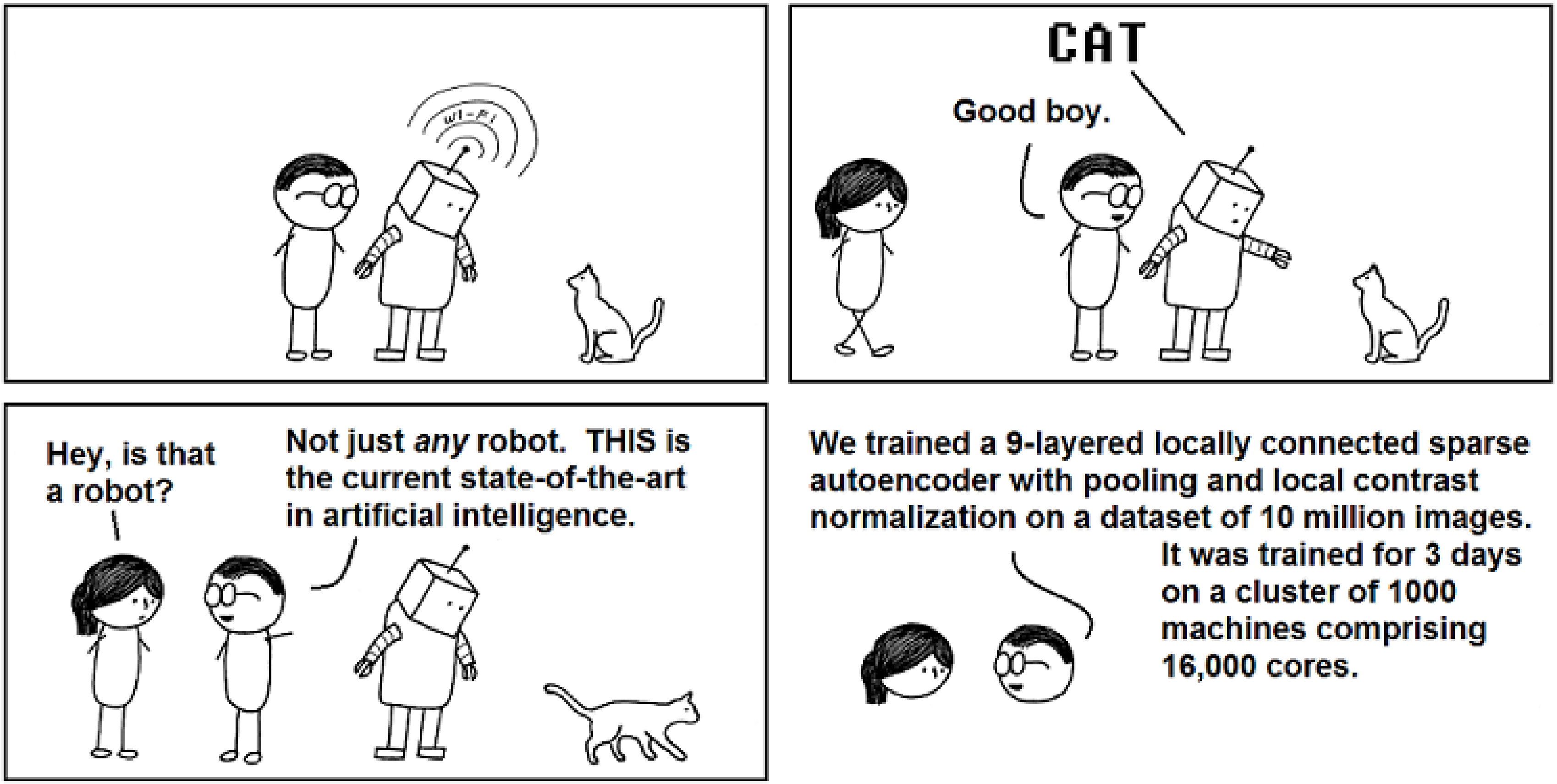
Introduction of Unsupervised Image Segmentation

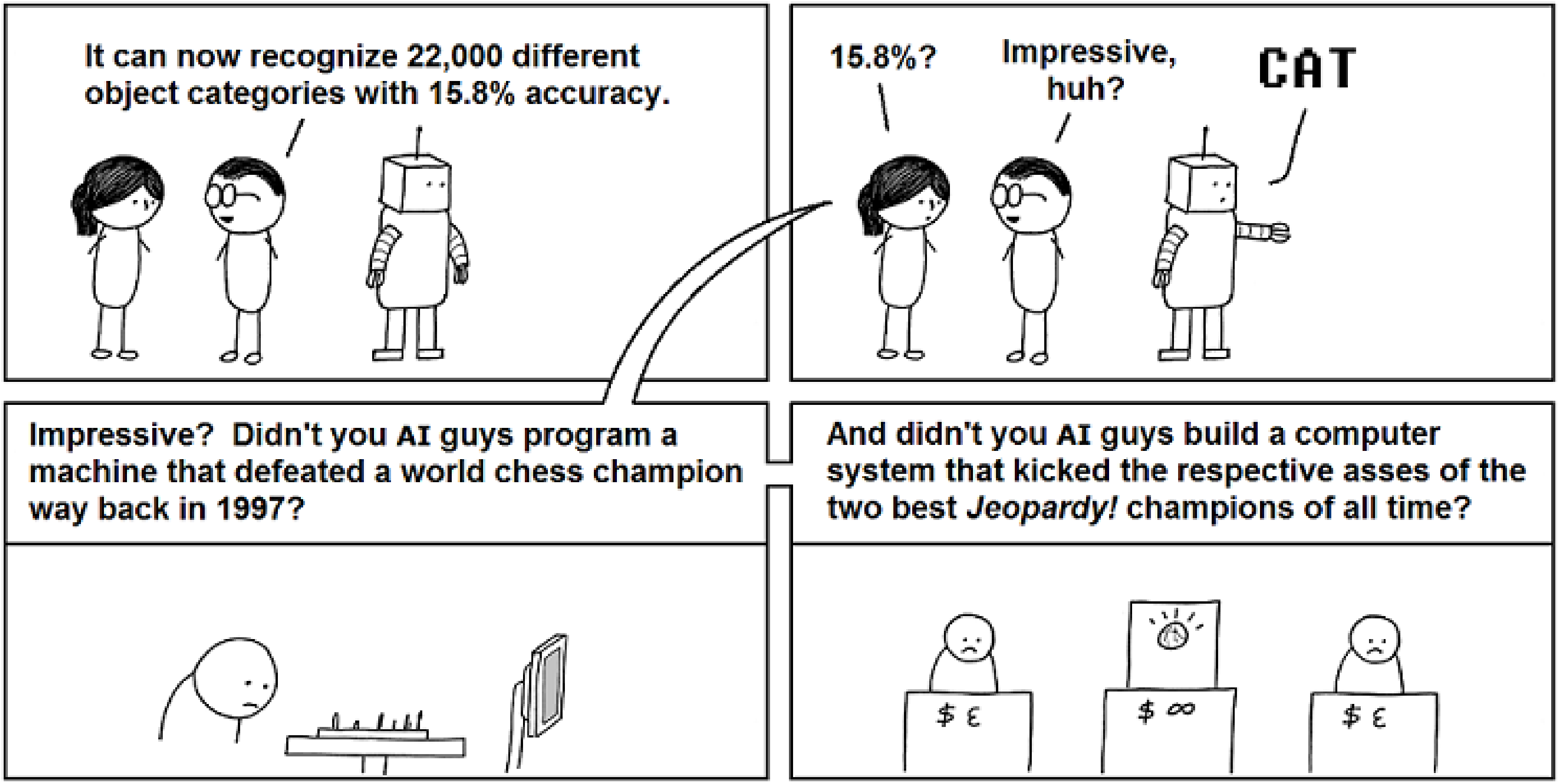
Presenter:
Haw-Shiuan Chang(張浩軒)

Normalized cuts and image segmentation, PAMI 2000
Mean shift: a robust approach toward feature space analysis, PAMI 2002
Efficient graph-based image segmentation, IJCV 2004
Contour Detection and Hierarchical Image Segmentation, PAMI 2011
Image Segmentation by Cascaded Region Agglomeration, CVPR 2013

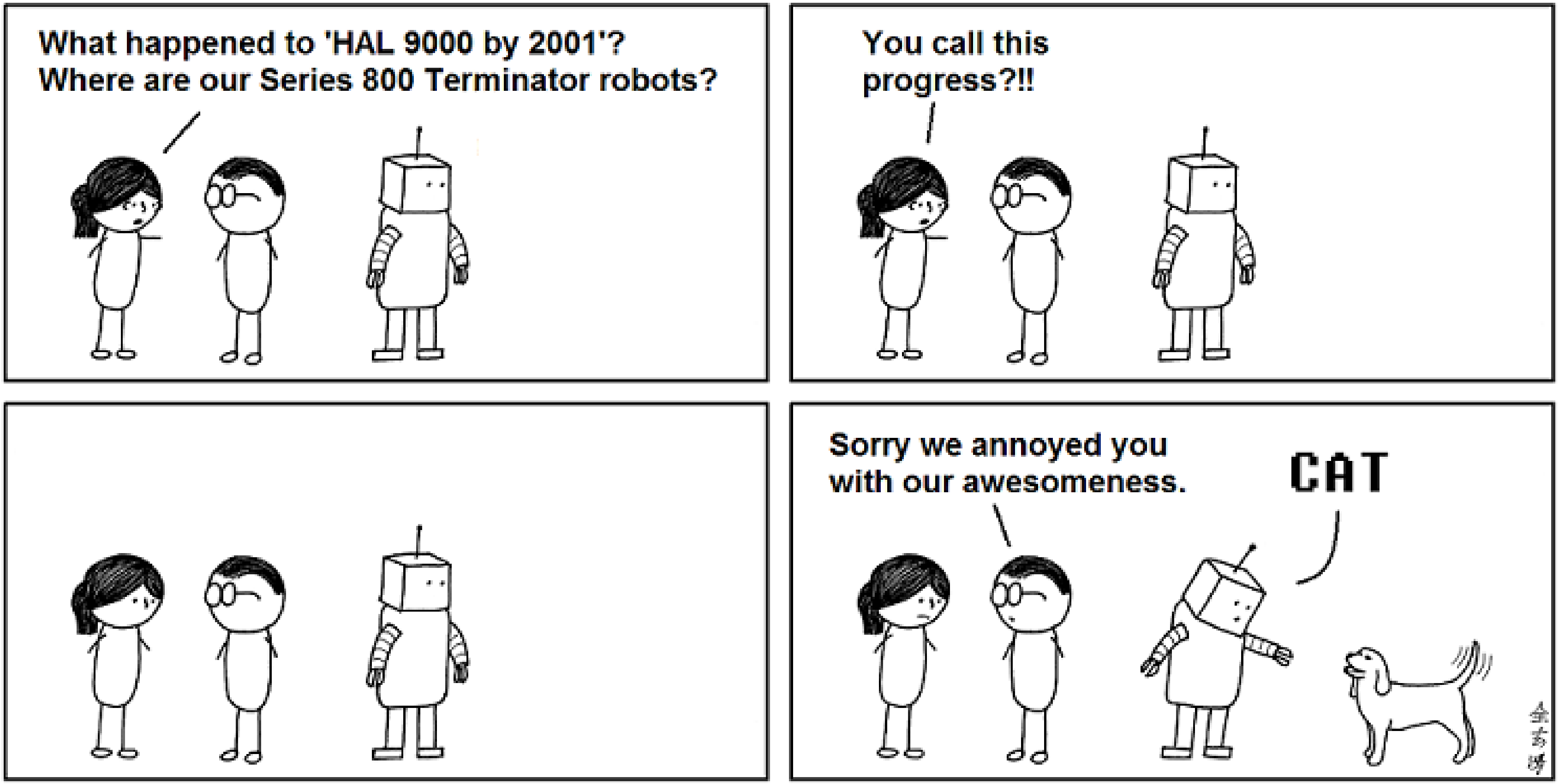
The Dream of Computer Vision

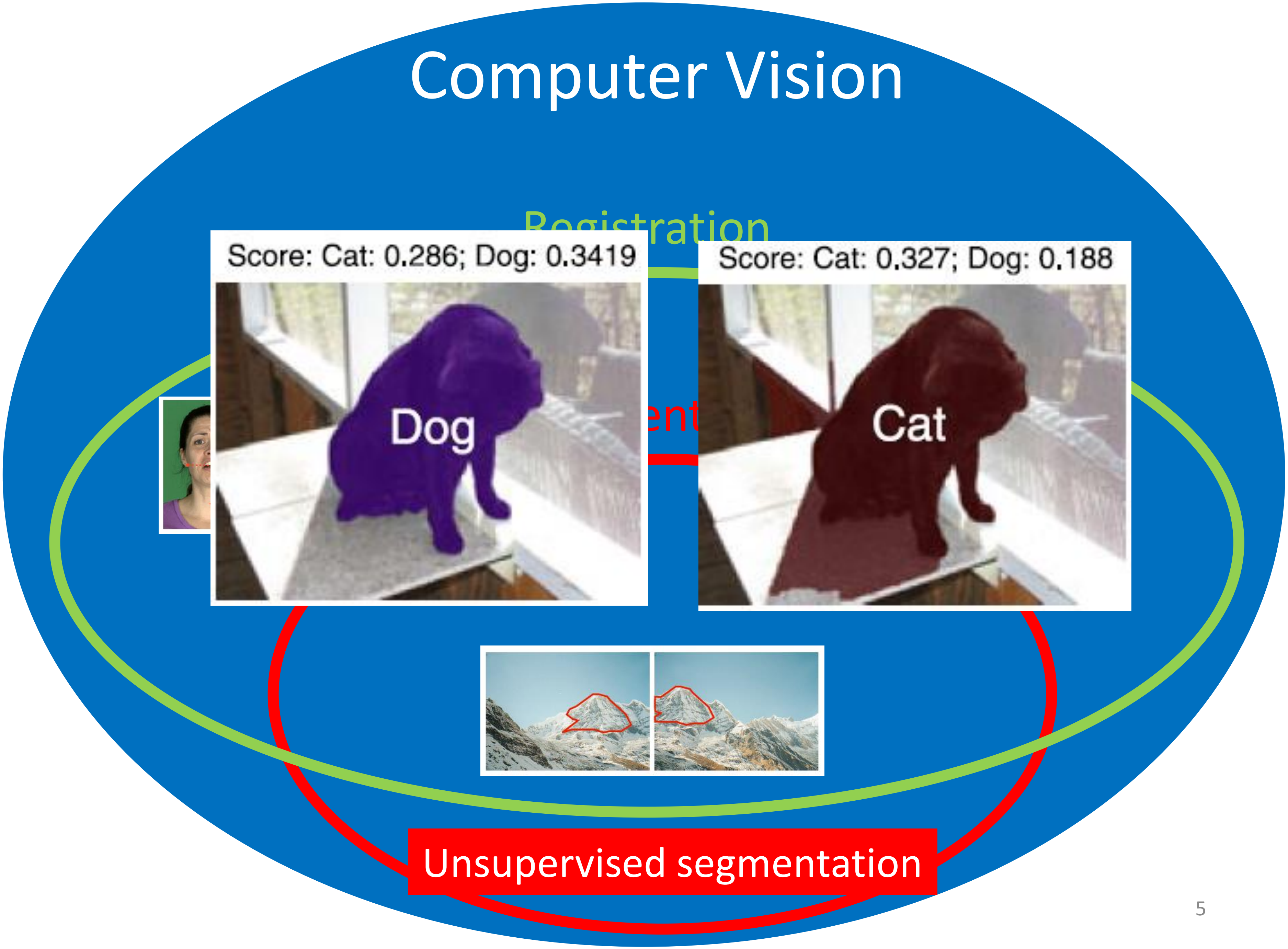


The Dream of Computer Vision



The Dream of Computer Vision




Computer Vision
Registration
Score: Cat: 0.286; Dog: 0.3419
Dog
Score: Cat: 0.327; Dog: 0.188
Cat
Unsupervised segmentation

Outline

- Introduction
- Methods
  - NCut
  - FH
  - Mean Shift
  - gPb-OWT-UCM
  - ISCRA
- Experimental Results
- Conclusion

Introduction

Supervised Image segmentation
Unsupervised Image segmentation
Application
Challenge
Methods

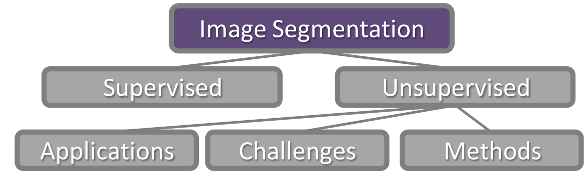


Image Segmentation

- *Image segmentation is one of the fundamental and most studied problems in computer vision.*

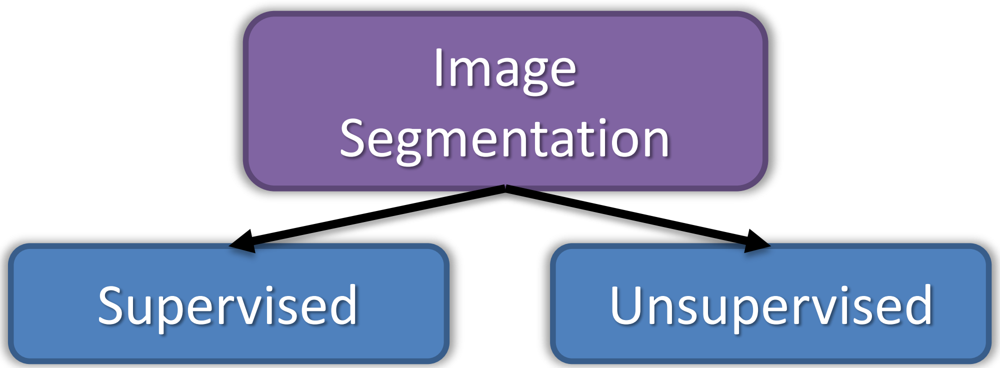
P. Arbelaez and L. Cohen

"Constrained Image Segmentation from Hierarchical Boundaries," CVPR 2008

- *Image segmentation is a fundamental low-level vision problem with a great potential in applications.*

Z. Li, X. Wu, and S. Chang

"Segmentation using superpixels: a bipartite graph partitioning approach," CVPR 2012

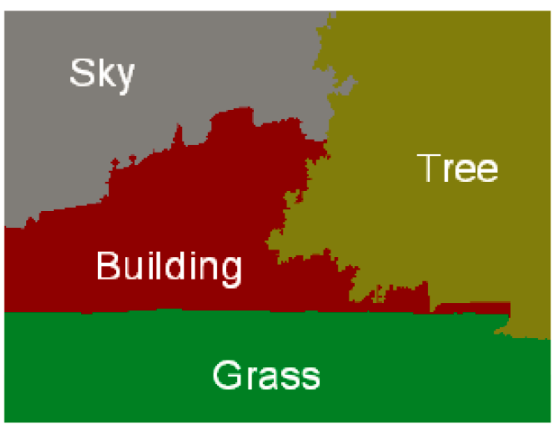


Supervised vs. Unsupervised

- Supervised image segmentation (semantic segmentation)

