Morphological Data Analysis

L. Najman

Joint work with Aditya Challa,
Sravan Danda and B.S. Daya Sagar

8/1/2019 − Labex Bézout − Optimization and Learning

Mathematical morphology

- is a theory and technique for the analysis and processing of geometrical structures,
  - based on set theory, lattice theory, topology, and random functions.
- most commonly applied to digital images,
  - but it can be employed as well on graphs, surface meshes, solids, and many other spatial structures.
- Basic morphological operators are erosion, dilation, opening and closing

Plan

- MorphMedian and semi-supervised clustering
  - The watershed as a classifier
- Some links with optimization framework
  - The Power Watershed framework
    - Random walker, spectral clustering

Part I: Morphological Median and the watershed

Morphological Median

Interpolation of shapes

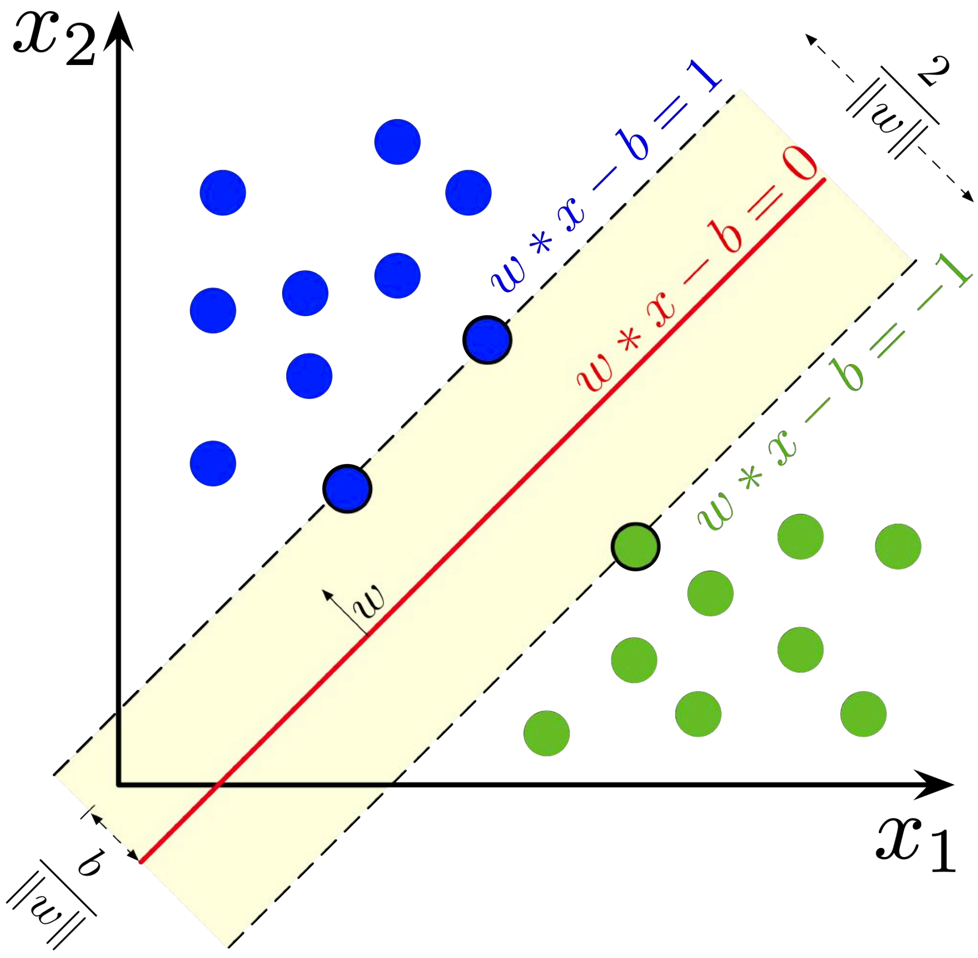
$$M(X,Y) = \bigcup_{\lambda \geq 0} \{(X \oplus \lambda B) \cap (Y \ominus \lambda B)\}$$

Rewriting Morphological Median

$$d(X, Z) = \inf\{\lambda \mid Z \subseteq X \oplus \lambda B\}$$

$$M(X, Y) = \{x | d(X, x) \leq d(Y^c, x)\} = IZ(X \mid Y^c)$$

Linear SVM: maximum margin



"Minimize $\|\vec{w}\|$ subject to $y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1$ for $i = 1, \ldots, n$."

The Maximum Margin Partition

$$
\begin{aligned}
V &= \text{Some set of points} \\
\rho(x, y) &:= \text{Dissimilarity between } x \text{ and } y \\
\rho(X, Y) &:= \inf_{x \in X, y \in Y} \rho(x, y) \\
X_0 = \text{Label 0 set} & \qquad X_1 = \text{Label 1 set} \\
\text{(Require)} V = M_0 \cup M_1 & \qquad X_0 \subset M_0, X_1 \subset M_1
\end{aligned}
$$

The Maximum Margin Partition

$$\begin{aligned} V &= \text{Some set of points} \\ \rho(x,y) &:= \text{Dissimilarity between } x \text{ and } y \\ \rho(X,Y) &:= \inf_{x\in X, y\in Y} \rho(x,y) \end{aligned}$$

$$X_0 = \text{Label 0 set} \qquad X_1 = \text{Label 1 set}$$

$$(\text{Require}) V = M_0 \cup M_1 \qquad X_0 \subset M_0, X_1 \subset M_1$$

Observe:

$$\mathit{Margin}(X_0, \text{boundary}) = \rho(X_0, M_1)$$

The Maximum Margin Partition

$$\begin{aligned} V &= \text{Some set of points} \\ \rho(x, y) &:= \text{Dissimilarity between } x \text{ and } y \\ \rho(X, Y) &:= \inf_{x \in X, y \in Y} \rho(x, y) \end{aligned}$$

$X_0 = \text{Label 0 set} \qquad X_1 = \text{Label 1 set}$

$(\text{Require}) V = M_0 \cup M_1 \qquad X_0 \subset M_0, X_1 \subset M_1$

Observe:

$$\begin{aligned} Margin(X_0, \text{boundary}) &= \rho(X_0, M_1) \\ Margin(X_1, \text{boundary}) &= \rho(X_1, M_0) \end{aligned}$$

The Maximum Margin Partition

$$
\begin{aligned}
V &= \text{Some set of points} \\
\rho(x,y) &:= \text{Dissimilarity between } x \text{ and } y \\
\rho(X,Y) &:= \inf_{x\in X, y\in Y} \rho(x,y) \\
X_0 = \text{Label 0 set} & \qquad X_1 = \text{Label 1 set} \\
\text{(Require)} V = M_0 \cup M_1 & \qquad X_0 \subset M_0, X_1 \subset M_1
\end{aligned}
$$

$$
Margin = \inf\{\rho(X_0, M_1), \rho(X_1, M_0)\}
$$

The Maximum Margin Partition

$$
\begin{aligned}
V &= \text{Some set of points} \\
\rho(x, y) &:= \text{Dissimilarity between } x \text{ and } y \\
\rho(X, Y) &:= \inf_{x \in X, y \in Y} \rho(x, y) \\
X_0 = \text{Label 0 set} & \qquad X_1 = \text{Label 1 set} \\
(\text{Require}) V = M_0 \cup M_1 & \qquad X_0 \subset M_0, X_1 \subset M_1
\end{aligned}
$$

Result (Maximum Margin Partition)

Given the definitions as above, a partition $V = M_0 \cup M_1$ is called the maximum partition if it is

$$\arg\max_{M_0, M_1} \inf \{\rho(X_0, M_1), \rho(X_1, M_0)\}$$

The Maximum Margin Partition

$$\arg\max_{M} \hat{\rho}(X_0, X_1, M) = \arg\max_{M} \left\{\inf \left\{\rho(X_0, \overline{M}), \rho(X_1, M)\right\}\right\}$$

$$\rho(X, Y) = \inf_{x \in X, y \in Y} \rho(x, y).$$

MorphMedian

Recall:

$$m(X, Y) = \{x | d(X, x) \leq d(Y^c, x)\}$$

MorphMedian

Result

Given (V, ρ)*, and* X_0, X_1*, every maximum margin partition is* MORPHMEDIAN *and vice versa.*

MorphMedian

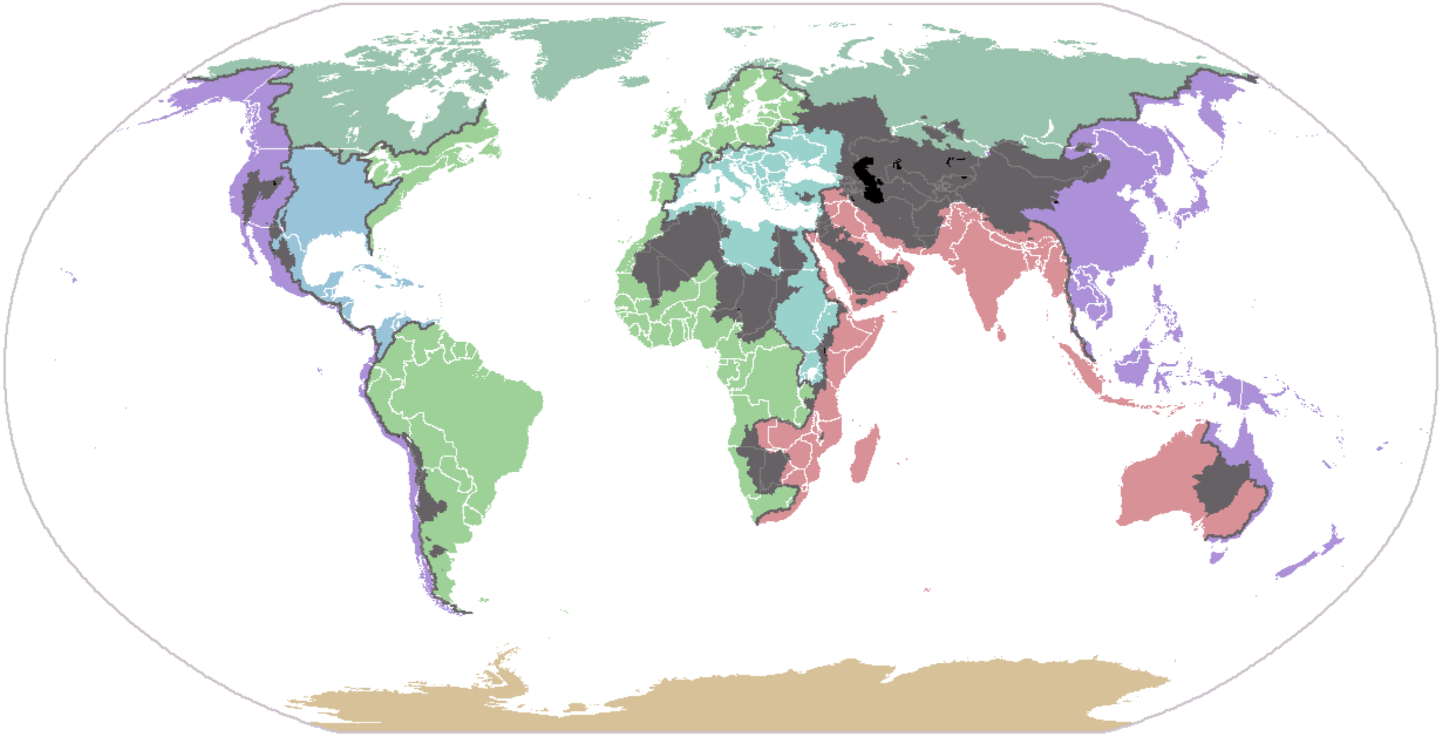
$V =$ Set of points
$\rho :=$ Dissimilarity Measure

