Opinions and conflict in social networks: models, computational problems and algorithms

Lecture 2: Mathematical background

Bertinoro International Spring School 2022

Aristides Gionis

argioni@kth.se

KTH Royal Institute of Technology

course overview

- lecture 1: introduction
  - polarization in social media; methods for detecting polarization
- lecture 2: mathematical background
  - submodular maximization; spectral graph theory
- lecture 3: methods for mitigating polarization
  - maximizing diversity, balancing information exposure
- lecture 4: signed networks; theory and applications
- lecture 5: opinion dynamics in social networks

why study these topics?

- spectral graph theory:
  - provides tools to reveal structure in graphs
  - e.g., used for community detection, a fundamental concept in social-network analysis
  - many other applications
  - mathematically interesting
- submodular maximization:
  - common type of optimization problem
  - many ML problems are modeled as submodular maximization

studying material

- spectral graph theory: von Luxburg. A tutorial on spectral clustering
- submodular maximization: several online tutorials, e.g., Krause and Guestrin 2009, Bilmes 2014, or Hassani and Karbasi 2020

spectral graph theory

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?

background: 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)

background: 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

background: 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 $u \in V$

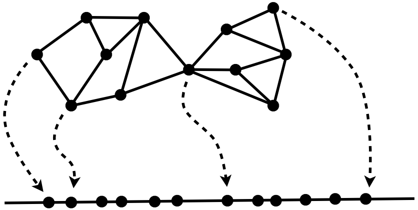
spectral graph analysis

- 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 $\lambda_n \geq 2$, **why ?**
  - because if G is bipartite, there exists S that cuts all edges

the largest eigenvalue

- on the other hand

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

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

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

- 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

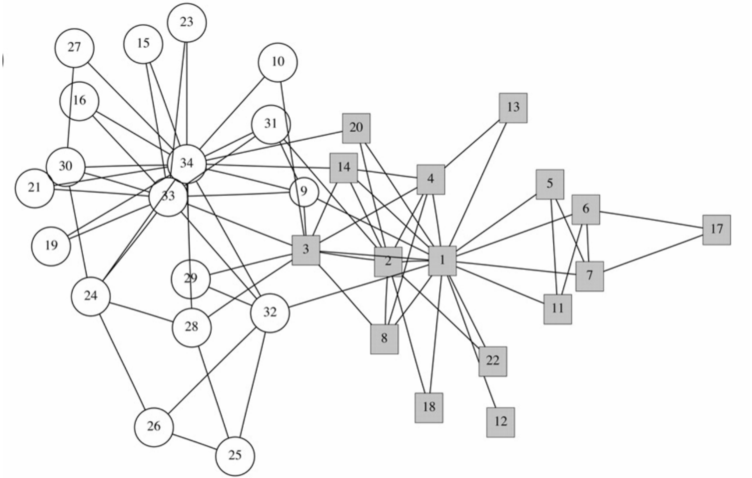
- deep connections between spectral properties of appropriately-defined matrices and structural properties of a graph
- spectral analysis provides a means to embed the graph nodes into low-dimensional Euclidean spaces
- spectral embeddings can be used to solve a number of different learning problems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

graph partitioning and community detection

motivation

- knowledge discovery
  - partition the web into sets of related pages (web graph)
  - find communities of friends in a social network (social graph)
  - find groups of related proteins (protein-protein interaction network)
- performance
  - partition the nodes of a large social network into different machines so that, to a large extent, friends are in the same machine (social networks)

graph partitioning



Zachary's karate-club network, figure from [Newman and Girvan, 2004]

community detection via conductance minimization

- recall conductance
  - the conductance of a set $S \subseteq V$ is defined as

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

  - the conductance of the graph G is defined as

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

  it expresses the probability to "move out" of S by following a random edge from S
  - a set S with small conductance value is a good community
- but there are exponentially many candidate sets S;
  - so, we need efficient and accurate methods
- basic definition works for partitioning the graph in two sets
  - for finding k communities we need to extend the definition

Cheeger's inequality

- Cheeger's inequality:

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

$$\Rightarrow \quad \text{conductance is small if and only if } \lambda_2 \text{ is small}$$