Image

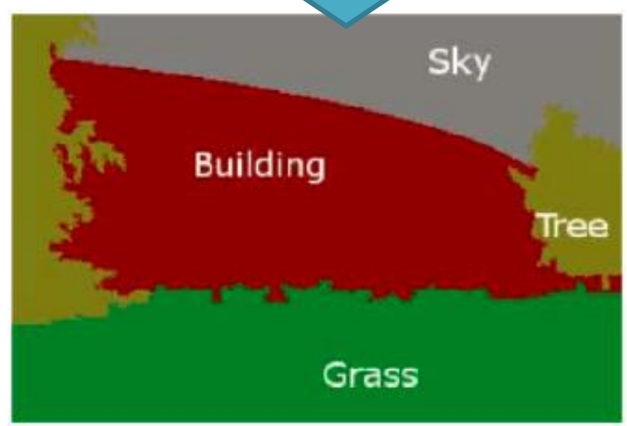
Labeled ground-truth



Training phase

Test Image

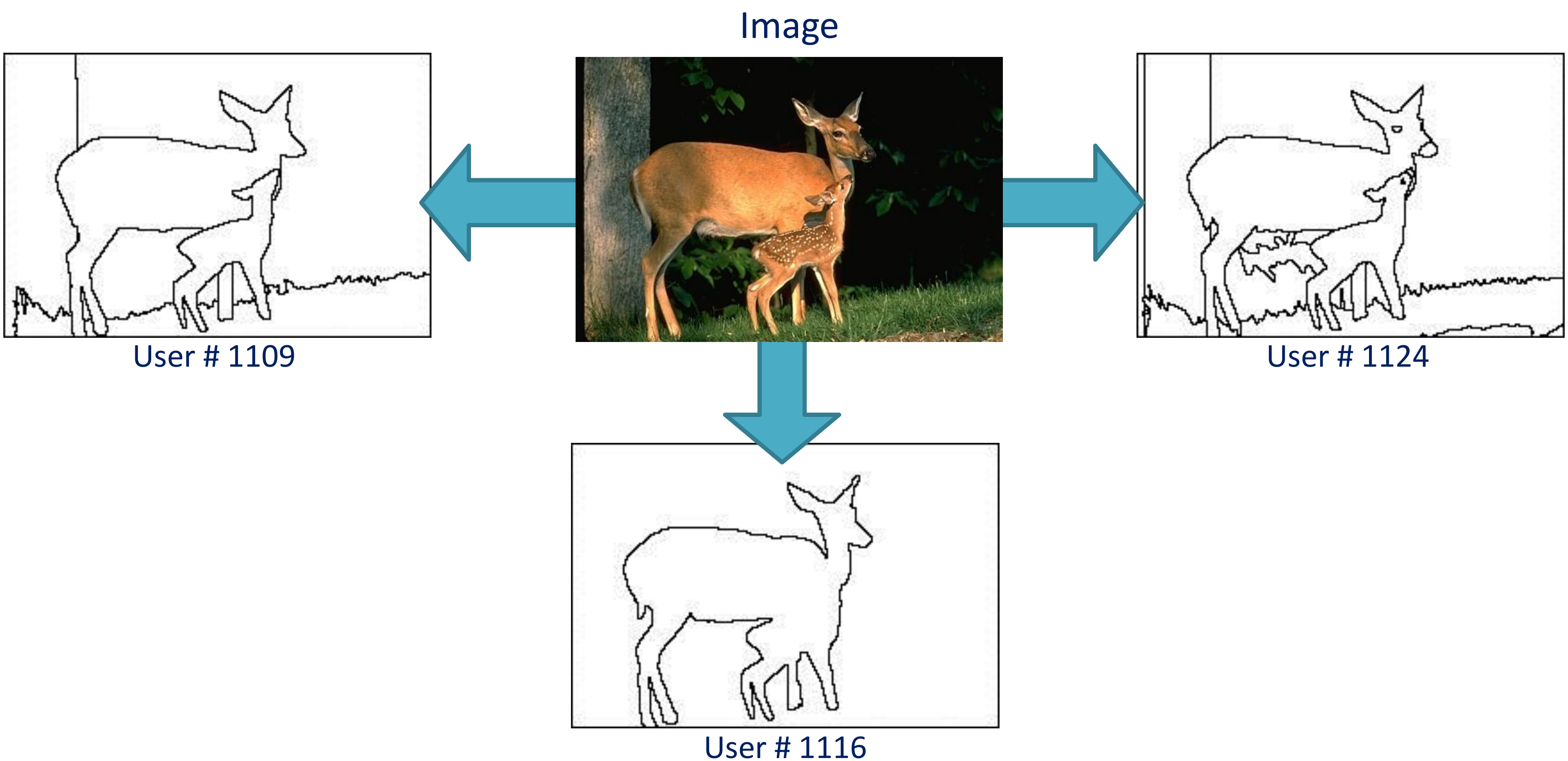
Goal



Testing phase

Supervised vs. Unsupervised

- Unsupervised image segmentation



Applications

- Why segmentation?
  - To increase the accuracy of registration
  - Allow the algorithm with higher time complexity

- E.g., object recognition

Object Recognition by Sequential Figure-Ground Ranking, IJCV 2012

Image

Traditional object recognition

Segmentation-based object recognition

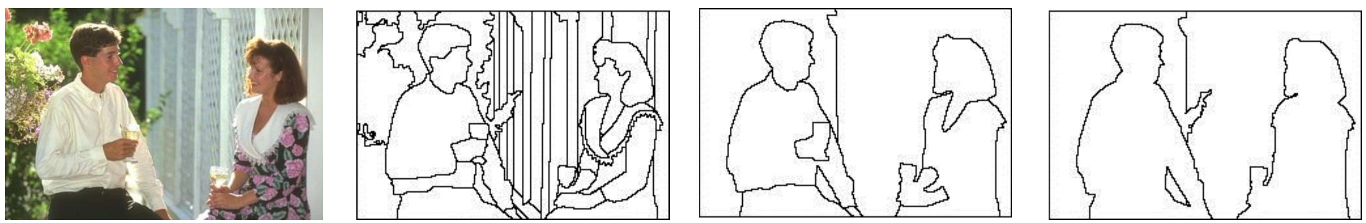
If time complexity of recognition is $O(p^2)$, we often can reduce to $O(sp^2)+O(p)$, where p is pixel number, sp is superpixel number (e.g. p=150000, sp=1000).

Applications

Applications	Main Components	
Object recognition / Semantic segmentation / Saliency	Classification	Unsupervised Image Segmentation
Image retrieval / Object retrieval / Co-segmentation	Matching	
Video segmentation	Motion estimation	
Image editing / Video editing	User interaction	
Video summary	Some of the above	

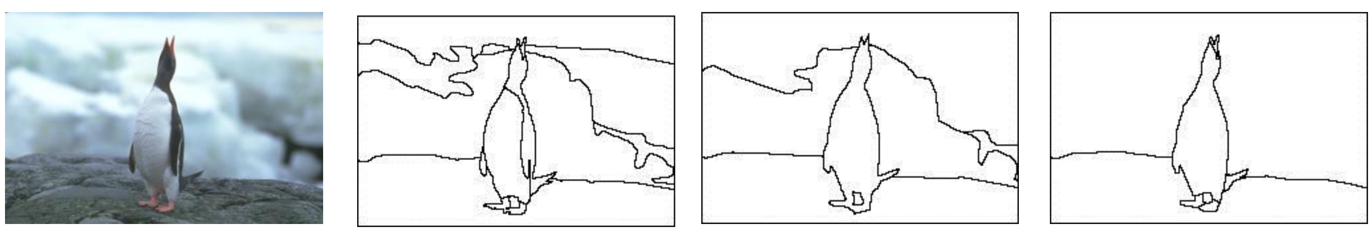
Challenges - objective function

- It is hard to define what good segmentation is because
  - Segmentation ambiguity



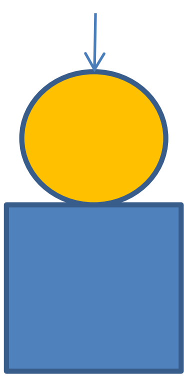
Fine → Coarse

  - Semantic gap
    - feature{Color, texture, shape, etc.} ?→ high-level concepts



  - Diverse image contexts (object types are unknown)

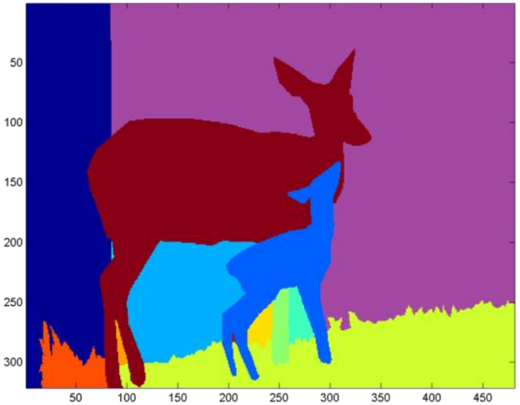
Head or ball?
Impossible
to know!



Challenges - optimization

- Basically, it is a NP-hard problem for most of the objective functions.
  - Like most of unsupervised clustering tasks
- Enormous search space
  - The upper bound of search space size (# possible solution) is $\frac{k^P}{k!}$ ($\sim 13^{150000}$)!
    - assuming that
      - the image is 300x500, # pixels (**P**) is 150000
      - ground truth have **k**=13 segments
      - regardless connectivity constraint
- Enumerate the search space is difficult.
  - Because of connectivity constraint

input



One output over $\sim 13^{150000}$ possibility!

The number of pixels has been saw by human:

640 x 480 x 30 x 60 x 60 x 24 x 365 x 6,000,000,000 x 200,000 ≈ 10^{30}

pixel fps secs mins hrs days people years

Methods

- I will introduce the methods which are widely used because of their
  - Simplicity
  - Efficiency
  - Good performance

	Performance	Efficiency	Complexity (# parameters)	Theoretical support
NCut (2000)	Bad (original version)	Reasonable	Simple	Best
FH (2004)	Reasonable (↑ for superpixels)	Fastest	Simple	Empirical
Mean shift (2002)	Reasonable (↑ for oversegments)	Reasonable	Simple	Good
gPb-OWT-UCM (need training) (2011)	Good	Slow	Complex	Empirical
ISCRA (2013+) (code not released)	Best (state of the art)	Slowest (need to use gPb)	Complex	Empirical

Methods

Ncut (Normalized Cut)

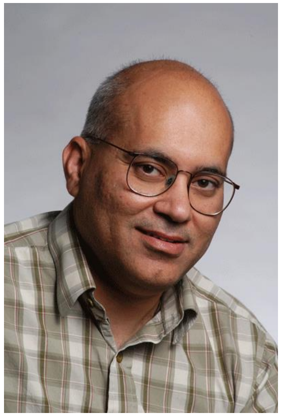
FH (Efficient graph-based image segmentation)

Mean Shift

gPb-OWT-UCM

ISCRA (Image Segmentation by Cascaded Region Agglomeration)

Citation: 7931
(since year 2000)

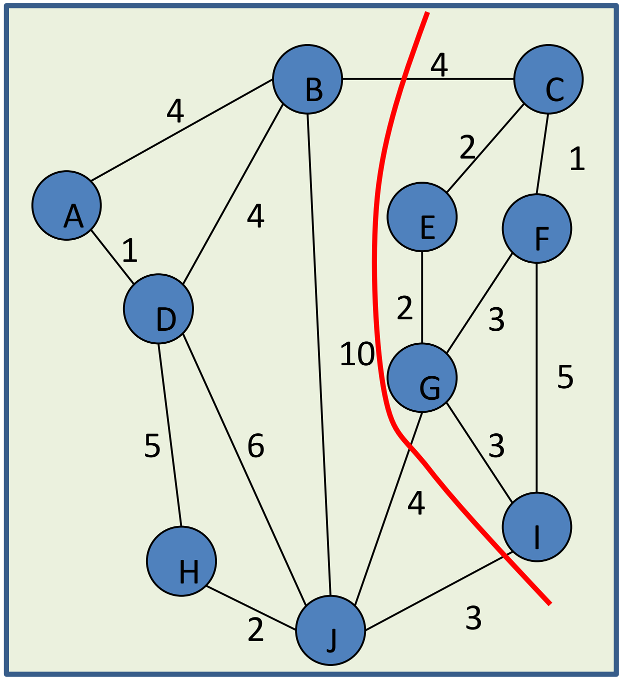


Normalized Cut

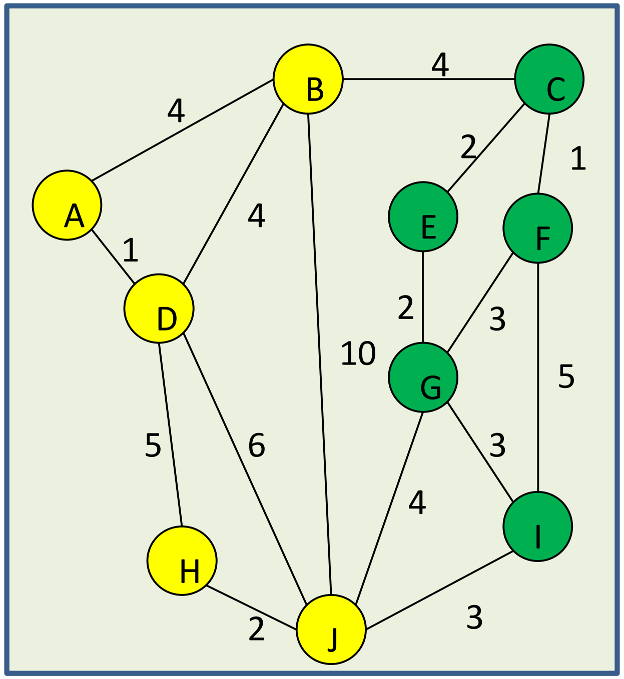
Normalized cuts and image segmentation, PAMI 2000

- NCut is one type of graph based clustering techniques.
  - similar to the spectral clustering
- NCut often becomes a component of other segmentation algorithms.

Assuming that we want to divide the image into **2** segments

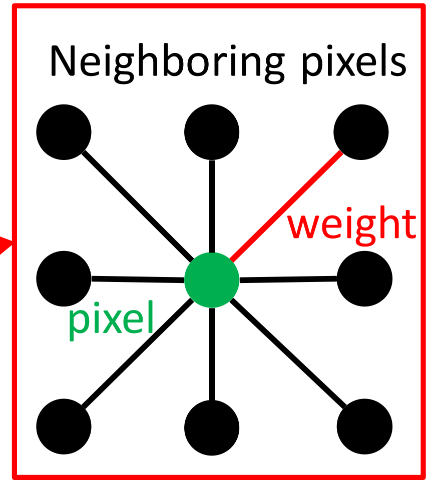


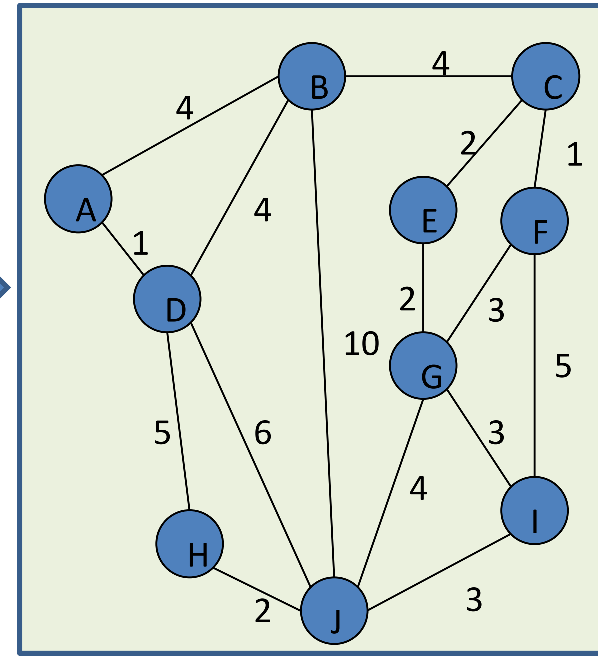
Graph



Clustered result

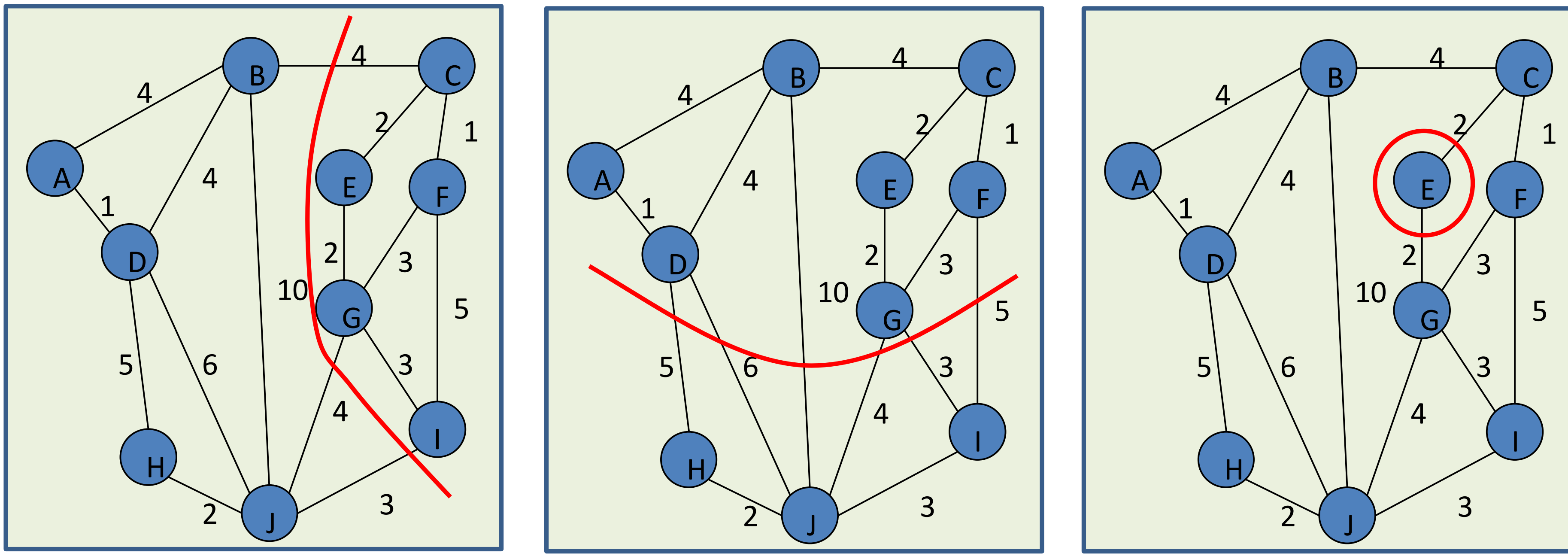
Images as Graphs





- Node for every pixel
- Edges between “neighbors”
- Edge weights/capacities are some measure of similarity.

What is a good segmentation?



(a) (b) (c)

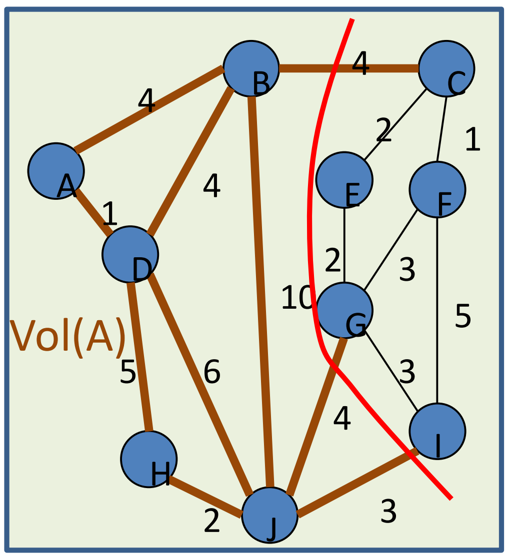
- (a) is better than (b) because it cuts edges with smaller weights.
- (a) is better than (c) because the clusters having similar sizes.

Objective function

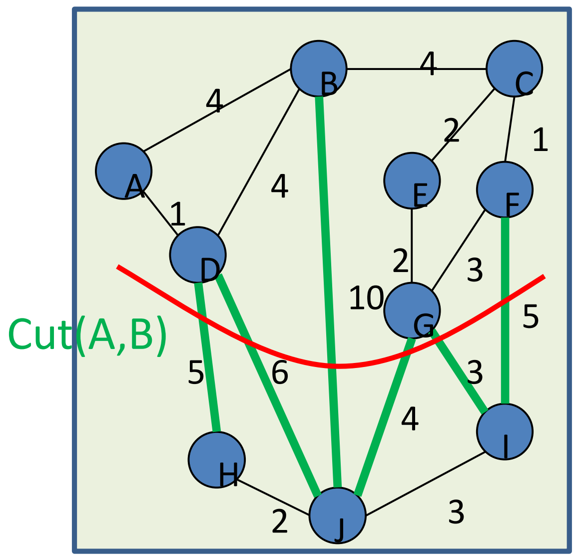
The weights between segment A and B

minimize → $$NCut(A,B) = Cut(A,B)\left(\frac{1}{vol(A)} + \frac{1}{vol(B)}\right)$$

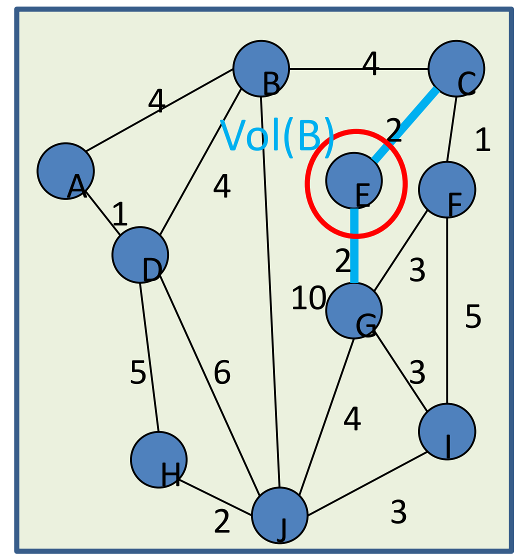
The weights within segment A



NCut=0.66 → best
(11/43+11/27)



NCut=1.48
(33/54+33/38)

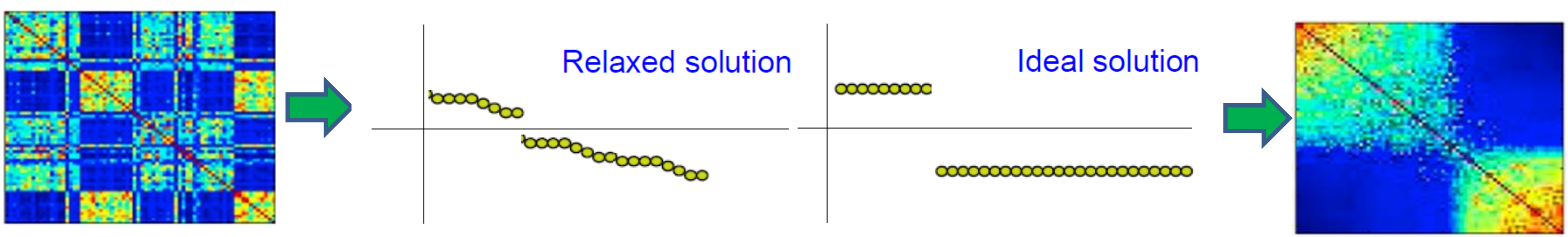


NCut=1.07
(4/59+4/4)

Optimization

$$NCut(A,B) = Cut(A,B)(\frac{1}{vol(A)} + \frac{1}{vol(B)})$$

- The problem is NP-hard.
  - Finding the global minimum needs to exam all possible cuts.
- Relax the problem (discrete → continual)
  - The pixel (node n) belonging to segments P(A|n)
    - e.g., P(A|n)=0.6 and P(B|n)=0.4
- After some derivations, NCut turns out to be an eigenvalue problem.



