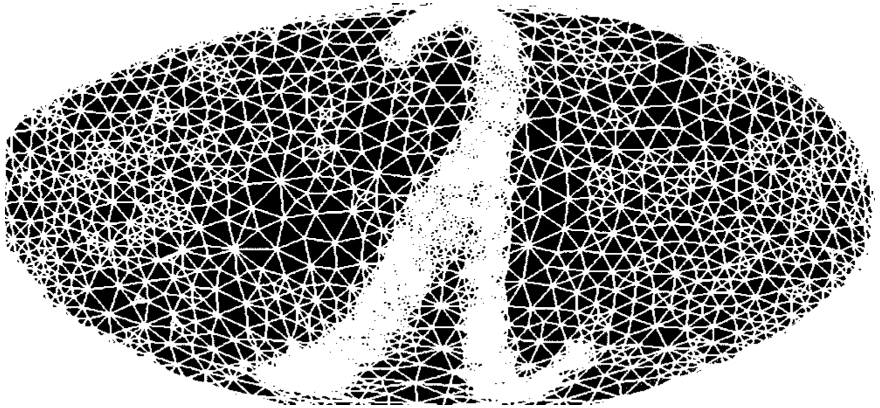


Elements of Spectral Graph Theory

Andreas Loukas

Signal Processing Laboratory (LTS2)

EPFL

Outline of lecture

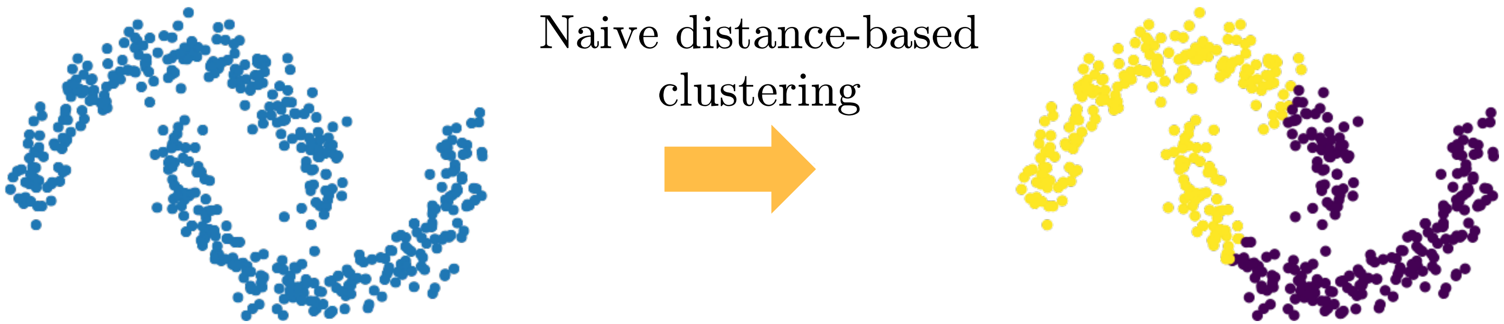
- Relational reasoning in high-dimensions
  - Distances in high-dimensions
  - The manifold assumption
  - Examples in machine learning

- Spectral graph theory 101
  - Gradient on a graph
  - Laplacian matrices
  - The Laplacian spectrum
  - Bottlenecks and diameter

Relational reasoning in high dimensions

In ML, one is often tasked with analyzing the relations between a set of items (e.g., cluster images).

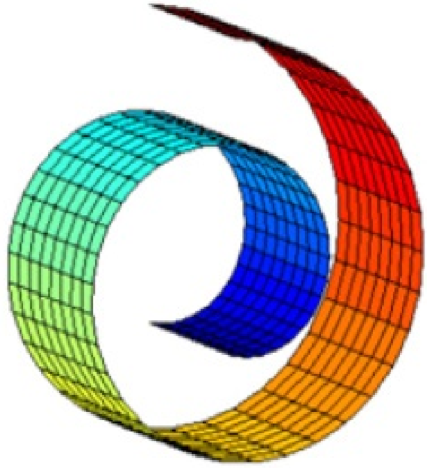
- Think of them as points in some high-dimensional feature space
- Solve problem by looking at distances between them

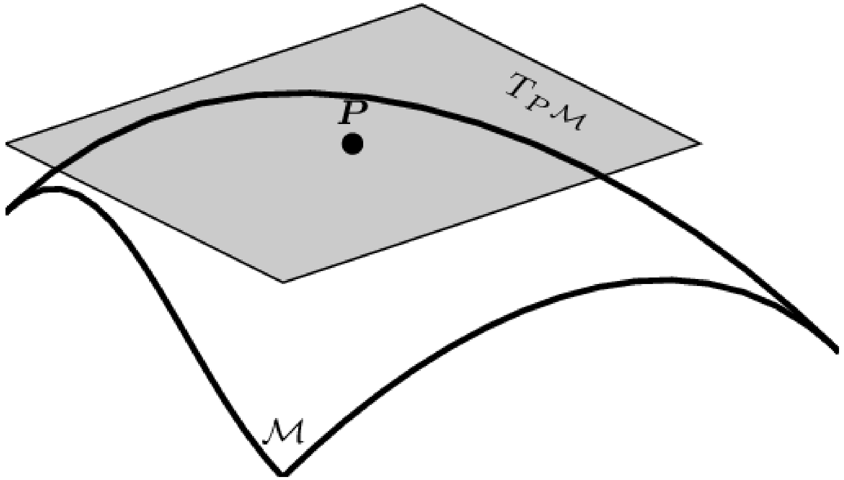


Key issue: In high dimensions, distances are not relevant!

The manifold assumption

- Think of points as living on a high-dimensional surface (manifold).
- Consider local geometry when reasoning about similarity.

