DD2434 Machine Learning, Advanced Course

Module 4 : learning with graphs

Aristides Gionis

argioni@kth.se

KTH Royal Institute of Technology

overview of module 4

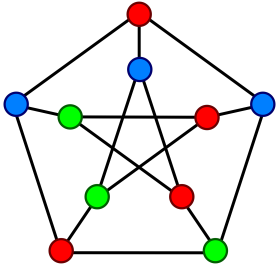
- introduction to mining and learning with graphs
- spectral graph theory
- spectral graph clustering
- graph embeddings

reading material

- von Luxburg. A tutorial on spectral clustering
- Hamilton. Graph representation learning (chapters 1, 2, 3, and 5)

graphs: a simple but universal model

- graph: $G = (V, E)$, set of vertices: V, set of edges: E
- vertices or nodes : represent entities
- edges or links: represent pairwise relations between entities
- basic model can be enriched with considering directions, weights, labels, timestamps, etc., on vertices and edges
- graphs are used to model many real-world datasets



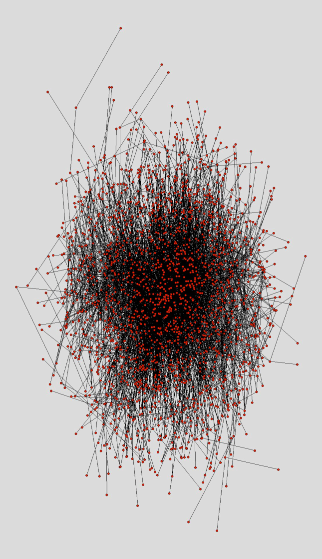
types of networks

- social networks
- knowledge and information networks
- technology networks
- biological networks

social networks

- nodes represent individuals

  links denote social interactions
  - networks of acquaintances
  - collaboration networks
    - actor networks
    - co-authorship networks
    - director networks
  - phone-call networks
  - e-mail networks
  - IM networks
  - sexual networks

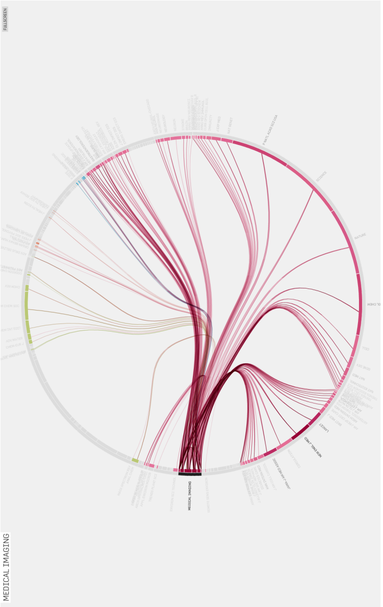


knowledge and information networks

- nodes store information items

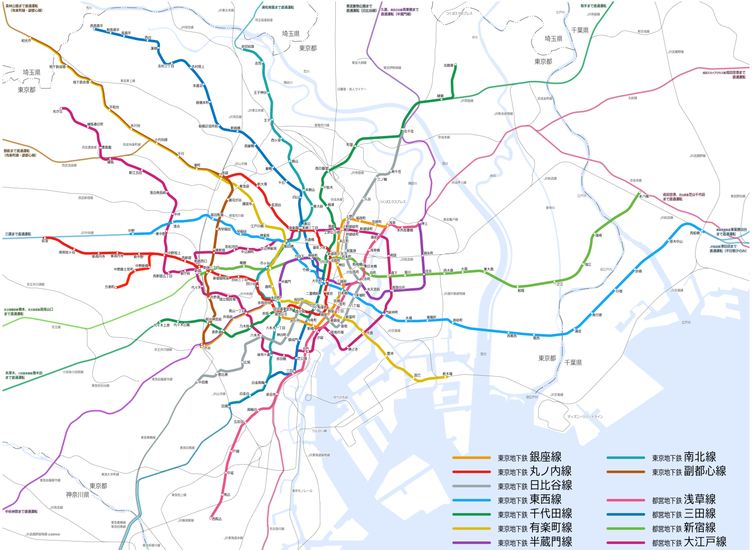
  links associate information between items

  - citation network (directed acyclic)
  - the web (directed)
  - word networks
  - networks of trust (directed)
  - software graphs (directed)
  - home page/blog networks (directed)
  - peer-to-peer networks
  - bluetooth networks