Affinity matrix

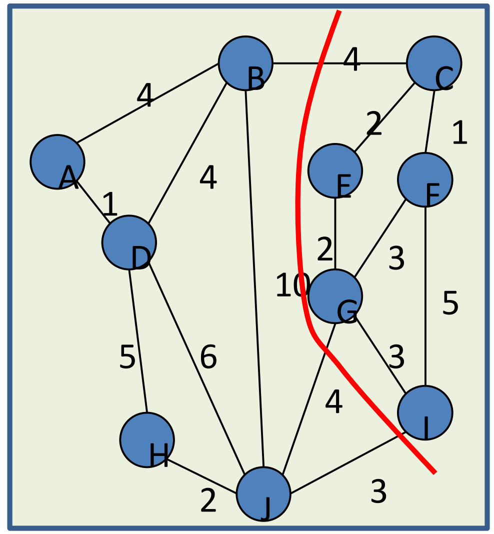
The eigen vector corresponding to P(A|n)

Clustering result

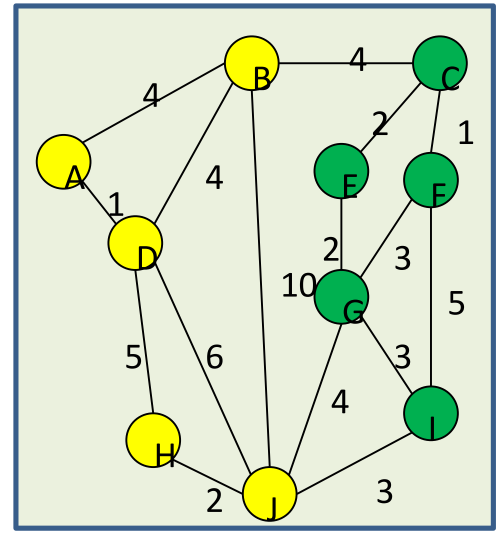
Conclusion - NCut

Normalized cuts and image segmentation, PAMI 2000

We repeat the process until the ideal # segments is reached.



Graph



Clustered results

	Performance	Efficiency	Complexity (# parameters)	Theoretical support
Ncut (2000)	Bad (original version)	Reasonable	Simple	Best

The main limitation is the assumption that the segments tend to have same sizes, but the drawback could be alleviated if we use more sophisticate methods to build the graph.

Methods

Ncut (Normalized Cut)

FH (Efficient graph-based image segmentation)

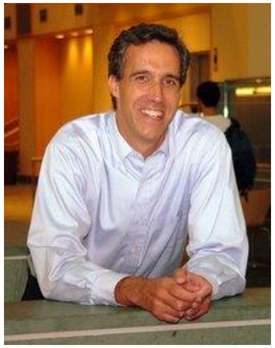
Mean Shift

gPb-OWT-UCM

ISCRA (Image Segmentation by Cascaded Region Agglomeration)

Citation: 2133
(since year 2004)

Pedro Felzenszwalb



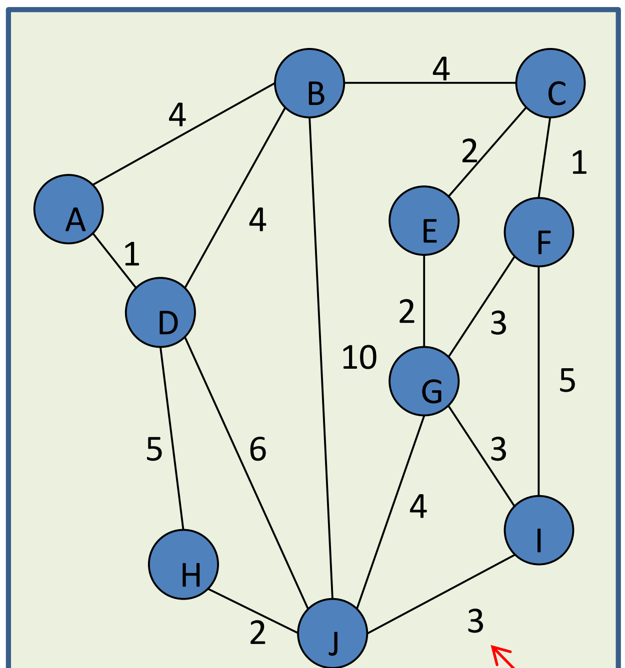
Daniel Huttenlocher

Efficient graph-based image segmentation

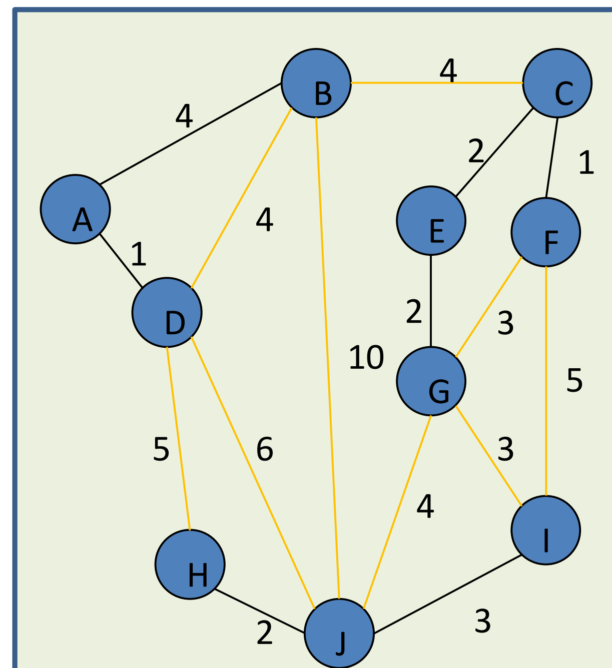
Efficient graph-based image segmentation, IJCV 2004

- Similar to single linkage clustering or minimal spanning tree
- Widely used in video segmentation

