another ($+$)

Compromise in +1/-1 restriction

- By relaxing the restriction of +1 and -1 in x to allow any real number, an $x^T L x$ smaller than the optimal under the restriction is often achieved
- The improvement can be guaranteed if x is orthogonal to $\mathbf{1}$ (or $-\mathbf{1}$) since by the min-max theorem, $\frac{\mu_{k-1}^T L \mu_{k-1}}{\mu_{k-1}^T \mu_{k-1}}$ is minimal among all $\frac{x^T L x}{x^T x}$ that are orthogonal to μ_k
 - However, in the present case, $x = [1 \ 1 \ 1 \ -1]$ and not orthogonal to $\mu_4 = [1 \ 1 \ 1 \ 1]$
 - Still, $\frac{\mu_3^T L \mu_3}{\mu_3^T \mu_3} = \lambda_3 = 1 = \min_{x \in \{1, -1\}^4} \frac{x^T L x}{x^T x}$
 - Though no guarantee, improvements are usual

The significance of μ_{k-1} and λ_{k-1}

- The heuristic for translating μ_{k-1} back into discrete values for a grouping of the vertices is an important topic
- μ_{k-1} is called the **Fiedler vector**
- λ_{k-1} is called the **Fiedler value**
 - The multiplicity of λ_{k-1} is always 1
 - Also called the **algebraic connectivity**
 - The further λ_{k-1} is from 0, the more connected is the graph

The case of $L' = D^{-1/2} L D^{-1/2}$

- While $x^T L x = \frac{1}{2} \sum_{1 \leq i, j \leq m} a_{ij} (x_i - x_j)^2$, $x^T L' x$ where $L' = D^{-1/2} L D^{-1/2}$ has no such significance
- But just for reference find $\min x^T L' x$

Group 1	Group 2	$x^T L' x$	$\frac{x^T L' x}{x^T x}$
A	B C D	5.7877	1.4469
B	A C D	3.8453	0.9613
C	A B D	3.8453	0.9613
D	A B C	2.5217	0.6304
A B	C D	6.1547	1.5387
A C	B D	6.1547	1.5387
A D	B C	3.4783	0.8696

The case of $L' = D^{-1/2}LD^{-1/2}$

□ Eigenvalues

λ_1	λ_2	λ_3	λ_4
1.7287	1.5000	0.7713	0.0000

□ Eigenvectors

μ_1	μ_2	μ_3	μ_4
-0.7355	0.0000	0.2899	0.6124
0.2444	-0.7071	-0.4362	0.5000
0.2444	0.7071	-0.4362	0.5000
0.5827	0.0000	0.7317	0.3536

□ The solution μ_3 places A, D in one group and B, C in another, which is only the second best solution

□ Rare instance where $\lambda_3 = 0.77 > 0.63 = \min_{x \in \{1, -1\}^4} \frac{x^T L' x}{x^T x}$