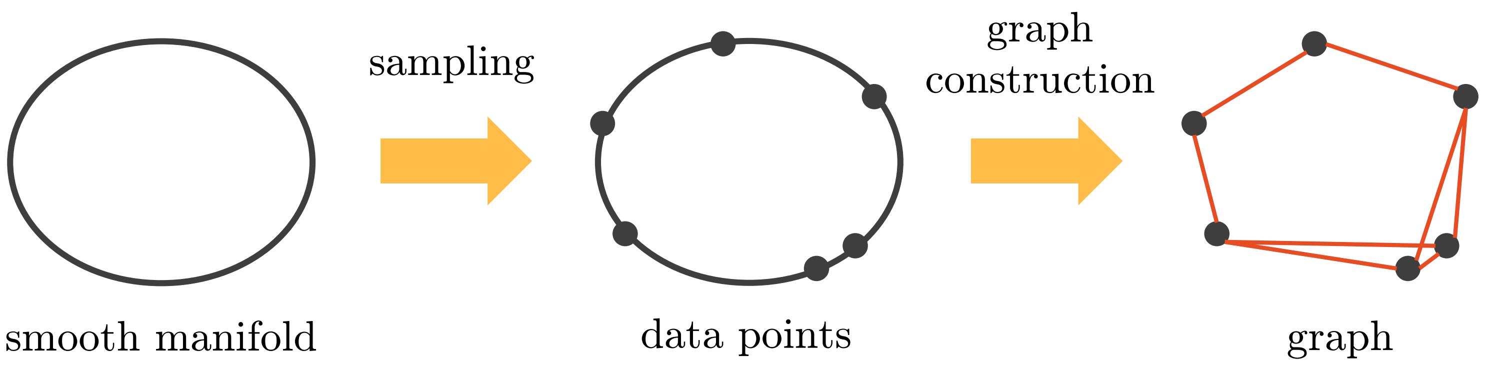




A manifold is a topological space that is locally (homeomorphic to) a Euclidean space.

From manifolds to graphs

Graphs can be seen as discrete approximations of the underlying manifold.

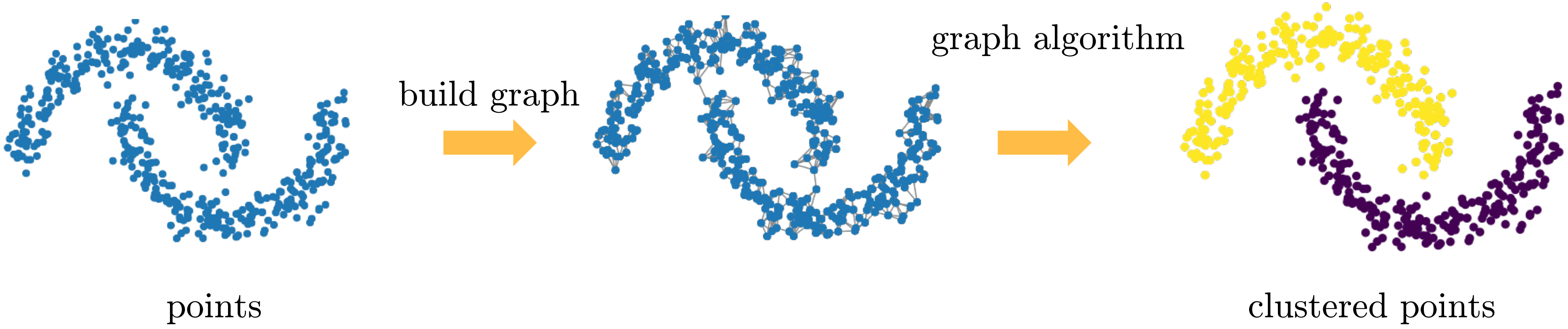


Approximation improves with more points (nodes N).

Relational reasoning with graphs

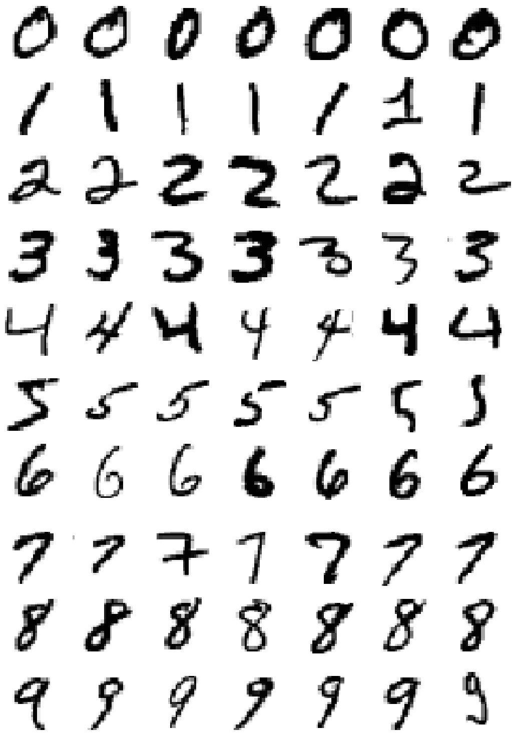
In practical terms:

- reason about relations of points by reasoning about the graph they form



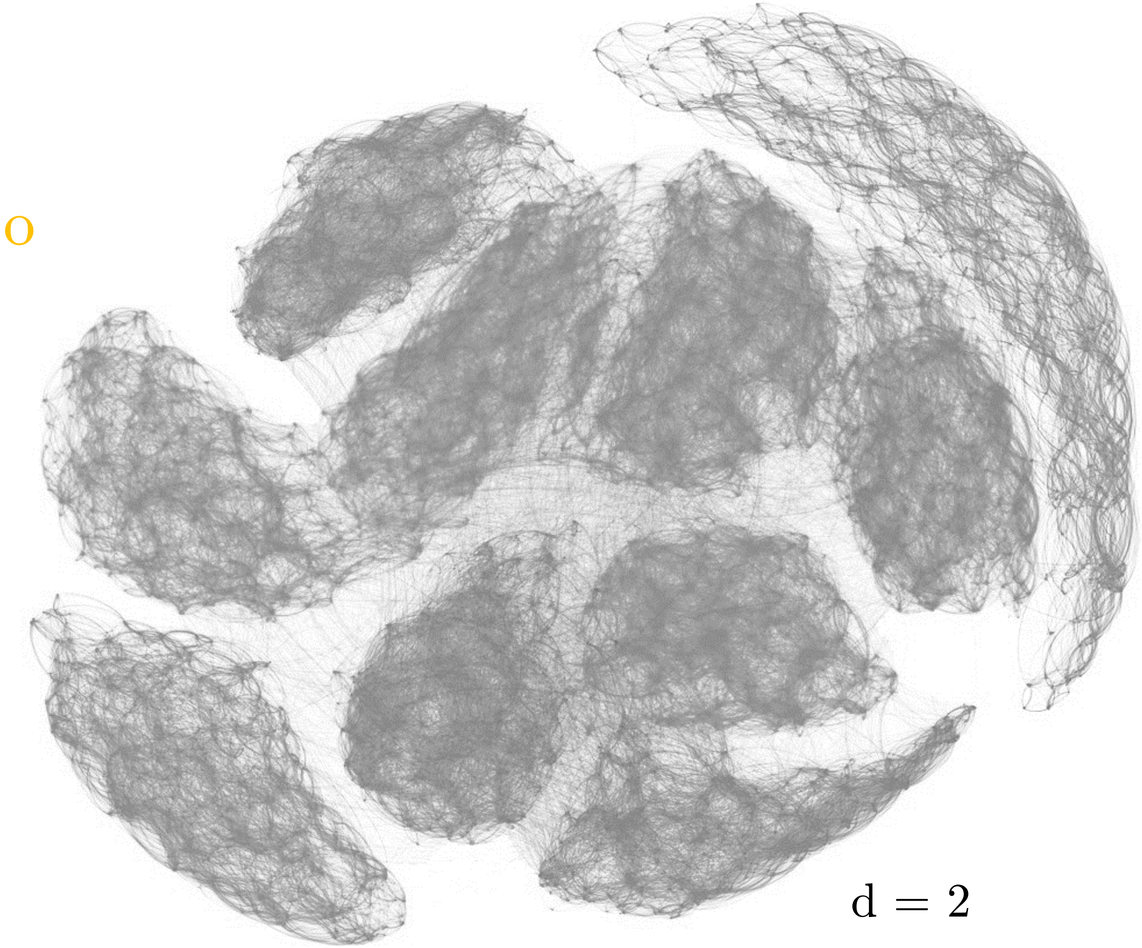
Example: unsupervised learning

Dimensionality reduction: visualize high-dimensional data in 2D or 3D



d = 28x28 pixels = 784

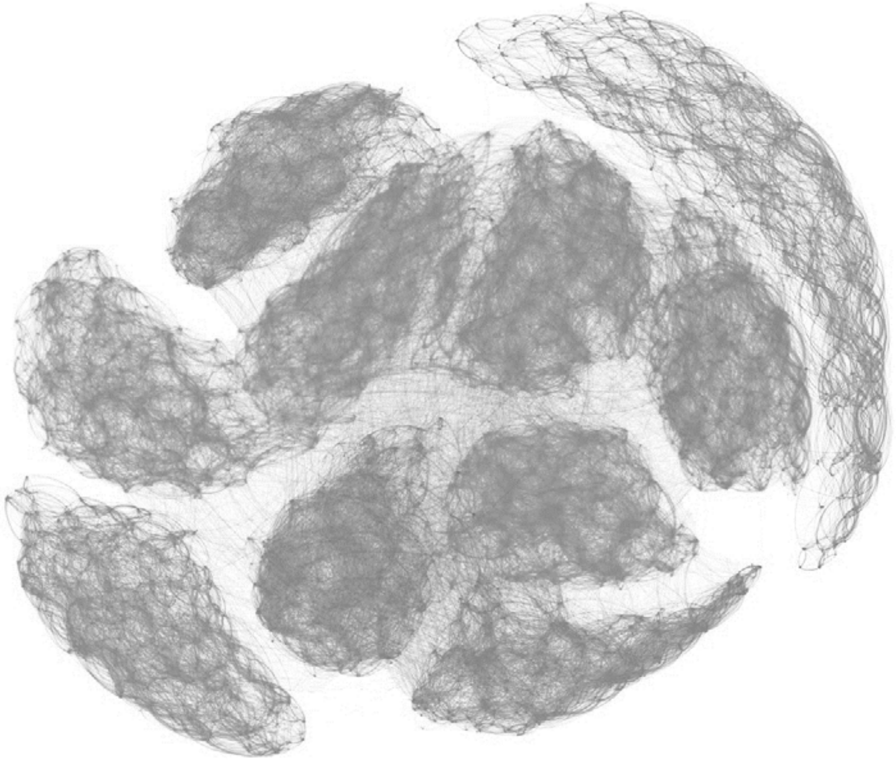
Find a way to draw graph nicely



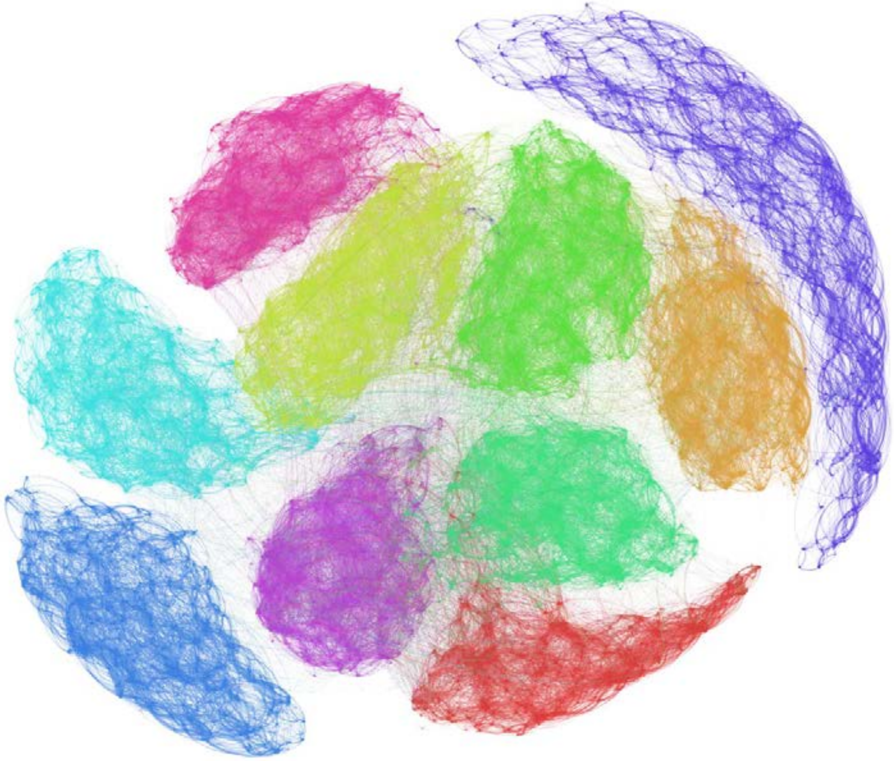
d = 2

Example: unsupervised learning

Cluster points into groups

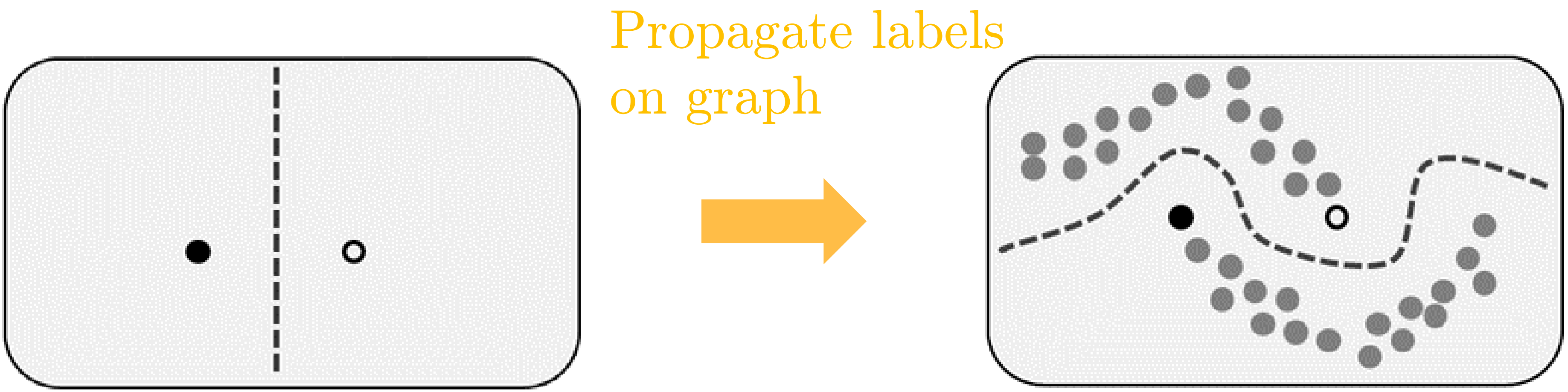


Partition graph



Example: semi-supervised learning

Classify the missing labels by taking into account unlabeled data.

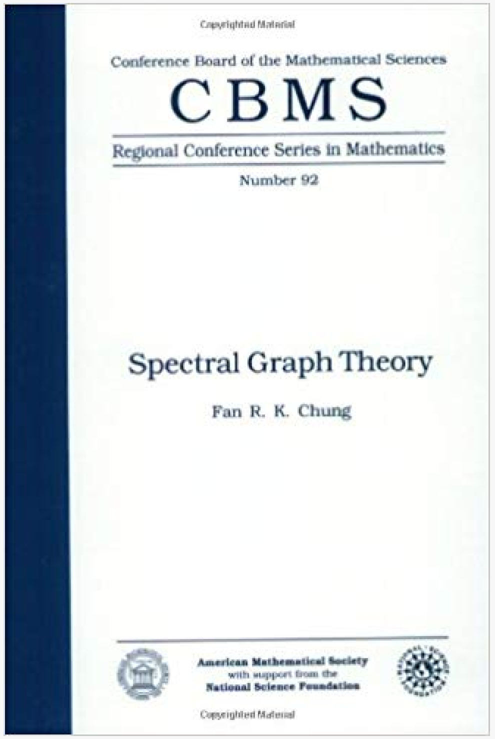


Supervised learning

Adding unlabeled data