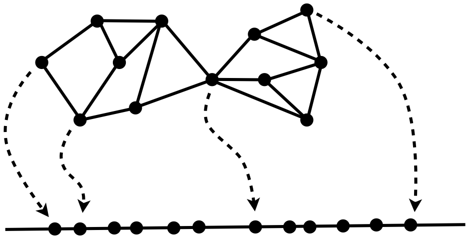
- right hand side inequality involves finding a cut based on the second smallest eigenvector

second smallest eigenvector $\mathbf{x}_2$

- $\mathbf{x}_2$ is also known as Fielder vector
- $\mathbf{x}_2$ is given by

$$\mathbf{x}_2 = \arg \min_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}^T\mathbf{x}_1=0}} \sum_{(u,v)\in E} (x_u - x_v)^2$$

- can be seen as 1-dimensional embedding that minimizes the above “potential function”



basic spectral-partition algorithm

partition nodes in two sets S and $V \setminus S$; measure of quality is conductance $\phi(S)$

1. form normalized Laplacian $\mathbf{L}' = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\,\mathbf{D}^{-1/2}$
2. compute eigenvector $\mathbf{x}_2$ (Fielder vector)
3. order vertices according their coefficient value on $\mathbf{x}_2$
4. consider only sweeping cuts: splits that respect the order
5. take the sweeping cut S that minimizes the conductance measure $\phi(S)$

theorem: the basic spectral-partition algorithm finds a cut S such that $\phi(S) \leq 2\sqrt{\phi(G)}$

proof: by Cheeger inequality, we know that the cut S induced by $\mathbf{x}_2$ satisfies $\phi(S) \leq \sqrt{2\,\lambda_2}$

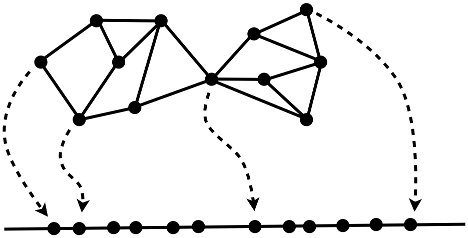
by Cheeger inequality again, we know that $\lambda_2 \leq 2\,\phi(G)$, thus

$$\phi(S) \leq \sqrt{2\,\lambda_2} \leq \sqrt{4\,\phi(G)} \leq 2\sqrt{\phi(G)}$$

spectral partitioning criteria

a number of different partitioning criteria can be used

1. conductance: find the partition that minimizes $\phi(G)$
2. bisection: split in two equal parts
3. sign: separate positive and negative values
4. gap: separate according to the largest gap



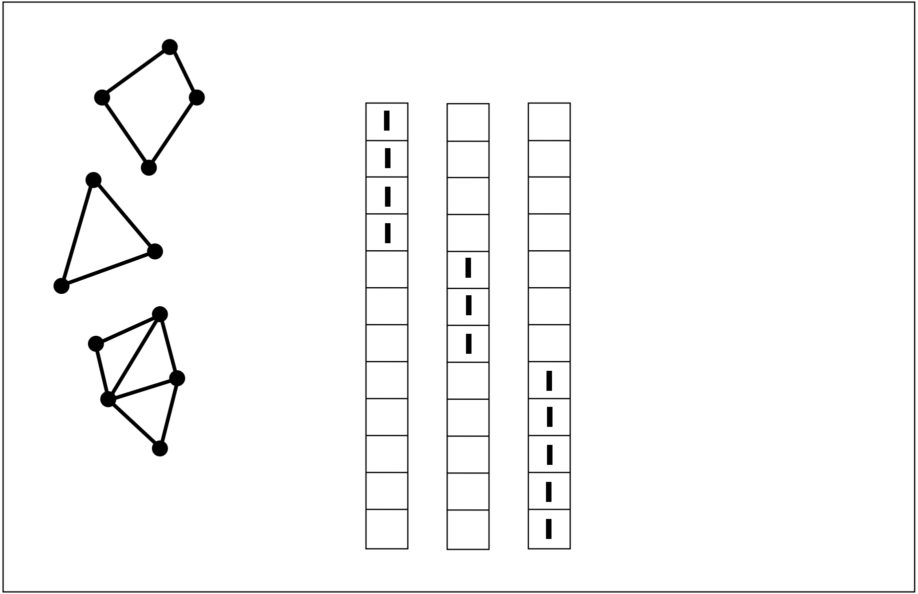
practical spectral-partitioning algorithms