Image



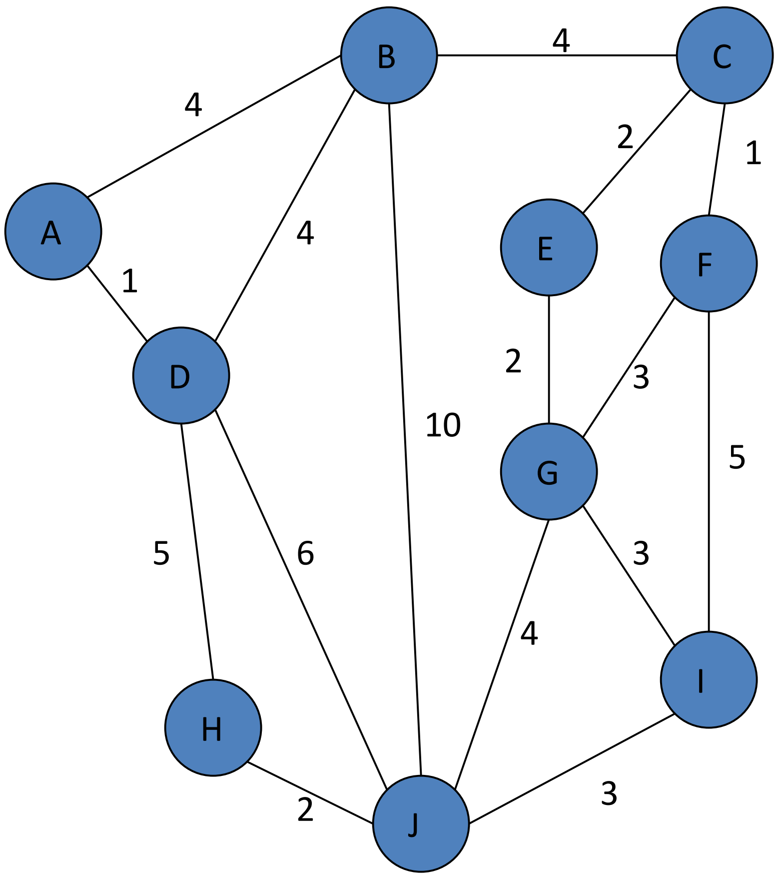
Graph

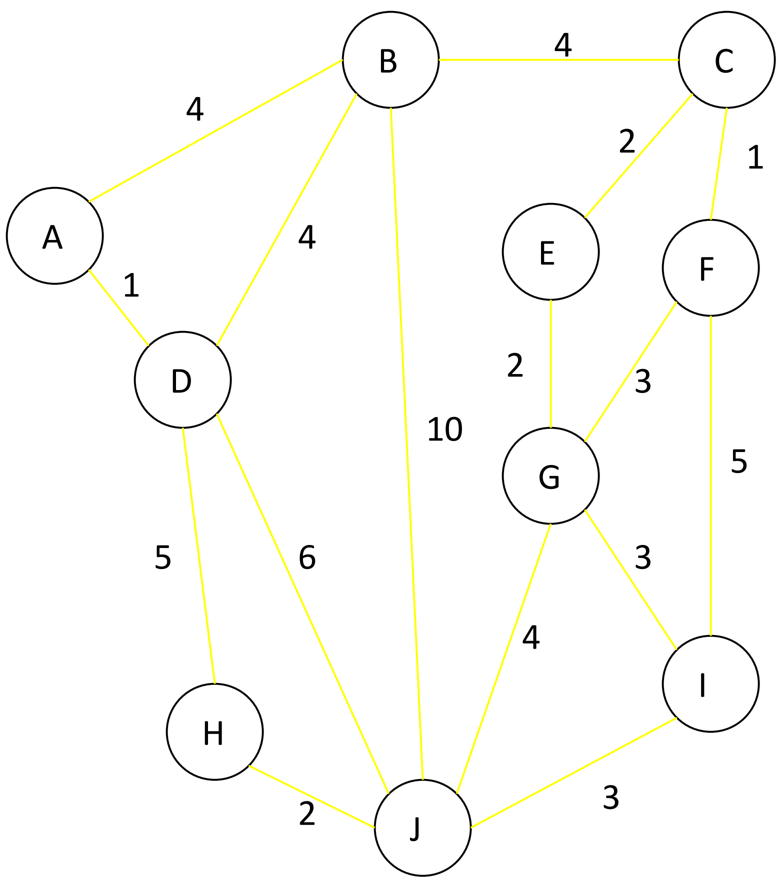


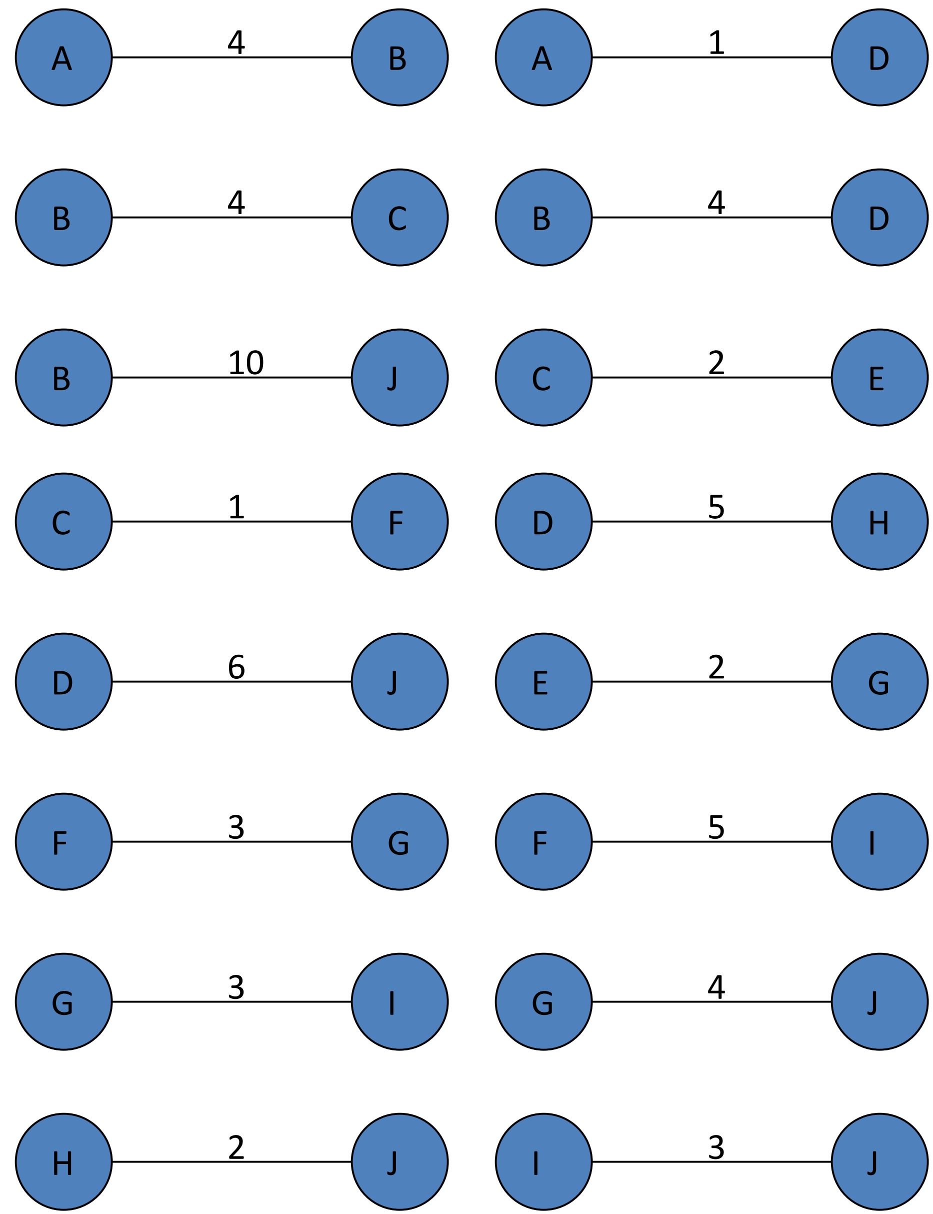
Clustered results

dissimilarity

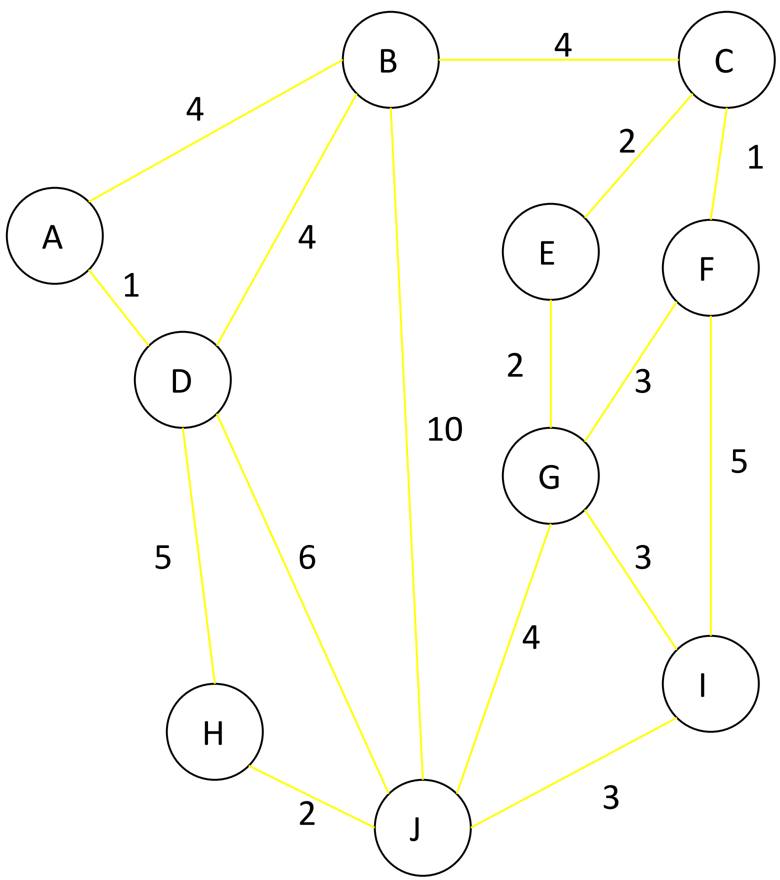
Image Graph

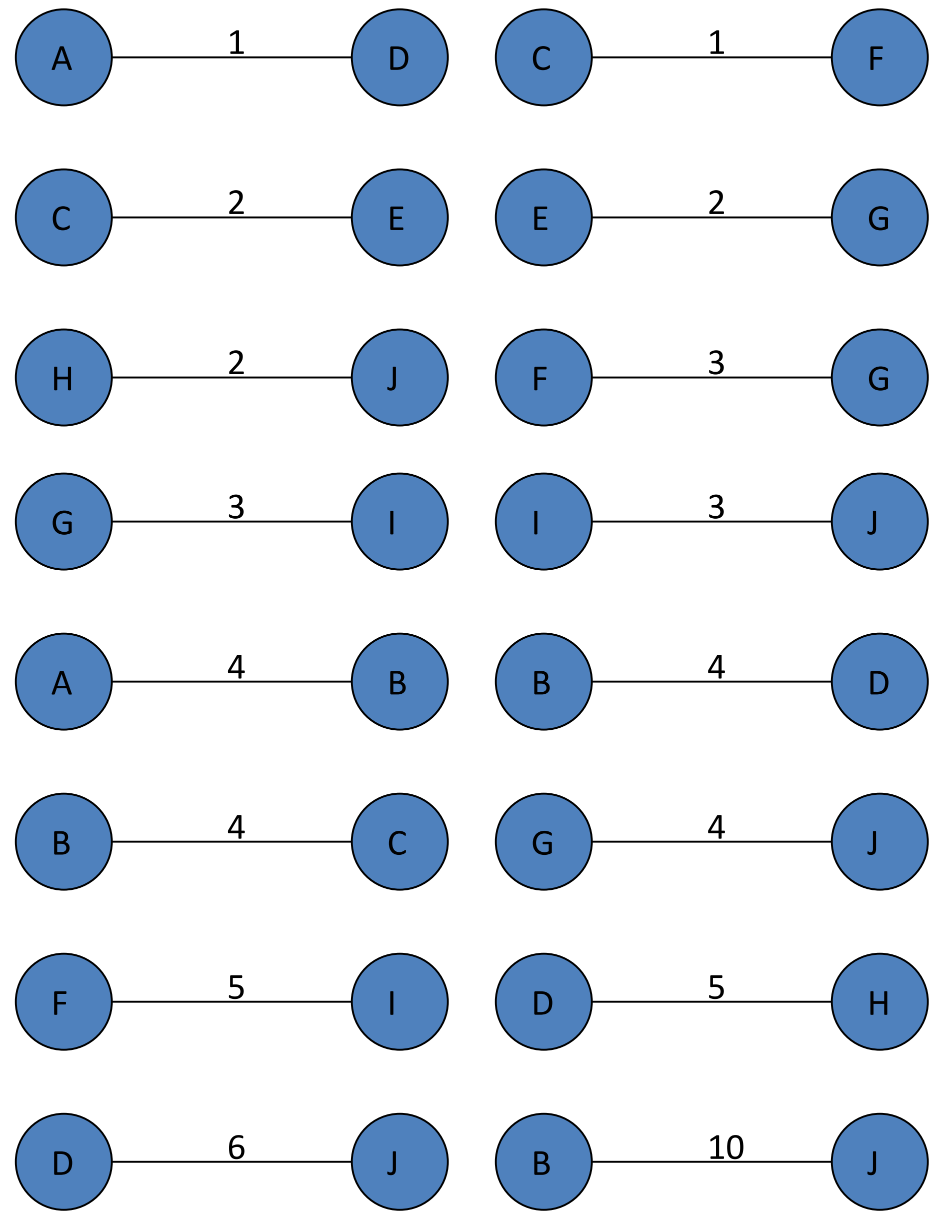



A
B
C
D
E
F
G
H
I
J
4
4
4
1
2
1
10
2
3
5
5
6
4
3
2
3

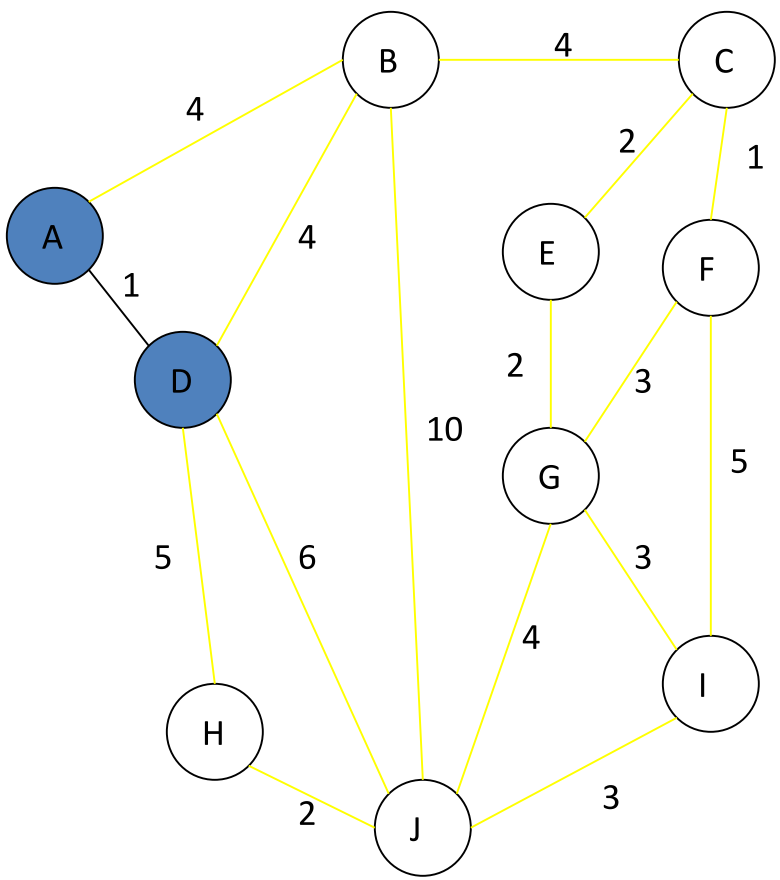

A 4 B
A 1 D
B 4 C
B 4 D
B 10 J
C 2 E
C 1 F
D 5 H
D 6 J
E 2 G
F 3 G
F 5 I
G 3 I
G 4 J
H 2 J
I 3 J

Sort Edges

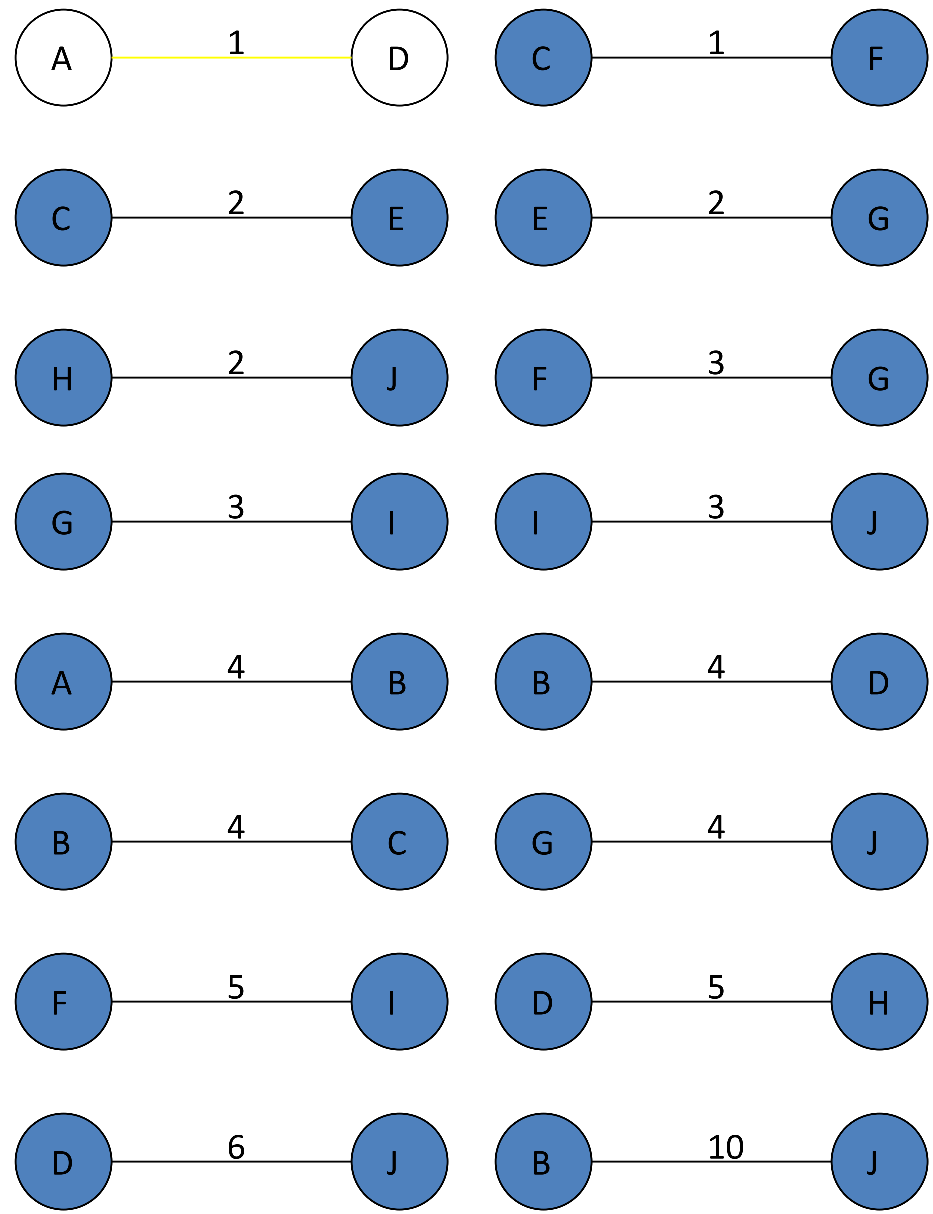

A
B
C
D
E
F
G
H
I
J
4
4
4
1
2
1
2
3
10
5
6
4
3
5
2
3


A 1 D
C 1 F
C 2 E
E 2 G
H 2 J
F 3 G
G 3 I
I 3 J
A 4 B
B 4 D
B 4 C
G 4 J
F 5 I
D 5 H
D 6 J
B 10 J

Add Edge

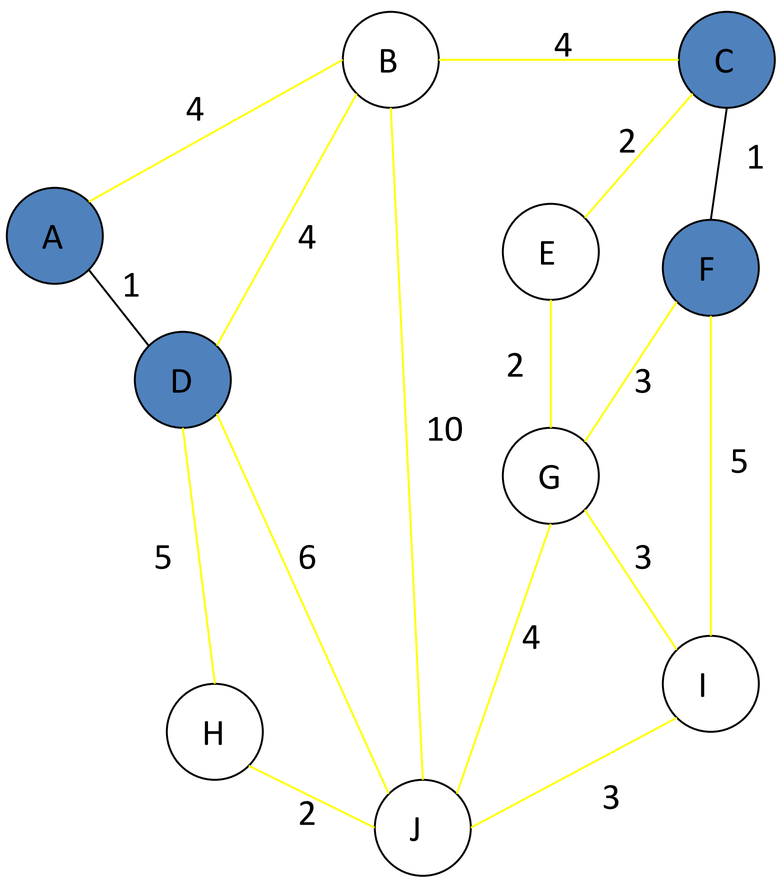


A
B
C
D
E
F
G
H
I
J
4
4
1
4
2
1
2
3
10
5
5
6
3
4
2
3

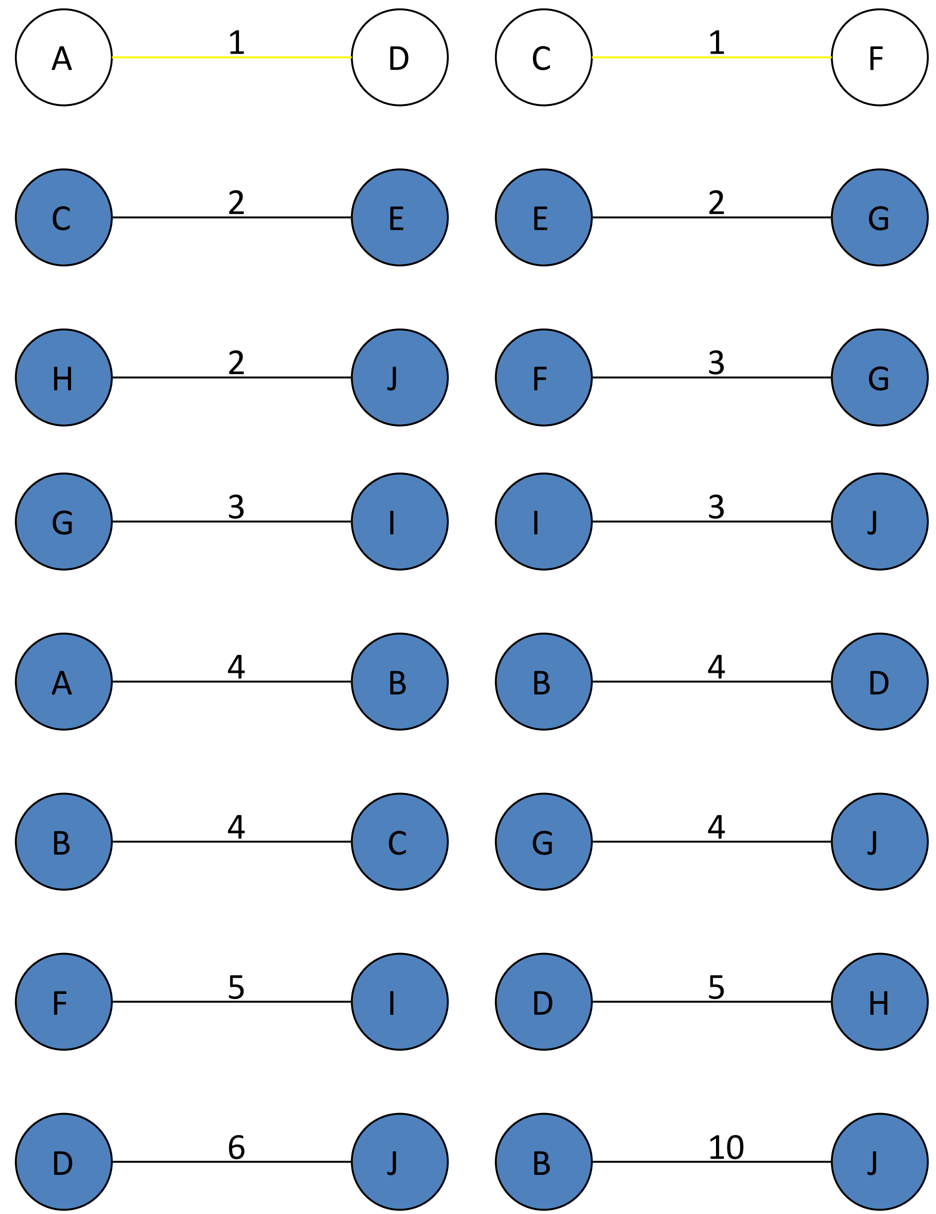


A 1 D
C 1 F
C 2 E
E 2 G
H 2 J
F 3 G
G 3 I
I 3 J
A 4 B
B 4 D
B 4 C
G 4 J
F 5 I
D 5 H
D 6 J
B 10 J

Add Edge

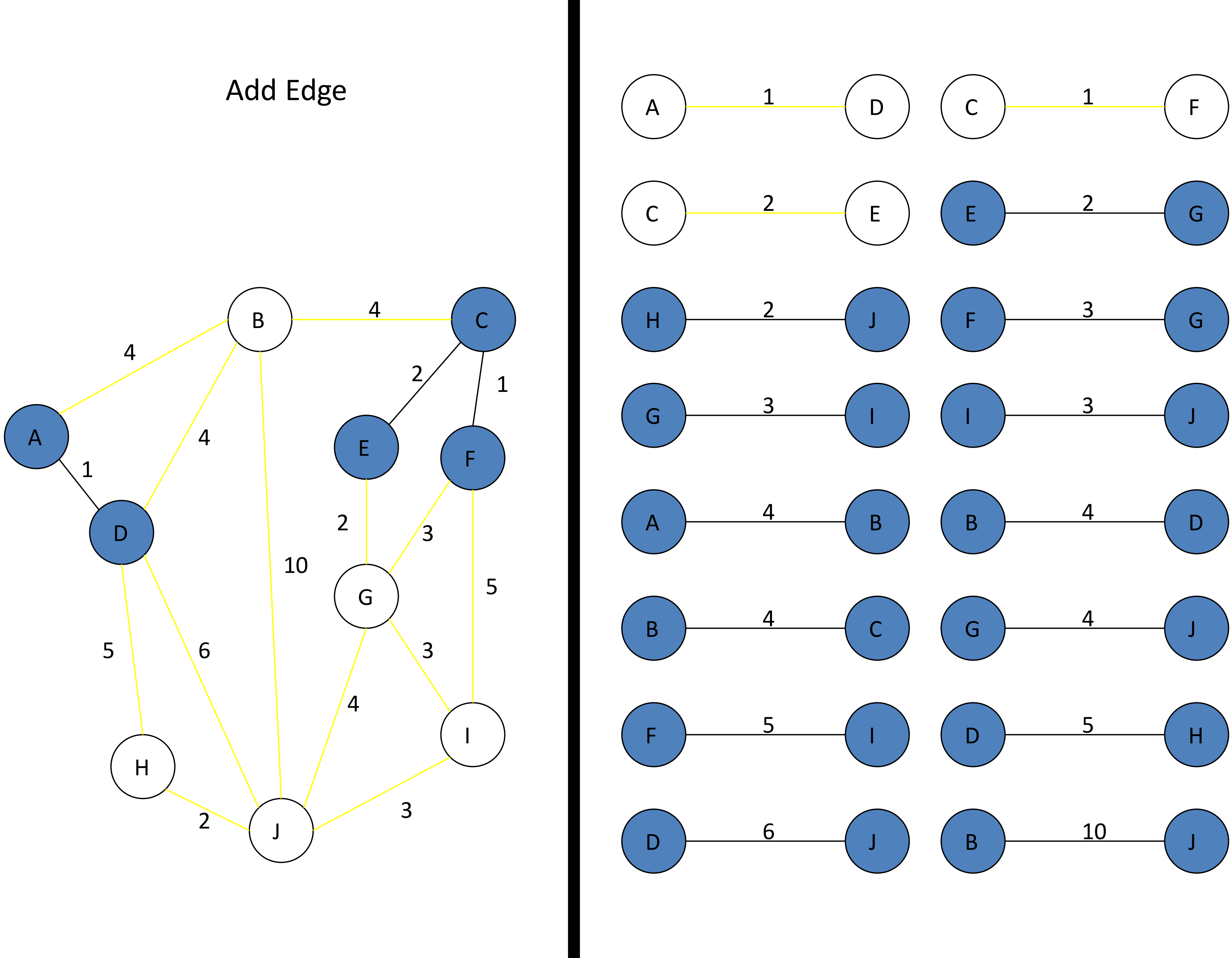


A
B
C
D
E
F
G
H
I
J
4
4
1
1
2
4
2
10
3
5
5
6
3
4
2
3



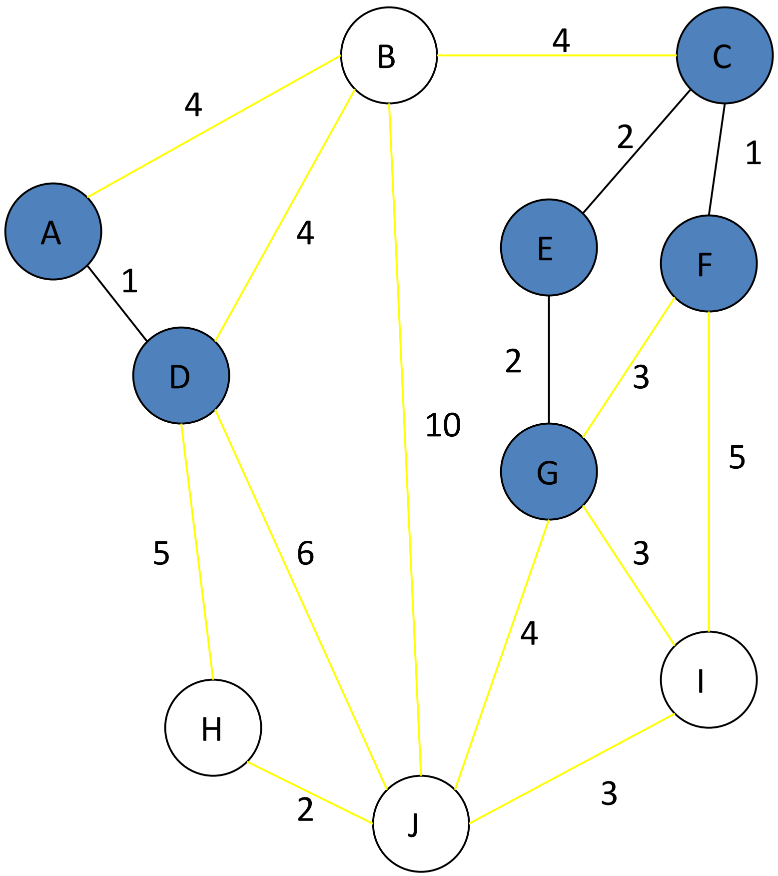
A 1 D
C 1 F
C 2 E
E 2 G
H 2 J
F 3 G
G 3 I
I 3 J
A 4 B
B 4 D
B 4 C
G 4 J
F 5 I
D 5 H
D 6 J
B 10 J