reveals local geometry

Outline of lecture

- Relational reasoning in high-dimensions
  - Distances in high-dimensions
  - The manifold assumption
  - Examples in machine learning
- Spectral graph theory 101
  - Gradient on a graph
  - Laplacian matrices
  - The Laplacian spectrum
  - Bottlenecks and diameter



Based on Chapters 1 and 2

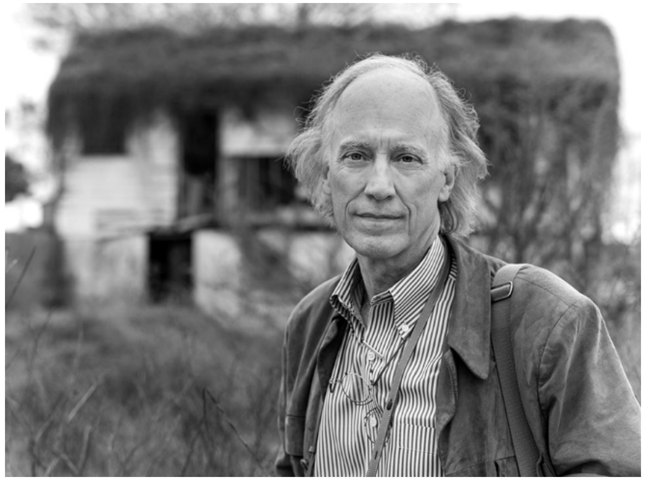
Spectral graph theory

- Emerged in the 50s and 60s in discrete mathematics.

- Combines graph theory and linear algebra

  Study of the properties of a graph in relationship to the eigenvalues and eigenvectors of matrices associated with the graph, such as its adjacency matrix or Laplacian matrix.

