Lecture 11: Graph clustering and random walks

John Sylvester Nicolás Rivera Luca Zanetti Thomas Sauerwald

Lent 2019

UNIVERSITY OF CAMBRIDGE

What is clustering?

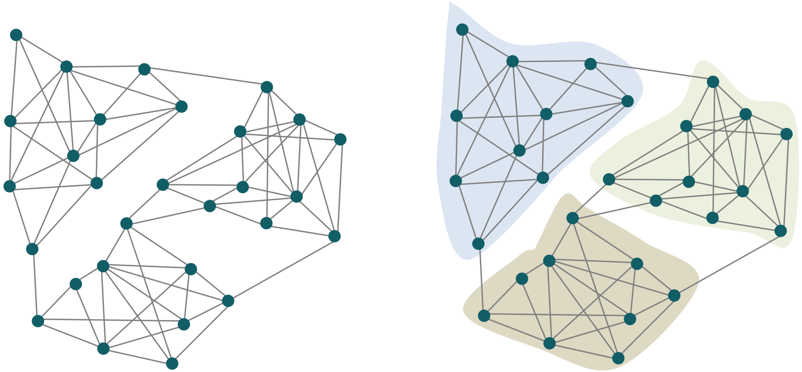
Clustering is the task of dividing objects in groups (clusters) so that similar objects are grouped together and dissimilar objects are separated in different groups

Different formalisations for different domains/applications:

- Geometric clustering: partition points in a Euclidean space
  - k-means, k-medians, k-centres, etc.
- Graph clustering: partition vertices in a graph
  - modularity, conductance, min-cut, etc.

Graph clustering

Partition the graph into pieces (clusters) so that vertices in the same piece have, on average, more connections among each other than with vertices in other clusters

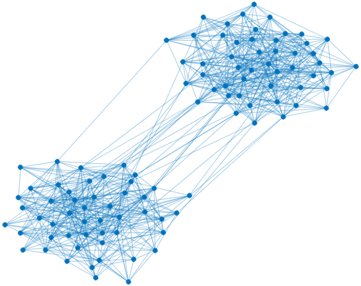


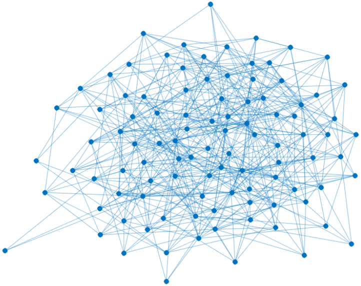
Why study graph clustering?

- Many practical applications, e.g.:
  - Community detection
  - Group webpages according to their topics
  - Find proteins performing the same function within a cell
  - Image segmentation
  - Identify bottlenecks in a network
- Connections with different areas of mathematics and TCS, e.g.:
  - Random walk theory
  - Combinatorics
  - Theory of metric spaces
  - Approximation algorithms
  - Complexity theory

Relation between clustering and mixing

- Which graph has a "cluster-structure"?
- Which graph mixes faster?





Weighted graphs and random walks

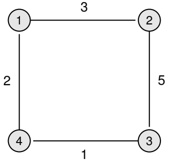
$G = (V, E, w)$ with weight function w, s.t.

- $w\colon V \times V \to \mathbb{R}_{\geq 0}$
- $w(x, y) > 0 \iff \{x, y\} \in E$
- $w(x, y) = w(y, x)$

The transition matrix of a lazy random walk on G is the n by n matrix P defined as

$$P(x, y) = \frac{w(x, y)}{2d(x)}, \qquad P(x, x) = \frac{1}{2}$$

where $d(x) = \sum_{z \in V} w(x, z)$.
It has stationary distribution π s.t. $\pi(x) = \frac{d(x)}{\sum_z d(z)}$.



$$P = \begin{pmatrix} \frac{1}{2} & \frac{3}{10} & 0 & \frac{1}{5} \\ \frac{3}{16} & \frac{1}{2} & \frac{5}{16} & 0 \\ 0 & \frac{5}{12} & \frac{1}{2} & \frac{1}{12} \\ \frac{1}{3} & 0 & \frac{1}{6} & \frac{1}{2} \end{pmatrix}$$

How do we formalise the concept of cluster/bottleneck?

Enter the conductance

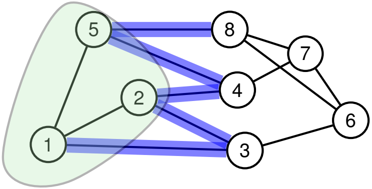
Let $G = (V, E, w)$ and $\emptyset \neq S \subset V$.
The conductance (edge expansion) of S is

$$\phi(S) := \frac{w(S, V \setminus S)}{\min\{\mathrm{vol}(S), \mathrm{vol}(V \setminus S)\}}$$

where $w(S, V \setminus S) = \sum_{x \in S, y \notin S} w(x, y)$ and $\mathrm{vol}(S) = \sum_{x \in S} d(x)$.
The conductance of G is

$$\phi(G) := \min_{\emptyset \neq S \subset V} \phi(S)$$

NP-hard to compute!



- $\phi(S) = \frac{5}{9}$
- $\phi(G) \in [0, 1]$ and $\phi(G) = 0$ iff G is disconnected
- If G is a complete graph, then $|E(S, V \setminus S)| = |S| \cdot (n - |S|)$ and $\phi(G) \approx 1/2$.

