Spectral Clustering

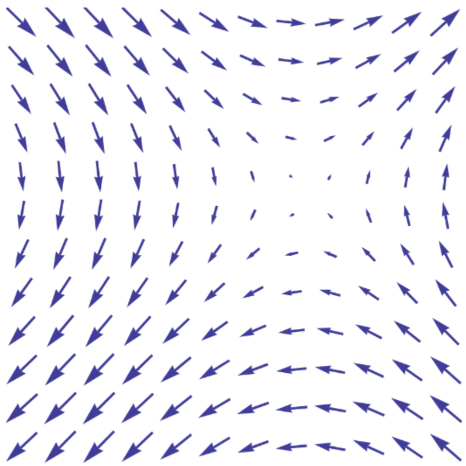
Part 1: The Graph Laplacian

Ng Yen Kaow

Laplacian of a function

- Given a multivariate function $f: \mathbb{R}^n \to \mathbb{R}$

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

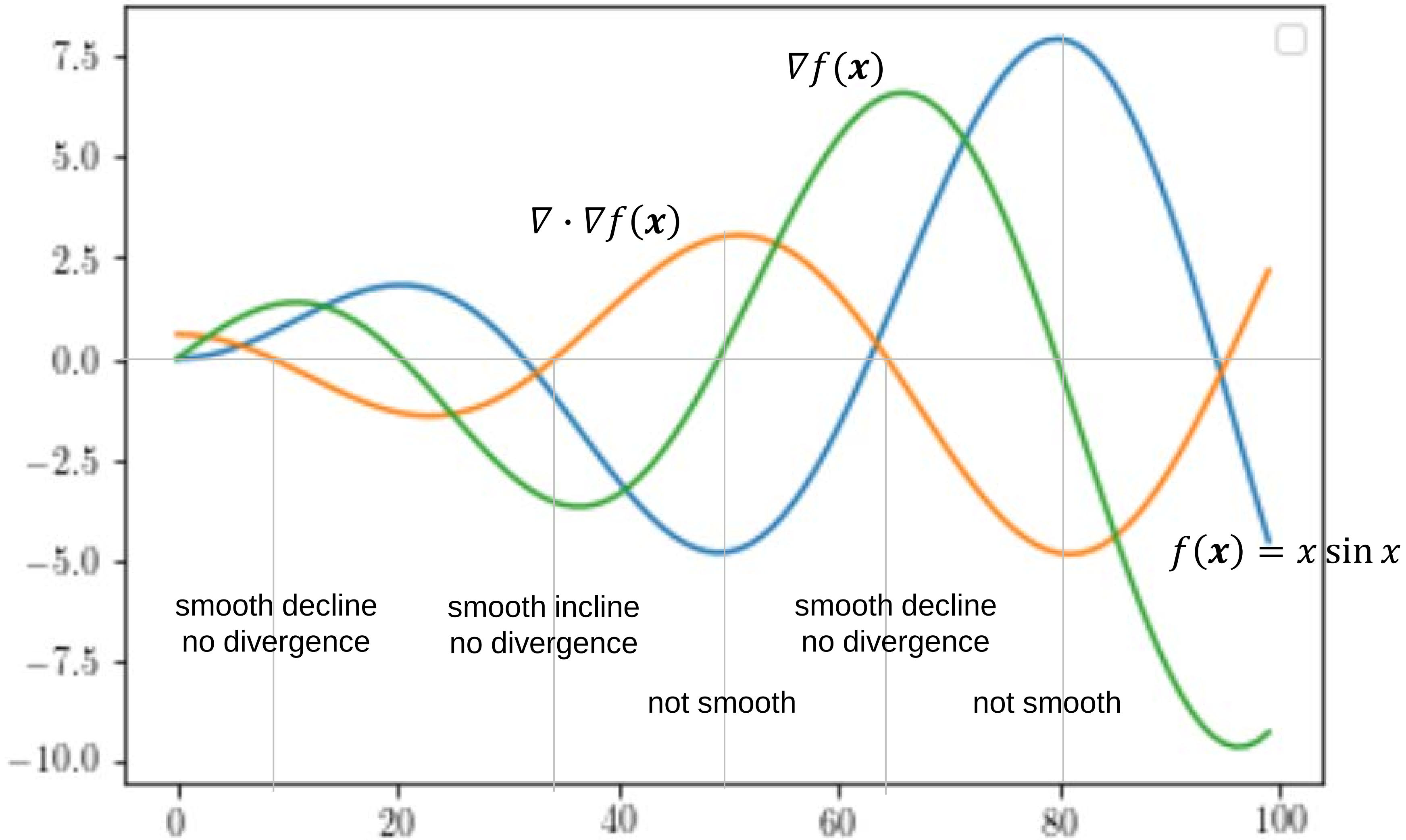


Vector field ∇f

- Δf, the Laplacian of f, is the divergence of ∇f, that is, $\Delta f(\boldsymbol{x}) = \nabla \cdot \nabla f(\boldsymbol{x})$
  - A scalar measurement of the smoothness in $\nabla f(\boldsymbol{x})$ about point $\boldsymbol{x}$