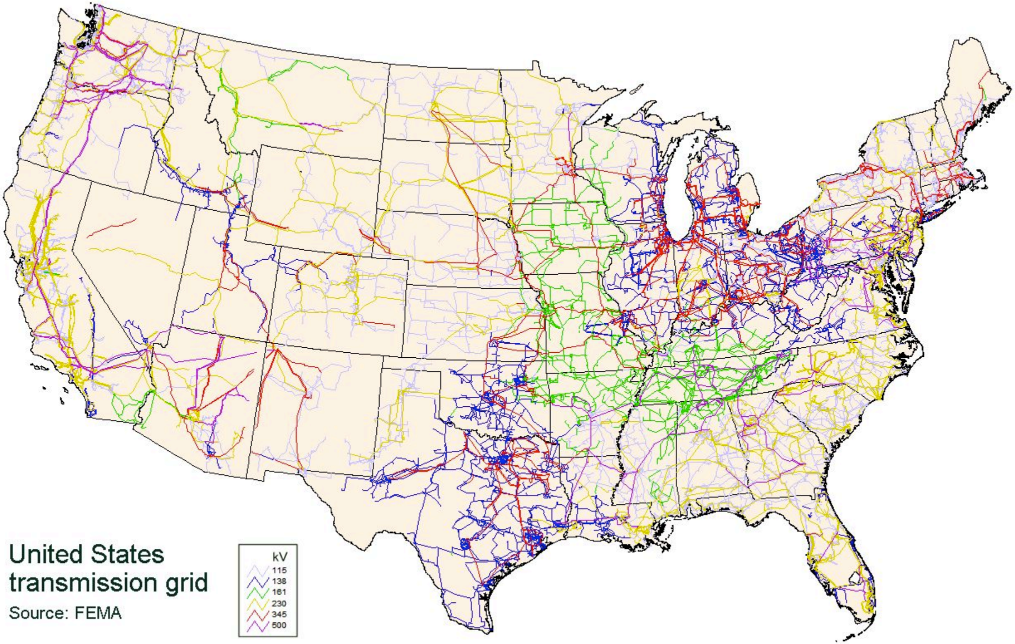


technological networks

- networks built for distribution of a commodity
  - the internet, power grids, telephone networks, airline networks, transportation networks

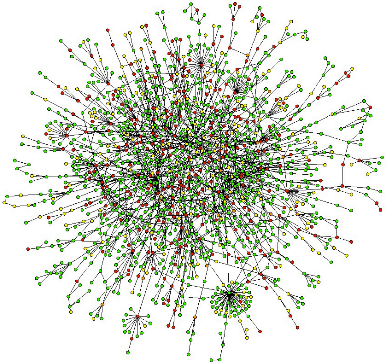


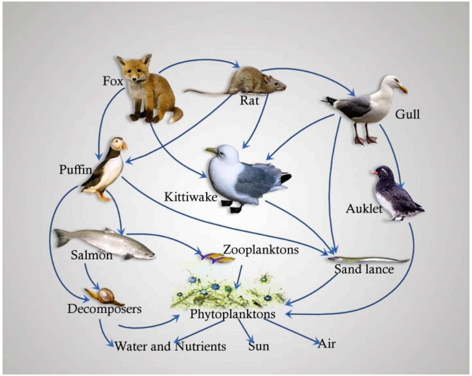
US power grid



biological networks

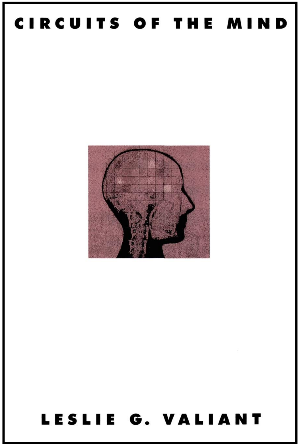
- biological systems represented as networks
  - protein-protein interaction networks
  - gene regulation networks
  - gene co-expression networks
  - metabolic pathways
  - the food web
  - neural networks





brain network

- human brain can be modeled as a complex network
  - how do neurons and neuronal structures connect and communicate ?