Cheeger's inequality

Cheeger's inequality

Let P be the transition matrix of a lazy random walk of a graph $G = (V, E, w)$ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Then,

$$\frac{1-\lambda_2}{2} \leq \phi(G) \leq \sqrt{2(1-\lambda_2)}.$$

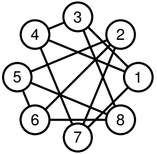
Spectral partitioning:

1. Let f_2 be the eigenvector corresponding to λ_2.
2. Order the vertices so that $f_2(u_1) \leq f_2(u_2) \leq \cdots \leq f_2(u_n)$
3. Try all $n-1$ sweep cuts $(\{u_1, u_2, \ldots, u_k\}, \{u_{k+1}, \ldots, u_n\})$ and return the one with smallest conductance

- It returns $S \subset V$ such that $\phi(S) \leq \sqrt{(1-\lambda_2)} \leq 2\sqrt{\phi(G)}$
- no constant factor approximation (in the worst case)
- mixing on G is $t_{mix} = O(\log(n)/\phi(G)^2)$.

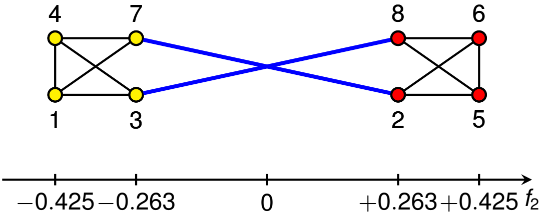
Illustration on a (very) small example

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} \qquad \mathbf{P} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 & \frac{1}{6} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\ \frac{1}{6} & 0 & \frac{1}{2} & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{6} \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{2} & 0 & 0 & \frac{1}{6} & 0 \\ 0 & \frac{1}{6} & 0 & 0 & \frac{1}{2} & \frac{1}{6} & 0 & \frac{1}{6} \\ 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{2} & 0 & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{2} \end{pmatrix}$$



$1 - \lambda_2 \approx 0.13$

$$f_2 = (-0.425, +0.263, -0.263, -0.425, +0.425, +0.425, -0.263, +0.263)^T$$



Best sweep: 4

Conductance: 0.166

Intuition

Let's start with a simplified example: $G = (V_1 \cup V_2, E)$

- disconnected with connected components supported on V_1 and V_2
- $|V_1| = |V_2| = n$
- regular (i.e., all vertices have the same degree)

We want to find $f\colon V \to \mathbb{R}$ such that $f \perp \mathbf{1}$ minimises

$$1 - \lambda_2 = \min_{\substack{f \in \mathbb{R}^n \setminus \{\mathbf{0}\} \\ f \perp \mathbf{1}}} \frac{\sum_{\{u,v\} \in E} (f(u) - f(v))^2}{\sum_{u \in V} 2d(u) f(u)^2}$$

- If f is constant on V_1 and V_2, $1 - \lambda_2 = 0$ (no edges between V_1 and V_2)
- We want $f \perp \mathbf{1} \implies \sum_u f(u) = 0$
- choose $f(u) = \begin{cases} 1 & \text{if } u \in V_1 \\ -1 & \text{if } u \notin V_2. \end{cases}$
- $f \perp \mathbf{1}$ and $1 - \lambda_2 = 0$

Hope: If $\phi(G)$ is small, a similar construction can give us a small spectral gap

Proof of the "easy" direction $(1-\lambda_2)/2 \le \phi(G)$

Proof: Recall that $1-\lambda_2 = \min_{\substack{f\in\mathbb{R}^n\setminus\{0\}\\ f\perp 1}} \frac{\sum_{\{u,v\}\in E} w(u,v)(f(u)-f(v))^2}{2\sum_{u\in V} d(u)f(u)^2}$

- Take $S\subset V$ minimising $\phi(G)$
- Construct $f\in\mathbb{R}^n$ s.t. $f(u) = \begin{cases} 1/\operatorname{vol}(S) & \text{if } u\in S \\ -1/\operatorname{vol}(V\setminus S) & \text{if } u\notin S. \end{cases}$
- $\langle f_2, 1\rangle_\pi = \sum_u \frac{f(u)d(u)}{2|E|} = (1/2|E|)\cdot\left(\sum_{u\in S}\frac{d(u)}{\operatorname{vol}(S)} - \sum_{u\notin S}\frac{d(u)}{\operatorname{vol}(V\setminus S)}\right) = 0$
- $\sum_{u\in V} d(u)f(u)^2 = \sum_{u\in S}\frac{d(u)}{\operatorname{vol}(S)^2} + \sum_{u\notin S}\frac{d(u)}{\operatorname{vol}(V\setminus S)^2} \ge \frac{1}{\operatorname{vol}(S)}$
- $\sum_{\{u,v\}\in E} w(u,v)(f(u)-f(v))^2 \le \sum_{\substack{\{u,v\}\in E\\ u\in S, v\notin S}} \frac{4w(u,v)}{\operatorname{vol}(S)^2} = \frac{4w(S,V\setminus S)}{\operatorname{vol}(S)^2}$
- $1-\lambda_2 \le \frac{4w(S,V\setminus S)}{2\operatorname{vol}(S)} = 2\phi(S) = 2\phi(G).$

$\square$

Spectral partitioning example