Result (morphMedian)

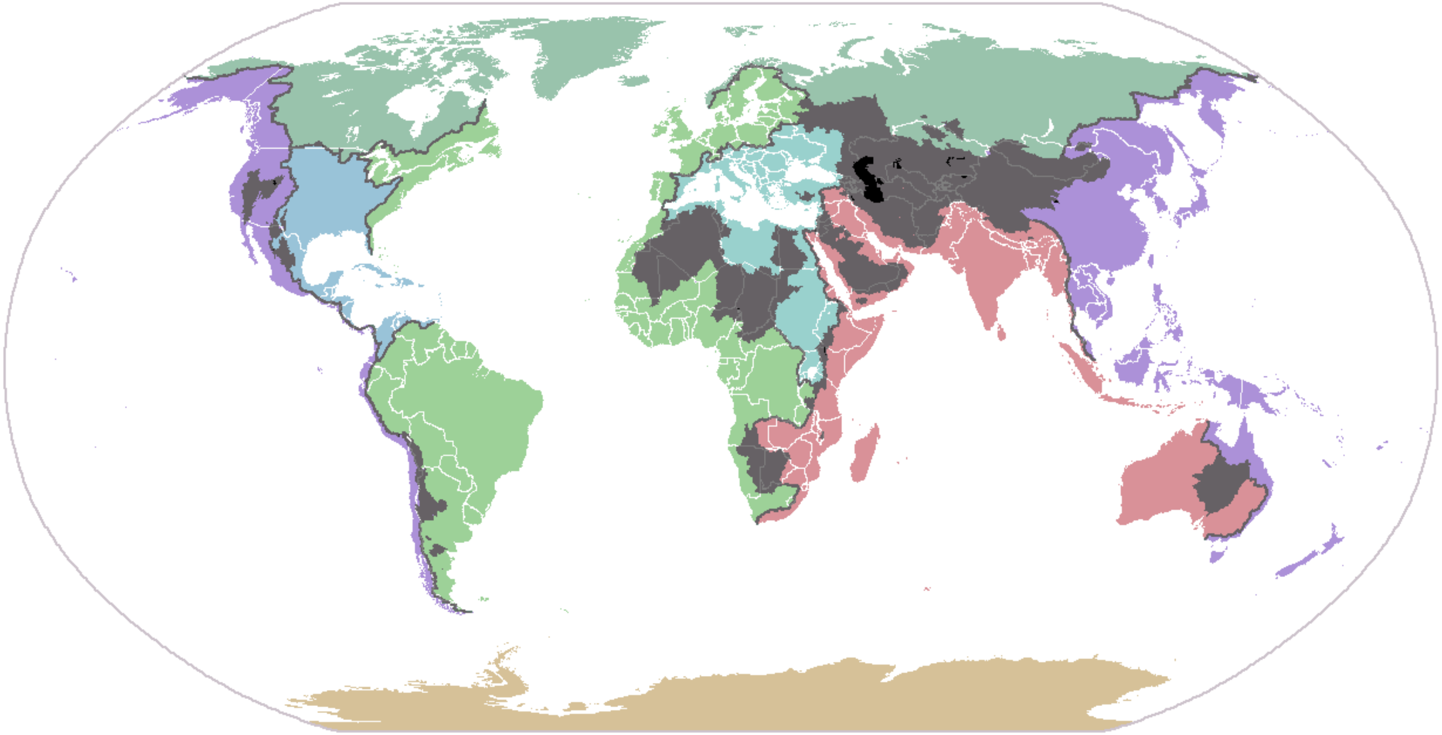
Let (V, ρ) be defined as above. X_0, X_1 denote the labelled sets. Define the morphMedian *partition as any partition which satisfies*

1. $x \in X_0$ *if* $\rho(X_0, x) < \rho(X_1, x)$
2. $x \in X_1$ *if* $\rho(X_1, x) < \rho(X_0, x)$

Watersheds



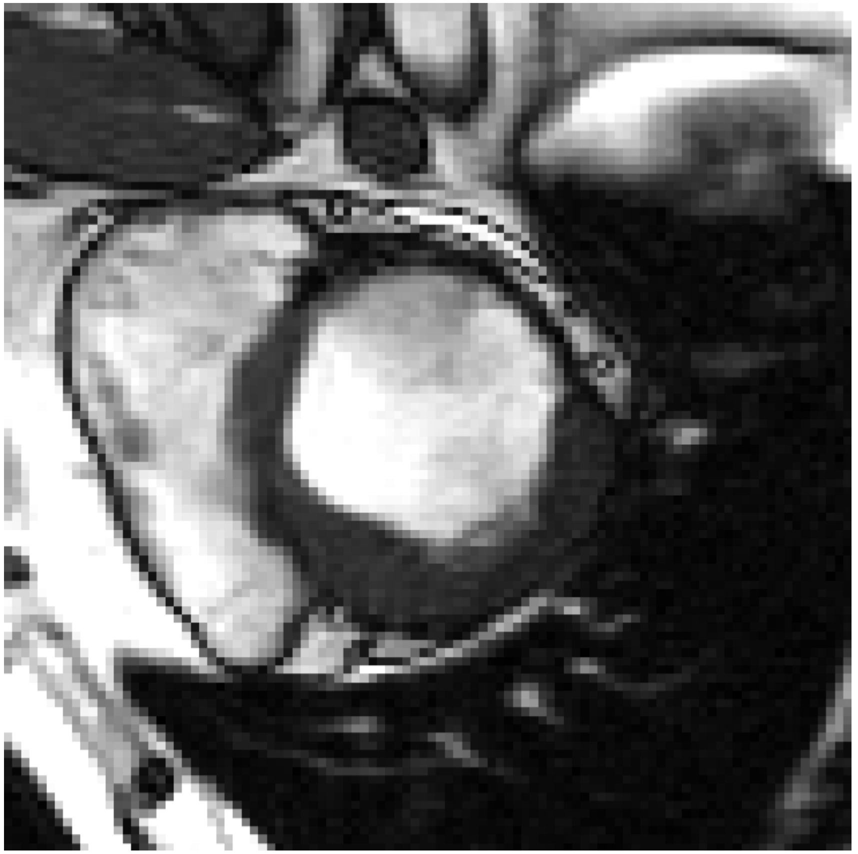
Watersheds



For topographics purposes, the watershed has been studied since the 19th century (Maxwell, Jordan, …)

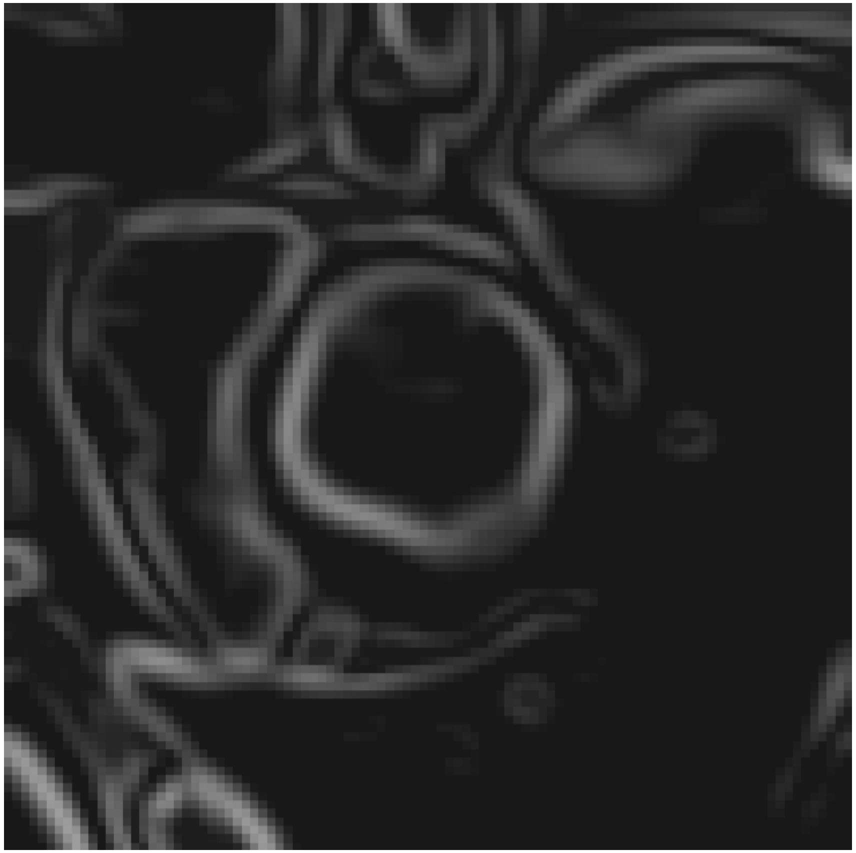
Watersheds

- One hundred years later (1978), it was introduced by Digabel and Lantuéjoul for image segmentation
- And popularized by L. Vincent and P. Soille in their celebrated 1991 PAMI paper



Watersheds

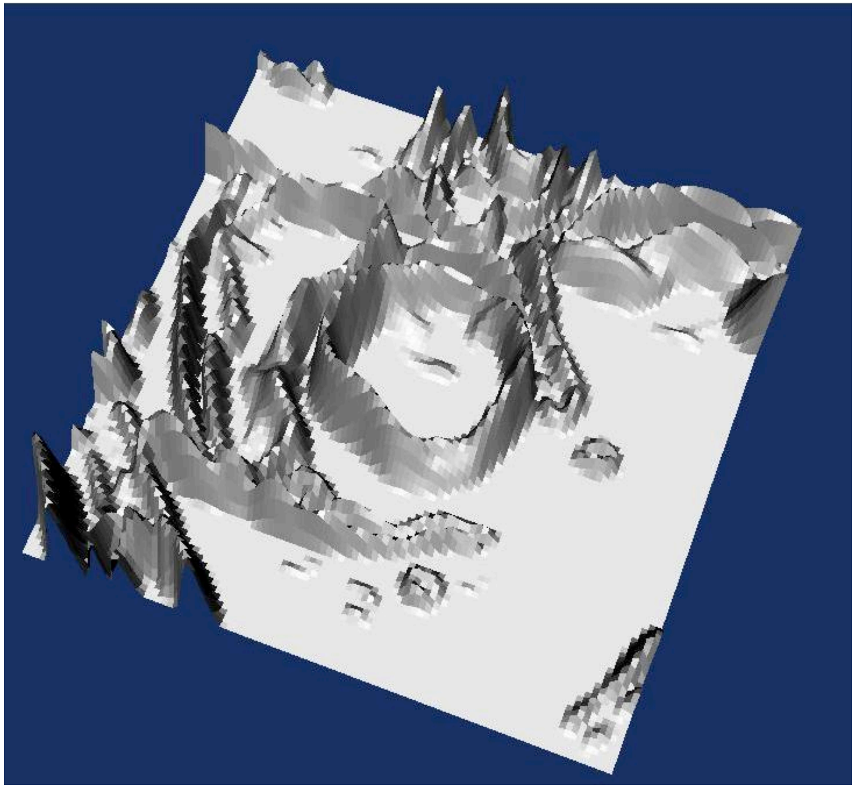
- One hundred years later (1978), it was introduced by Digabel and Lantuéjoul for image segmentation



- And popularized by L. Vincent and P. Soille in their celebrated 1991 PAMI paper