network science

- the world is full with networks
- what are the research questions to consider?
  - understand their topology and measure their properties
  - study evolution and dynamics
  - create realistic models
  - design algorithms for learning and making predictions on network data

properties of real-world networks

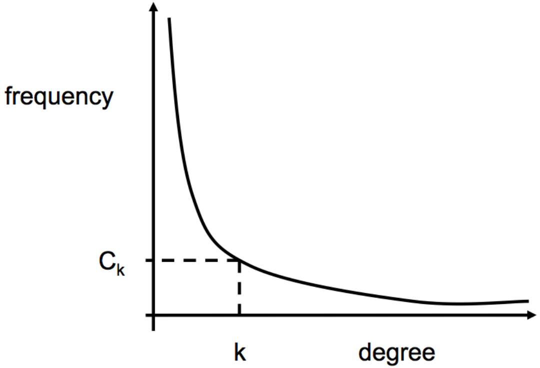
diverse collections of graphs arising from different phenomena

are there typical patterns ? **yes !**

- static networks
  - heavy tails
  - clustering coefficients
  - communities
  - small diameters
- time-evolving networks
  - densification
  - shrinking diameters

degree distribution

- C_k = number of vertices with degree k



- problem : find the probability distribution that fits best the observed data

power-law degree distribution

- C_k = number of vertices with degree k, then

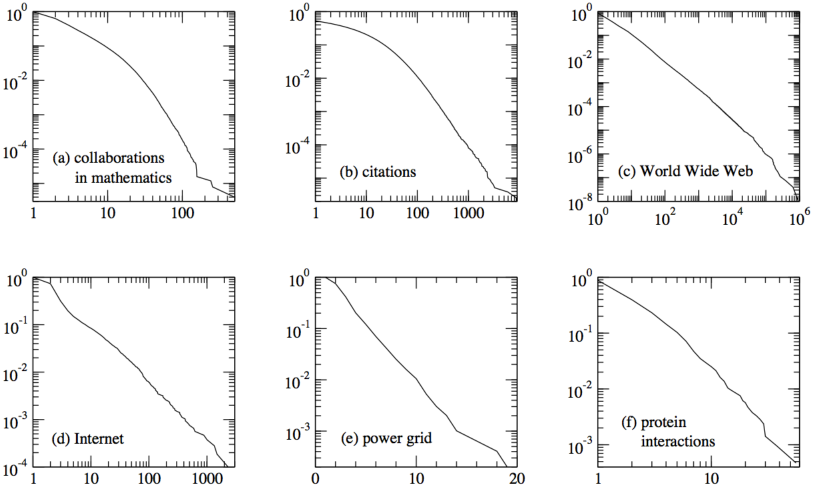
$$C_k = c\, k^{-\gamma}$$

with $\gamma > 1$, or

$$\ln C_k = \ln c - \gamma \ln k$$

- plotting $\ln C_k$ versus $\ln k$ gives a straight line with slope $-\gamma$
- heavy-tail distribution : there is a non-negligible fraction of nodes that have very high degrees (hubs)
- scale free : average is not informative

power-law degree distribution



power-laws in a wide variety of networks

very different from Erdős-Rényi random graphs

clustering coefficients

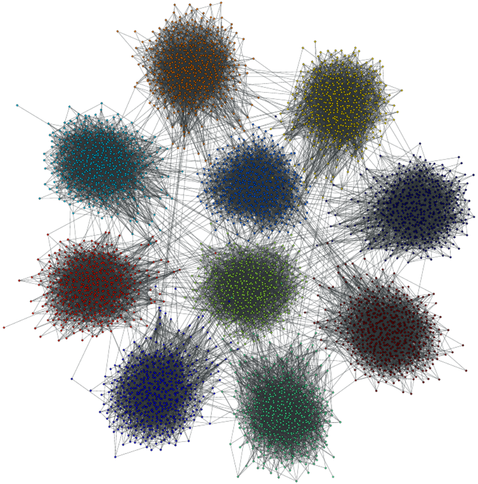
- a proposed measure to capture local clustering is graph transitivity

$$T(G) = \frac{3 \times \{\text{number of triangles in the network}\}}{\{\text{number of connected triples of vertices}\}}$$

- captures "transitivity of clustering"
- if u is connected to v and
  v is connected to w, it is also likely that
  u is connected to w

community structure

informal definition of community: a set of vertices densely connected to each other and sparsely connected to the rest of the graph



artificial communities: `http://projects.skewed.de/graph-tool/`

community structure

- Leskovec et al. study community structure in an collection of real-world networks
- introduce the network community profile (NCP) plot
- characterizes the best possible community over a range of scales

community structure

important findings of Leskovec et al.