Add Edge

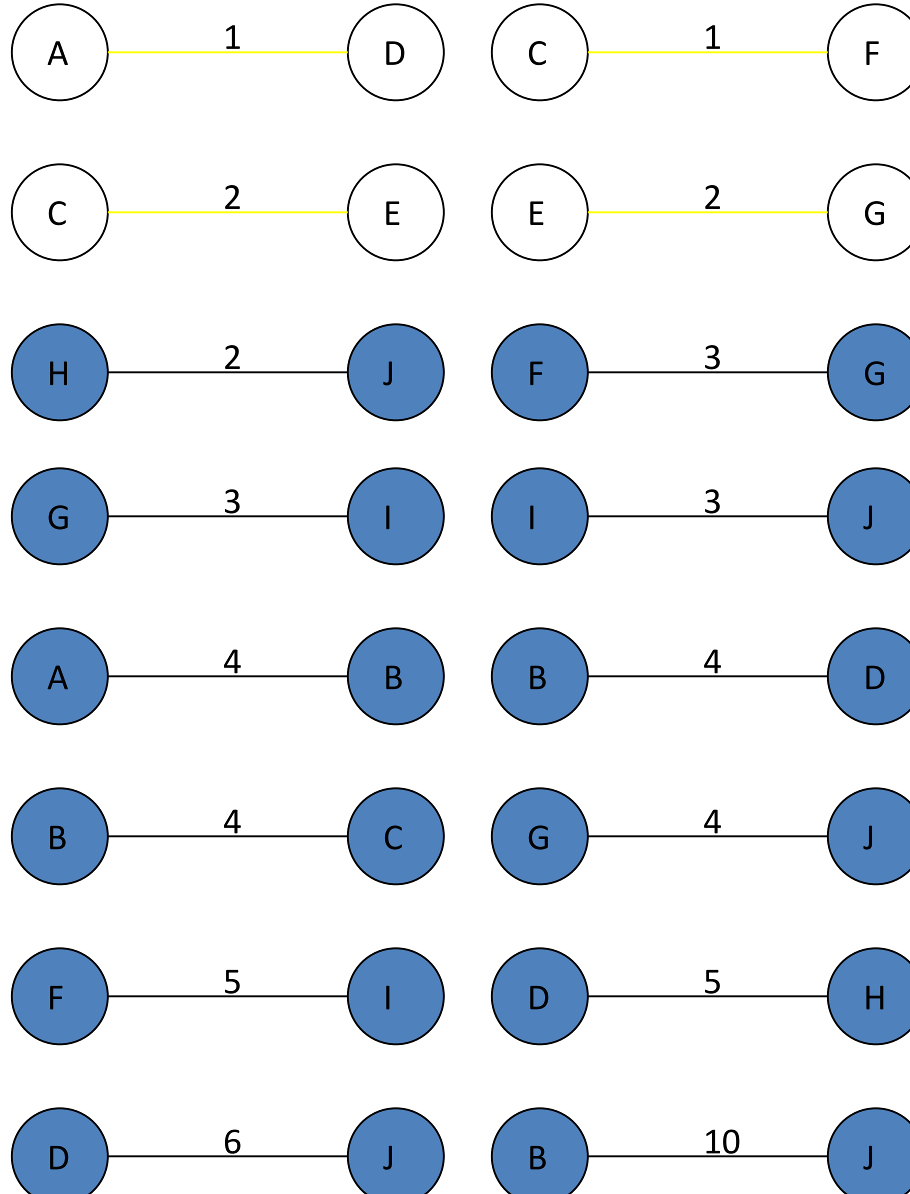


A
B
C
D
E
F
G
H
I
J
4
4
4
1
2
1
2
3
5
10
5
6
4
3
2
3
A 1 D
C 1 F
C 2 E
E 2 G
H 2 J
F 3 G
G 3 I
I 3 J
A 4 B
B 4 D
B 4 C
G 4 J
F 5 I
D 5 H
D 6 J
B 10 J

Add Edge

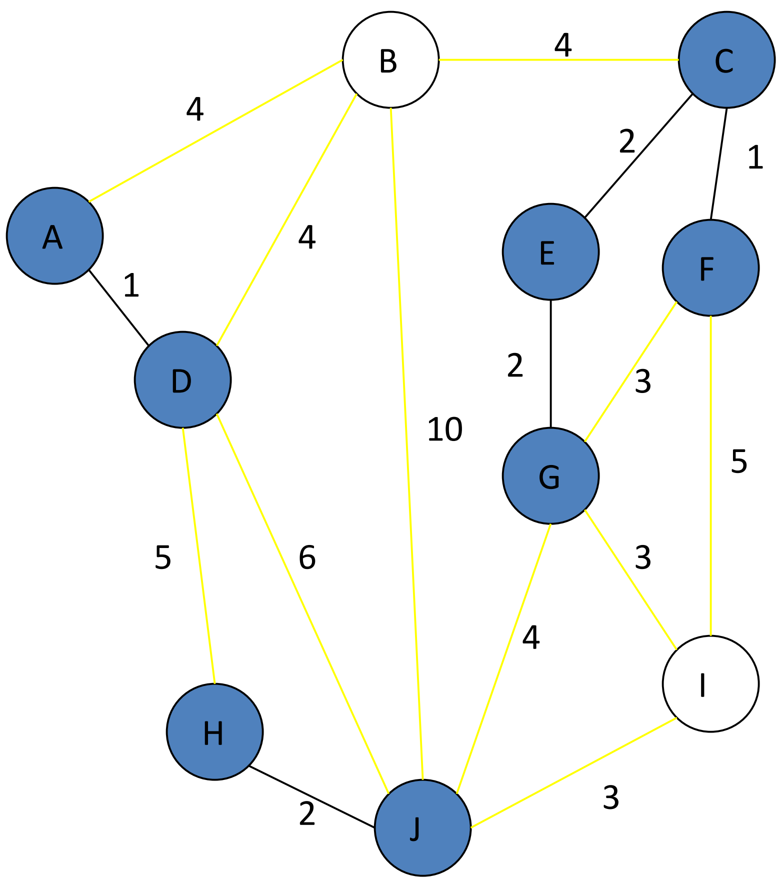


A
B
C
D
E
F
G
H
I
J
4
4
1
4
2
1
2
3
5
10
5
6
3
4
2
3

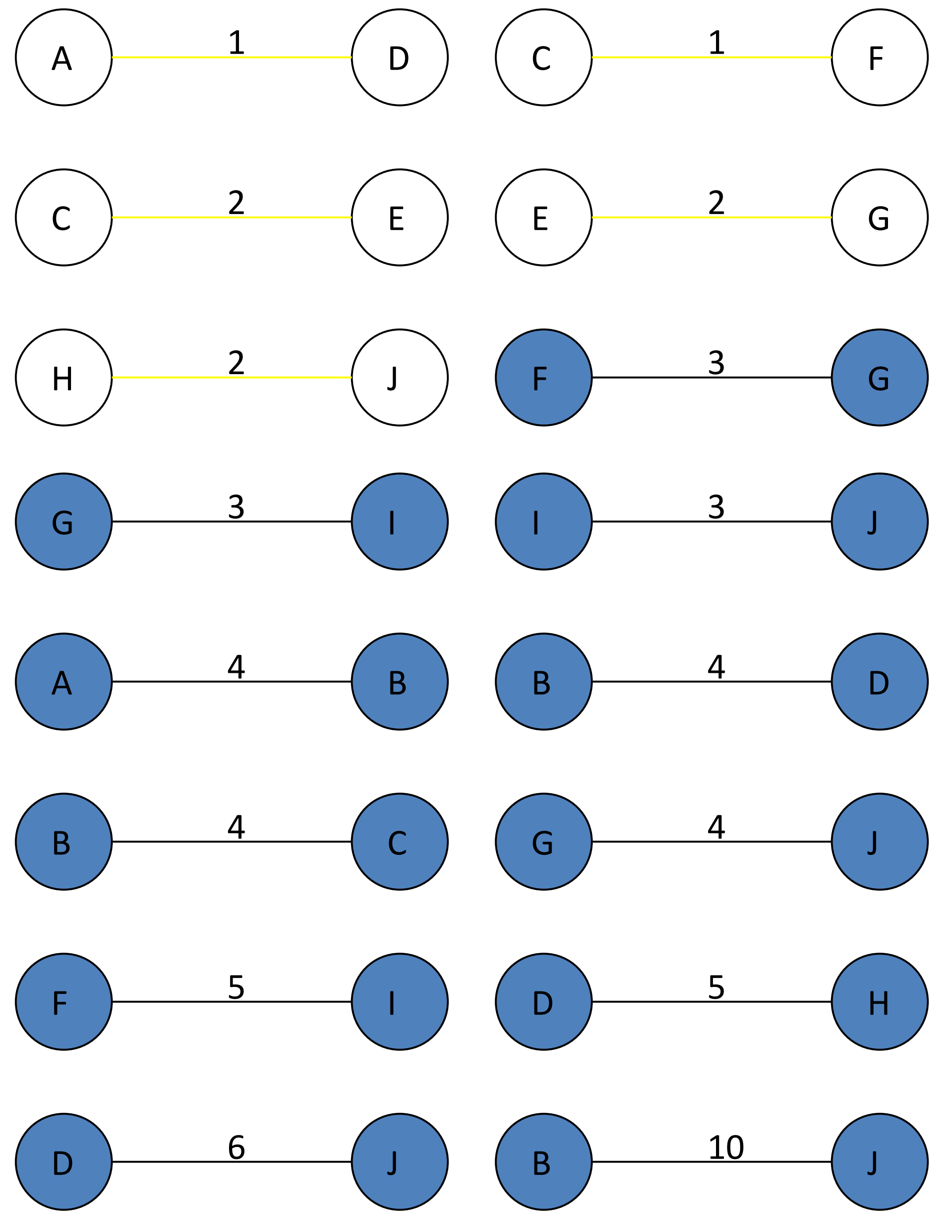


A 1 D
C 1 F
C 2 E
E 2 G
H 2 J
F 3 G
G 3 I
I 3 J
A 4 B
B 4 D
B 4 C
G 4 J
F 5 I
D 5 H
D 6 J
B 10 J

Add Edge



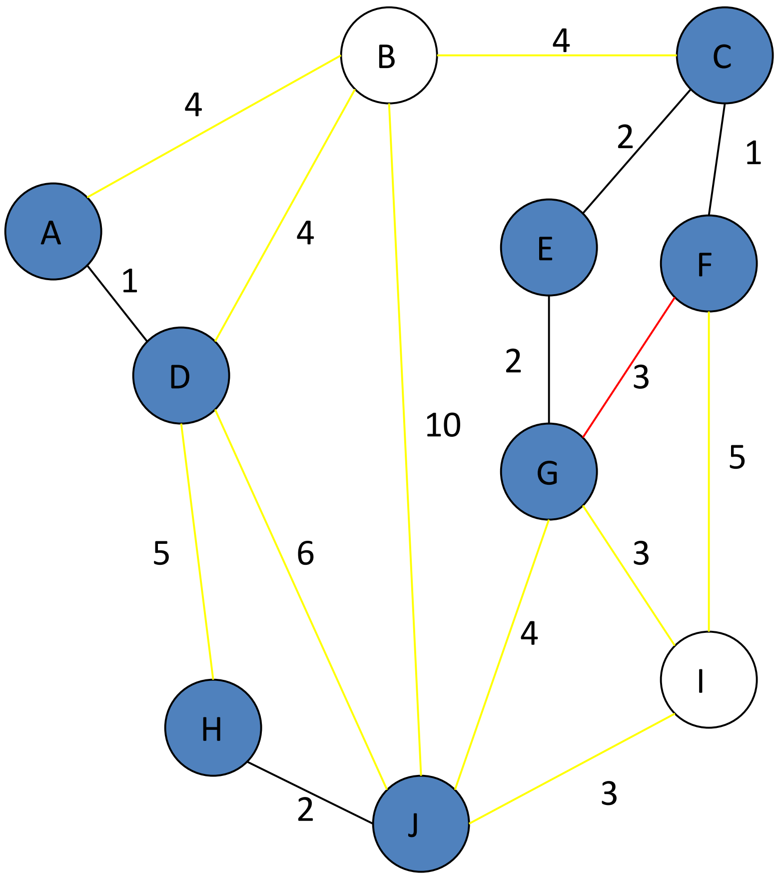
A
B
C
D
E
F
G
H
I
J
4
4
1
4
2
1
10
2
3
5
5
6
4
3
2
3