1. utilize more eigenvectors than just the Fielder vector
   - use k eigenvectors
2. different versions of the graph Laplacian matrix

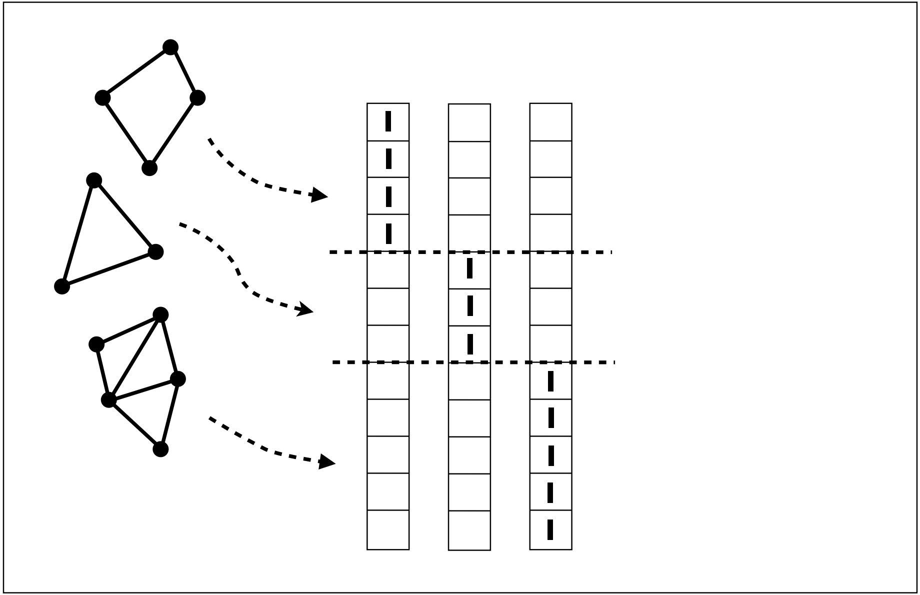
using k eigenvectors

- ideal scenario: the graph consists of k disconnected components (perfect clusters)
- in this case, the multiplicity of the eigenvalue 0 of the graph Laplacian is k
  - the eigenspace of eigenvalue 0 is spanned by indicator vectors of the graph components