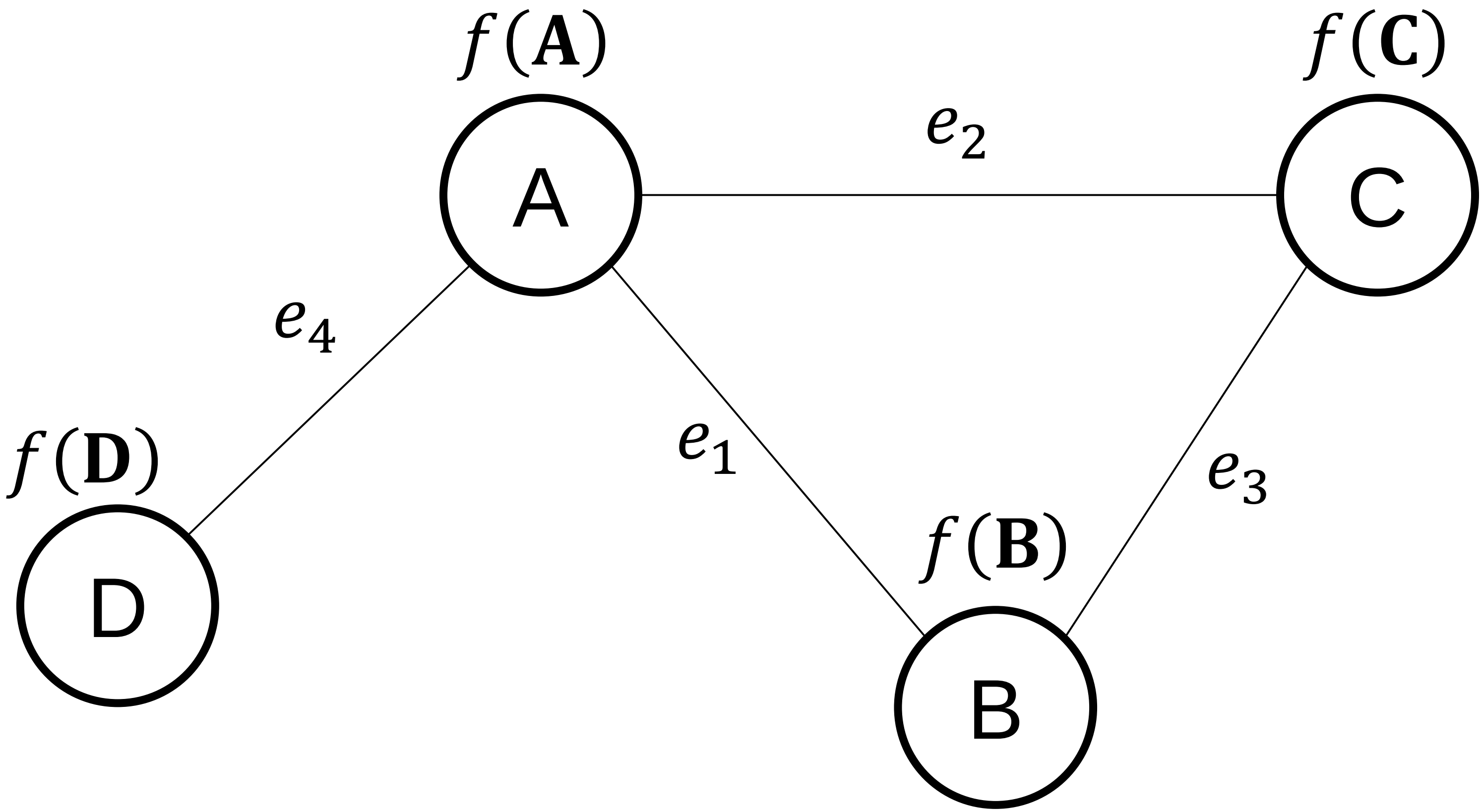
Laplacian of a function

- 1-D example



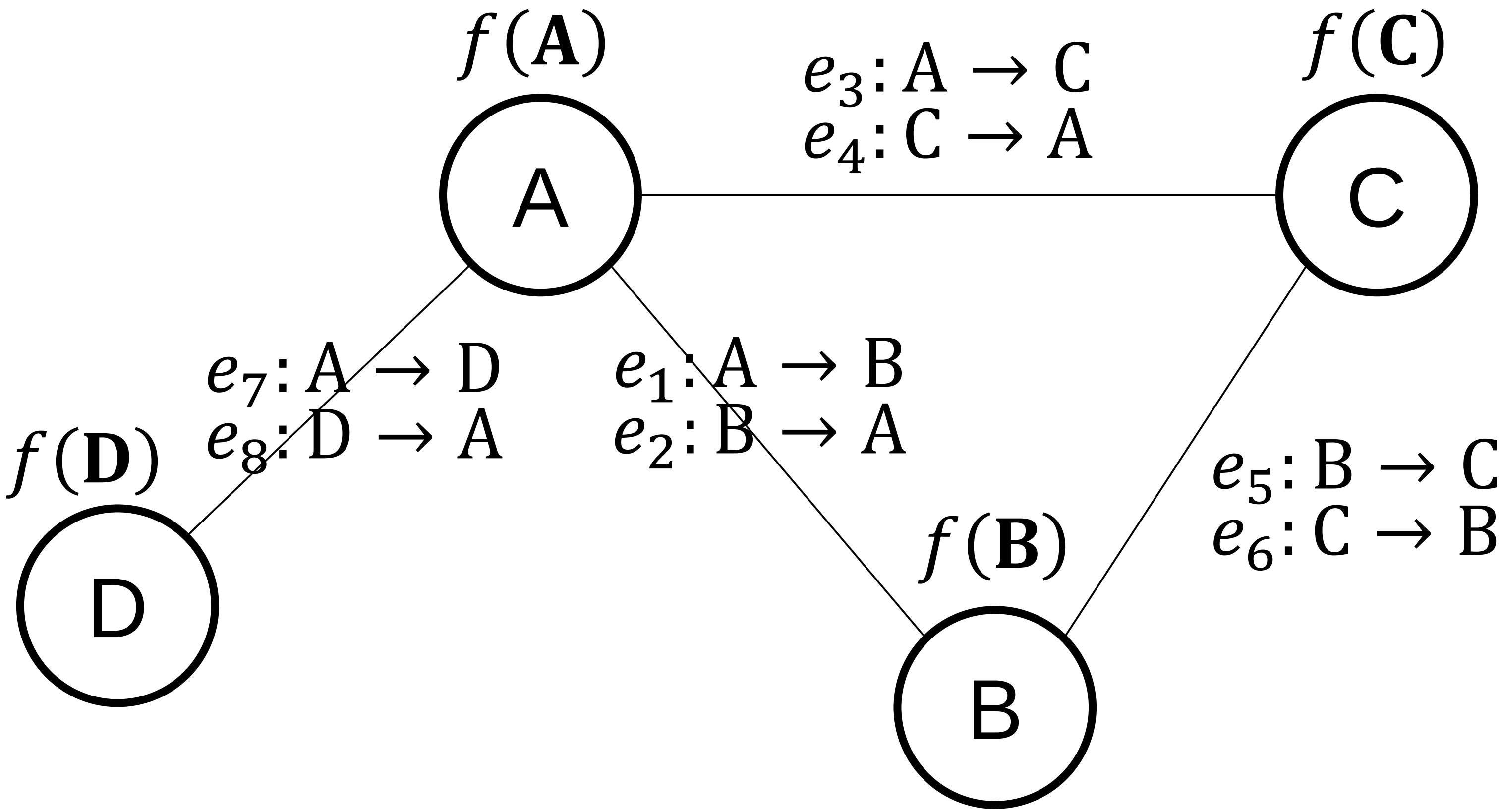
Incidence matrix

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

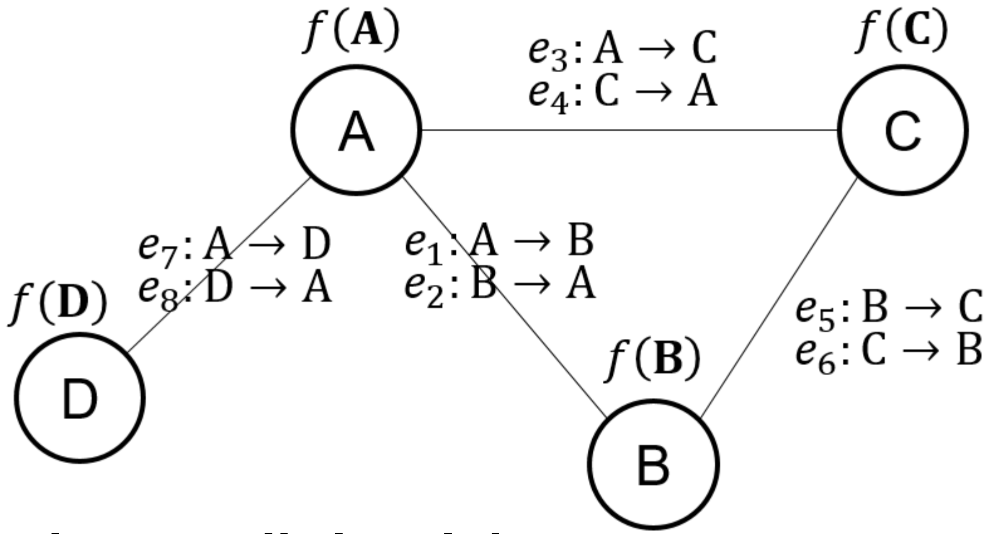


Incidence matrix

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



Incidence matrix



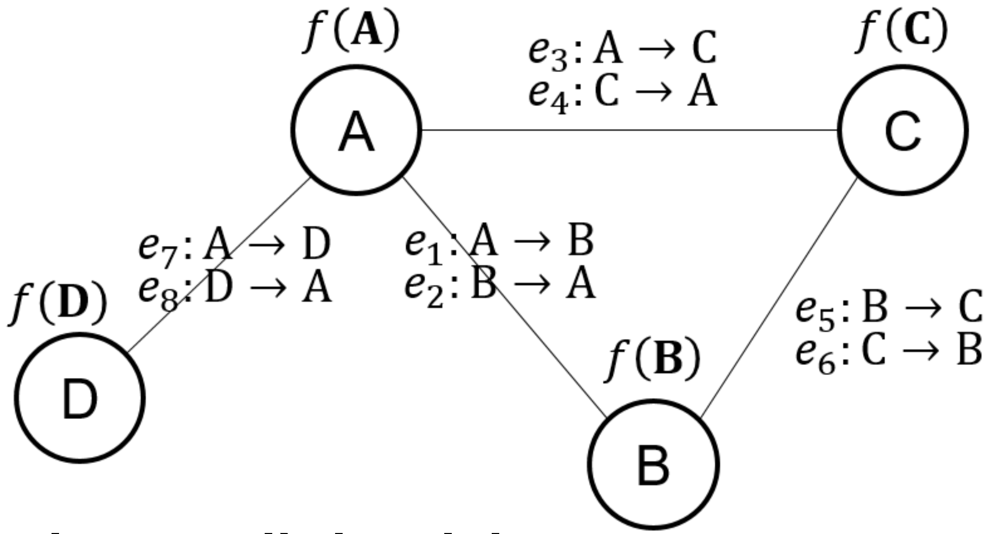
- (Oriented) incidence matrix

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

- Every column in the incidence matrix describes an edge

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