- Connection to differential geometry, but for irregular and discrete spaces.



Graph as a matrix

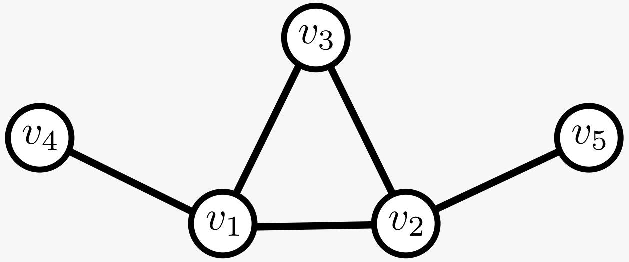
A graph $G = (V, E, W)$, where

- $V = \{v_1, v_2, \dots, v_N\}$ is the set of N nodes (vertices)
- $E = \{e_1, e_2, \dots, e_M\}$ is the set of M edges and $e_l = (v_i, v_j)$.
- W is the $N \times N$ weighted adjacency matrix, defined as:

$$W(i,j) := \begin{cases} 0 & \text{if } v_i = v_j \\ w_{ij} & \text{if } (v_i, v_j) \in E \\ 0 & \text{if } (v_i, v_j) \notin E. \end{cases}$$

Properties:

- $w_{ij} > 0$ is the weight of the edge between v_i and v_j,
- if G is undirected, we have $W = W^\top$
- If G is unweighted (i.e., $w_{ij} = 1$), we write A and call it adjacency matrix (what does $A^k(i,j)$ compute?)



$$W = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The gradient on a graph

Let vector $f = (f(1), f(2), \dots, f(N)) \in \mathbb{R}^N$ be a function on V.

The gradient of f is a function defined on the edges: $Sf = \begin{bmatrix} \frac{\partial f}{\partial e_1} \\ \vdots \\ \frac{\partial f}{\partial e_M} \end{bmatrix}$

Along $e_l = (v_i, v_j)$, the gradient is

$$\frac{\partial f}{\partial e_l} = \begin{cases} \sqrt{w_{ij}}\, (f(i) - f(j)) & \text{if} \quad i < j \\ -\sqrt{w_{ij}}\, (f(i) - f(j)) & \text{if} \quad i > j. \end{cases}$$

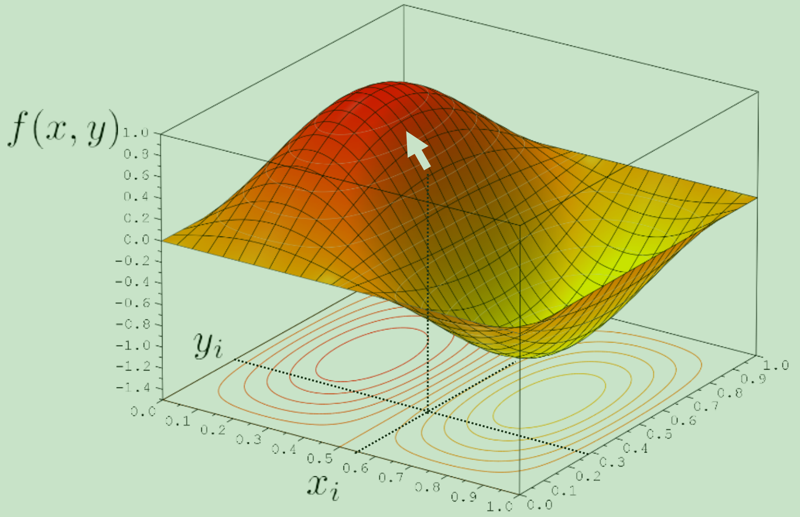
The computation is done using the $M \times N$ incidence matrix

$$S(l,:) := \begin{array}{cccccccccc} [\cdots & 0 & \sqrt{w_{ij}} & 0 & \cdots & 0 & -\sqrt{w_{ij}} & 0 & \cdots] \\ & & \text{i} & & & & \text{j} & & \end{array}$$

such that $(Sf)(l) = S(l,:)f = \frac{\partial f}{\partial e_l}$

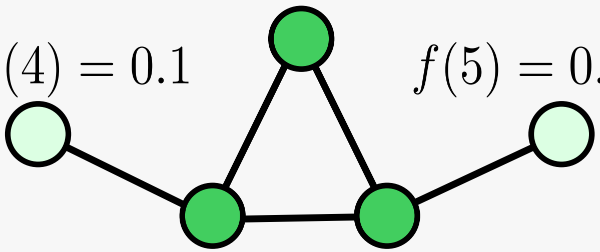
The gradient of a function $f : \mathbb{R}^2 \to \mathbb{R}$ is

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \end{bmatrix} \approx \begin{bmatrix} \frac{f(x_i,y_i) - f(x_j,y_i)}{x_i - x_j} \\ \\ \frac{f(x_i,y_i) - f(x_i,y_i)}{y_i - y_j} \end{bmatrix}$$



The gradient on a graph

A worked-out example:



$$S = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{bmatrix} \qquad f = \begin{bmatrix} f(1) \\ f(2) \\ f(3) \\ f(4) \\ f(5) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0.1 \\ 0.1 \end{bmatrix}$$

$$Sf = \begin{bmatrix} f(1) - f(2) \\ f(1) - f(3) \\ f(2) - f(3) \\ f(1) - f(4) \\ f(2) - f(5) \end{bmatrix} = \begin{bmatrix} 1 - 1 \\ 1 - 1 \\ 1 - 1 \\ 1 - 0.1 \\ 1 - 0.1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.9 \\ 0.9 \end{bmatrix}$$