1. up to certain size k_0 (~ 100 vertices) there are good cuts
   - as the size increases so does the quality of the community
2. at the size k_0 we observe the best possible community
   - such communities are typically connected to the remainder with a single edge
3. above the size k_0 the community quality decreases
   - this is because they blend in and gradually disappear

small-world phenomena

small worlds : graphs with short paths

- Stanley Milgram (1933-1984)
- conducted a number of famous sociology experiments
- obedience to authority (1963)
- small-world experiment (1967)
  - we live in a small-world

for criticism on the small-world experiment, see *"Could It Be a Big World After All? What the Milgram Papers in the Yale Archives Reveal About the Original Small World Study"* by Judith Kleinfeld

small-world experiments

- letters were handed out to people in Nebraska to be sent to a target in Boston
- people were instructed to pass on the letters to someone they knew on first-name basis
- the letters that reached the destination (64 / 296) followed paths of length around 6
- *six degrees of separation* (term due to John Guare)

  see also :

  – the Kevin Bacon game

  – the Erdős number

small diameter

proposed measures

- diameter : largest shortest-path over all pairs
- effective diameter : upper bound of the shortest path of 90% of the pairs of vertices
- average shortest path : average of the shortest paths over all pairs of vertices
- characteristic path length : median of the shortest paths over all pairs of vertices
- hop-plots : plot of $|N_h(u)|$, the number of neighbors of u at distance at most h, as a function of h

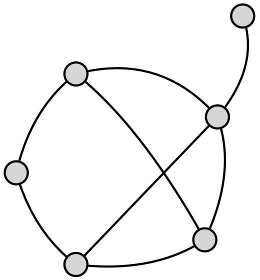
mining and learning with graphs

typical problems

- graph clustering and community detection
- node classification
- link prediction
- link classification

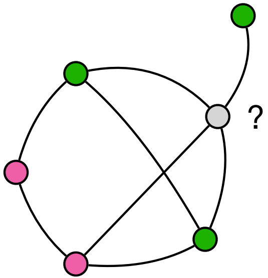
node classification

- use information about graph attributes and graph structure to make predictions about node labels



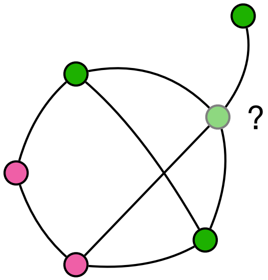
node classification

- use information about graph attributes and graph structure to make predictions about node labels



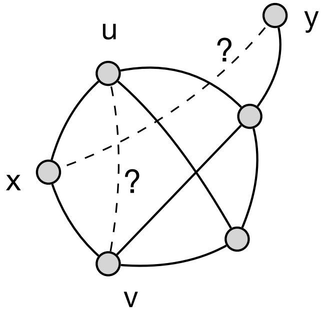
node classification

- use information about graph attributes and graph structure to make predictions about node labels



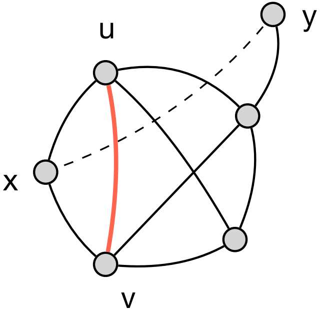
link prediction

- use information about graph attributes and graph structure to make predictions about new links



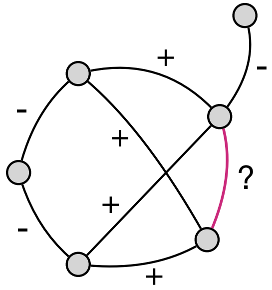
node classification

- use information about graph attributes and graph structure to make predictions about new links



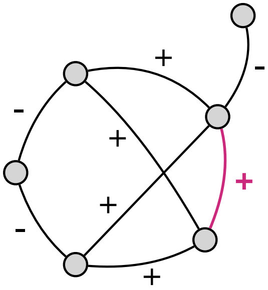
link classification

- use information about graph attributes and graph structure to make predictions about link labels



node classification

- use information about graph attributes and graph structure to make predictions about link labels



summary and discussion

- many real-world data can be modeled as graphs, simple or with attributes
- graphs across different application domains have structural similarities
- many machine learning problems can be formulated in the graph setting

spectral graph theory

objective :

- view the adjacency (or related) matrix of a graph with a linear algebra lens
- identify connections between spectral properties of such a matrix and structural properties of the graph
  - connectivity
  - bipartiteness
  - cuts
  - ...
- spectral properties = eigenvalues and eigenvectors
- in other words, what does the eigenvalues and eigenvectors of the adjacency (or related) matrix tell us about the graph?