Watersheds

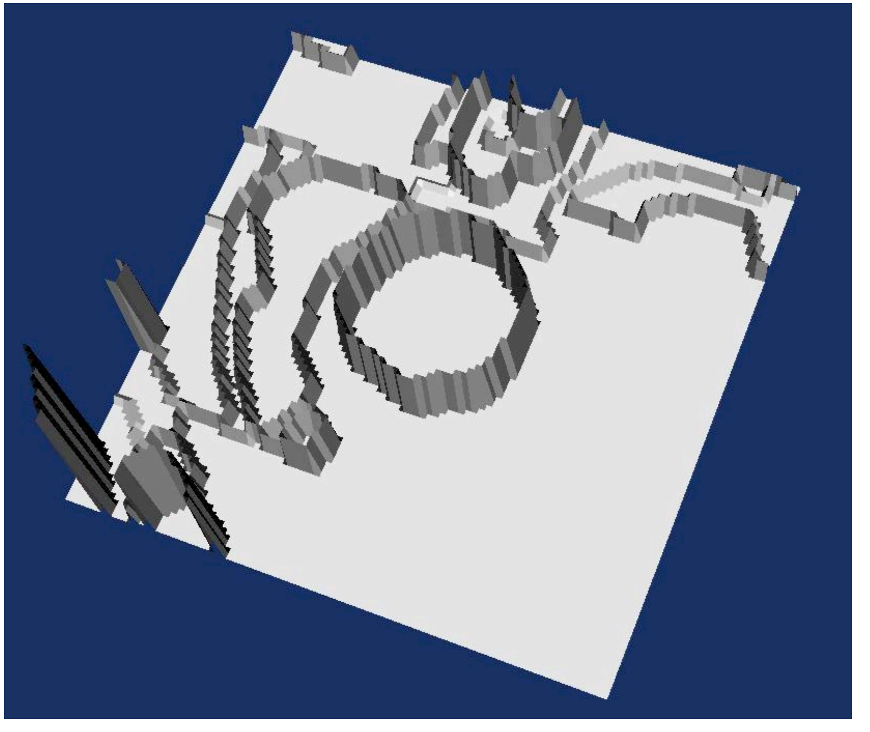
- One hundred years later (1978), it was introduced by Digabel and Lantuéjoul for image segmentation



- And popularized by L. Vincent and P. Soille in their celebrated 1991 PAMI paper

Watersheds

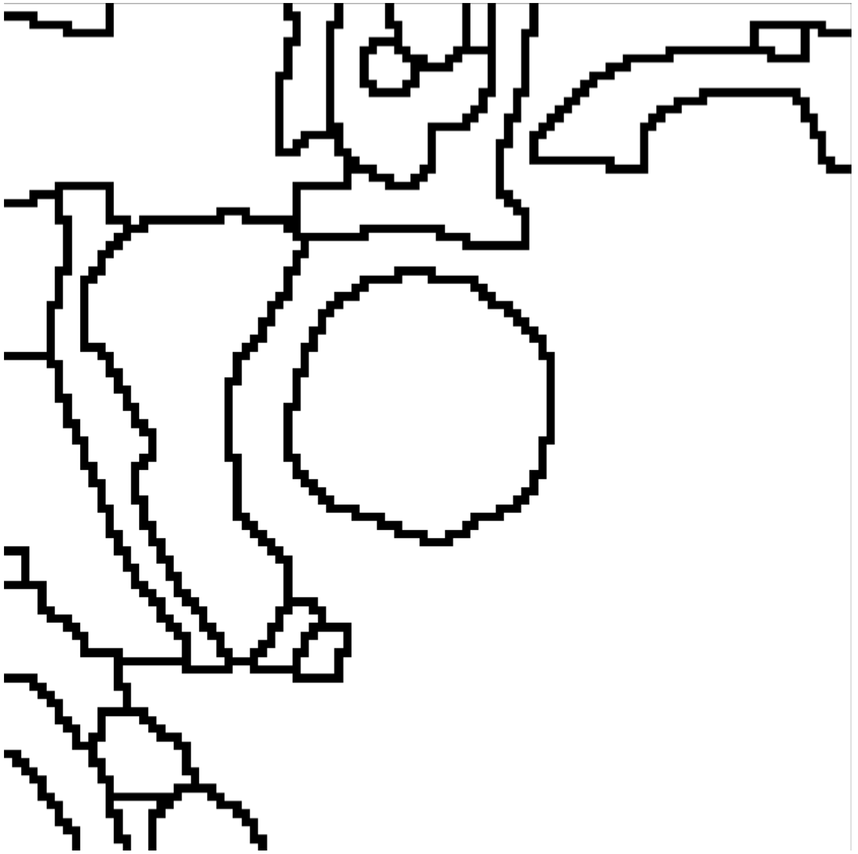
- One hundred years later (1978), it was introduced by Digabel and Lantuéjoul for image segmentation



- And popularized by L. Vincent and P. Soille in their celebrated 1991 PAMI paper

Watersheds

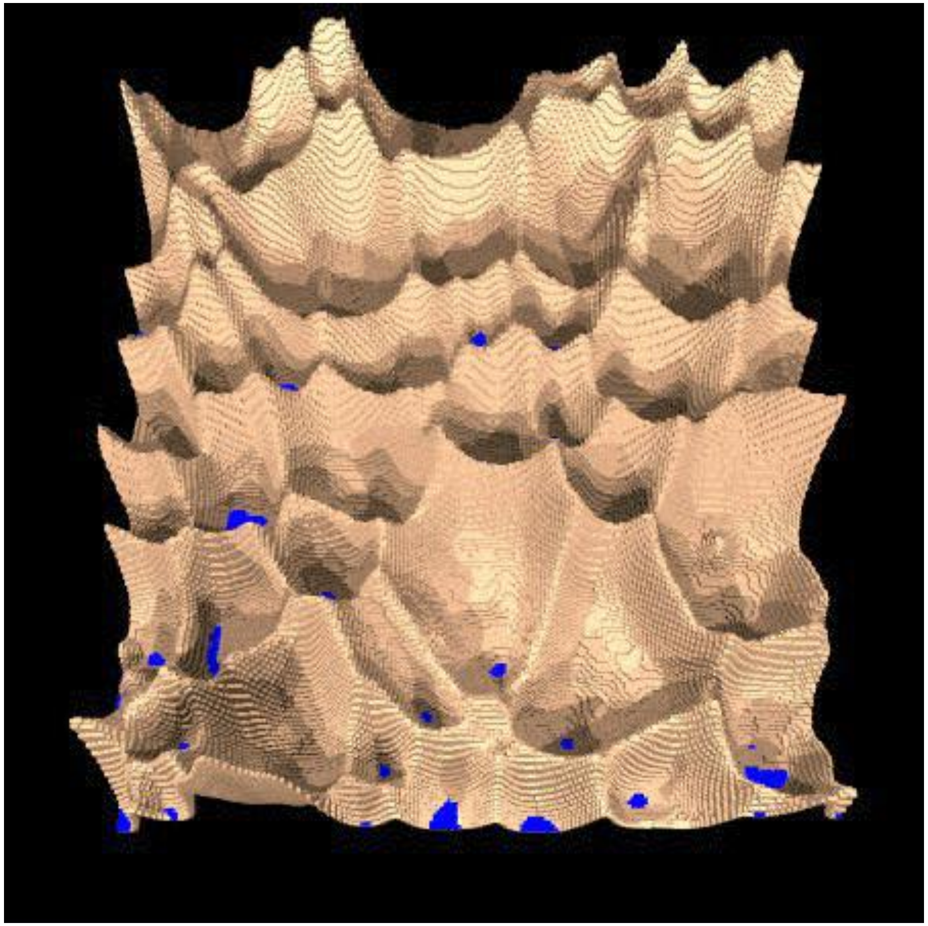
- One hundred years later (1978), it was introduced by Digabel and Lantuéjoul for image segmentation



- And popularized by L. Vincent and P. Soille in their celebrated 1991 PAMI paper

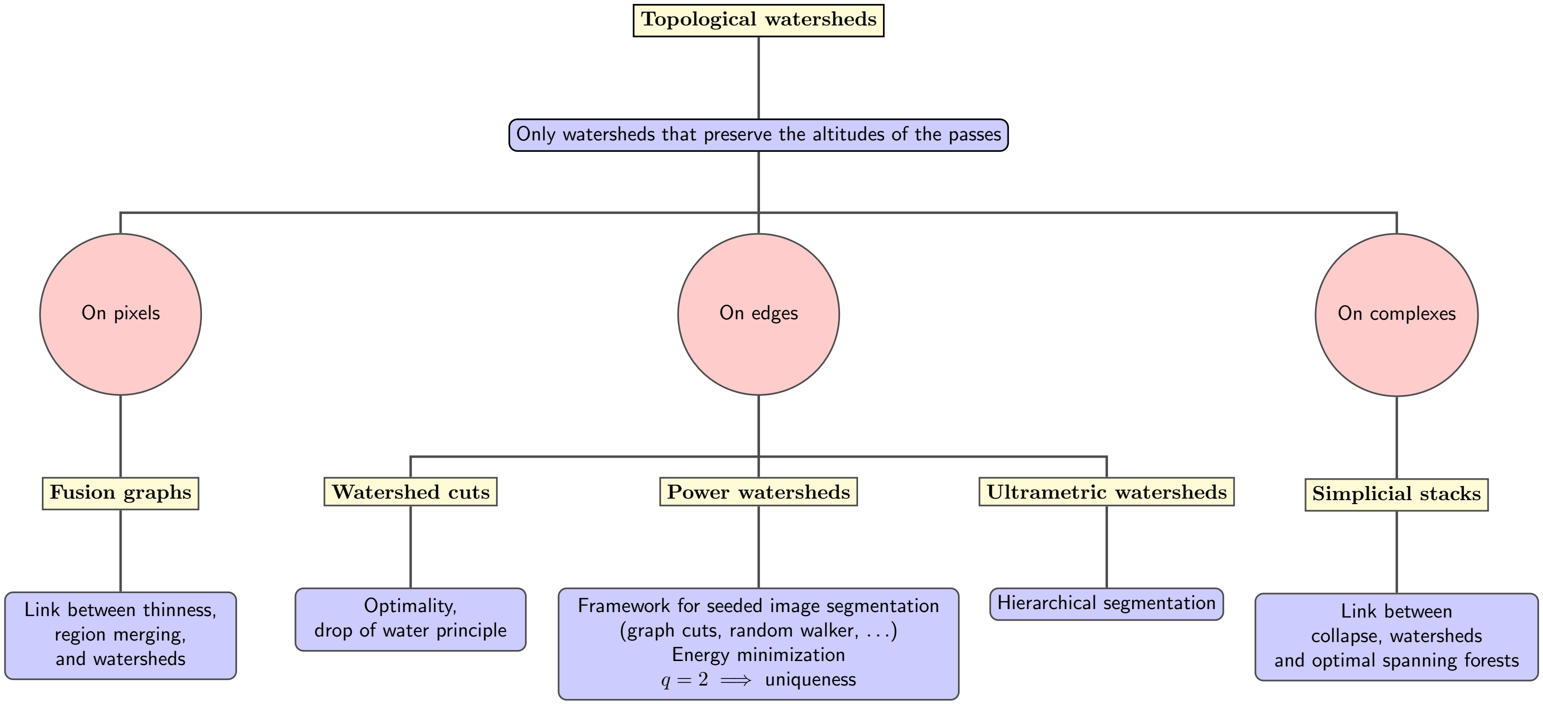
Watersheds

- One hundred years later (1978), it was introduced by Digabel and Lantuéjoul for image segmentation