The combinatorial Laplacian

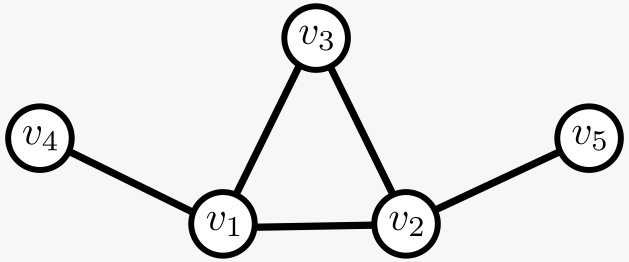
Similar to the continuous case, we can use the gradient to define the $N \times N$ combinatorial Laplacian matrix:

$$L := S^\top S \qquad (\Delta f := \nabla \cdot \nabla f)$$

The entries of L are given by

$$L(i,j) = \begin{cases} d_i & \text{if } v_i = v_j \\ -w_{ij} & \text{if } (v_i, v_j) \in E \\ 0 & \text{if } (v_i, v_j) \notin E. \end{cases} \qquad (d_i := \textstyle\sum_{(v_i,v_j)\in E} w_{ij})$$

- L is symmetric ($L = L^\top$)
- In matrix form, we have $L = D - W$, where $D = \mathrm{diag}(d_1, d_2, \ldots, d_N)$.
- L converges to the Laplace-Beltrami operator on manifold as $N \to \infty$.



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

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

The Laplacian quadratic form

Let vector $f \in \mathbb{R}^N$ be a function with one entry per node.

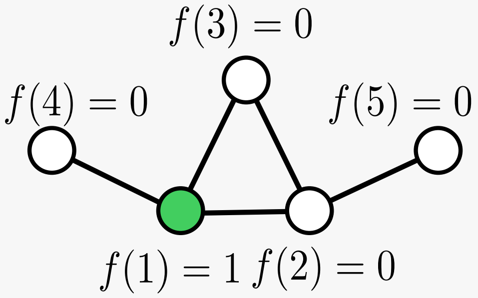
The quadratic form of the Laplacian is given by:

$$f^\top L f = \sum_{l=1}^{M} \left(\frac{\partial f}{\partial e_l} \right)^2 = \sum_{(v_i, v_j) \in E} w_{ij} \left(f(i) - f(j) \right)^2 \quad \text{(proof)}$$

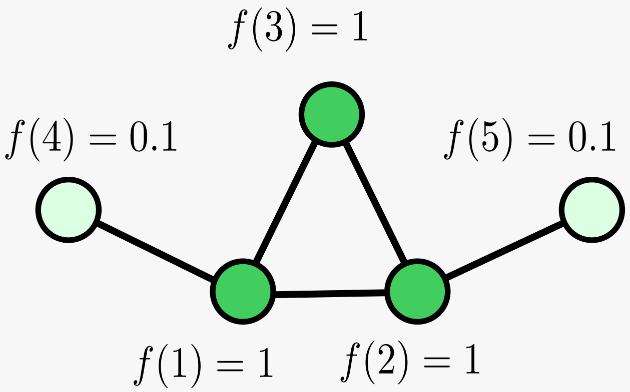
Question:

- What does $f^\top L f$ tell us about function f w.r.t. G?

  It tells us how smooth f is!



$$f^\top L f = 3 \times (1 - 0)^2 = 3$$



$$\begin{aligned} f^\top L f &= (f(1) - f(2))^2 \\ &+ (f(1) - f(3))^2 \\ &+ (f(2) - f(3))^2 \\ &+ (f(1) - f(4))^2 \\ &+ (f(2) - f(5))^2 \\ &= 2 \times (0.9)^2 \\ &= 1.62 \end{aligned}$$

The combinatorial Laplacian spectrum

Eigenvalue equation $Lu_k = \lambda_k u_k$, with

- eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ (non-negative scalars)
- eigenvectors $u_1, u_2, \ldots, u_N$ (each is a function in $\mathbb{R}^N$)

In matrix form, we have

$$L = U \Lambda U^\top = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_N \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix} \begin{bmatrix} - & u_1^\top & - \\ & \cdots & \\ - & u_N^\top & - \end{bmatrix}.$$

Properties:

- $\lambda_1 = 0$ and $u_1 = c\,1_N$ for some constant c (proof)
- Number of connected components of G = number of zero eigenvalues (proof)

Eigenvector interpretation

The min-max theorem (a.k.a. Courant-Fischer-Weyl theorem) says:

$$u_1 = \underset{f \in \mathbb{R}^N, \|f\|=1}{\arg\min} \; f^\top L f \quad \text{and} \quad \lambda_1 = u_1^\top L u_1 = 0$$

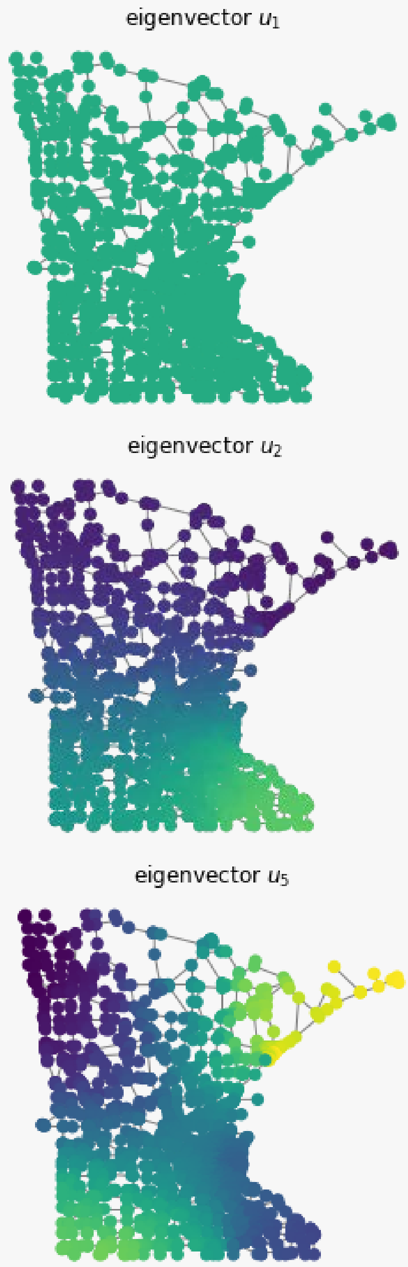
$$u_2 = \underset{f \in \mathbb{R}^N, \|f\|=1, f \perp u_1}{\arg\min} \; f^\top L f \quad \text{and} \quad \lambda_2 = u_2^\top L u_2$$

$$\vdots$$

$$u_k = \underset{f \in \mathbb{R}^N, \|f\|=1, f \perp u_1, \cdots, u_{k-1}}{\arg\min} \; f^\top L f \quad \text{and} \quad \lambda_k = u_k^\top L u_k$$

Observations:

- eigenvectors form an orthonormal basis that goes from smoothest to least-smooth on G
- eigenvalues tell you how smooth eigenvectors are.