once again, eigenvalues and eigenvectors

- consider a real $n \times n$ matrix $\mathbf{A}$, i.e., $\mathbf{A} \in \mathbb{R}^{n \times n}$
- $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{A}$ if there exists $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x} \neq \mathbf{0}$ such that

$$\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}$$

  such a vector $\mathbf{x}$ is called eigenvector of $\mathbf{A}$ corresponding to eigenvalue λ

- alternatively,

$$(\mathbf{A} - \lambda \mathbf{I})\,\mathbf{x} = \mathbf{0} \quad \text{or} \quad det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

  it follows that $\mathbf{A}$ has n eigenvalues (possibly complex and possibly with multiplicity > 1)

once again, eigenvalues and eigenvectors

- consider a real and symmetric $n \times n$ matrix $\mathbf{A}$

  (e.g., the adjacency matrix of an undirected graph)

  then

  – eigenvalues are real and eigenvectors are orthogonal

- $\mathbf{A}$ is positive semi-definite if $\mathbf{x}^T \mathbf{A}\, \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$

- a symmetric positive semi-definite real matrix has real and non negative eigenvalues

variational characterization of eigenvalues

- consider a real and symmetric $n \times n$ matrix $\mathbf{A}$
  - the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ can be ordered

$$\lambda_1 \leq \ldots \leq \lambda_n$$

- the eigenvalues satisfy the following minimax principles w.r.t. Rayleigh quotient

$$\begin{aligned}
\lambda_n &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{A}\, \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \\
\lambda_1 &= \min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{A}\, \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \\
\lambda_2 &= \min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{x}^T \mathbf{x}_1 = 0}} \frac{\mathbf{x}^T \mathbf{A}\, \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \quad \text{and “so on” for the other eigenvalues}
\end{aligned}$$

- very useful way to think about eigenvalues

background: eigenvalues and eigenvectors

- the inverse holds, i.e.,

$$\lambda_1 = \min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \min_{\mathbf{x} \neq \mathbf{0}} \frac{\sum_{ij} A_{ij} x_i x_j}{\sum_i x_i^2}$$

  and if $\mathbf{x}$ is an "optimal vector", then $\mathbf{x}$ is eigenvector of λ_1

  - optimal vector: $\arg\min$ of the expression above

- similarly

$$\lambda_2 = \min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{x}^T \mathbf{x}_1 = 0}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{x}^T \mathbf{x}_1 = 0}} \frac{\sum_{ij} A_{ij} x_i x_j}{\sum_i x_i^2}$$

  and if $\mathbf{x}$ is an "optimal vector", then $\mathbf{x}$ is eigenvector of λ_2

  - optimal vector: $\arg\min$ of the expression above

spectral graph analysis

- apply the eigenvalue characterization for graphs
- consider $G = (V, E)$ an undirected and d-regular graph
  - regular graph is used w.l.o.g. for simplicity of expositions
- question : which matrix to consider ?
  - the adjacency matrix $\mathbf{A}$ of the graph
  - some matrix $\mathbf{B}$ so that $\mathbf{x}^T \mathbf{B}\, \mathbf{x}$ is related to a structural property of the graph
- let $\mathbf{A}$ be the adjacency matrix of G; define the *Laplacian matrix* of G as

$$\mathbf{L} = \mathbf{I} - \frac{1}{d}\mathbf{A} \qquad \text{or} \qquad \mathbf{L}_{ij} = \begin{cases} 1 & \text{if } i = j \\ -1/d & \text{if } (i,j) \in E, i \neq j \\ 0 & \text{if } (i,j) \notin E, i \neq j \end{cases}$$

spectral graph analysis

- for the Laplacian matrix $\mathbf{L} = \mathbf{I} - \frac{1}{d}\mathbf{A}$, we can show that (proof left as exercise)

$$\mathbf{x}^T \mathbf{L}\, \mathbf{x} = \frac{1}{d} \sum_{(u,v)\in E} |x_u - x_v|^2$$

here, x_u is the coordinate of the vector $\mathbf{x}$ that corresponds to vertex $u \in V$