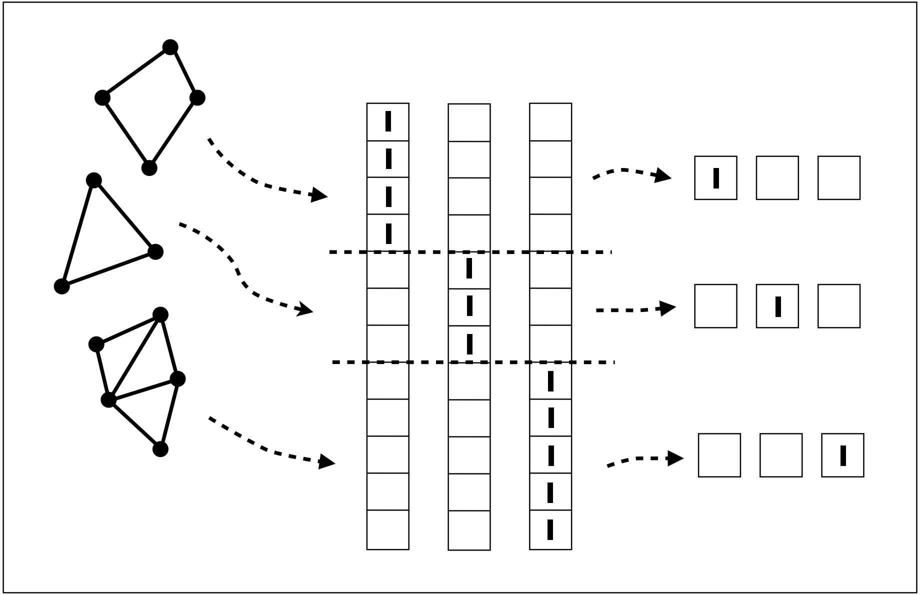
using k eigenvectors



using k eigenvectors



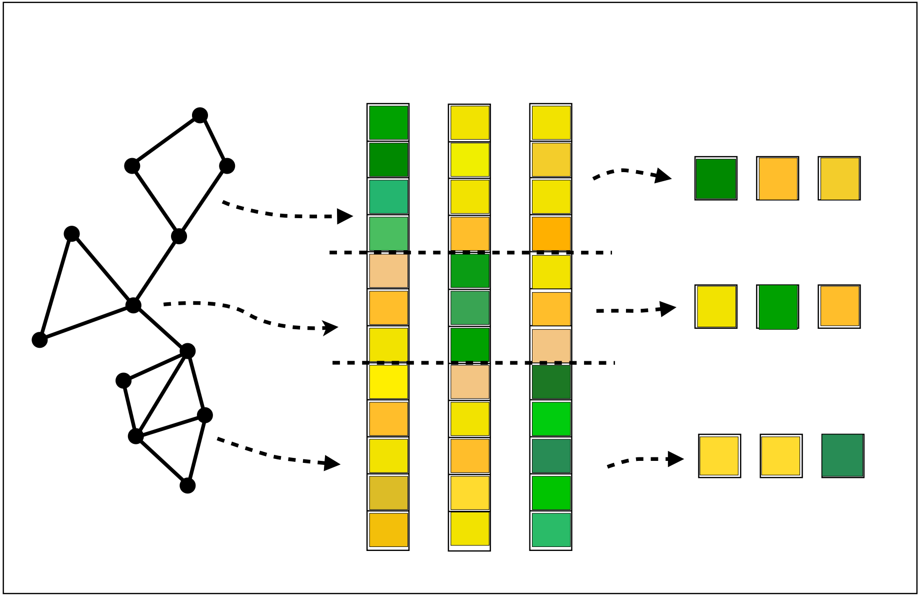
using k eigenvectors



using k eigenvectors

- robustness under perturbations : if the graph has less well-separated components the previous structure holds approximately
- clustering of Euclidean points can be used to separate the components

using k eigenvectors



Laplacian matrices

- normalized Laplacian : $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\,\mathbf{D}^{-1/2}$
- unnormalized Laplacian : $\mathbf{L}_u = \mathbf{D} - \mathbf{A}$
- normalized “random-walk” Laplacian : $\mathbf{L}_{\mathrm{rw}} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}$

all Laplacian matrices are related

- unnormalized Laplacian : $\lambda_2 = \min_{\substack{||\mathbf{x}||=1 \\ \mathbf{x}^T\mathbf{u}_1=0}} \sum_{(i,j)\in E} (x_i - x_j)^2$

- normalized Laplacian :

$$\lambda_2 = \min_{\substack{||\mathbf{x}||=1 \\ \mathbf{x}^T\mathbf{u}_1=0}} \sum_{(i,j)\in E} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2$$

- $(\lambda, \mathbf{u})$ is an eigenvalue/vector of $\mathbf{L}_{\mathrm{rw}}$ if and only if $(\lambda, \mathbf{D}^{1/2}\,\mathbf{u})$ is an eigenvalue/vector of $\mathbf{L}$

- $(\lambda, \mathbf{u})$ is an eigenvalue/vector of $\mathbf{L}_{\mathrm{rw}}$ if and only if $(\lambda, \mathbf{u})$ is a solution to the generalized eigen-problem $\mathbf{L}_u\,\mathbf{u} = \lambda\,\mathbf{D}\,\mathbf{u}$

algorithm 1: unnormalized spectral clustering

input : graph adjacency matrix $\mathbf{A}$, number k

1. form diagonal matrix $\mathbf{D}$
2. form unnormalized Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$
3. compute the first k eigenvectors $u_1, \ldots, u_k$ of $\mathbf{L}$
4. form matrix $\mathbf{U} \in \mathbb{R}^{n \times k}$ with columns $u_1, \ldots, u_k$
5. consider the i-th row of $\mathbf{U}$ as point $y_i \in \mathbb{R}^k, i = 1, \ldots, n,$
6. cluster the points $\{y_i\}_{i=1,\ldots,n}$ into clusters $C_1, \ldots, C_k$