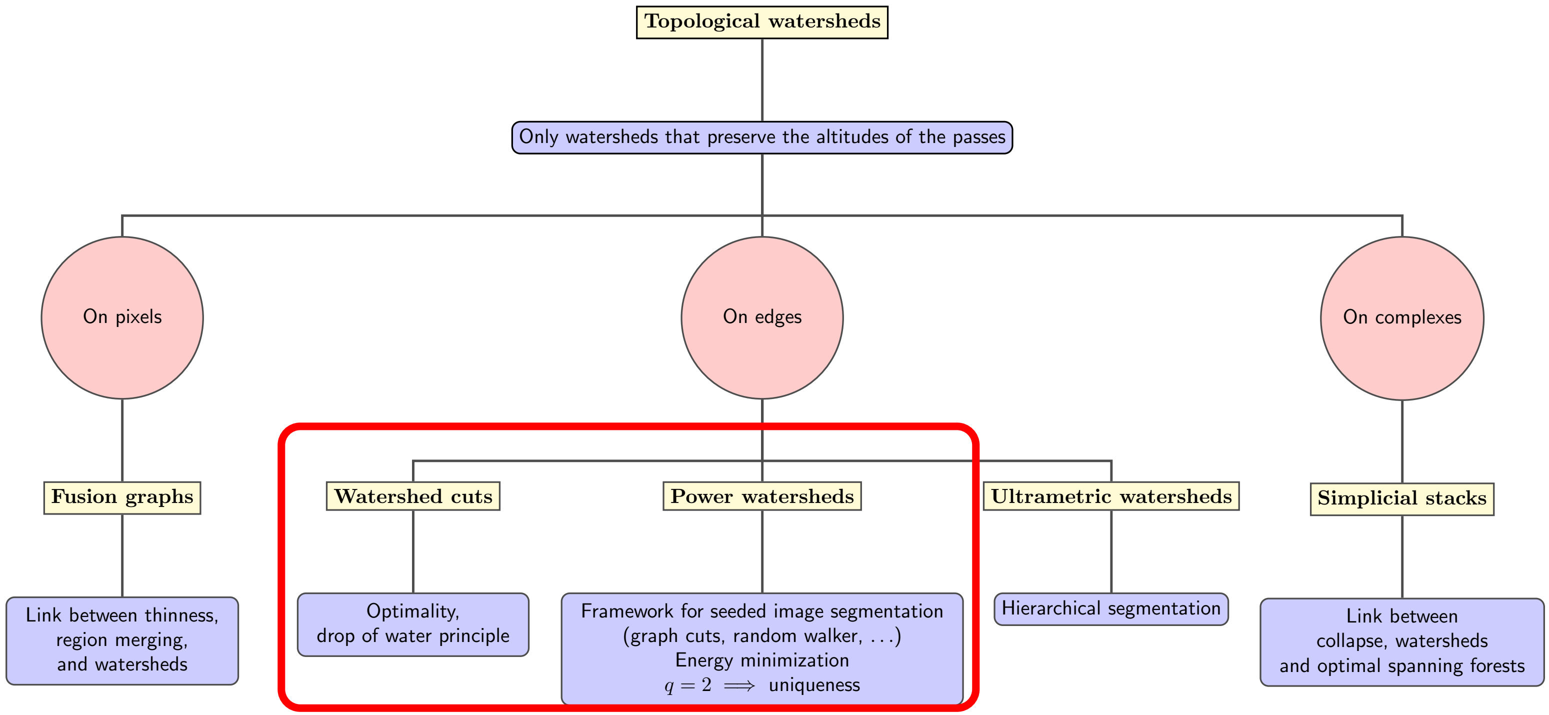


- And popularized by L. Vincent and P. Soille in their celebrated 1991 PAMI paper

The family of discrete watersheds



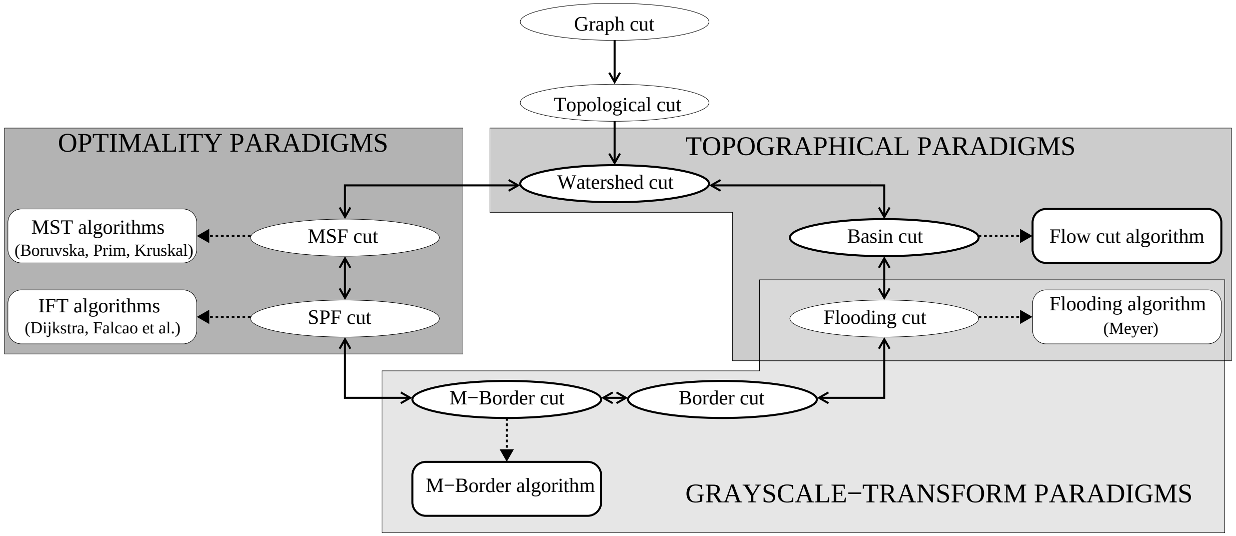
The family of discrete watersheds



Watershed cuts

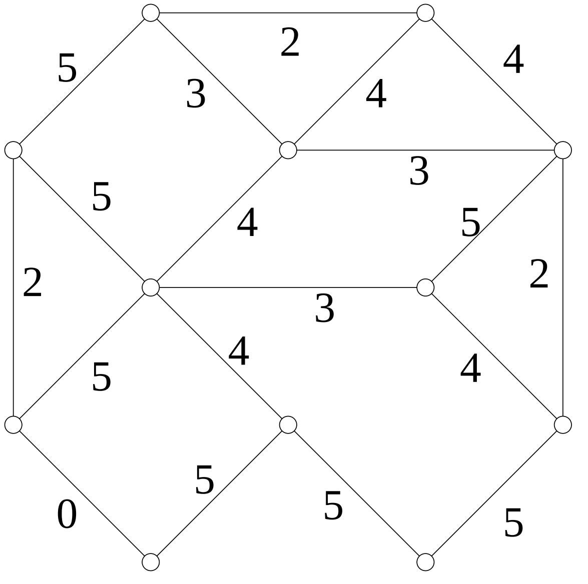
Important idea (Cousty et al., PAMI 2009, 2010)

- *Defined by the drop of water principle*
- *Equivalent to a catchment basins principle*
- *Optimal – an equivalence with Minimum Spanning Trees*



Notations

- Let $G = (V, E)$ be a graph.
- Let F be a map from E to $\mathbb{R}$.



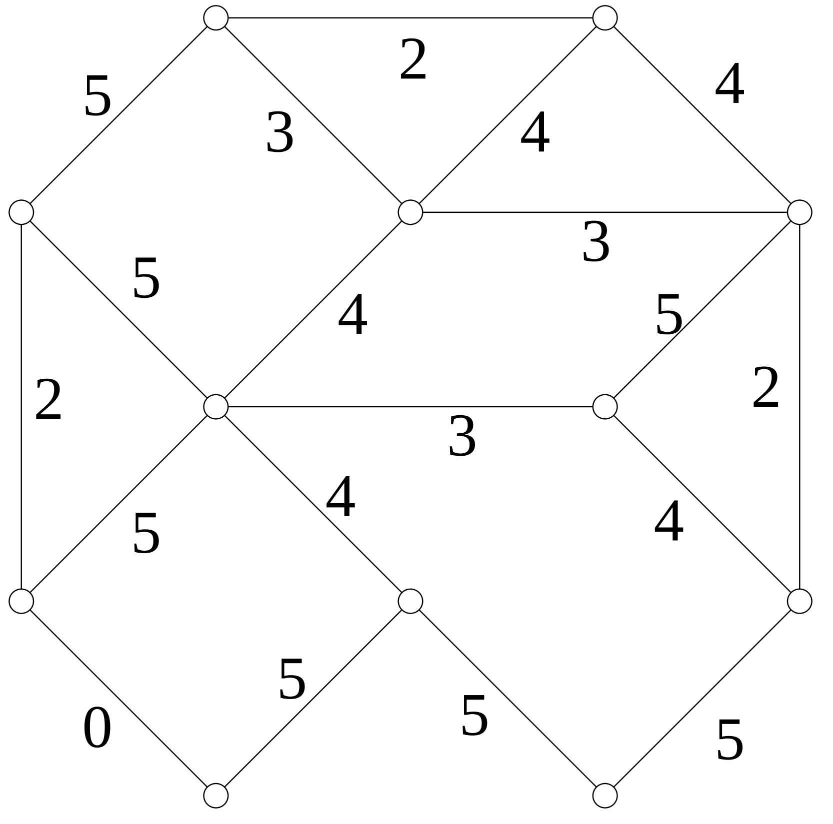
Minimum spanning forest

- The *weight of a forest* Y is the sum of its edge weights *i.e.*, $\sum_{u \in E(Y)} F(u)$.

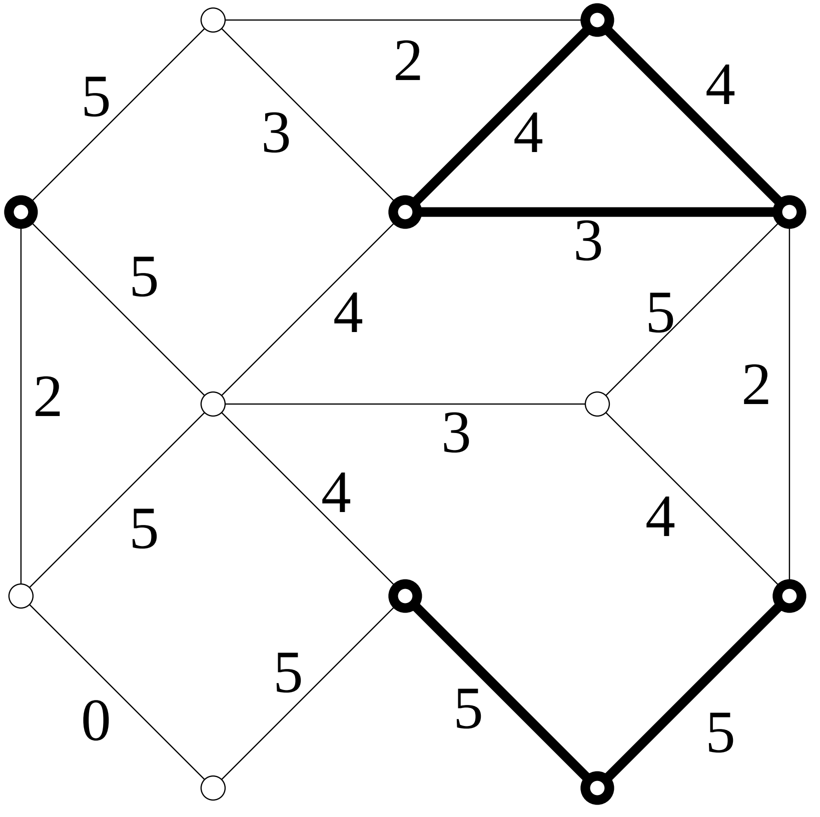
Definition

We say that Y is a minimum spanning forest (MSF) relative to X *if* Y *is a spanning forest relative to* X *and if the weight of* Y *is less than or equal to the weight of any other spanning forest relative to* X.