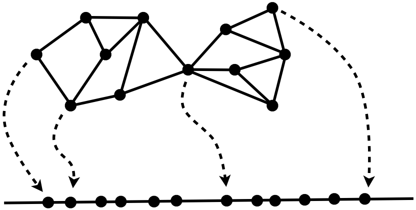
spectral graph analysis

- for the Laplacian matrix $\mathbf{L} = \mathbf{I} - \frac{1}{d}\mathbf{A}$, we can show that (proof left as exercise)

$$\mathbf{x}^T \mathbf{L}\, \mathbf{x} = \frac{1}{d} \sum_{(u,v)\in E} |x_u - x_v|^2$$

  here, x_u is the coordinate of the vector $\mathbf{x}$ that corresponds to vertex $u \in V$

- vector $\mathbf{x}$ is seen as a 1-dimensional embedding
  - i.e., mapping the vertices of the graph onto the real line



the smallest eigenvalue

apply the eigenvalue characterization theorem for $\mathbf{L}$

- what is λ_1 ?

$$\lambda_1 = \min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{L}\, \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \min_{\mathbf{x} \neq \mathbf{0}} \frac{\sum_{(u,v)\in E} |x_u - x_v|^2}{d \sum_{u \in V} x_u^2}$$

- observe that $\lambda_1 \geq 0$
- can it be $\lambda_1 = 0$?
- **yes** : take $\mathbf{x}$ to be the constant vector

the second smallest eigenvalue

apply the eigenvalue characterization theorem for $\mathbf{L}$

- what is λ_2 ?

$$\lambda_2 = \min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{x}^T \mathbf{x}_1 = 0}} \frac{\mathbf{x}^T \mathbf{L} \, \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{x}^T \mathbf{x}_1 = 0}} \frac{\sum_{(u,v) \in E} |x_u - x_v|^2}{d \sum_{u \in V} x_u^2}$$

- can it be $\lambda_2 = 0$?
- $\lambda_2 = 0$ if and only if the graph is disconnected
  - map the vertices of each connected component to a different constant

the k-th smallest eigenvalue

- alternative characterization for λ_k

$$\lambda_k = \min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{x} \in \mathbb{S}_{k-1}^{\perp}}} \max \frac{\sum_{(u,v)\in E} |x_u - x_v|^2}{d \sum_{u \in V} x_u^2}$$

- $\lambda_k = 0$ if and only if the graph has at least k connected components

  where

  - $\mathbb{S}_k$: space spanned by k independent vectors
  - $\mathbb{S}_k^{\perp}$: space orthogonal to $\mathbb{S}_k$

the largest eigenvalue

- what about λ_n ?

$$\lambda_n = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{L}\, \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\sum_{(u,v) \in E} |x_u - x_v|^2}{d \sum_{u \in V} x_u^2}$$

- consider a boolean version of this problem
  - restrict mapping to $\{-1, +1\}$
  - mapping of vertices to $\{-1, +1\}$ corresponds to a graph cut S

$$\lambda_n \geq \max_{\mathbf{x} \in \{-1,+1\}^n} \frac{\sum_{(u,v) \in E} |x_u - x_v|^2}{d \sum_{u \in V} x_u^2}$$

'$\geq$' because $\mathbf{x} \in \{-1, +1\}^n$ is more restricted case

the largest eigenvalue

- consider the graph cut S defined by the mapping of vertices to $\{-1,+1\}$, then :

$$
\begin{aligned}
\lambda_n &\geq \max_{\mathbf{x}\in\{-1,+1\}^n} \frac{\sum_{(u,v)\in E} |x_u - x_v|^2}{d \sum_{u\in V} x_u^2} \\
&= \max_{S\subseteq V} \frac{4\, E(S, V\setminus S)}{d\, n} \\
&= \max_{S\subseteq V} \frac{4\, E(S, V\setminus S)}{2\,|E|} \\
&= \frac{2\, maxcut(G)}{|E|}
\end{aligned}
$$

where, $E(S,T)$ is the number of edges between $S, T \subseteq V$

- it follows that if G is bipartite then $\lambda_n \geq 2$
  - because if G is bipartite, there exists S that cuts all edges

the largest eigenvalue

- on the other hand

$$\begin{aligned}\lambda_n &= \max_{\mathbf{x}\neq\mathbf{0}} \frac{\sum_{(u,v)\in E} |x_u - x_v|^2}{d\sum_{u\in V} x_u^2} \\ &= \max_{\mathbf{x}\neq\mathbf{0}} \frac{2d\sum_{u\in V} x_u^2 - \sum_{(u,v)\in E}(x_u + x_v)^2}{d\sum_{u\in V} x_u^2} \\ &= 2 - \min_{\mathbf{x}\neq\mathbf{0}} \frac{\sum_{(u,v)\in E}(x_u + x_v)^2}{d\sum_{u\in V} x_u^2}\end{aligned}$$