    # e.g., with k-means clustering

output : clusters $A_1, \ldots, A_k$ with $A_i = \{j \mid y_j \in C_i\}$

algorithm 2: normalized spectral clustering

[Shi and Malik, 2000]

input : graph adjacency matrix $\mathbf{A}$, number k

1. form diagonal matrix $\mathbf{D}$
2. form unnormalized Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$
3. compute the first k eigenvectors $u_1, \ldots, u_k$ of the generalized eigen-problem $\mathbf{L}\,\mathbf{u} = \lambda\,\mathbf{D}\,\mathbf{u}$

   # i.e., eigenvectors of $\mathbf{L}_{\mathrm{rw}}$
4. form matrix $\mathbf{U} \in \mathbb{R}^{n \times k}$ with columns $u_1, \ldots, u_k$
5. consider the i-th row of $\mathbf{U}$ as point $y_i \in \mathbb{R}^k, i = 1, \ldots, n$,
6. cluster the points $\{y_i\}_{i=1,\ldots,n}$ into clusters $C_1, \ldots, C_k$

   # e.g., with k-means clustering

output : clusters $A_1, \ldots, A_k$ with $A_i = \{j \mid y_j \in C_i\}$

algorithm 3: normalized spectral clustering

[Ng et al., 2001]

input : graph adjacency matrix $\mathbf{A}$, number k

1. form diagonal matrix $\mathbf{D}$
2. form normalized Laplacian $\mathbf{L}' = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\,\mathbf{D}^{-1/2}$
3. compute the first k eigenvectors $u_1, \ldots, u_k$ of $\mathbf{L}'$
4. form matrix $\mathbf{U} \in \mathbb{R}^{n \times k}$ with columns $u_1, \ldots, u_k$
5. normalize $\mathbf{U}$ so that rows have norm 1
6. consider the i-th row of $\mathbf{U}$ as point $y_i \in \mathbb{R}^k, i = 1, \ldots, n$
7. cluster the points $\{y_i\}_{i=1,\ldots,n}$ into clusters $C_1, \ldots, C_k$

   \# e.g., with k-means clustering

output : clusters $A_1, \ldots, A_k$ with $A_i = \{j \mid y_j \in C_i\}$

notes on the spectral algorithms

- quite similar except for using different Laplacians
- can be used to cluster any type of data, not just graphs
  - form all-pairs similarity matrix and use as adjacency matrix
- computation of the first eigenvectors of sparse matrices can be done efficiently using the Lanczos method

which Laplacian to use?

[von Luxburg, 2007]

- when graph vertices have about the same degree all Laplacians are about the same
- for skewed degree distributions normalized Laplacians tend to perform better
- normalized Laplacians are associated with conductance, which is a good objective
  - conductance involves $vol(S)$ rather than $|S|$ and captures better the community structure

summary

- spectral analysis reveals structural properties of a graph
- used for graph partitioning, but also for other problems
- well-studied area, many results and techniques
- for graph partitioning and community detection many other methods are available

submodular maximization

submodular set functions

- consider a ground set U
- a function $f : 2^U \to \mathbb{R}$ is submodular if

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$

  for all $A, B \subseteq U$

- equivalently ("diminishing returns")

$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B)$$

  for all $A \subseteq B \subseteq U$ and $x \in U \setminus B$

submodular set functions

may or not satisfy the following properties

- non-negative : $f(A) \geq 0$ for all $A \subseteq U$
- monotone : $f(A) \leq f(B)$ for all $A \subseteq B \subseteq U$
- symmetric : $f(A) = f(U \setminus A)$ for all $A \subseteq U$

example: coverage in set systems

- $S_1, \ldots, S_n$ subsets of U
- function $f : 2^{\{1,\ldots,n\}} \to \mathbb{R}_+$
- coverage :

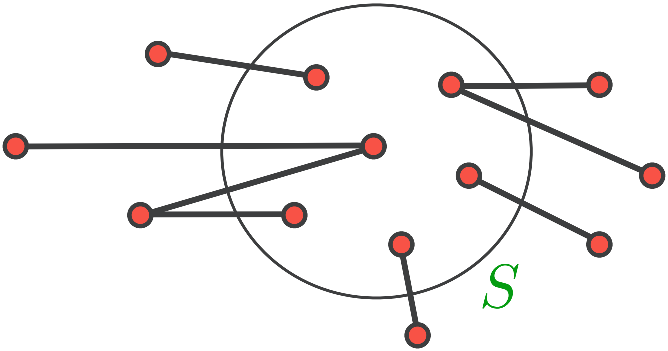
$$f(A) = |\cup_{i \in A} S_i|$$

- weighted coverage :