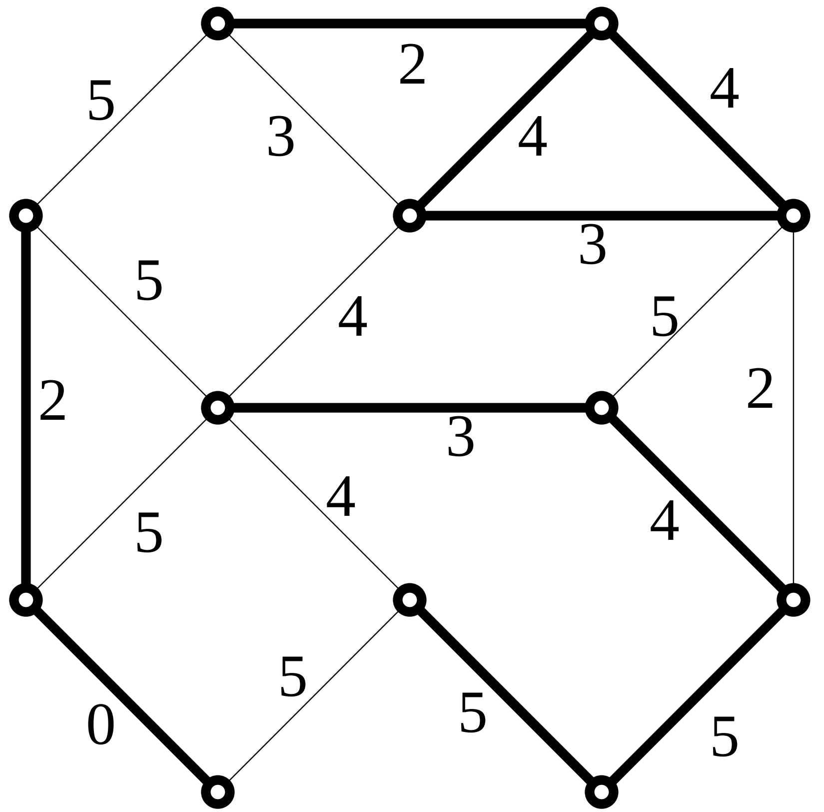
MSF - Example



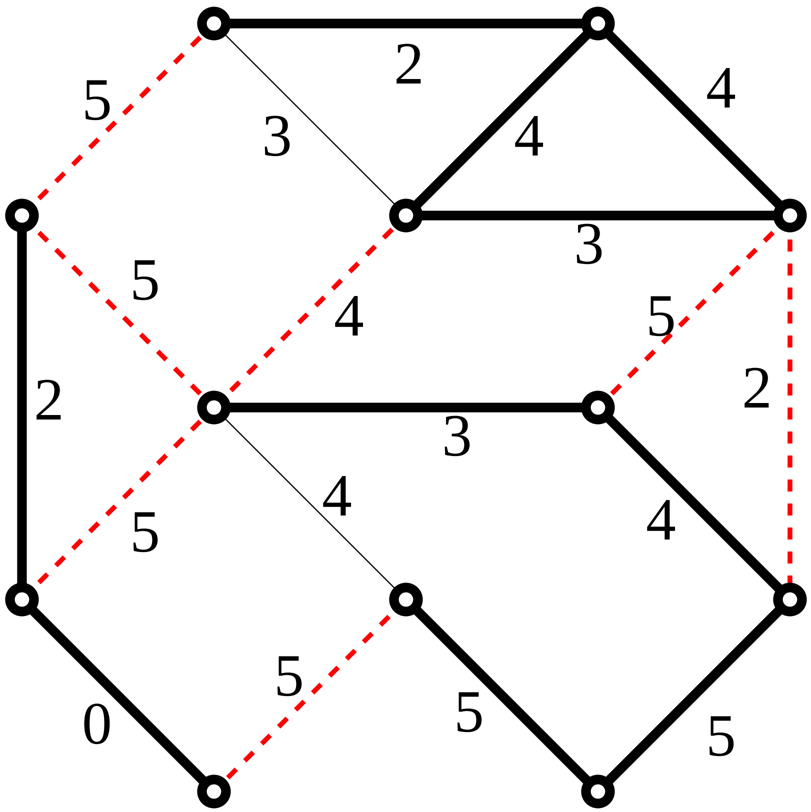
MSF - Example



MSF - Example



MSF - Example



Watershed and MSF equivalence

Theorem

An edge-set $S \subseteq E$ is a MSF cut for the minima of F if and only if *S is a watershed cut of F.*

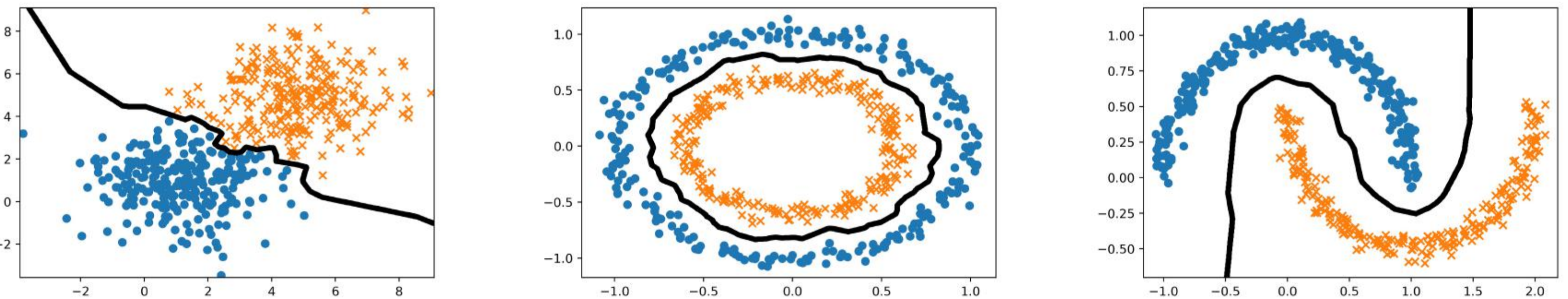
Watershed cuts as classifiers

Result

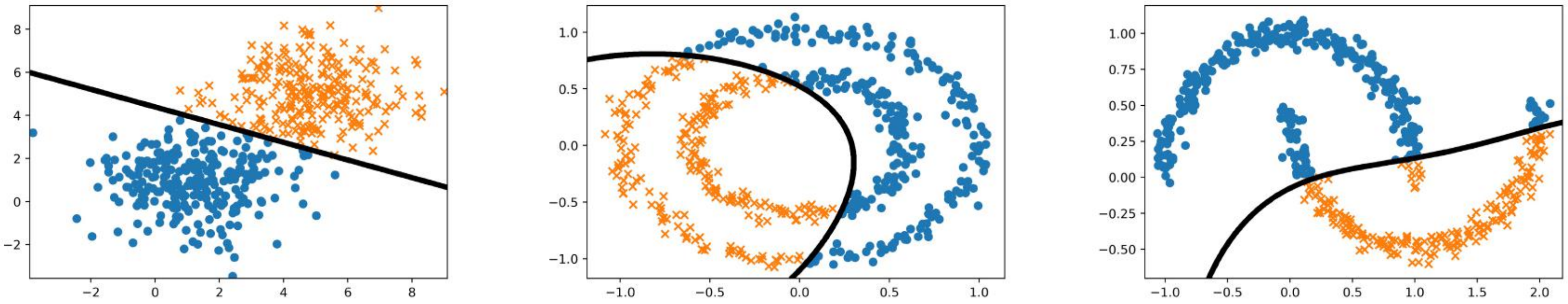
Given an edge weighted graph $G = (V, E, W)$, and a set of seeds $S = X_0 \cup X_1$, MSF-watershed returns a maximum margin partition with set of points as V and

$$\rho(x, y) = \inf_{\pi \in \Pi(x,y)} \sup_{e \in \pi} W(e)$$

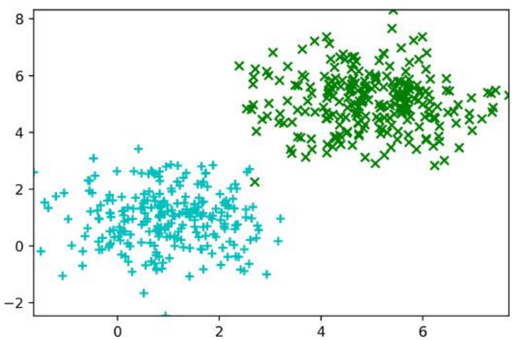
Watershed-cut as classifiers for semi-supervised learning



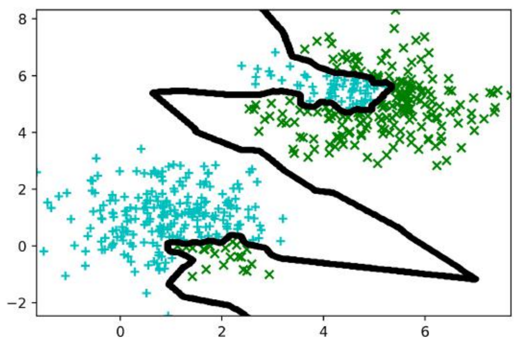
Results for SVM



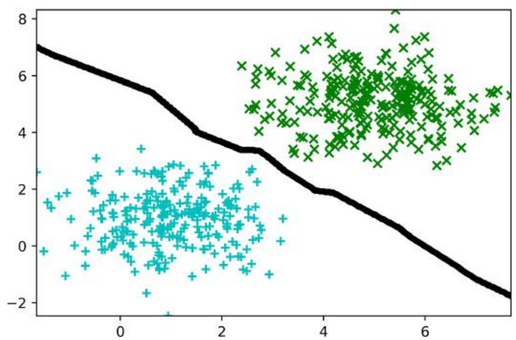
Morphological regularization



(a)



(b)



(c)

(a) Data with some wrongly labeled points
(b) MSF partition (Watershed-cut)
(c) Area-filtered watershed

Work in progress

- The watershed is a classifier
  - Hence, what if we “ensemble” watersheds?
  - Hence, what if we combine them with Neural nets