- first note that $\lambda_n \leq 2$
- $\lambda_n = 2$ if and only if there is $\mathbf{x}$ such that $x_u = -x_v$ for all $(u, v) \in E$
- $\lambda_n = 2$ if and only if G has a bipartite connected component

summary so far

- eigenvalues and structural properties of G :
  - $\lambda_2 = 0$ if and only if G is disconnected
  - $\lambda_k = 0$ if and only if G has at least k connected components
  - $\lambda_n = 2$ if and only if G has a bipartite connected component

robustness

- how robust are these results ?
- for instance, what if $\lambda_2 = \epsilon$?
  - is the graph G almost disconnected ?

    i.e., does it have small cuts ?
- or, what if $\lambda_n = 2 - \epsilon$?
  - does it have a component that is “close” to bipartite ?

the second eigenvalue

we can rewrite the expression for λ_2 as

$$\lambda_2 = \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}^T\mathbf{x}_1=0}} \frac{\sum_{(u,v)\in E}(x_u - x_v)^2}{d\sum_{u\in V} x_u^2} = \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}^T\mathbf{x}_1=0}} \frac{\frac{1}{d}\sum_{(u,v)\in E}(x_u - x_v)^2}{\frac{1}{2n}\sum_{(u,v)\in V^2}(x_u - x_v)^2}$$

where V^2 is the set of ordered pairs of vertices

why?

$$\sum_{(u,v)\in V^2}(x_u - x_v)^2 = 2n\sum_v x_v^2 - 2\sum_{u,v} x_u x_v = 2n\sum_v x_v^2 - 2\left(\sum_u x_u\right)^2$$

$$\text{and} \quad \sum_u x_u = 0 \quad \text{since} \quad \mathbf{x}^T\mathbf{x}_1 = 0$$

the second eigenvalue

we can further rewrite the expression for λ_2 as follows:

note that we change the argument over which we take min

$$\lambda_2 = \min_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}^T\mathbf{x}_1=0}} \frac{\frac{1}{d}\sum_{(u,v)\in E}(x_u - x_v)^2}{\frac{1}{2n}\sum_{(u,v)\in V^2}(x_u - x_v)^2} = \min_{\mathbf{x}\text{ non const}} \frac{\frac{2}{d}\sum_{(u,v)\in E}(x_u - x_v)^2}{\frac{1}{n}\sum_{(u,v)\in V^2}(x_u - x_v)^2}$$

why?

because for any $\mathbf{y}$ that is not a constant vector,

and having $\mathbf{y}^T\mathbf{x}_1 = \sum_u y_u = s \neq 0$,

the vector $\mathbf{x} = \mathbf{y} - s/n$ achieves the same value

and it satisfies $\mathbf{x}^T\mathbf{x}_1 = \sum_u x_u = 0$ and $\mathbf{x} \neq \mathbf{0}$

the second eigenvalue

we can further rewrite the expression for λ_2 as follows:

note that we change the argument over which we take min

$$\lambda_2 = \min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{x}^T \mathbf{x}_1 = 0}} \frac{\frac{1}{d} \sum_{(u,v) \in E} (x_u - x_v)^2}{\frac{1}{2n} \sum_{(u,v) \in V^2} (x_u - x_v)^2} = \min_{\mathbf{x} \text{ non const}} \frac{\frac{2}{d} \sum_{(u,v) \in E} (x_u - x_v)^2}{\frac{1}{n} \sum_{(u,v) \in V^2} (x_u - x_v)^2}$$

why?

so, for any non-constant $\mathbf{y}$ that minimizes the r.h.s. ratio

there is an $\mathbf{x}$ (with $\mathbf{x}^T \mathbf{x}_1 = 0$ and $\mathbf{x} \neq \mathbf{0}$) that minimizes the l.h.s. ratio