$$w : U \to \mathbb{R}_+ \quad \text{and} \quad f(A) = \sum_{x \in \cup_{i \in A} S_i} w(x)$$

example: cut in graphs

- consider undirected graph $G = (V, E)$
- cut function $f : 2^V \to R_+$ defined as $f(S) = |E(S, V \setminus S)|$



in the previous examples

- coverage in set systems

  $\Rightarrow$ monotone and non-negative

- cut functions in undirected graphs and hypergraphs

  $\Rightarrow$ symmetric and non-negative

- cut functions in directed graphs

  $\Rightarrow$ non-negative

the maximization problem

- given submodular function $f : 2^U \to \mathbb{R}$

  find $S \subseteq X$ to maximize $f(S)$

  subject to constraints

- value-oracle model
- generalizes many interesting problems **NP**-hard problems
- minimization problem is polynomial (e.g., min-cut)

monotone functions

- $f(U)$ trivial maximizer

- more interesting to maximize under cardinality constraints
- find $S \subseteq U$ subject to $|S| \leq k$ that maximizes $f(S)$

- MAX k-COVER is a special case
- greedy gives $(1 - 1/e)$ approximation [Nemhauser et al., 1978]
- no better approximation unless **P**=**NP**

the greedy algorithm

1. $S \leftarrow \emptyset$
2. while $|S| < k$
3. $\quad i \leftarrow \arg\max_j f(S \cup \{j\})$
4. $\quad S \leftarrow S \cup \{i\}$
5. return S

analysis of the greedy

let $S_j = \{e_1, \ldots, e_j\}$ be first j elements picked by the greedy

$$
\begin{aligned}
f(S^*) &\leq f(S_{i-1}) + \sum_{e \in S^* \setminus S_{i-1}} f_{S_{i-1}}(e) && \text{(monotonicity and submodularity)} \\
&\leq f(S_{i-1}) + \sum_{e \in S^* \setminus S_{i-1}} f(S_i) - f(S_{i-1}) && \text{(greediness)} \\
&\leq f(S_{i-1}) + k(f(S_i) - f(S_{i-1})) && (|S^*| \leq k)
\end{aligned}
$$

it follows

$$f(S_i) - f(S^*) \geq \frac{k-1}{k}(f(S_{i-1}) - f(S^*))$$

which by induction implies

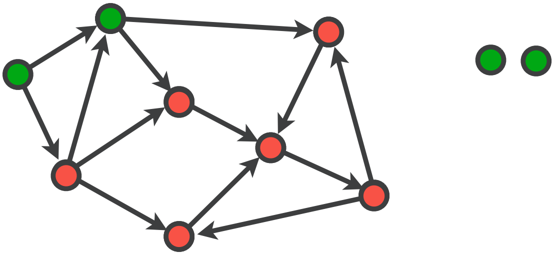
$$f(S_i) \geq \left(1 - (1 - 1/k)^i\right) f(S^*)$$

and so

$$f(S_{greedy}) = f(S_k) \geq \left(1 - (1 - 1/k)^k\right) f(S^*) \geq \left(1 - \frac{1}{e}\right) f(S^*)$$

widely applicable in data mining

- example : maximize the spread of influence in social networks [Kempe et al., 2003]
- assume that an action is spread in a social network
- assume a spreading model such as independent cascade
- find a set of k initial seeds to maximize the spread
- spreading model is randomized, so we want to maximize expected spread



non-monotone functions

- unconstrainted version becomes interesting
- find $S \subseteq X$ to maximize $f(S)$

- a particularly interesting case is the MAX-CUT problem
- what do we know about the approximation of MAX-CUT?
- picking a random set of vertices gives $1/2$ ($1/4$ for MAX-DICUT)
- SDP relaxation technique gives 0.878 (0.796 for MAX-DICUT)

  (major breakthrough) [Goemans and Williamson, 1995]
- 0.53 by spectral approach [Trevisan, 2012]

unconstrainted problem

[Feige et al., 2011]

- first constant-factor approximations for non-negative submodular functions
- $1/2$ approximation for symmetric functions
- $2/5 = 0.4$ approximation for the non-negative functions
- lower bound: better than $1/2$ approximation requires exponential number of value queries

unconstrainted problem

[Feige et al., 2011]