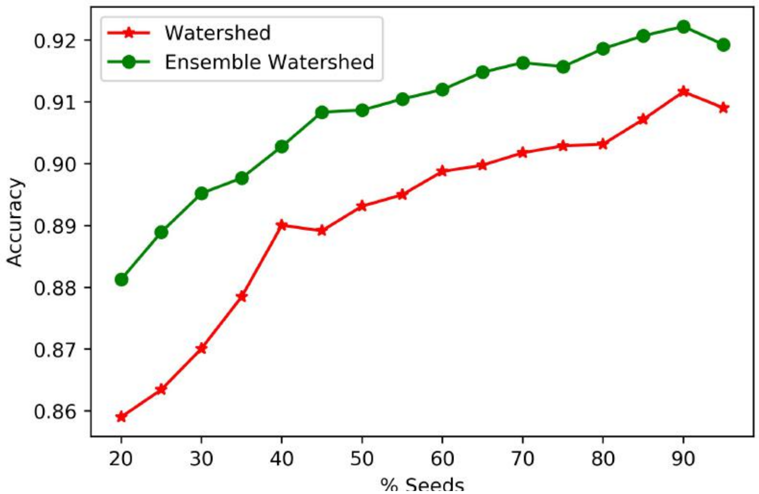
Ensemble watersheds

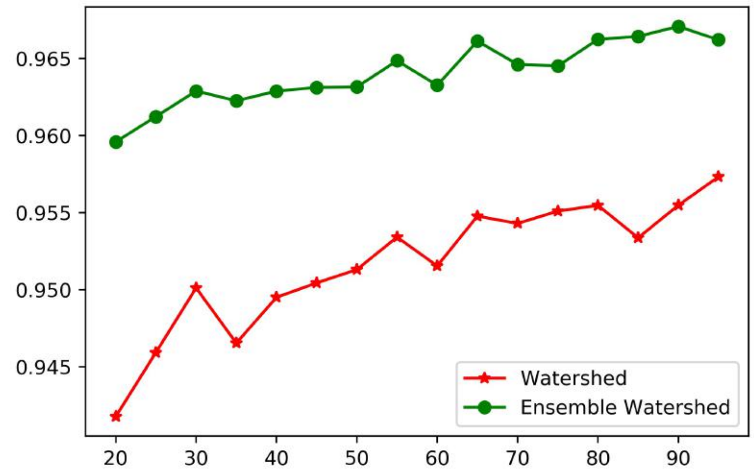
TABLE I
RESULTS OBTAINED USING DIFFERENT METHODS ON DATASETS FROM CHAPPELLE

Method	SSL1	SSL2	SSL3	SSL4	SSL5	SSL6	SSL7
watershed	96.53±0.70	95.66±0.87	99.77±0.22	51.35±3.64	55.19±1.48	95.70±0.45	54.84±1.32
IFT-SUM	96.96±0.53	95.17±0.24	95.06±1.22	53.92±2.95	61.15±0.76	90.06±0.85	64.60±1.96
RW	98.16±0.34	91.41±0.92	95.68±1.42	54.27±2.56	67.75±5.59	91.70±1.30	75.56±4.16
PW	97.99±0.49	89.42±0.78	95.68±1.42	52.22±2.29	67.75±5.59	91.68±1.36	75.56±4.16
SVM	93.78±0.67	90.81±0.57	56.87±0.81	60.00±2.81	83.57±0.80	22.16±0.49	84.49±1.22
1NN	96.96±0.53	95.18±0.24	95.06±1.22	53.92±2.95	61.15±0.75	90.06±0.85	64.61±1.97
RFC	95.36±0.87	87.74±0.58	91.42±0.64	55.76±2.33	72.75±0.89	90.51±1.06	70.45±1.75
Ensemblewatershed	98.17±0.35	92.71±1.17	99.38±0.90	53.16±3.14	64.39±3.11	95.09±0.94	68.29±1.77

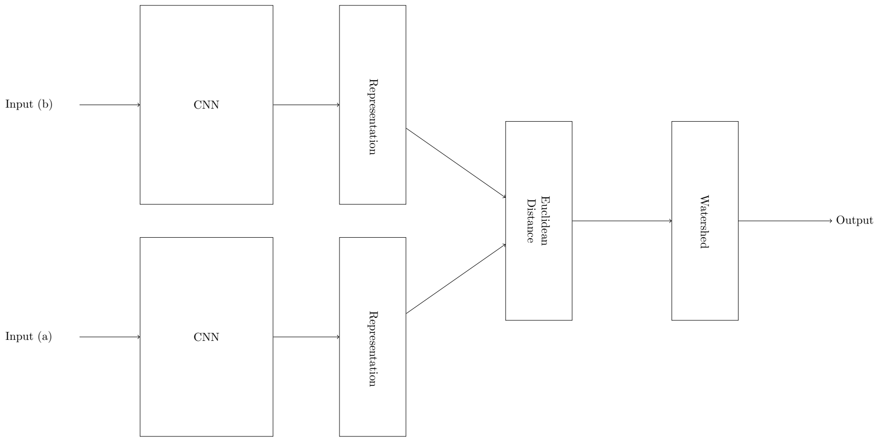
This morning results!

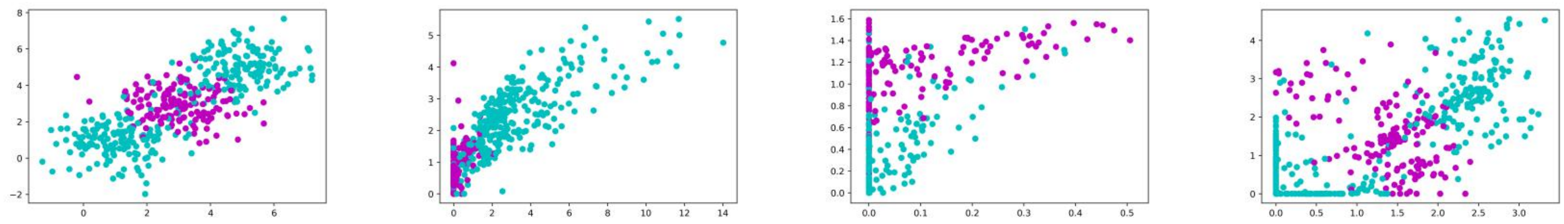
Ensemble watersheds





Using watershed as a layer of Neural Network





(a) (b) (c) (d)

(b), (c) and (d) are representation of the data by the NN

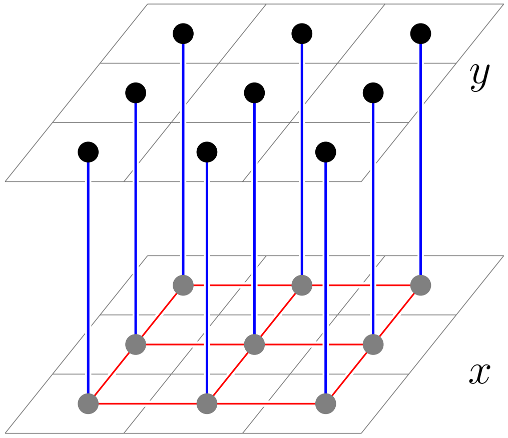
(a) Data (b) NN (c) Siamese Network (d) Siamese + WS

Preserves the structure of the data!

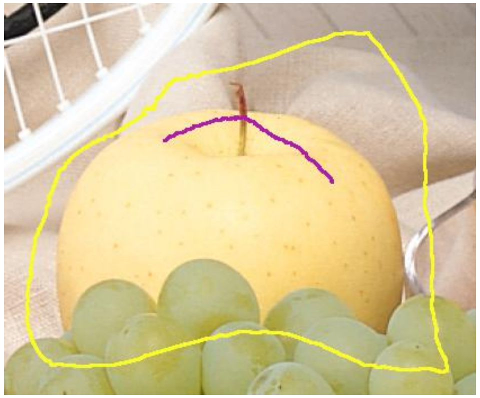
Part II: Watershed and optimization

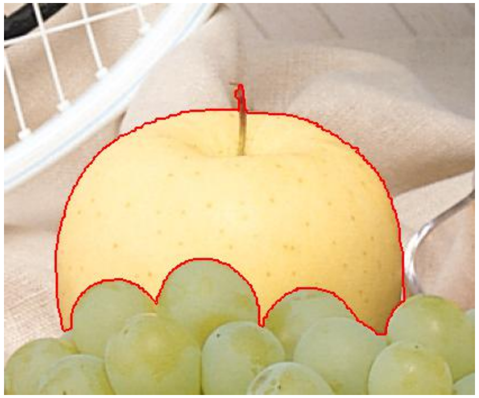
Some notations

- A simple, finite graph $G = (V, E)$ with nodes v_i and $|V| = m$
- Edge: e_{ij} spanning two vertices v_i and v_j
- Pairwise weight: w_{ij} for an edge e_{ij},
- Unary weight: w_i unary weights penalizing the (observed) configuration at node v_i.
- We are looking for x, a regularized version of the observed configuration y



A generic formulation





Power-watershed with $q \geq 0$

Let $q \geq 0$, we set

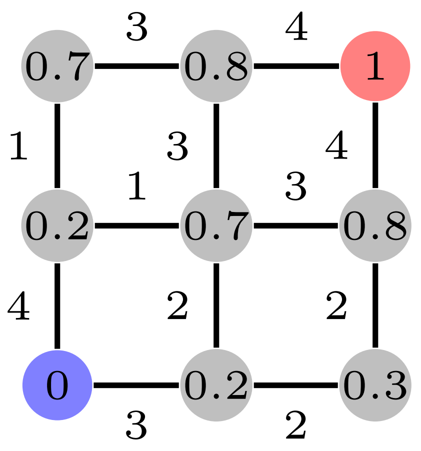
$$W^p(x) = \underbrace{\sum_{e_{ij} \in E} {w_{ij}}^p |x_i - x_j|^q}_{\text{Smoothness term}} + \underbrace{\sum_{v_i \in V} {w_i}^p |x_i - f_i|^q}_{\text{Data term}} \tag{3}$$

Random walker

- Combinatorial Dirichlet problem. [Grady 2006] (q=2)
- Resolution of system of linear equations.

Advantages

- Energy formulation $\rightarrow$ extends to a large class of problems
- No blocking artefacts



Random walker

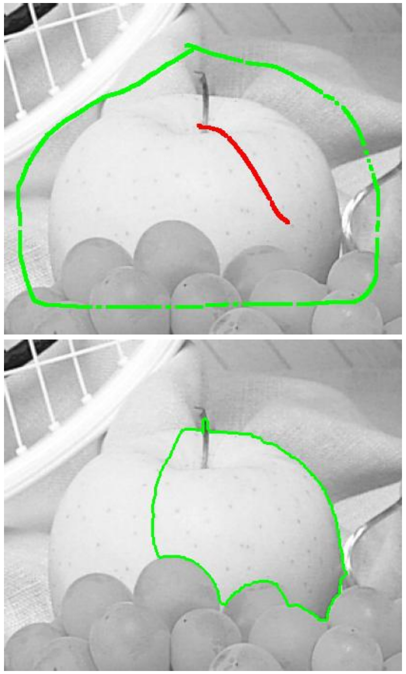
- Combinatorial Dirichlet problem. [Grady 2006] (q=2)
- Resolution of system of linear equations.

Advantages

- Energy formulation $\rightarrow$ extends to a large class of problems
- No blocking artefacts

Drawbacks

- Requires a more centered markers placement
- Super-linear complexity



And the watershed?

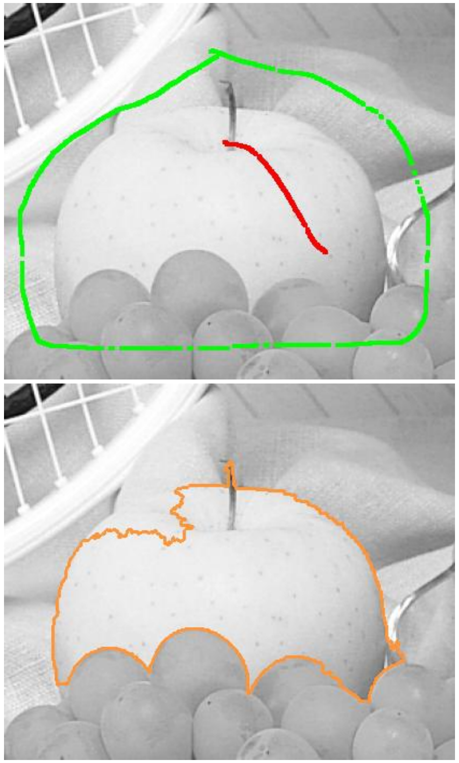
- Watershed [Beucher-Lantuéjoul 1979, Vincent-Soille 1991]

Advantages

- Fast
- Multilabel
- Robust to markers size

Drawbacks

- Leaking effect



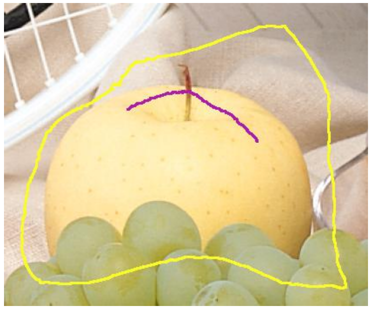
The Power Watershed framework

$$x^*_{p,q} = \arg\min_x \underbrace{\sum_{e_{ij} \in E} {w_{ij}}^p |x_i - x_j|^q}_{\text{Smoothness term}} + \underbrace{\sum_{v_i \in V} {w_i}^p |x_i - l_i|^q}_{\text{Data term}}$$