(notice that since $\mathbf{x}$ subtracts the constant s/n from $\mathbf{y}$ the ratios are the same, as the constant s/n cancels out inside the differences)

the second eigenvalue

we can now write

$$\lambda_2 = \min_{\mathbf{x} \text{ non const}} \frac{\frac{2}{nd}\sum_{(u,v)\in E}(x_u - x_v)^2}{\frac{1}{n^2}\sum_{(u,v)\in V^2}(x_u - x_v)^2} = \min_{\mathbf{x} \text{ non const}} \frac{\mathbb{E}_{(u,v)\in E}[(x_u - x_v)^2]}{\mathbb{E}_{(u,v)\in V^2}[(x_u - x_v)^2]}$$

consider again discrete version of the problem, $x_u \in \{0, 1\}$

$$\min_{\substack{\mathbf{x}\in\{0,1\}^n \\ \mathbf{x} \text{ non const}}} \frac{\mathbb{E}_{(u,v)\in E}[(x_u - x_v)^2]}{\mathbb{E}_{(u,v)\in V^2}[(x_u - x_v)^2]} = \min_{S\subseteq V} \frac{n}{d} \frac{E(S,\overline{S})}{|S|\,|\overline{S}|} = usc(G)$$

where $usc(G)$ is the uniform sparsest cut of G defined as

$$usc(G) = \min_{S\subseteq V} \frac{n}{d} \frac{E(S,\overline{S})}{|S|\,|\overline{S}|}$$

uniform sparsest cut

- it can be shown that

$$\lambda_2 \leq usc(G) \leq \sqrt{8\lambda_2}$$

- the first inequality holds because λ_2 is a relaxation to *usc*
- the second inequality is constructive :
- if $\mathbf{x}$ is an eigenvector of λ_2
  - then there is some $t \in V$ such that the cut $(S, V \setminus S) = (\{u \in V \mid x_u \leq x_t\}, \{u \in V \mid x_u > x_t\})$ has cost $usc(S) \leq \sqrt{8\lambda_2}$

conductance

- conductance : another popular measure for cuts
- the conductance of a set $S \subseteq V$ is defined as

$$\phi(S) = \frac{E(S, V \setminus S)}{d|S|}$$

- it expresses the probability to "move out" of S by following a random edge from S
- we are interested in sets of small conductance
- the conductance of the graph G is defined as

$$\phi(G) = \min_{\substack{S \subseteq V \\ 0 \leq S \leq |V|/2}} \phi(S)$$

Cheeger's inequality

- Cheeger's inequality:

$$\frac{\lambda_2}{2} \leq \frac{usc(G)}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

$\Rightarrow$ conductance is small if and only if λ_2 is small

- the two leftmost inequalities are "easy" to show
- the first follows by the definition of relaxation
- the second follows by

$$\frac{usc(S)}{2} = \frac{n}{2d} \frac{E(S, V \setminus S)}{|S||V \setminus S|} \leq \frac{E(S, V \setminus S)}{d|S|} = \phi(S)$$

since $|V \setminus S| \geq n/2$

generalization to non-regular graphs

- $G = (V, E)$ is undirected and non-regular
- let d_u be the degree of vertex u
- define $\mathbf{D}$ to be a diagonal matrix whose u-th diagonal element is d_u
- the *normalized Laplacian matrix* of G is defined

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\,\mathbf{A}\,\mathbf{D}^{-1/2}$$

or

$$\mathbf{L}_{uv} = \begin{cases} 1 & \text{if } u = v \\ -1/\sqrt{d_u\,d_v} & \text{if } (u,v) \in E, u \neq v \\ 0 & \text{if } (u,v) \notin E, u \neq v \end{cases}$$

generalization to non-regular graphs

- with the *normalized Laplacian* the eigenvalue expressions become (e.g., λ_2)

$$\lambda_2 = \min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \langle \mathbf{x}, \mathbf{x}_1 \rangle_D = 0}} \frac{\sum_{(u,v) \in E} (x_u - x_v)^2}{\sum_{u \in V} d_u x_u^2}$$

  where we use weighted inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_D = \sum_{u \in V} d_u x_u y_u$$

summary

- eigenvalues and structural properties of G :
  - $\lambda_2 = 0$ if and only if G is disconnected
  - $\lambda_k = 0$ if and only if G has at least k connected components
  - $\lambda_n = 2$ if and only if G has a bipartite connected component
  - small λ_2 if and only if G is "almost" disconnected (small conductance)

discussion

- deep connections between spectral properties of appropriately-defined matrices and structural properties of a graph
- spectral analysis provides a means to embed the graph nodes into low-dimensional Euclidean spaces
- spectral embeddings can be used to solve a number of different learning problems on graphs
- many other state-of-the-art embeddings are known, going beyond spectral analysis