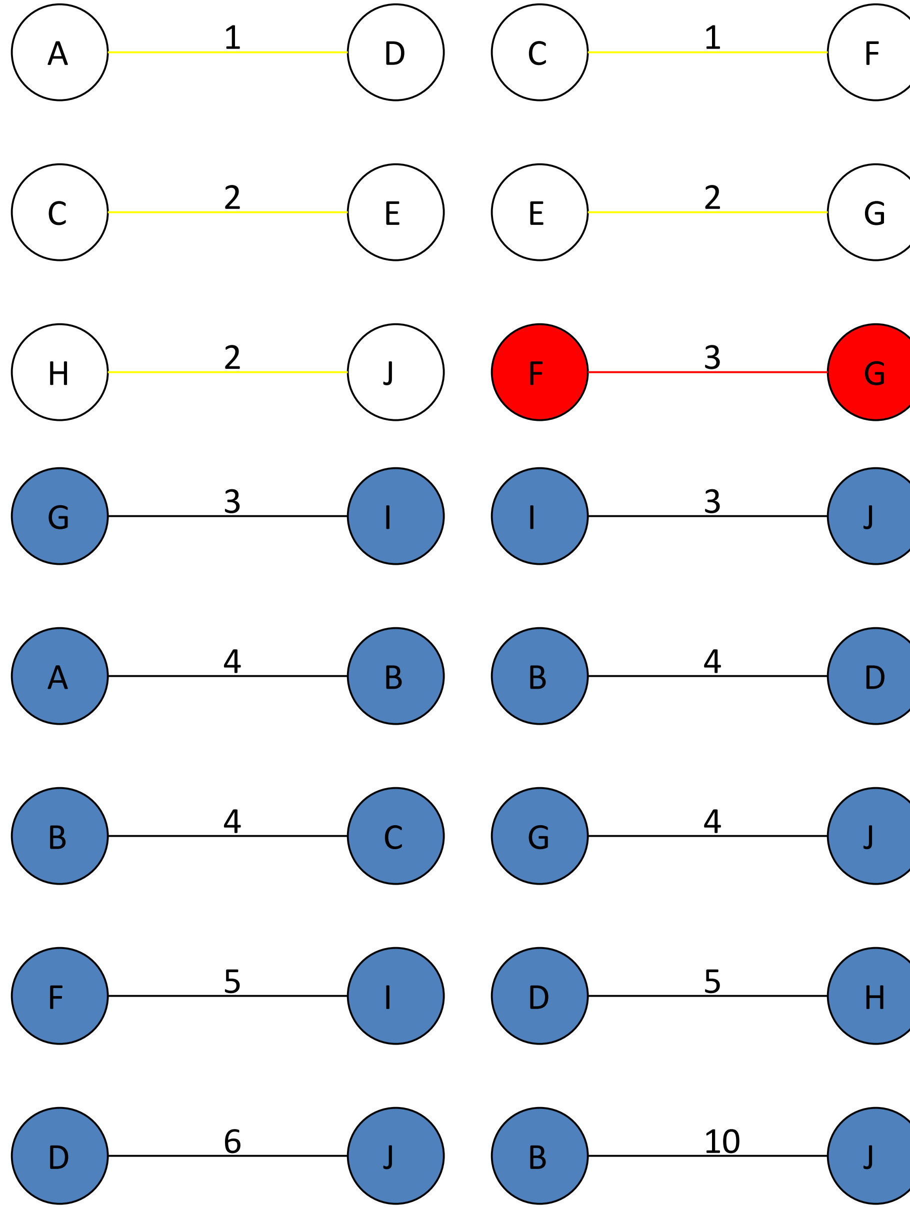


A 1 D
C 1 F
C 2 E
E 2 G
H 2 J
F 3 G
G 3 I
I 3 J
A 4 B
B 4 D
B 4 C
G 4 J
F 5 I
D 5 H
D 6 J
B 10 J

Cycle

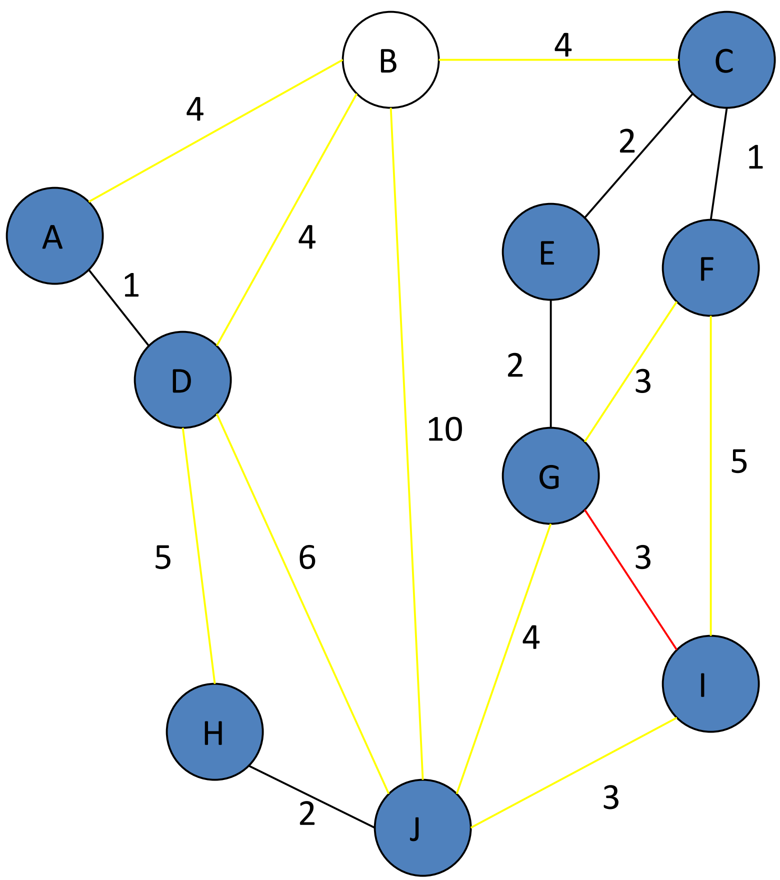
Don’t Add Edge

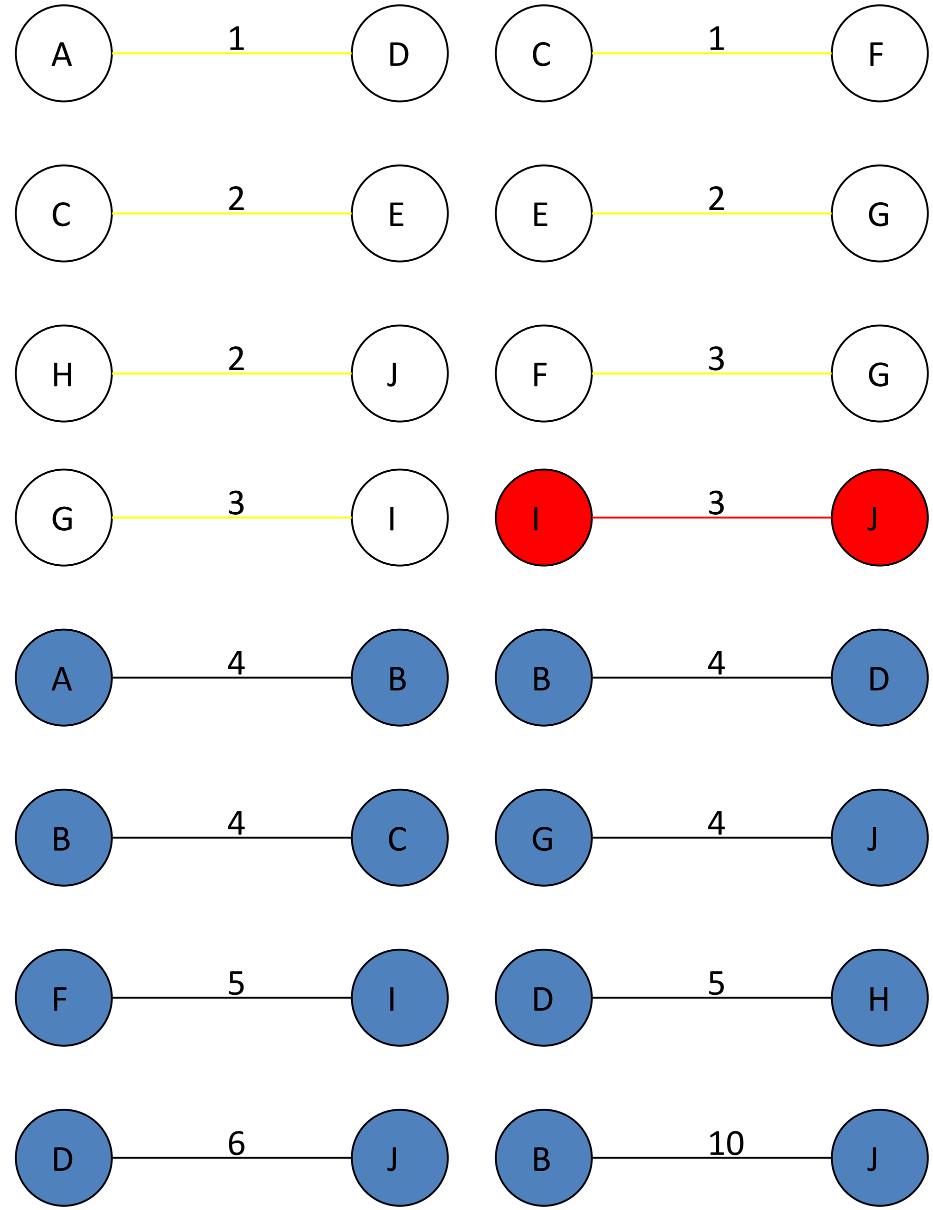




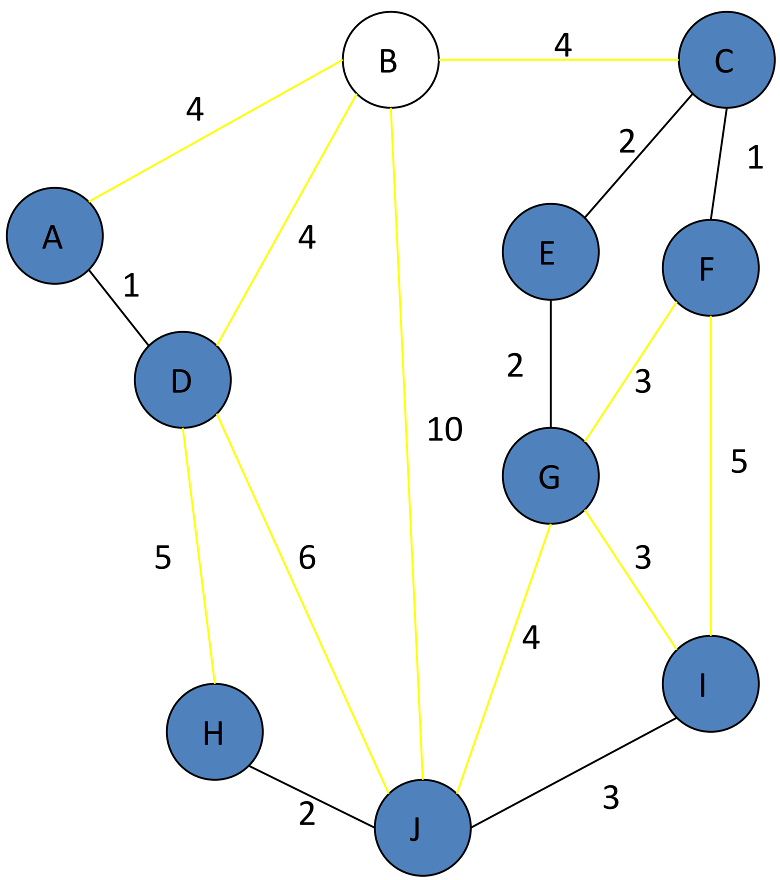
Higher than a dynamic threshold!

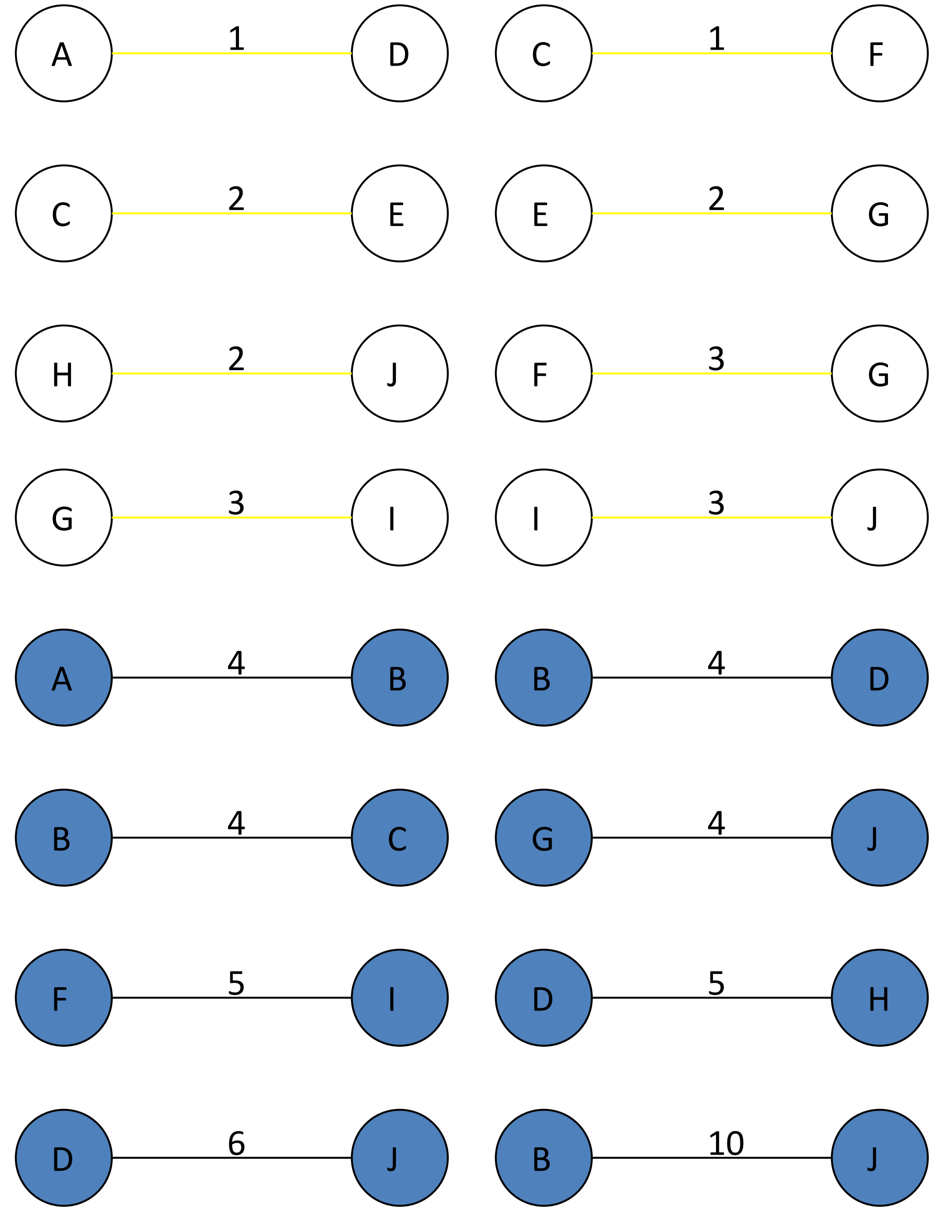
Don't Add Edge



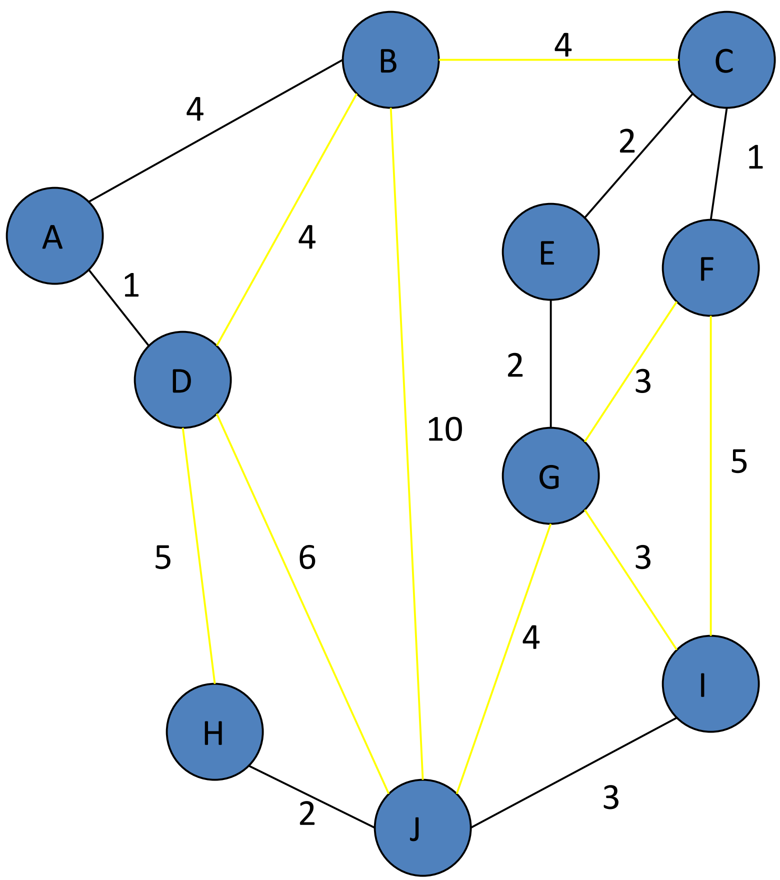


Add Edge

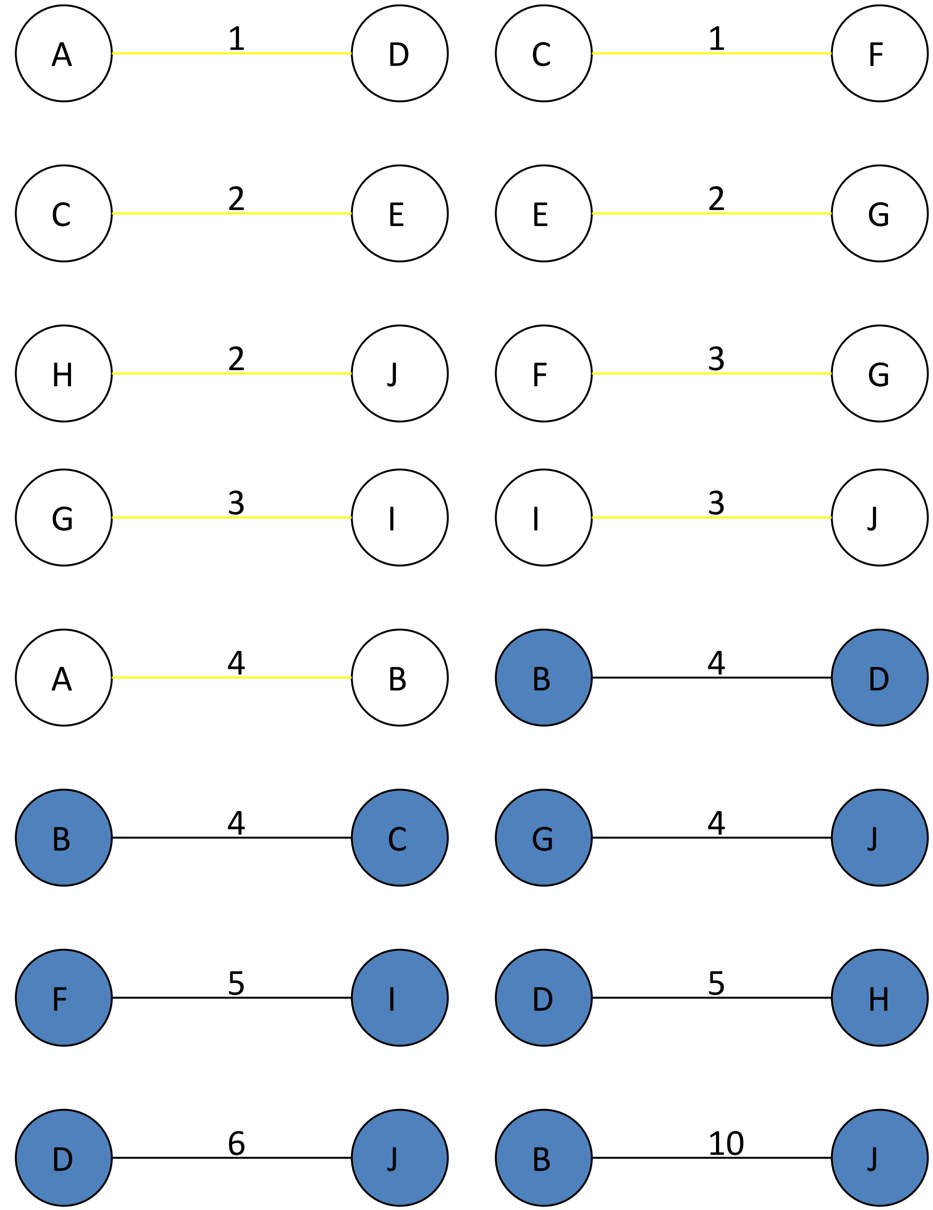

B
4
C
4
A
2
1
E
F
1
4
D
2
3
G
10
5
5
6
3
4
H
I
2
J
3


A 1 D
C 1 F
C 2 E
E 2 G
H 2 J
F 3 G
G 3 I
I 3 J
A 4 B
B 4 D
B 4 C
G 4 J
F 5 I
D 5 H
D 6 J
B 10 J

Add Edge



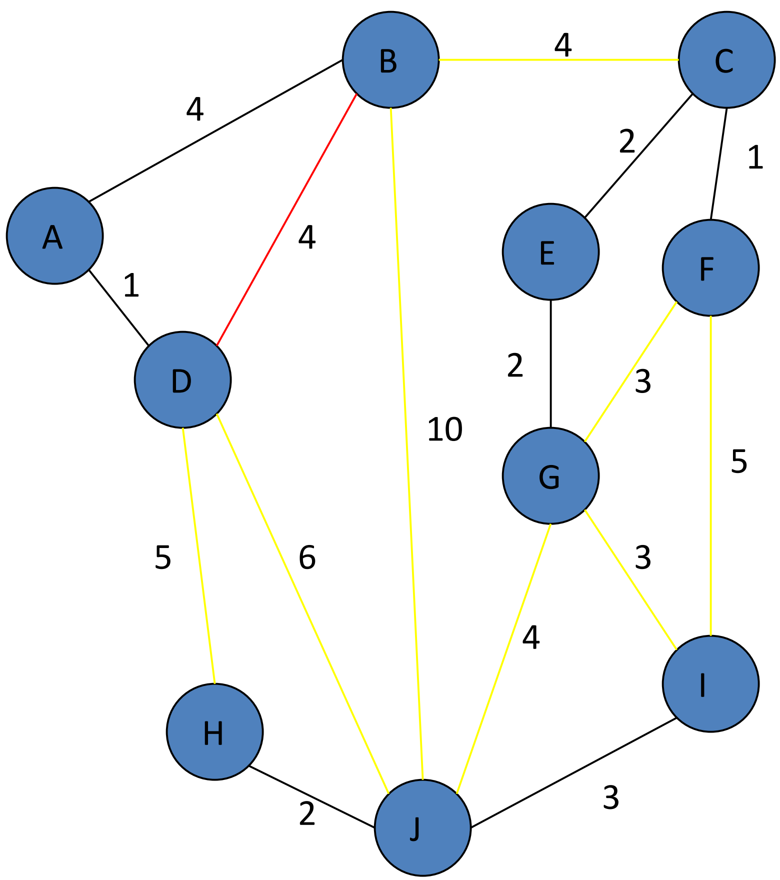
A
B
C
D
E
F
G
H
I
J
4
4
1
4
2
1
2
3
10
5
5
6
3
4
2
3