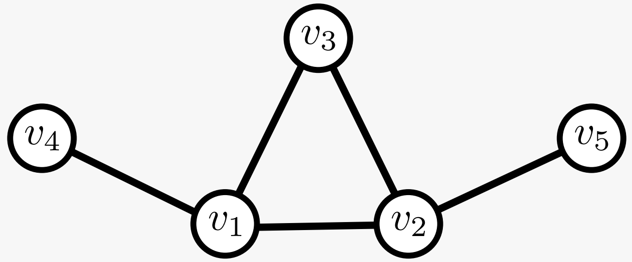
The normalized Laplacian

The $N \times N$ normalized Laplacian matrix is given by:

$$L_n(i,j) := \begin{cases} 1 & \text{if } v_i = v_j \\ -\frac{w_{ij}}{\sqrt{d_i d_j}} & \text{if } (v_i, v_j) \in E \\ 0 & \text{if } (v_i, v_j) \notin E. \end{cases}$$

- $L_n = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$
- $L_n = S_n^\top S_n$, where $S_n = S D^{-\frac{1}{2}}$
- Also, we have:

$$f^\top L_n f = \|S_n f\|_2^2 = \|S D^{-1/2} f\|_2^2$$
$$= \sum_{(v_i,v_j)\in E} w_{ij} \left(\frac{f(i)}{\sqrt{d_i}} - \frac{f(j)}{\sqrt{d_j}} \right)^2 \geq 0$$



$$L_n = \begin{bmatrix} 1 & \frac{-1}{3} & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{3}} & 0 \\ \frac{-1}{3} & 1 & \frac{-1}{\sqrt{6}} & 0 & \frac{-1}{\sqrt{3}} \\ \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & 1 & 0 & 0 \\ \frac{-1}{\sqrt{3}} & 0 & 0 & 1 & 0 \\ 0 & \frac{-1}{\sqrt{3}} & 0 & 0 & 1 \end{bmatrix}$$

The normalized Laplacian spectrum

Eigenvalue equation $L_n u_k = \lambda_k u_k$ is equivalent to $L u_k = \lambda_k D u_k$.

- eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N \leq 2$
- eigenvectors $u_1, u_2, \ldots, u_N$ (these are functions in $\mathbb{R}^N$)

Properties:

- $\lambda_1 = 0$ and $u_1 = \dfrac{D^{1/2}1}{\|D^{1/2}1\|_2} = \begin{bmatrix} \sqrt{\frac{d_1}{\mathrm{vol}(G)}} \\ \vdots \\ \sqrt{\frac{d_N}{\mathrm{vol}(G)}} \end{bmatrix}$, with $\mathrm{vol}(G) := \sum_i d_i$ (proof)
- Number of connected components of G = number of zero eigenvalues (same proof as before)
- $\lambda_N \leq 2$ and equality is met when G is bipartite (partial proof)

The Cheeger constant

Cuts play a fundamental role in graph theory:

- Let $T \subset V$ a set of nodes and $\bar{T} := V - T$ its complement.
- The cut induced by T is defined as $w(T, \bar{T}) := \sum_{v_i \in T, v_j \in \bar{T}} w_{ij}$
- The volume of a set is $\mathrm{vol}(T) := \sum_{v_i \in T} d_i$

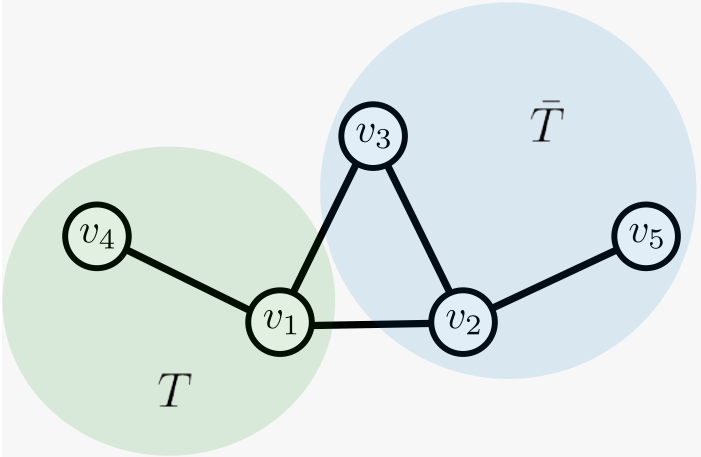
The conductance of a set T on G is:

$$h_G(T) := \frac{w(T, \bar{T})}{\min\{\mathrm{vol}(T), \mathrm{vol}(\bar{T})\}}$$