- pick a random set

  $1/4$ for non-negative function (on expectation)

  $1/2$ for symmetric function (on expectation)

- local search
  - initialize S to best singleton
  - $S =$ local optimum (add or delete elements)
  - return the best of S and $U \setminus S$

  $1/3$ approximation for non-negative function

  $1/2$ for non-negative symmetric function

random set analysis

- for $A \subseteq U$, $A(p)$ is a random set where each element of A is selected with prob p
- algorithm returns $R = U(1/2)$
- lemma I

$$E[f(A(p))] \geq (1-p)\, f(\emptyset) + p\, f(A)$$

  can proven by induction on the size of A and using submodularity
- lemma II

$$\begin{aligned} E[f(A(p) \cup B(q))] \quad \geq \quad & (1-p)(1-q)\, f(\emptyset) + \\ & p(1-q)\, f(A) + \\ & (1-p)q\, f(B) + \\ & pq\, f(A \cup B) \end{aligned}$$

  to prove use lemma I

random set analysis

- algorithm returns

$$R = U(1/2) = S^*(1/2) \cup \overline{S^*}(1/2)$$

- by applying lemma II

$$\begin{aligned} E[f(R)] &= E[f(S^*(1/2) \cup \overline{S^*}(1/2))] \\ &= \frac{1}{4}f(\emptyset) + \frac{1}{4}f(S^*) + \frac{1}{4}f(\overline{S^*}) + \frac{1}{4}f(U) \end{aligned}$$

- gives $1/4$ for non-negative and $1/2$ for symmetric function

unconstrained problem

[Feige et al., 2011]

- local search
  - initialize S to best singleton
  - $S =$ local optimum (add or delete elements)
  - return the best of S and $U \setminus S$

$1/3$ approximation for non-negative function

$1/2$ for non-negative symmetric function

analysis of local search

- lemma III: if S is a local optimum then $f(S) \geq f(T)$ for all $S \subseteq T$ and $T \subseteq S$
- proof

  take $S \subseteq T$ and consider $S = X_0 \subseteq \ldots \subseteq X_\ell = T$

  by submodularity and local optimality

$$0 \geq f(S \cup \{x_i\}) - f(S) \geq f(X_i) - f(X_{i-1})$$

  summing up gives $0 \geq f(X_\ell) - (X_0)$ or $f(S) \geq (T)$
- corollary

  for optimum S^* and local optimum S it is $f(S) \geq f(S \cup S^*)$ and $f(S) \geq f(S \cap S^*)$

analysis of local search (cont)

- it is

$$f(S) \geq f(S \cup S^*) \quad \text{and} \quad f(S) \geq f(S \cap S^*)$$

- by submodularity and non-negativity

$$f(S \cup S^*) + f(U \setminus S) \geq f(S^* \setminus S) + f(U) \geq f(S^* \setminus S)$$

$$f(S \cap S^*) + f(S^* \setminus S) \geq f(S^*) + f(\emptyset) \geq f(S^*)$$

- combining we get

$$2f(S) + f(U \setminus S) \geq f(S^*)$$

- and so

$$\max\{f(S), f(U \setminus S)\} \geq \frac{1}{3} f(S^*)$$

unconstrainted problem

[Buchbinder et al., 2015]

- tight $1/2$ approximation for general non-negative submodular function
- randomized algorithm, approximation $1/2$
- deterministic algorithm, approximation $1/3$

deterministic algorithm

[Buchbinder et al., 2015]

Algorithm 1: DeterministicUSM$(f, \mathcal{N})$

$X_0 \leftarrow \emptyset$, $Y_0 \leftarrow \mathcal{N}$.
for i = 1 *to* n **do**
    $a_i \leftarrow f(X_{i-1} \cup \{u_i\}) - f(X_{i-1})$.
    $b_i \leftarrow f(Y_{i-1} \setminus \{u_i\}) - f(Y_{i-1})$.
    if $a_i \geq b_i$ **then** $X_i \leftarrow X_{i-1} \cup \{u_i\}$, $Y_i \leftarrow Y_{i-1}$.
    else $X_i \leftarrow X_{i-1}$, $Y_i \leftarrow Y_{i-1} \setminus \{u_i\}$.
return X_n (or equivalently Y_n).

