Spectral Graph Clustering

Martin Vejmelka

19.3.2009

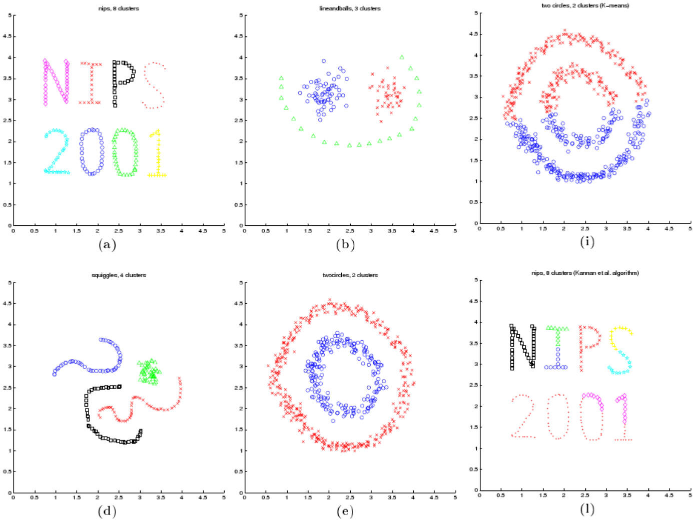
Outline

1. Introduction
   - Graphs
   - Graph Laplacian
2. Approaches to SGC
   - Graph cut methods
   - Matrix perturbation theory
   - Random walk approach
   - SGC Algorithms
3. Advanced approaches
   - Manor-Zelnik and Perona
   - Azran and Ghaharmani
4. Summary

Spectral Graph Clustering

- **Very** successful technique for partitioning data
- Based on relatively “old” results
- Recently surged in popularity (computational advances)
- Already used in many research fields
- A lot of theoretical results, different approaches

Why all the fuss ? (image Ng et al. NIPS, 2001)



Graphs

- Undirected graph is a tuple $G = (V, E)$
- V are vertices $v_i \in V$
- E are edges $e_i = \{v_j, v_k\}$
- Weight matrix $W = w_{ij}$, $w_{ij} \geq 0$ (weighted graphs)
- degree of vertex $d_i = \sum_j w_{ij}$
- volume of a set of vertices $vol(A) = \sum_{i \in A} d_i$
- vertex set complement $\bar{A} = V \backslash A$

Graphs from data

- No principled or formalized way to generate affinity matrix
- Selection of affinity, typically

$$w_{ij} = \exp\left(-\frac{||x_i - x_j||}{\sigma^2}\right)$$

- Sparsification:
  - k-nearest neighbor graph
  - ϵ-distance graph (thresholded)
  - fully connected graph
- Prior knowledge integration

Graph Laplacian

- Name derived from similarity with Laplace operator
- No consistent definition of Graph Laplacian
- D is diagonal degree matrix, vertex degrees d_i on diagonal
- unnormalized Laplacian $L = D - W$
- normalized Laplacian $L_{\mathrm{sym}} = I - D^{-1/2}WD^{-1/2}$
- normalized Laplacian $L_{\mathrm{rw}} = I - D^{-1}W$

Graph Laplacian — Properties

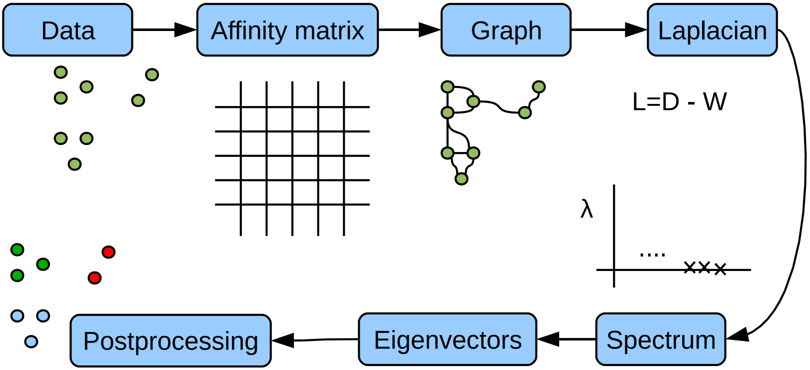
- Quadratic form

$$x^T L x = \sum_{i,j} w_{ij}(x_i - x_j)^2$$

- L is symmetric, positive semi-definite
- Smallest eigenvalue is $\lambda_1 = 0$ with eigenvector $\mathbf{1}$
- All eigenvalues are non-negative

$$0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$$

Overview


Data
Affinity matrix
Graph
Laplacian
L=D - W
λ
....
Spectrum
Eigenvectors
Postprocessing

Summary

- Examples
- Graphs — basic notions
- Graphs built from data
- Laplacians of Graphs
- Basic properties of Laplacians