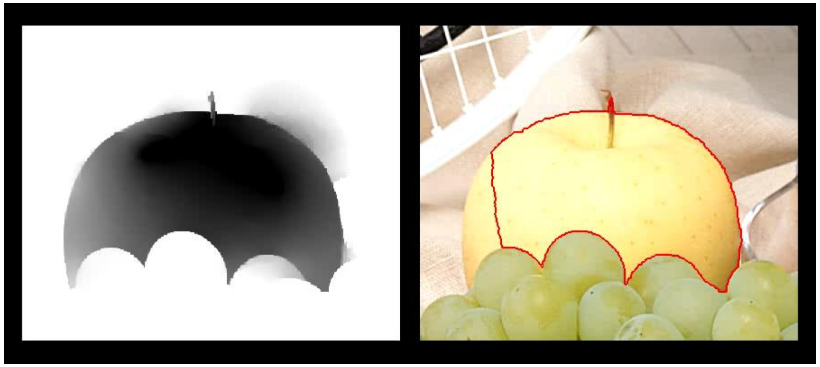
$$\bar{x} = \lim_{p \to \infty} x^*_{p,q}$$

Convergence of RW when p→∞

Input seeds



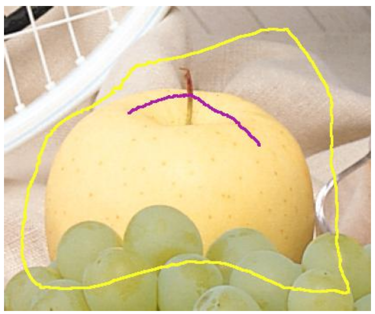
$$x^*_1 = \arg\min_x \underbrace{\sum_{e_{ij}\in E} w_{ij}^{\ 1} |x_i - x_j|^2}_{\text{Smoothness term}} + \underbrace{\mathcal{D}(x)}_{\text{Data fidelity}}$$



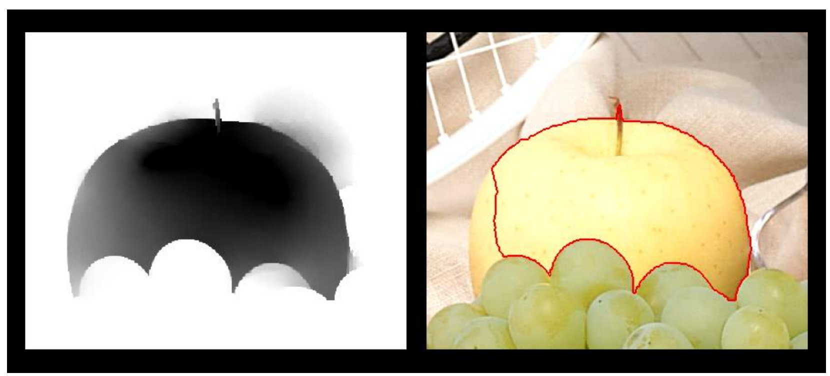
solution x^*_1 cut: threshold of x^*_1

Convergence of RW when p➔∞

Input seeds



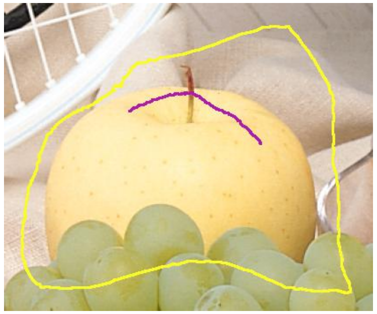
$$x^*_2 = \arg\min_x \underbrace{\sum_{e_{ij}\in E} w_{ij}^{\;2}|x_i - x_j|^2}_{\text{Smoothness term}} + \underbrace{\mathcal{D}(x)}_{\text{Data fidelity}}$$



solution x^*_2 cut: threshold of x^*_2

Convergence of RW when p→∞

Input seeds

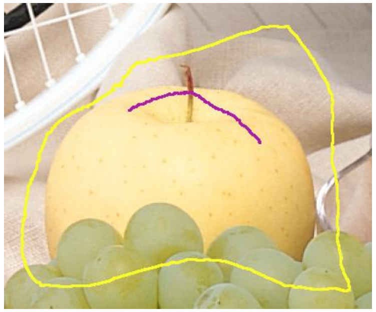


$$x^*_3 = \arg\min_x \underbrace{\sum_{e_{ij}\in E} w_{ij}^{\ 3} |x_i - x_j|^2}_{\text{Smoothness term}} + \underbrace{\mathcal{D}(x)}_{\text{Data fidelity}}$$

solution x^*_3 cut: threshold of x^*_3

Convergence of RW when p➔∞

Input seeds

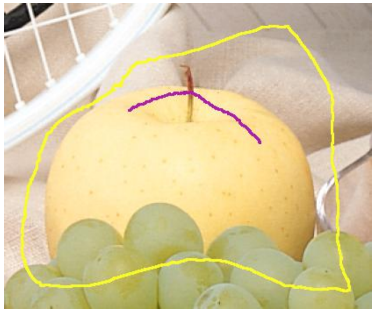


$$x^*_4 = \arg\min_x \underbrace{\sum_{e_{ij}\in E} w_{ij}^{\ 4} |x_i - x_j|^2}_{\text{Smoothness term}} + \underbrace{\mathcal{D}(x)}_{\text{Data fidelity}}$$

solution x^*_4 cut: threshold of x^*_4

Convergence of RW when p➔∞

Input seeds

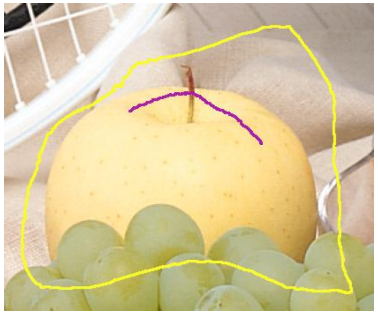


$$x^*_6 = \arg\min_x \underbrace{\sum_{e_{ij}\in E} w_{ij}^{\;6}|x_i - x_j|^2}_{\text{Smoothness term}} + \underbrace{\mathcal{D}(x)}_{\text{Data fidelity}}$$

solution x^*_6 cut: threshold of x^*_6

Convergence of RW when p→∞

Input seeds

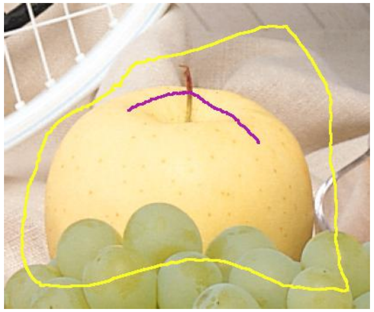


$$x^*_{9} = \arg\min_x \underbrace{\sum_{e_{ij}\in E} w_{ij}^{\ 9} |x_i - x_j|^2}_{\text{Smoothness term}} + \underbrace{\mathcal{D}(x)}_{\text{Data fidelity}}$$

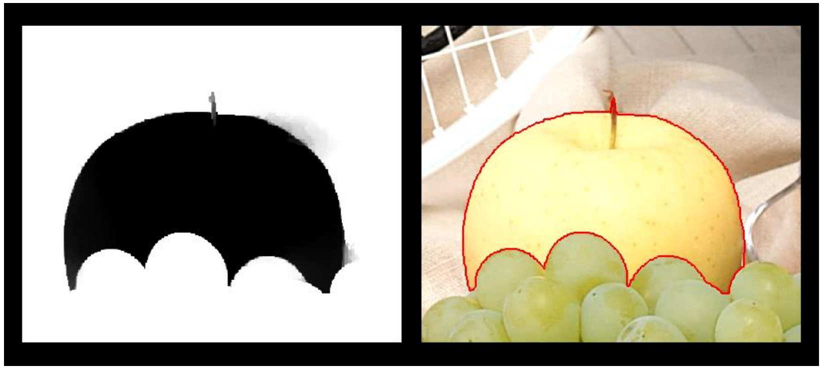
solution x^*_9 cut: threshold of x^*_9

Convergence of RW when p→∞

Input seeds



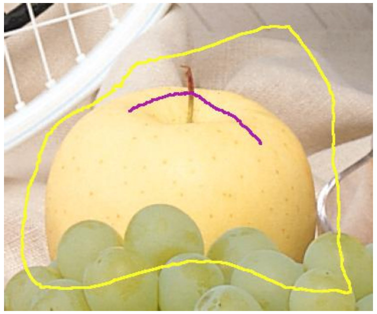
$$x_{13}^* = \arg\min_x \underbrace{\sum_{e_{ij}\in E} w_{ij}^{13}|x_i - x_j|^2}_{\text{Smoothness term}} + \underbrace{\mathcal{D}(x)}_{\text{Data fidelity}}$$



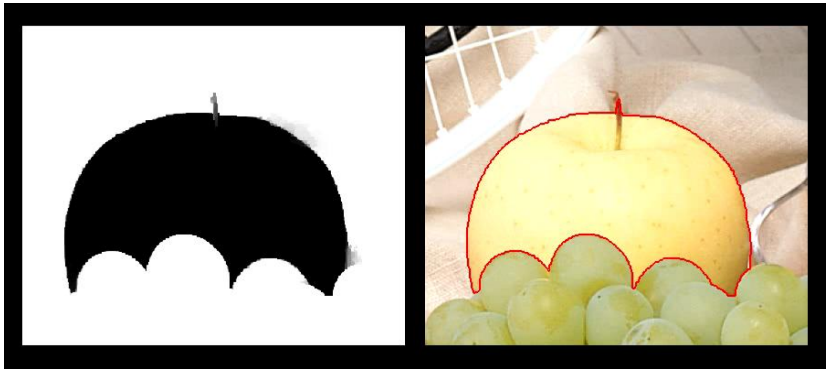
solution x_{13}^* cut: threshold of x_{13}^*

Convergence of RW when p→∞

Input seeds



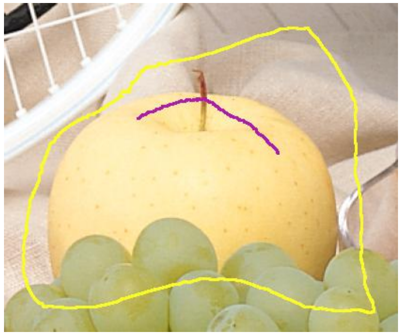
$$x_{18}^* = \arg\min_x \underbrace{\sum_{e_{ij} \in E} w_{ij}^{18} |x_i - x_j|^2}_{\text{Smoothness term}} + \underbrace{\mathcal{D}(x)}_{\text{Data fidelity}}$$



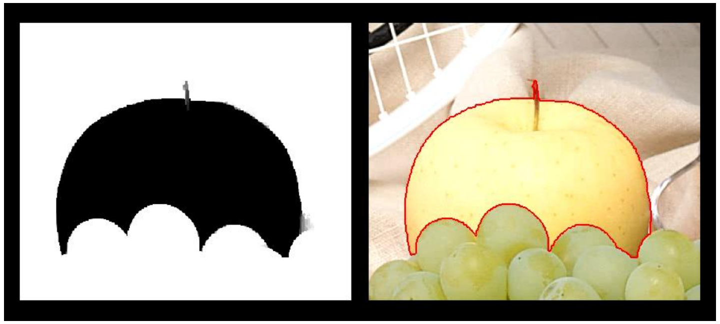
solution x_{18}^* cut: threshold of x_{18}^*

Convergence of RW when p➔∞

Input seeds



$$x_{p}^{*} = \arg\min_{x} \underbrace{\sum_{e_{ij}\in E} w_{ij}^{p} |x_i - x_j|^q}_{\text{Smoothness term}} + \underbrace{\mathcal{D}(x)}_{\text{Data fidelity}}$$



$\bar{x} = \lim_{p\to\infty} x_p^*$ cut: threshold of $\bar{x}$

Theorems

When $p \to \infty$,

- the obtained cut is an MSF cut.
- when $q > 1$, the solution $\bar{x}$ is unique.