randomized algorithm

[Buchbinder et al., 2015]

Algorithm 2: RandomizedUSM$(f, \mathcal{N})$

$X_0 \leftarrow \emptyset$, $Y_0 \leftarrow \mathcal{N}$.
for i = 1 *to* n **do**
  $a_i \leftarrow f(X_{i-1} \cup \{u_i\}) - f(X_{i-1})$.
  $b_i \leftarrow f(Y_{i-1} \setminus \{u_i\}) - f(Y_{i-1})$.
  $a'_i \leftarrow \max\{a_i, 0\}$, $b'_i \leftarrow \max\{b_i, 0\}$.
  with probability $a'_i/(a'_i + b'_i)^*$ **do**: $X_i \leftarrow X_{i-1} \cup \{u_i\}$, $Y_i \leftarrow Y_{i-1}$.
  else (with the compliment probability $b'_i/(a'_i + b'_i)$) **do**: $X_i \leftarrow X_{i-1}$, $Y_i \leftarrow Y_{i-1} \setminus \{u_i\}$.
return X_n (or equivalently Y_n).

* If $a'_i = b'_i = 0$, we assume $a'_i/(a'_i + b'_i) = 1$.

max-sum diversification

[Borodin et al., 2012]

- U is a ground set
- $d : U \times U \rightarrow \mathbb{R}$ is a metric distance function on U
- $f : 2^U \rightarrow \mathbb{R}$ is a submodular function

- we want to find $S \subseteq U$ such that
  $\phi(S) = f(S) + \lambda \sum_{u,v \in S} d(u, v)$ is maximized and
  $|S| \leq k$

max-sum diversification

[Borodin et al., 2012]

- consider $S \subseteq U$ and $x \in U \setminus S$
- define the following types of marginal gain

  $d_x(S) = \sum_{v \in S} d(x, v)$

  $f_x(S) = f(S \cup \{x\}) - f(S)$

  $\phi_x(S) = \frac{1}{2} f_x(S) + \lambda d_x(S)$

- greedy algorithm on marginal gain $\phi_x(S)$ gives factor 2 approximation

max-sum diversification – the greedy

[Borodin et al., 2012]

1. $S \leftarrow \emptyset$
2. while $|S| < k$
3. $\quad i \leftarrow \arg\max_{\{j \in U \setminus S\}} \phi_j(S)$
4. $\quad S \leftarrow S \cup \{i\}$
5. return S

conclusions

- maximization of submodular functions
- monotone, constraints, symmetric, ...
- recent developments in theory community
- simple algorithms
- neat analysis
- many applications in machine learning

references I

Borodin, A., Lee, H. C., and Ye, Y. (2012).
Max-sum diversification, monotone submodular functions and dynamic updates.
In *Proceedings of the 31st symposium on Principles of Database Systems*, pages 155–166. ACM.

Buchbinder, N., Feldman, M., Seffi, J., and Schwartz, R. (2015).
A tight linear time (1/2)-approximation for unconstrained submodular maximization.
SIAM Journal on Computing, 44(5):1384–1402.

Feige, U., Mirrokni, V. S., and Vondrak, J. (2011).
Maximizing non-monotone submodular functions.
SIAM Journal on Computing, 40(4):1133–1153.

Goemans, M. X. and Williamson, D. P. (1995).
Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming.
Journal of the ACM (JACM), 42(6):1115–1145.

Kempe, D., Kleinberg, J., and Tardos, E. (2003).
Maximizing the spread of influence through a social network.
In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM Press.

references II

Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978).
An analysis of approximations for maximizing submodular set functions I.
Mathematical Programming, 14(1):265–294.

Newman, M. E. J. and Girvan, M. (2004).
Finding and evaluating community structure in networks.
Physical Review E, 69(2).

Ng, A., Jordan, M., and Weiss, Y. (2001).
On spectral clustering: Analysis and an algorithm.
NIPS.

Shi, J. and Malik, J. (2000).
Normalized cuts and image segmentation.
IEEE transactions on Pattern Analysis and Machine Intelligence, 22(8).

Trevisan, L. (2012).
Max cut and the smallest eigenvalue.
SIAM Journal on Computing, 41(6):1769–1786.

references III

von Luxburg, U. (2007).
A Tutorial on Spectral Clustering.
arXiv.org.