Some History

- M. Fiedler, 1973–1975
  relationship between connected components and the Laplacian of graph
- Donnath and Hoffman, 1973
  suggested use of eigenvalues of adjacency matrix for graph partitioning
- Pothen, Simon and Liou, 1990
  2-way partitioning using Fiedler vector

Outline

1. Introduction
   - Graphs
   - Graph Laplacian
2. Approaches to SGC
   - Graph cut methods
   - Matrix perturbation theory
   - Random walk approach
   - SGC Algorithms
3. Advanced approaches
   - Manor-Zelnik and Perona
   - Azran and Ghaharmani
4. Summary

Graph cuts — MinCut

- For disjoint $A, B \subset V$,

$$\mathrm{cut}(A, B) = \sum_{v_i \in A, v_j \in B} w_{ij}$$

- MinCut

$$\mathrm{MINCUT}(A_1, \ldots, A_k) = \sum_{i=1}^{k} cut(A_i, \bar{A}_i)$$

- Can be solved efficiently but partitioning unsatisfactory
- Isolated vertices often become clusters

Graph cuts — RatioCut

- RatioCut (Hagen and Kahng, 1992)

$$\text{RATIOCUT}(A_1, ..., A_k) = \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{|A_i|}$$

- Clusters more balanced (small clusters are penalized)
- Leads to NP-hard problem, for 2-way

$$\min_x x^T Lx, x \perp \mathbf{1}, x \text{ piecewise constant}, ||x|| = \sqrt{|V|}$$

- Solved by relaxing the piecewise constant constraint — then solution is the 2nd smallest eigenvector
- This eigenvector must be thresholded

Graph cuts — NCut

- NCUT (Normalized Cut), **Shi and Malik, 2000**

$$\mathrm{NCUT}(A_1, \ldots, A_k) = \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{vol(A_i)}$$

- Intuitive explanation: maximizes connectedness inside clusters and disconnectedness between clusters at the same time
- Leads to NP-hard problem, for 2-way

$$\min_x x^T Lx, Dx \perp \mathbf{1}, x \text{ piecewise constant}, x^T Dx = |V|$$

- Again relaxing the piecewise continuity constraint — solution is 2nd smallest generalized eigenvector of

$$Lx = \lambda Dx$$

Graph cuts — Problems

- Only NCut has a “nice” justification
- Relaxed solution has no theoretical relationship to original solution except in special cases
- Other relaxations possible $\rightarrow$ SDP programming
- Eigenvector must be postprocessed
  - thresholding on c (usually 0)
  - thresholding on median (to get equal-sized groups)
  - thresholding at largest gap

Matrix perturbation theory

- Ng, Jordan and Weiss, NIPS, 2001.
- Block-like structure of ideal connectivity matrix
- Guaranteed to work under ideal conditions
- Davis-Kahan perturbation theorem (of eigenvalues)

$$\begin{bmatrix} A & \cdot & \cdot \\ \cdot & B & \cdot \\ \cdot & \cdot & C \end{bmatrix}$$

Random walk matrix

- Meila and Shi, 2001
- Markov random walk theory basis
- A good cluster is such that a random walk originating in it stays in it for a long time
- $P = D^{-1}W$ is a stochastic (transition) matrix
- P_{ij} represents probability of moving from node i to j
- Eigenvalues λ of $Px = \lambda x$ correspond to $1 - \lambda$ of $Lx = \lambda Dx$

Random walk matrix — theoretical analysis

- $NCUT(A, B)$ criterion corresponds to transition probability between sets A and B under stationary distribution of states.
- Guaranteed to work if transition probabilities between partitions only depend on partition index
- Connected with the *lumpability* of Markov chains
- Cf. with Ng, Jordan, Weiss ideal block matrix
- This directly extends the usefulness of NCUT algorithm

Algorithms for SGC

For purposes of this talk we will divide the algoritms into three generations.

- 1st generation — individual works, ad-hoc approaches
- 2nd generation — incorporation of other people's work (mixing)
- 3rd generation — beyond basic clustering strategies

Generation I

- Pothen, Simon, Liou, 1990.
- Scott, Longuet-Higgins, 1990 — use of multiple eigenvectors, image segmentation.
- Costeira and Kanade, 1995 — Shape analysis from motion
- Perona and Freeman, 1998 — thresholding of eigenvector, image segmentation.
- Shi and Malik, 2000 — normalized cut.
- Work with criterion and its solution primarily

Generation II

- Ng, Jordan, Weiss, 2001 — k-means clustering if eigenvector basis.
- Meila, 2000 — multiway cut.
- Meila and Shi, 2001 — random walk approach.
- Ding et al, 2001 — MinMaxCut.
- Work end-to-end (data $\rightarrow$ clusters, postprocessing).