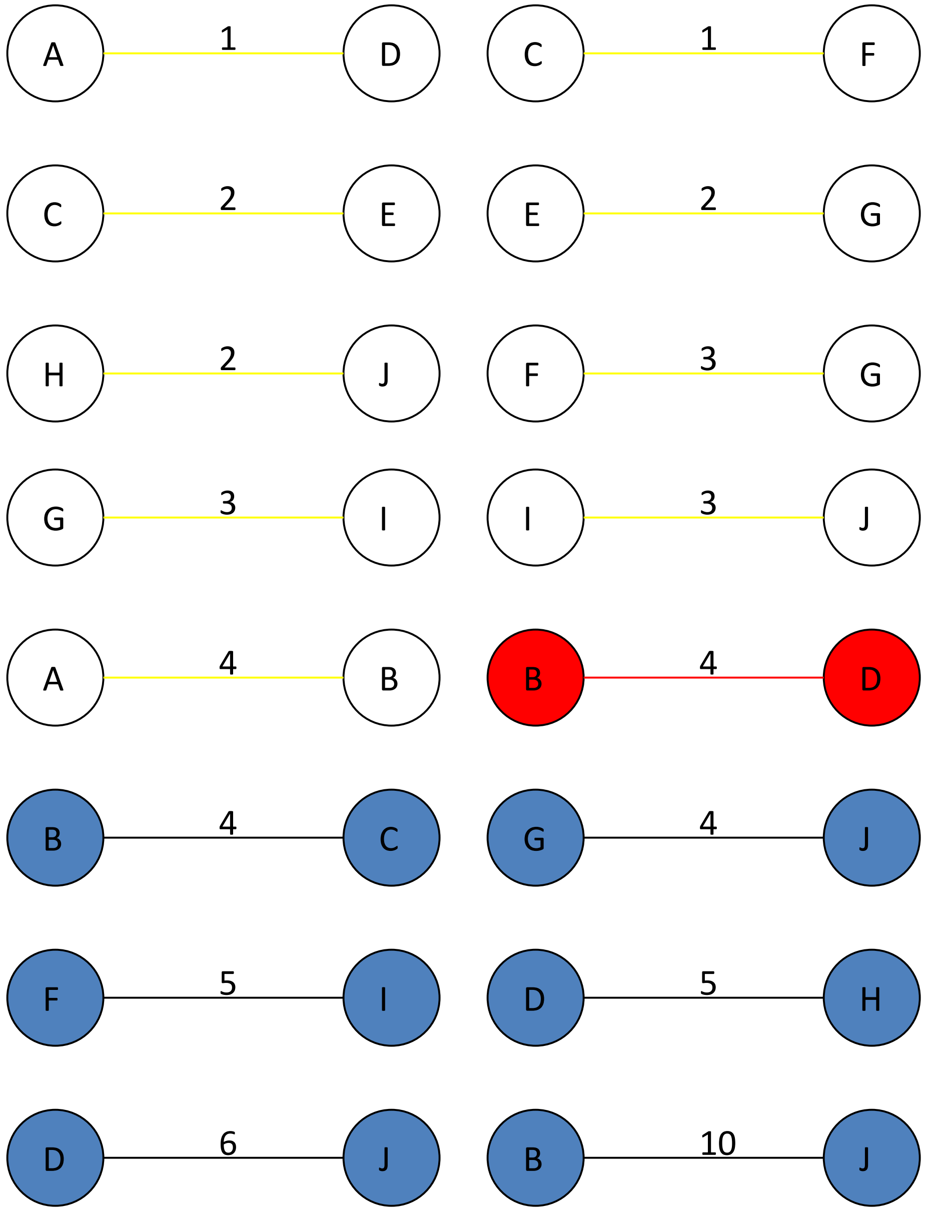


A 1 D
C 1 F
C 2 E
E 2 G
H 2 J
F 3 G
G 3 I
I 3 J
A 4 B
B 4 D
B 4 C
G 4 J
F 5 I
D 5 H
D 6 J
B 10 J

Cycle

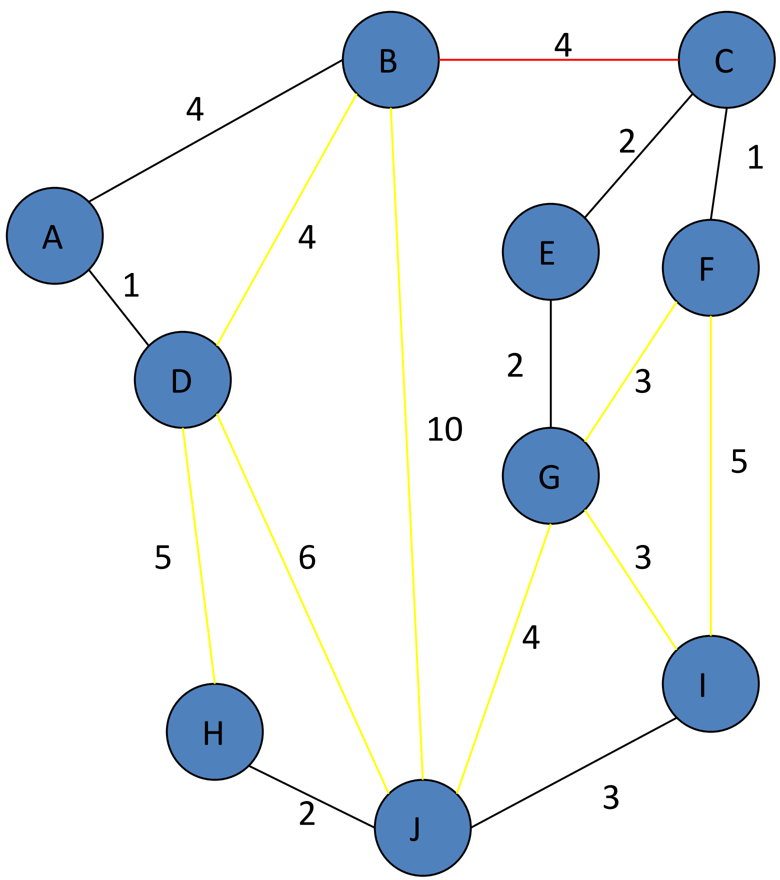
Don’t Add Edge

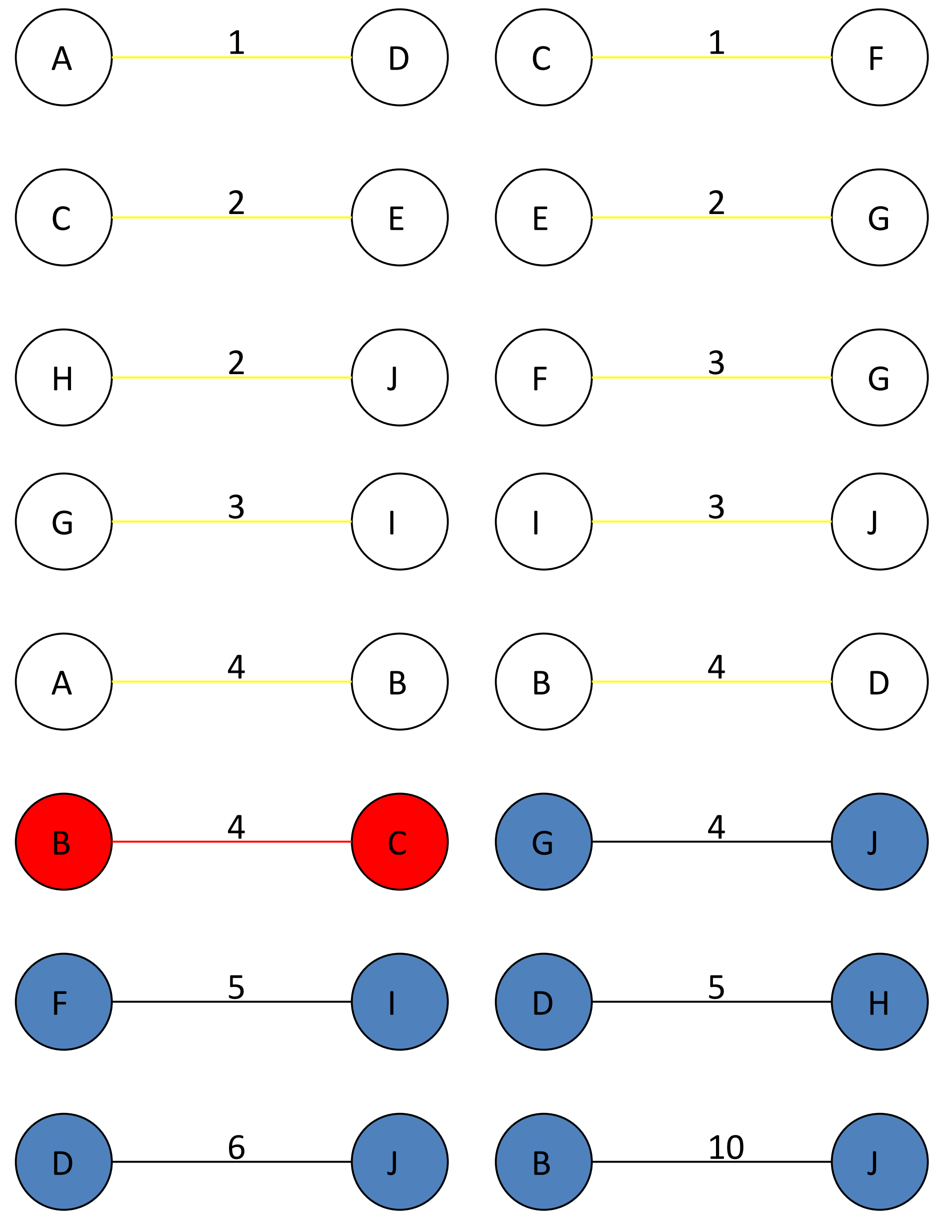




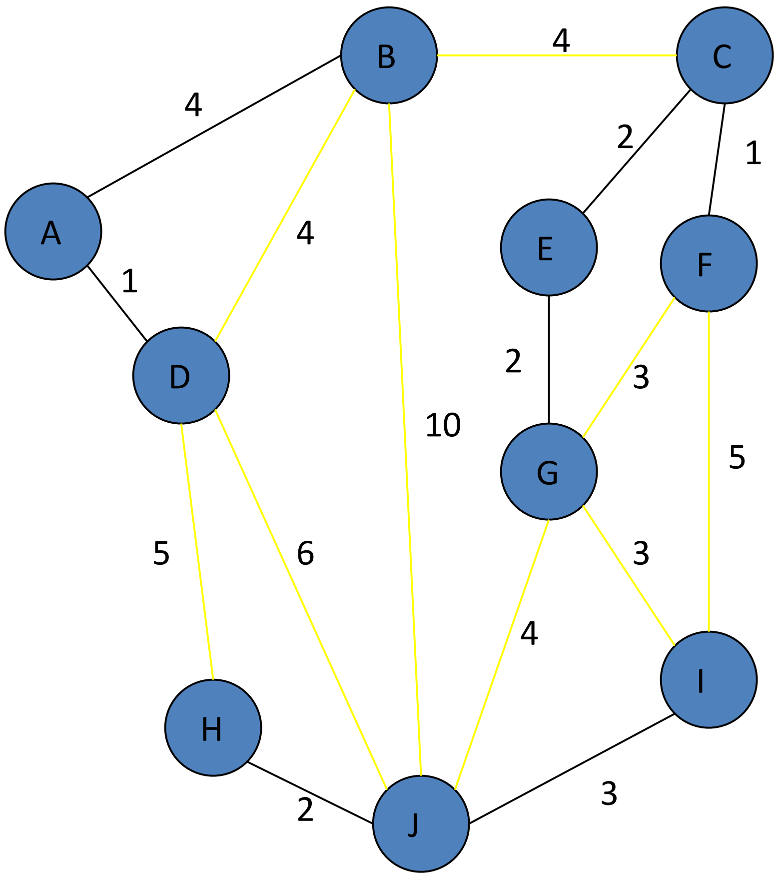
Higher than a dynamic threshold!

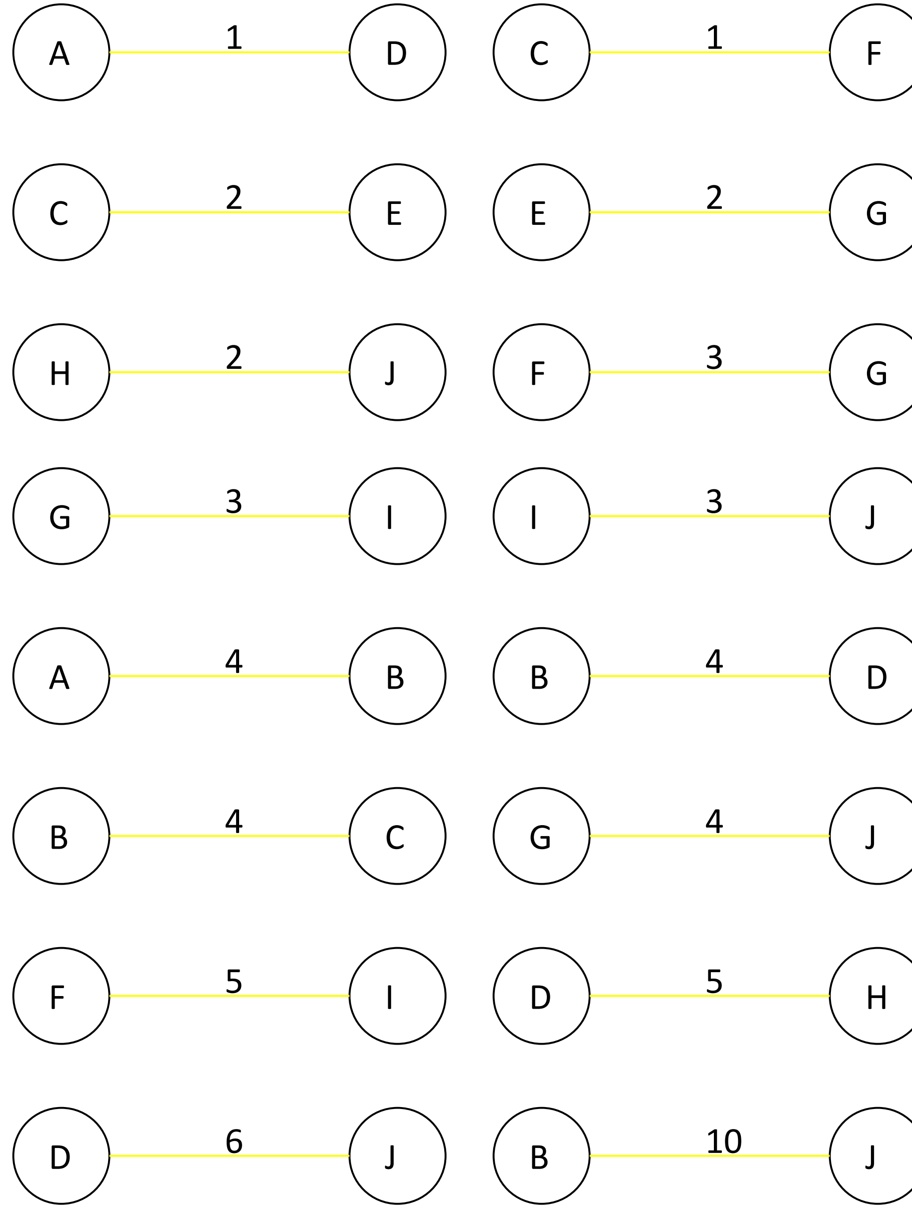
Don’t Add Edge





Final -> 3 clusters



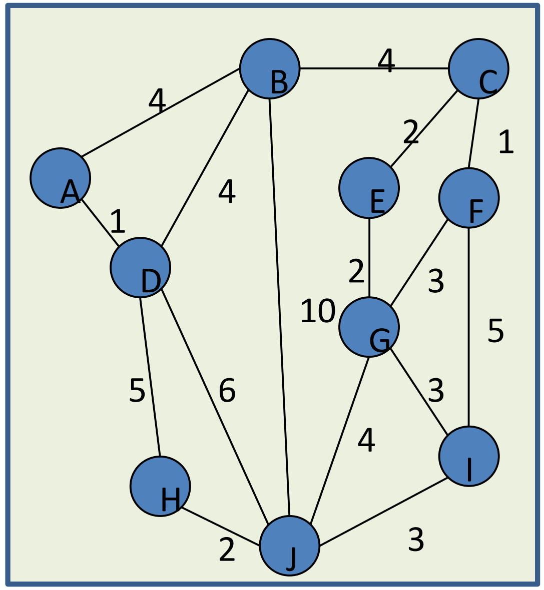


Example modified from Jonathan Davis's slides

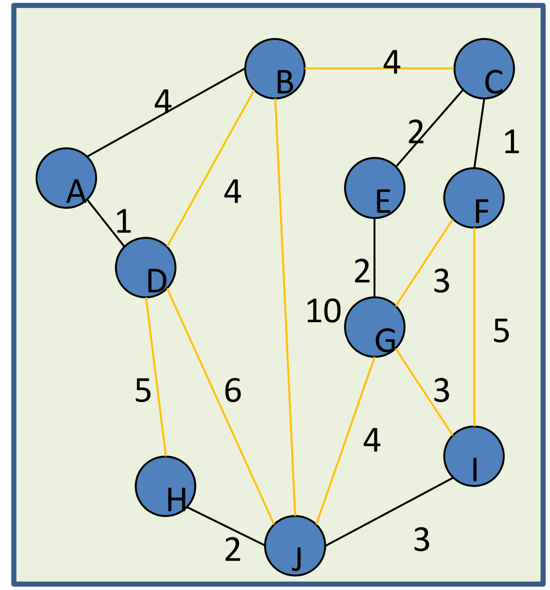
Conclusion - FH

Efficient graph-based image segmentation, IJCV 2004

Image



Graph



Clustered result

	Performance	Efficiency	Complexity (# parameters)	Theoretical support
FH	Reasonable (good for superpixels)	Fastest	Simple	Empirical

It is a good choice for general purpose applications if you don’t mind the following drawbacks: 1. Sensitive to noise in images. 2. The shape of the segments might be strange.

Methods

Ncut (Normalized Cut)
FH (Efficient graph-based image segmentation)
Mean Shift
gPb-OWT-UCM
ISCRA (Image Segmentation by Cascaded Region Agglomeration)

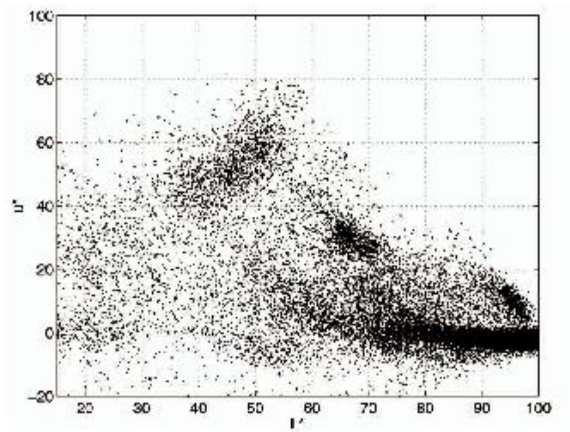
Citation: 6311
(since year 2002)

Mean Shift

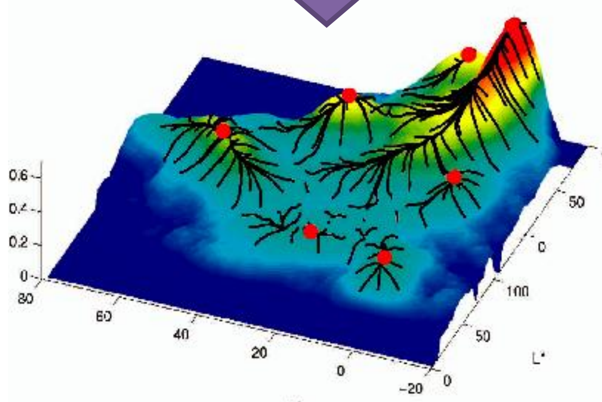
Mean shift: a robust approach toward feature space analysis, PAMI 2002

Image

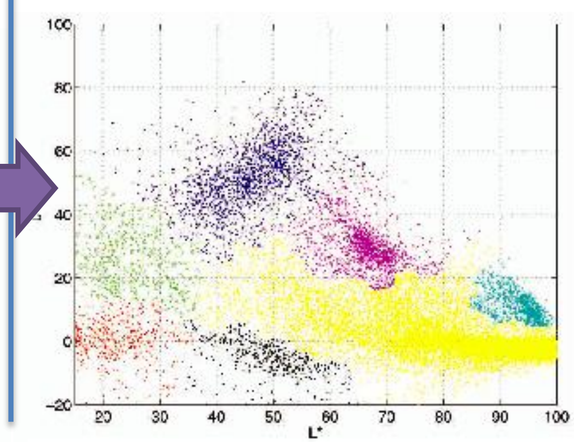
Projection on (L^*, u^*) space



In fact, mean shift usually work on 5 dimensions feature space (including 3 color channels and x, y for spatial location).



Mean shift (mode seeking)