The (extended) Power Watershed framework

- Let $p > 0$, $m > 0$, $n > 0$ and
- n real numbers $1 \geq \lambda_0 > \lambda_1 > \ldots \lambda_{n-1} > 0$

$$Q^p(x) = \sum_{0 \leq k < n} \lambda_k^p Q_k(x) \tag{1}$$

where, for all $0 \leq k < n$, $Q_k : \mathbb{R}^m \to \mathbb{R}$ is a continuous function. We search $x^\star \in \mathbb{R}^m$ such that

$$x^\star \in \lim_{p \to \infty} \operatorname*{arg\,min}_{x \in \mathbb{R}^m} Q^p(x) \tag{2}$$

Main PW theorem

Theorem

Set

$$M_0 = \arg\min_{x \in \mathbb{R}^m} Q_0(x) \tag{4}$$

$$\forall\ 1 \leq k < n,\ \ M_k = \arg\min_{x \in M_{k-1}} Q_k(x) \tag{5}$$

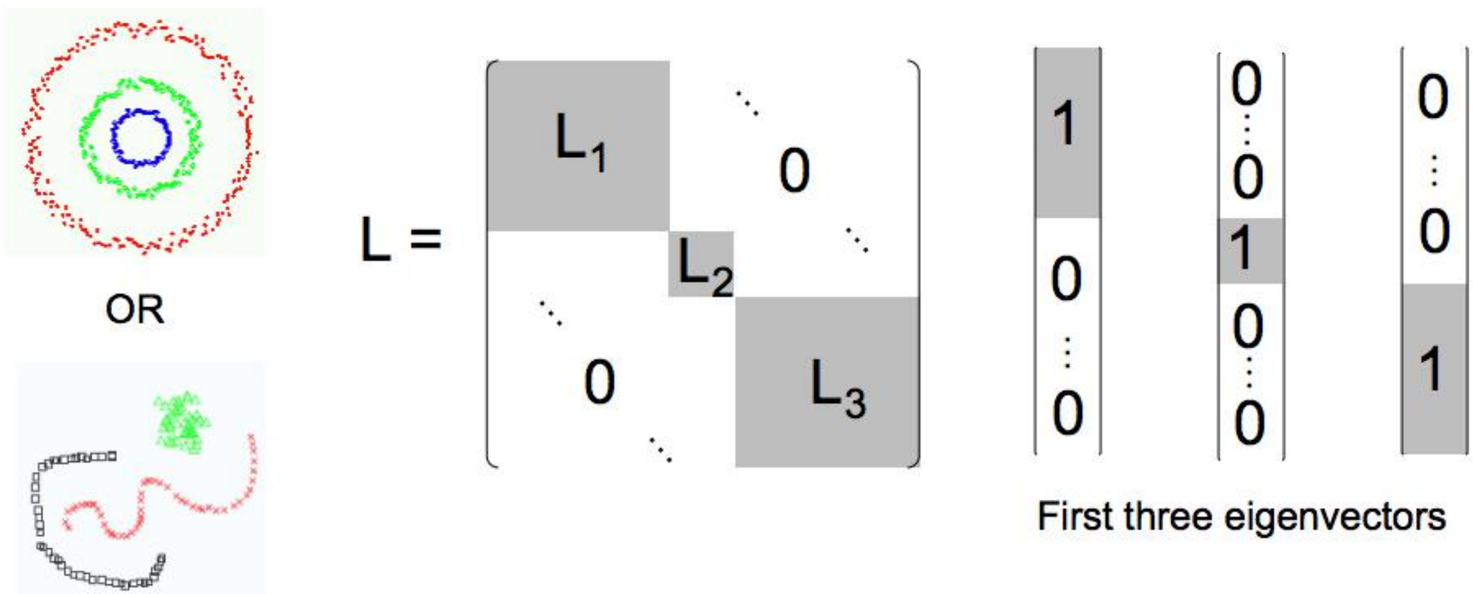
Any convergent sequence $(x_p)_{p>0}$ *of minimizers of* Q^p *converges to some point of* M_{n-1}.
Furthermore, we can estimate the minimum of Q^p *as follows:*

$$\min_{x \in \mathbb{R}^m} Q^p(x) = \sum_{0 \leq k < n} \lambda_k^p m_k + o(\lambda_{n-1}^p) \tag{6}$$

where $m_k = \min_{x \in M_k} Q_k(x)$.

Spectral clustering: intuitive explanation

Let L be one of the (many) graph-Laplacians.



Spectral clustering: ratio-cut

Problem (Ratio-cut algorithm)

For finding k cluster, solve

$$\underset{H \in \mathbb{R}^{m \times k}}{\text{minimize}} \; Tr(H^t L H)$$

$$\textit{subject to } H^t H = \mathbb{I}$$

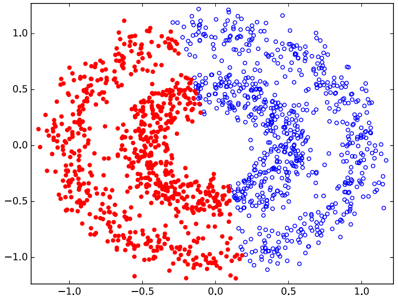
where L is the graph-laplacian

D is the diagonal matrix $diag(d_1, \ldots, d_n)$ with $d_i = \sum_j w_{ij}$, and $L = D - W$.

Ratio-cut

Let L_k as the graph-laplacian of the subgraph induced by the edges whose weights are exactly w_k.

$$\underset{H \in \mathbb{R}^{m \times k}}{\text{minimize}} \sum_{k=1}^{j} w_k \, Tr(H^t L_k H)$$
$$\text{subject to } H^t H = \mathbb{I}$$

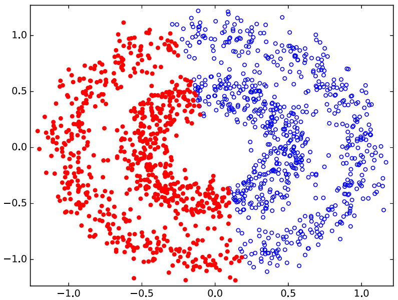


Ratio-cut

Power Ratio-cut

Let L_k as the graph-laplacian of the subgraph induced by the edges whose weights are exactly w_k.

$$\underset{H \in \mathbb{R}^{m \times k}}{\text{minimize}} \sum_{k=1}^{j} w_k^p \, Tr(H^t L_k H)$$
$$\text{subject to } H^t H = \mathbb{I}$$

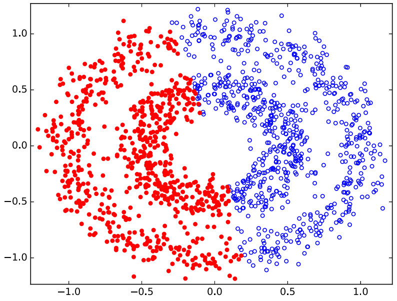


Ratio-cut

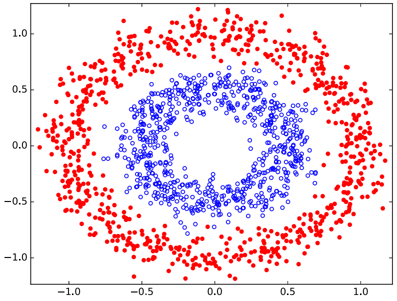
Power Ratio-cut

Let L_k as the graph-laplacian of the subgraph induced by the edges whose weights are exactly w_k.

$$\lim_{p \to \infty} \underset{H \in \mathbb{R}^{m \times k}}{\text{minimize}} \sum_{k=1}^{j} w_k^p \, Tr(H^t L_k H)$$
$$\text{subject to } H^t H = \mathbb{I}$$



Ratio-cut

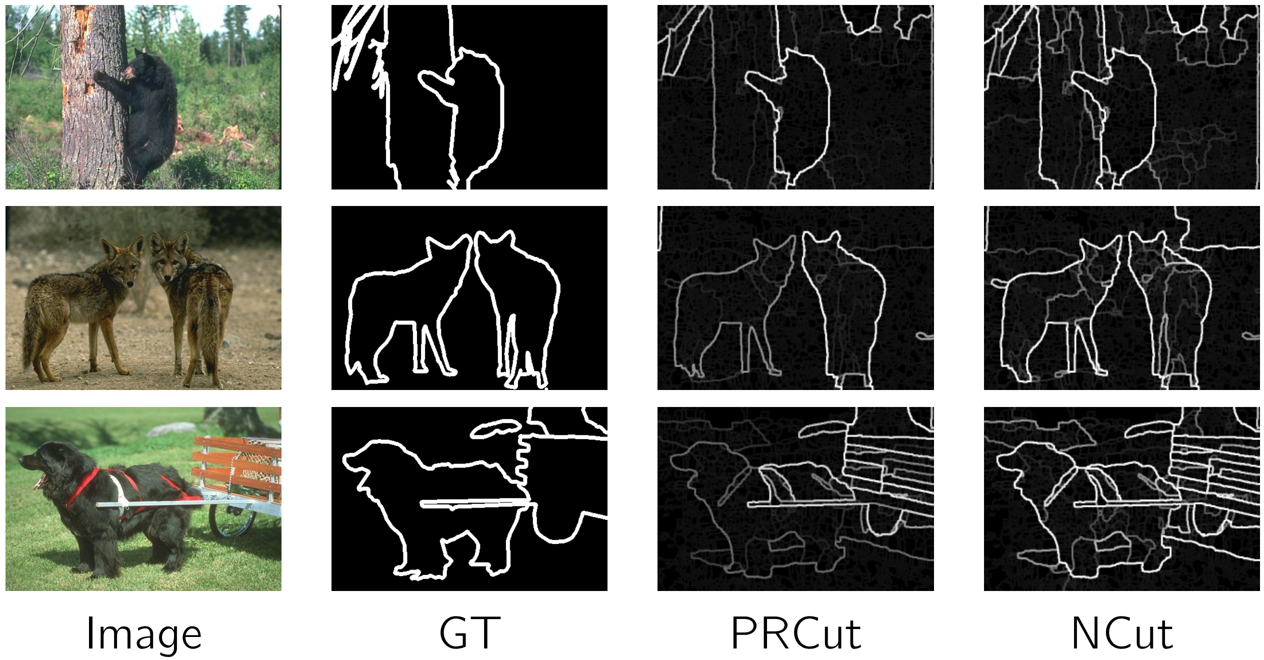


Power ratio-cut

PR-cut in practice

- Cluster the weights with (for example) K-means
- Apply a MST-like algorithm on the clustered weights to get a rough clustering
- Refine the (weighted) borders of the clusters with Ratio-cut

Replacing NCut with PRCut in MCG



Remark

Same quality of results obtained much faster
replacing Normalized Cut by Power Ratio cut
in the Multiscale Combinatorial Grouping technique

Main message

- The center of the clusters are easy to cluster
- Borders are more difficult
- Hence, apply an easy and fast algorithm on the centers (such as a MST), and do something more fancy on the borders

Question

How do we identify the borders and the centers of the cluster?

Main message

- The center of the clusters are easy to cluster
- Borders are more difficult
- Hence, apply an easy and fast algorithm on the centers (such as a MST), and do something more fancy on the borders

Question

How do we identify the borders and the centers of the cluster?

Use a MST!

MM in data science

- Is usefull ☺
- Need to revisit everything we have done
  - From a new perspective
- Much work to do!