Incidence matrix



- (Oriented) incidence matrix

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

- Every row in the incidence matrix describes a vertex

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

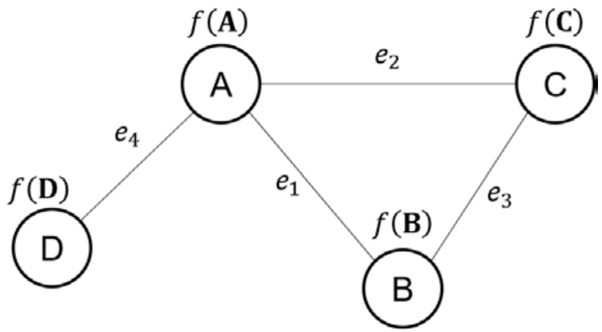
$w(e)$ is the weight of edge e

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

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

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

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

(Oriented) incidence matrix

As shown earlier, this is

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

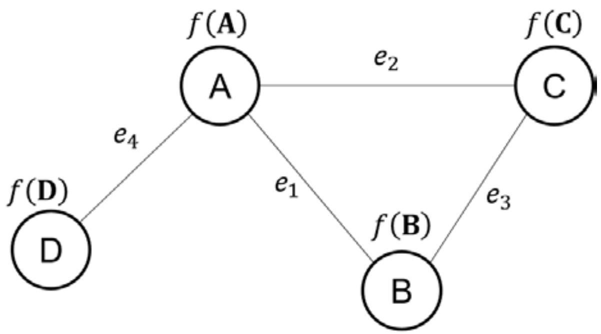
$$\begin{array}{c}\\ \mathrm{A}\\ \mathrm{B}\\ \mathrm{C}\\ \mathrm{D}\end{array}\begin{array}{c}\begin{array}{cccccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8\end{array}\\ \begin{bmatrix} 1 & -1 & 1 & -1 & 0 & 0 & 1 & -1\\ -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0\\ 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0\\ 0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