Generation III

- Manor-Zelnik and Perona, 2004.
- Azran and Ghaharmani, 2006.
- Concentrate on postprocessing or preprocessing, new perspectives on previous algorithms.

Outline

1 Introduction
- Graphs
- Graph Laplacian

2 Approaches to SGC
- Graph cut methods
- Matrix perturbation theory
- Random walk approach
- SGC Algorithms

3 Advanced approaches
- Manor-Zelnik and Perona
- Azran and Ghaharmani

4 Summary

Manor-Zelnik and Perona

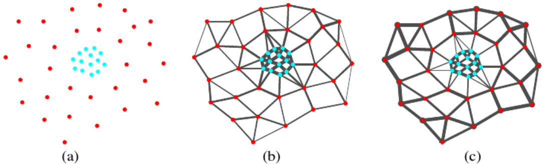
- Based on the Ng,Jordan,Weiss approach
- Multicut method (k-way clustering)
- Automatic tuning of affinity matrix
- Determination of clusters from eigenvectors
- Eigenvalue analysis: clusters from λ not clear.

Affinity matrix

- Sometimes clusters have different densities
- Construct modified affinity matrix

$$W_{ij} = \exp - \frac{||x_i - x_j||}{\sigma_i \sigma_j}$$

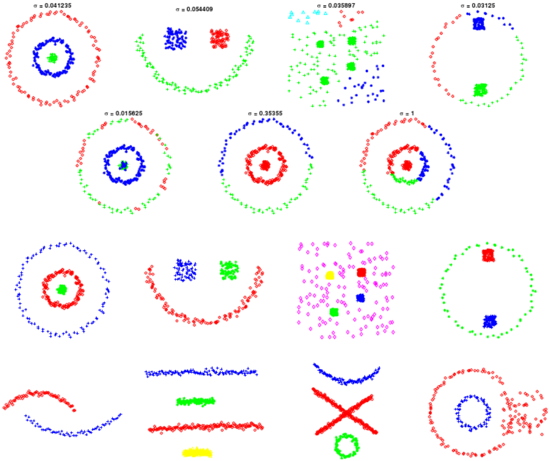
- With $\sigma_i = ||x_i - x_{iK}||$ distance to the k-th neighbor



Cluster count determination

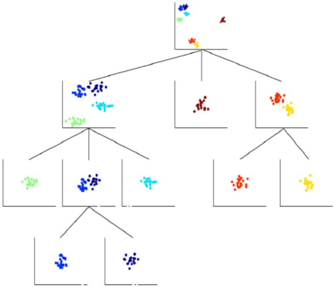
- Using Givens rotations, try to recover block structure of matrix
$$V = (v_1, ..., v_k)$$
with v_i the 2nd, 3rd, ... eigenvectors
- If cluster count is k then a rotation matrix is guaranteed to exist which will ensure each row of V has only 1 non-zero element
- Minimize cost function — “diagonality” of V
- Incremental approach — fast implementation

Examples

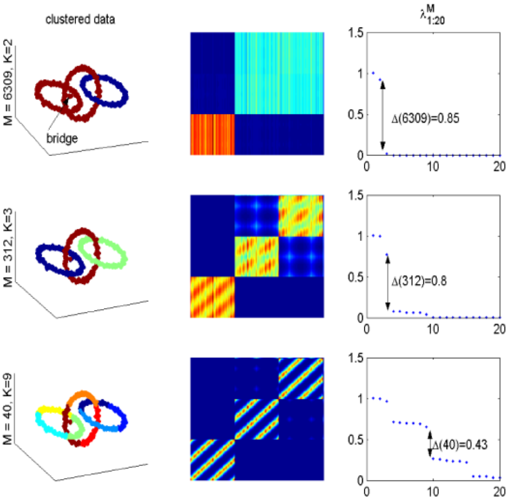


Azran and Ghaharmani

- Based on the random-walk approach
- Multicut method (k-way clustering)
- Noticed that P^m represents m-step random walks
- P^m for increasing m reveals coarser clusters
- Computing P^m is expensive, but computing λ_i^m is fast
- Search for maximal eigengap $\lambda_i^m - \lambda_{i-1}^m$



Examples


clustered data
M = 6309, K=2
bridge
Δ(6309)=0.85
M = 312, K=3
Δ(312)=0.8
M = 40, K=9
Δ(40)=0.43
$\lambda^{M}_{1:20}$

Outline

1. Introduction
   - Graphs
   - Graph Laplacian
2. Approaches to SGC
   - Graph cut methods
   - Matrix perturbation theory
   - Random walk approach
   - SGC Algorithms
3. Advanced approaches
   - Manor-Zelnik and Perona
   - Azran and Ghaharmani
4. Summary

Things left out

- Derivations of results
- Some algorithms variants
- Some theoretical connections
- Manifold learning
- Probably lots of other concepts ...

Thank you

Thank you for your attention !