The Cheeger constant

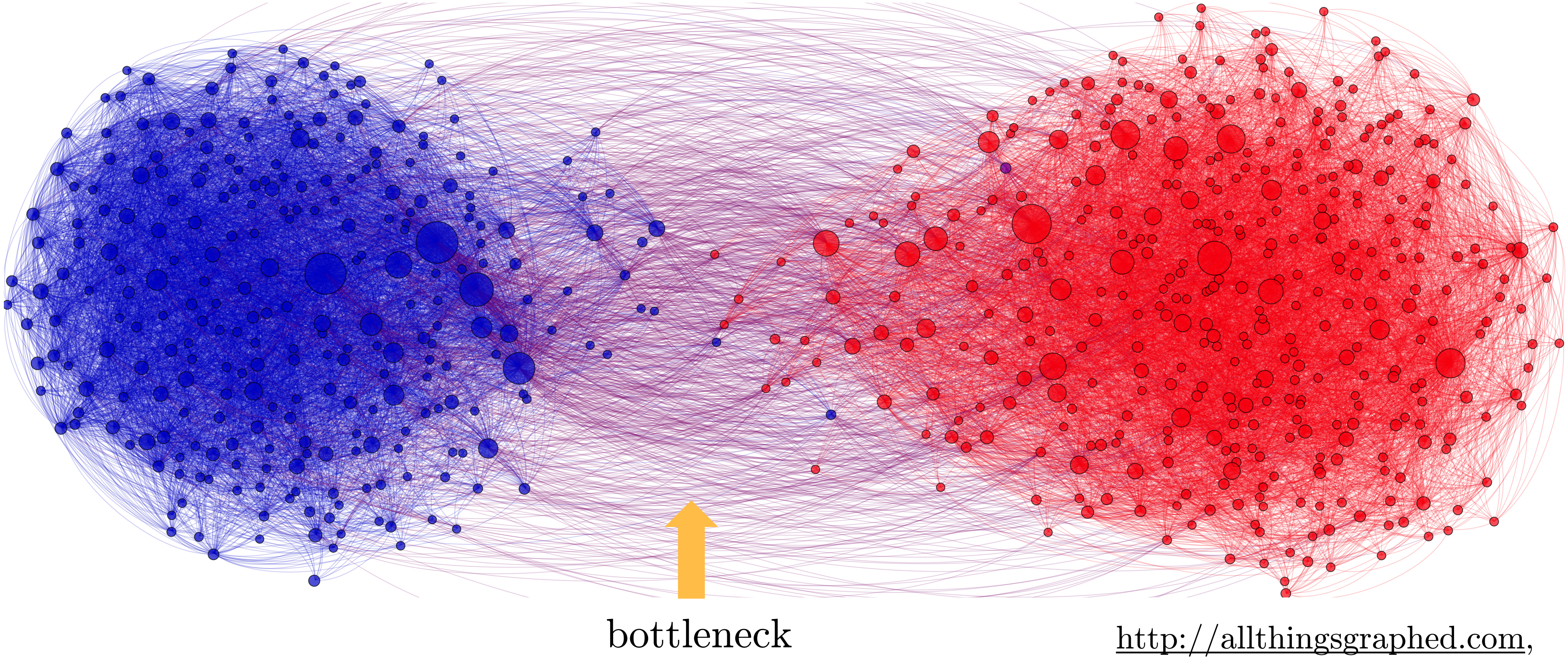
$$h_G := \min_{T \subset V} h_G(T).$$

measures the presence of a bottleneck.



$$h_G(T) = \frac{w(T, \bar{T})}{\mathrm{vol}(T)} = \frac{w_{12} + w_{13}}{d_1 + d_4} = \frac{2}{4} = \frac{1}{2}$$

Cuts and bottlenecks



http://allthingsgraphed.com,
data by Lada Adamic 2004

Bottlenecks and algebraic connectivity

Theorem 1 (Cheeger inequality (Polya, Szego)).

$$\frac{\lambda_2}{2} \leq h_G \leq \sqrt{2\lambda_2}$$

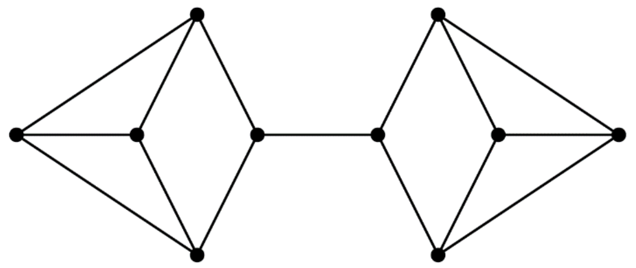
Theorem 2. *If $w_{ij} = 1$ for all $(v_i, v_j) \in E$, we have*

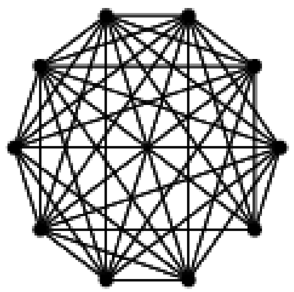
$$diam(G) \geq \frac{1}{\lambda_2 \, vol(G)}$$

graph almost disconnected,
large bottlenecks,
large diameter

$0 \longleftarrow \lambda_2 \longrightarrow 1$

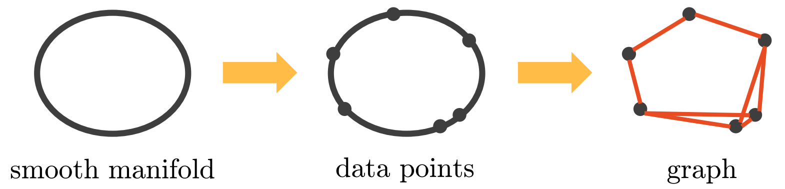
graph almost fully-connected,
no bottlenecks,
small diameter

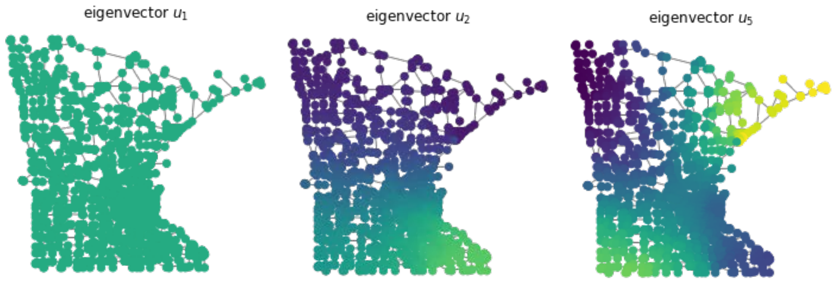




Summary

- In high dimensions, distances are not relevant!
- Thinking of points as living on a high-dimensional surface.
- Graphs can be seen as discrete approximations of the underlying manifold.





Spectral graph theory

- Differential geometry for graphs
- Define graph gradient, Laplacian matrices
- Eigenvectors are the smoothest functions
- Spectrum tells us something about our graph:
  - bottlenecks
  - diameter