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

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

(Oriented) incidence matrix++

Ordering $\mathrm{A} < \mathrm{B} < \mathrm{C} < \mathrm{D}$



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

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 $\mathrm{e} = (v, u)$,

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

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

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

Graph Laplacian

- Extend the Laplacian $\Delta f(\boldsymbol{x}) = \nabla \cdot \nabla f$ on $f: \mathbb{R}^n \to \mathbb{R}$ to one for $f: V \to \mathbb{R}$
- We have
  - $(M_{ev}) f(x)$ (i.e. $(M_{ve})^{\mathrm{T}} f(x)$) 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 $L = (M_{ve})(M_{ve})^{\mathrm{T}}$

Normalized Graph Laplacian

- The graph Laplacian L of an undirected graph is defined as $L = (M_{ve})(M_{ve})^{\mathrm{T}}$ or $(M_{ev})^{\mathrm{T}}(M_{ev})$
  - The oriented incidence matrix++ is typically implied
- Define the normalized version of a Laplacian L as $D^{-1/2}LD^{-1/2}$, where D is the diagonal matrix indicating the degree of each vertex
  - The reason for such a normalization will only become apparent in Part 3

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++

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

- Note that normalization unified 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

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

$$x^{\mathrm{T}} L x = x^{\mathrm{T}} D x - x^{\mathrm{T}} A x = \sum_{i=1}^{m} d_i x_i^2 - \sum_{i,j=1}^{m} a_{ij} x_i x_j$$

$$= \frac{1}{2} \left(\sum_{i=1}^{m} d_i x_i^2 - 2 \sum_{i,j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m} d_i x_i^2 \right)$$

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

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

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

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

$$\frac{1}{2} \sum_{i,j=1}^{m} a_{ij} (x_i - x_j)^2 = \sum_{i,j=1,i<j}^{m} a_{ij} (x_i - x_j)^2$$
$$= 4 \sum_{1 \le i < j \le m, x_i \ne x_j} a_{ij}$$

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

Intuitions of the graph Laplacian

- Compute $x^T Lx$ for all x
  - e.g. $x = [1, -1, -1, -1]$ gives $x^T Lx = 12$
- 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 $x^T Lx = 0$ are trivial solutions
  - Best x is $[1\ 1\ 1\ -1]$, that is, A, B, C in one group and D in another
- The optimal x can be approximately found

Group 1	Group 2	$x^T Lx$
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

Rayleigh Quotient

- Minimize $x^{\mathrm{T}}Lx$
  - Consider instead problem of minimizing $\dfrac{x^{\mathrm{T}}Lx}{x^{\mathrm{T}}x}$
    - x is of only +1 and -1 $\Rightarrow$ $x^{\mathrm{T}}x = |x| =$ const

Group 1	Group 2	$x^T Lx$	$\frac{x^T Lx}{x^T x}$
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

- $\frac{x^{\mathrm{T}}Lx}{x^{\mathrm{T}}x}$ is known as the Rayleigh quotient
  - By the min-max theorem of Rayleigh quotient,
$$\min_{x}\frac{x^{\mathrm{T}}Lx}{x^{\mathrm{T}}x} = \lambda_k$$
  - where λ_k is the smallest eigenvalue in the decomposition of $Lx = \lambda x$, and
  - $\mu_k = \underset{x}{\operatorname{argmin}}\frac{x^{\mathrm{T}}Lx}{x^{\mathrm{T}}x}$
- 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

More precisely, -9.51E-17

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

- $\lambda_3 = 1 =$ optimal value for $\frac{1}{2}\sum_{1\le i,j\le m} a_{ij}\left(x_i - x_j\right)^2$
- 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 $x^{\mathrm{T}}Lx$ 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, $\frac{{\mu_{k-1}}^{\mathrm{T}}L\mu_{k-1}}{{\mu_{k-1}}^{\mathrm{T}}\mu_{k-1}}$ is minimal among all $\frac{x^{\mathrm{T}}Lx}{x^{\mathrm{T}}x}$ 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, $\frac{{\mu_3}^{\mathrm{T}}L\mu_3}{{\mu_3}^{\mathrm{T}}\mu_3} = \lambda_3 = 1 = \min\limits_{\boldsymbol{x}\in\{1,-1\}^4} \frac{x^{\mathrm{T}}Lx}{x^{\mathrm{T}}x}$
    - 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