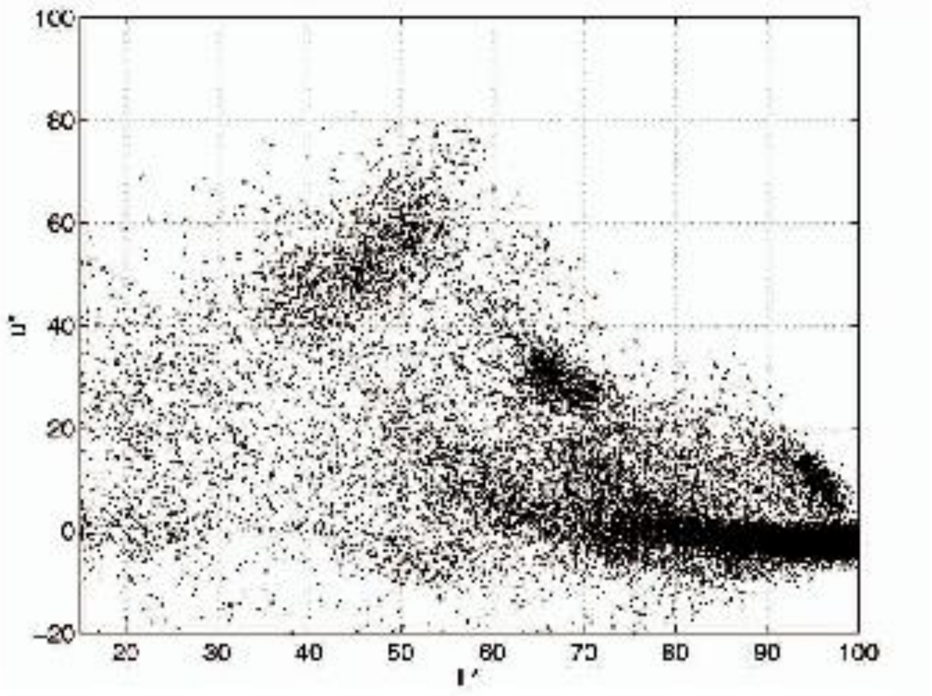


Clustered result

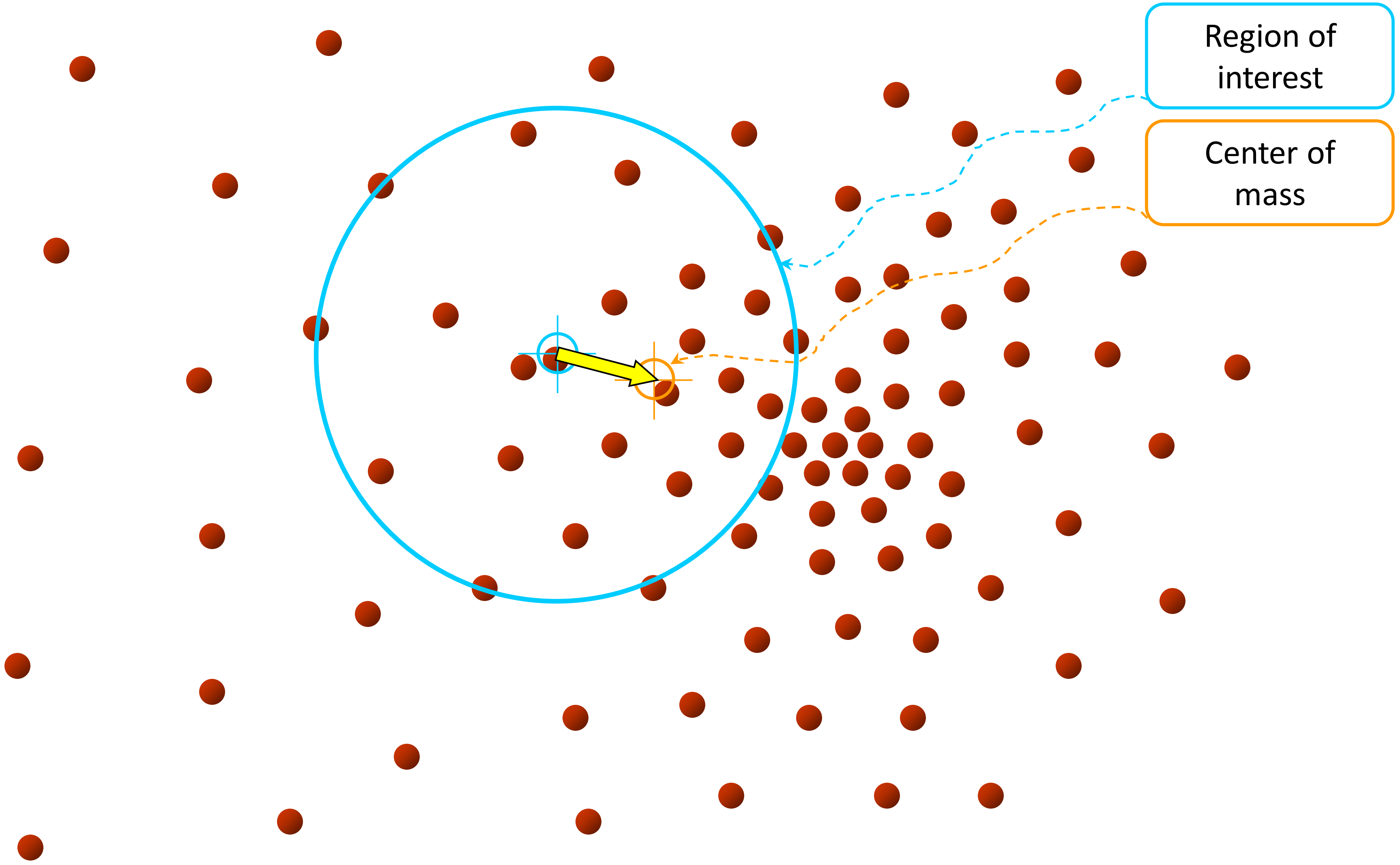
Mean Shift Example

Projection on (L^*, u^*) space

Image

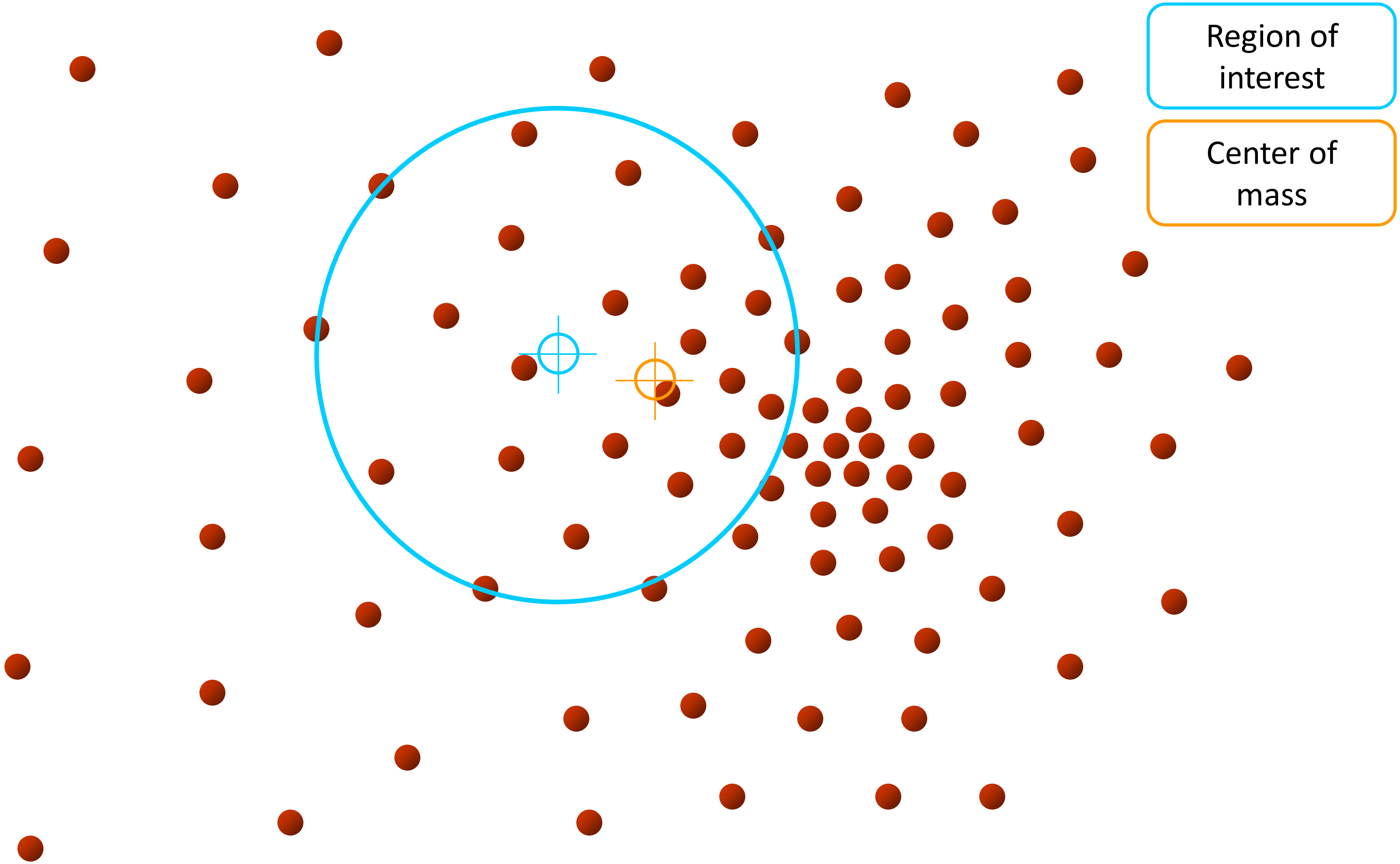


Intuitive Description



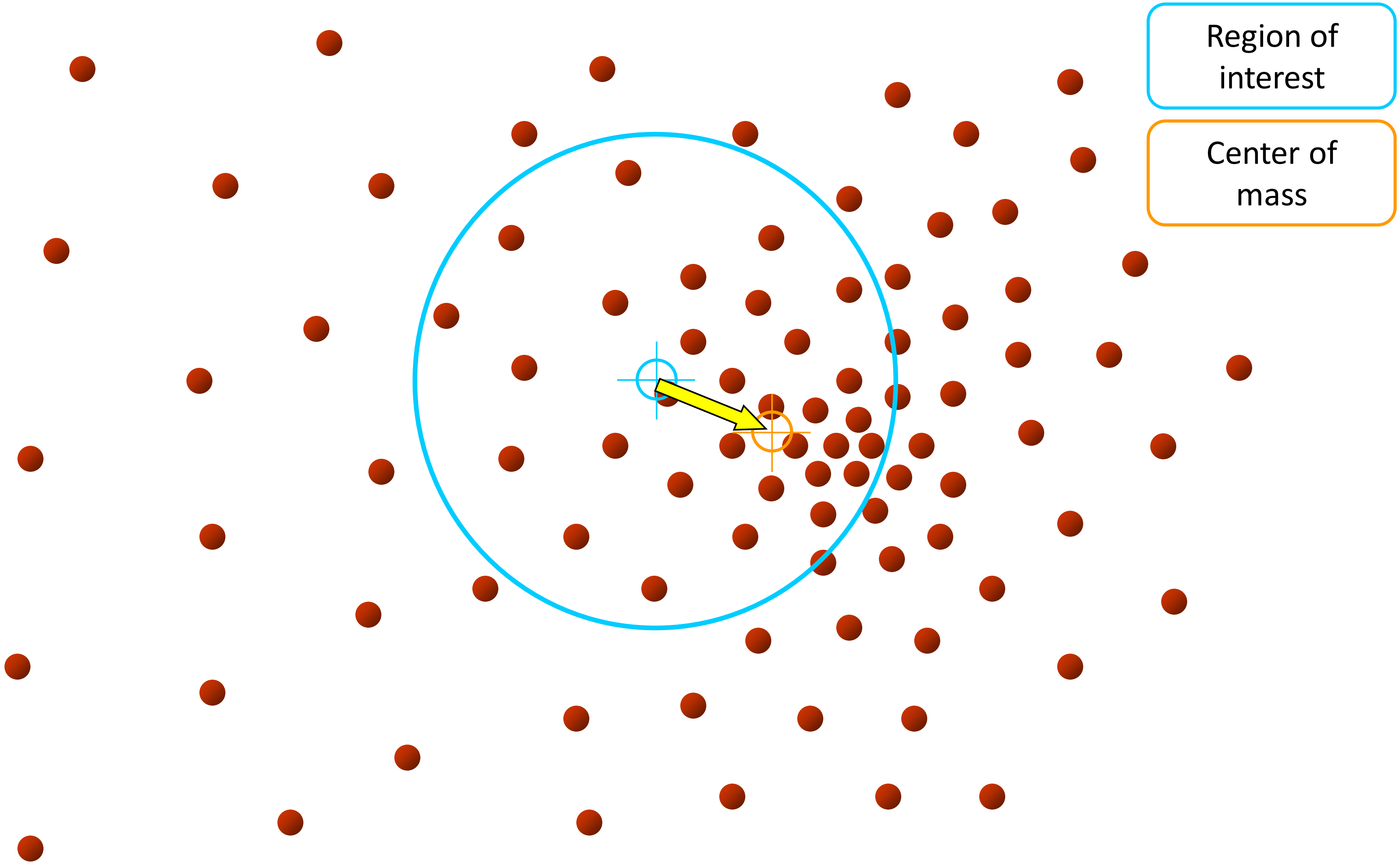
Objective : Find the densest region

Intuitive Description



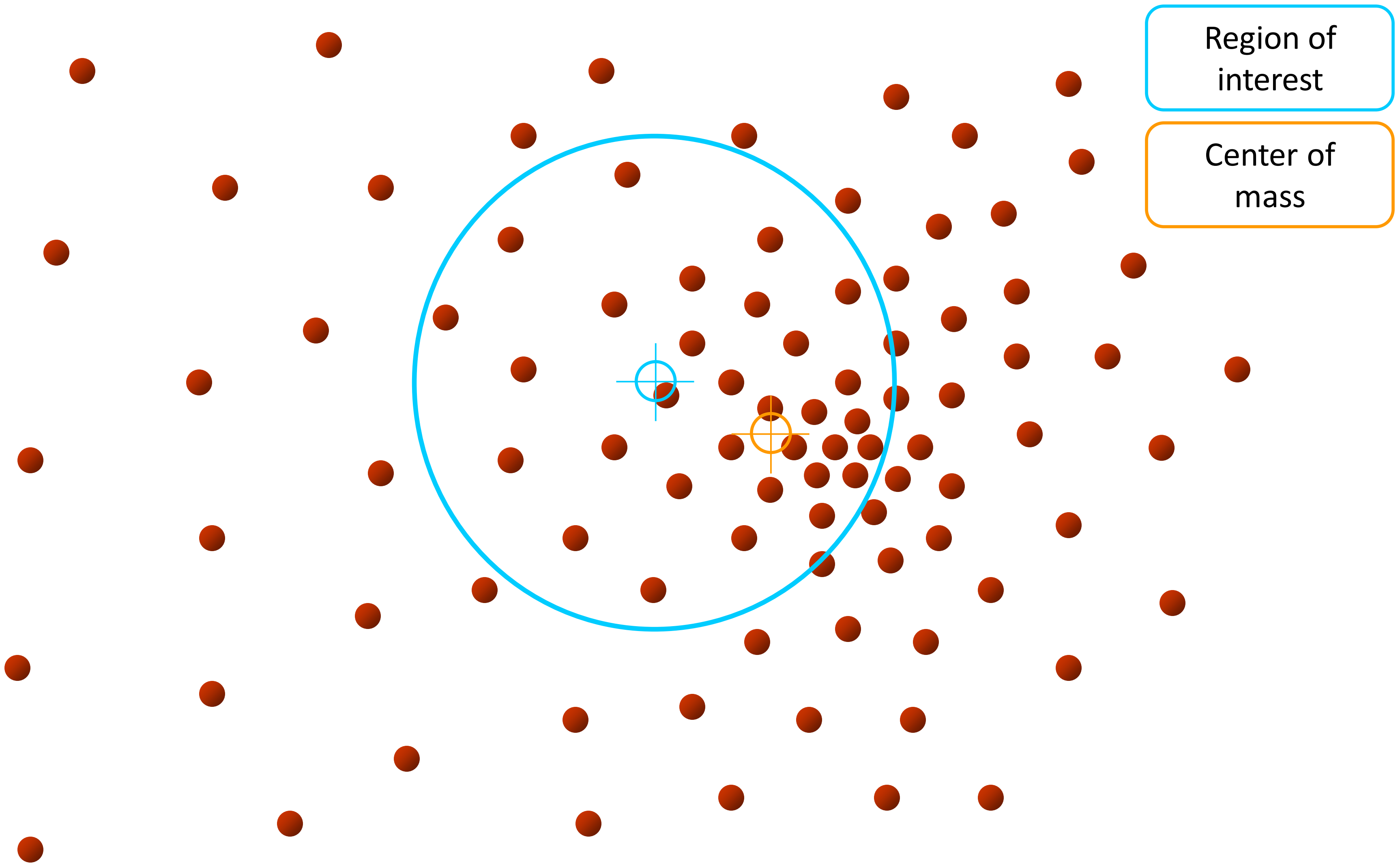
Objective : Find the densest region

Intuitive Description



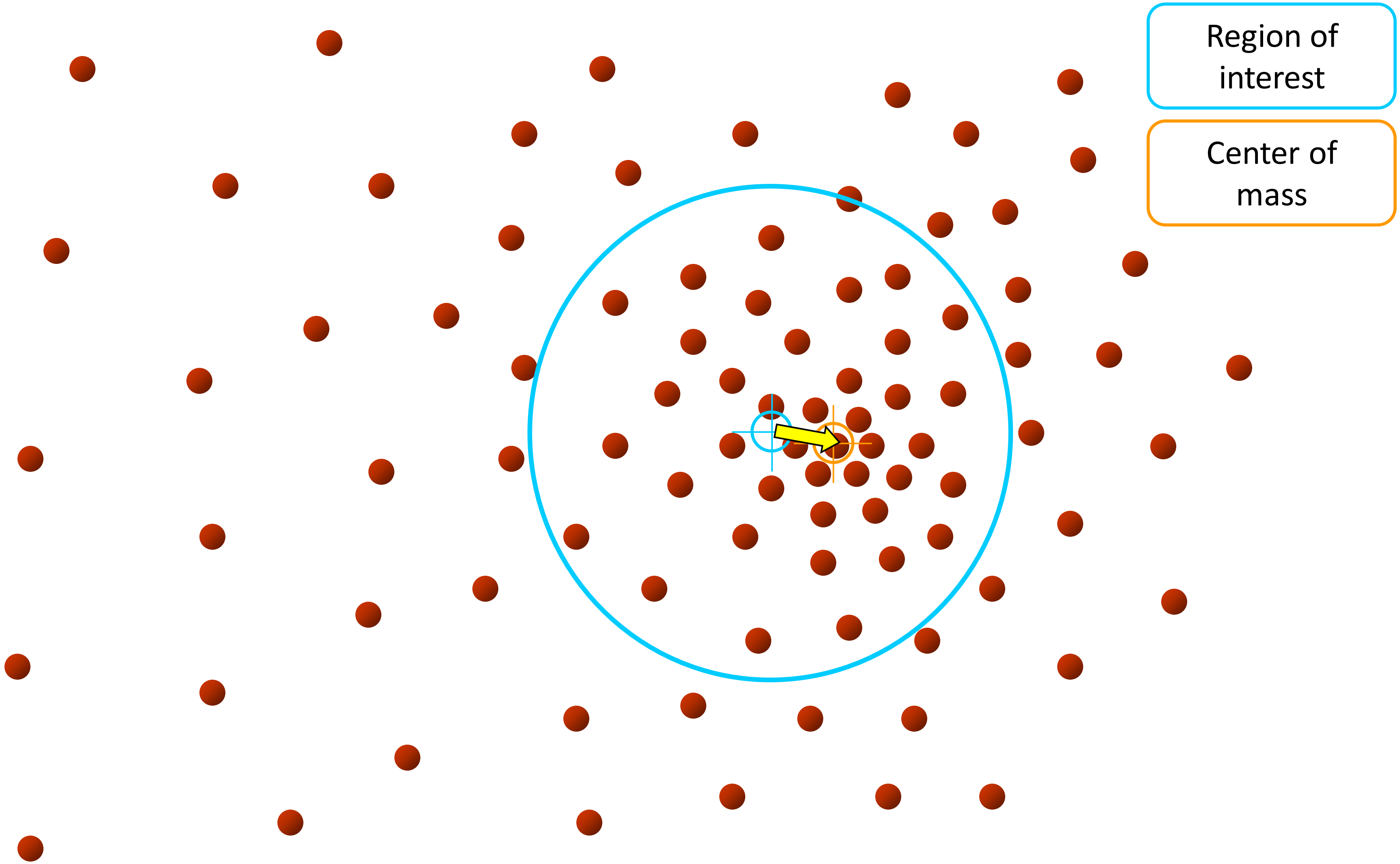
Objective : Find the densest region

Intuitive Description



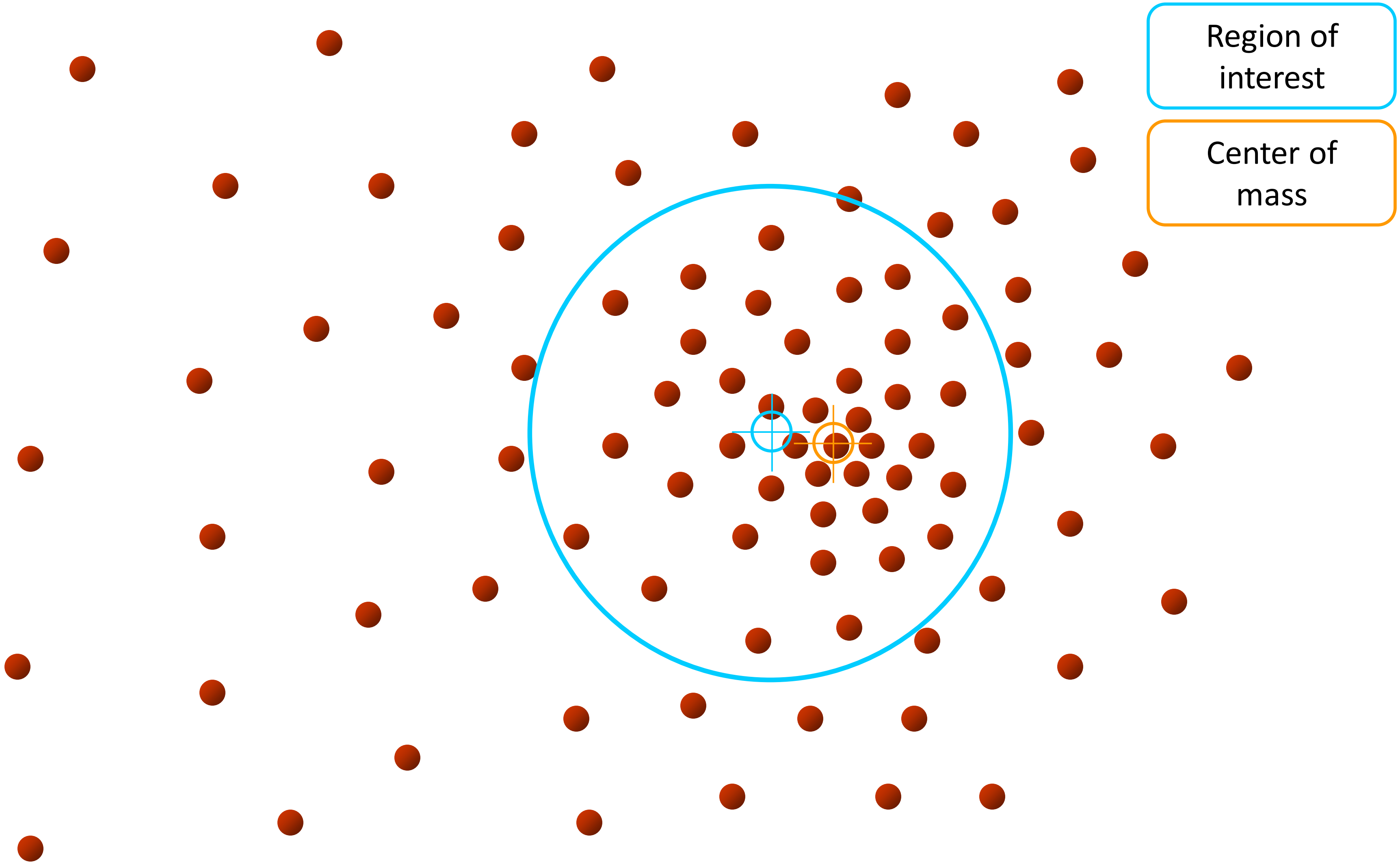
Objective : Find the densest region

Intuitive Description



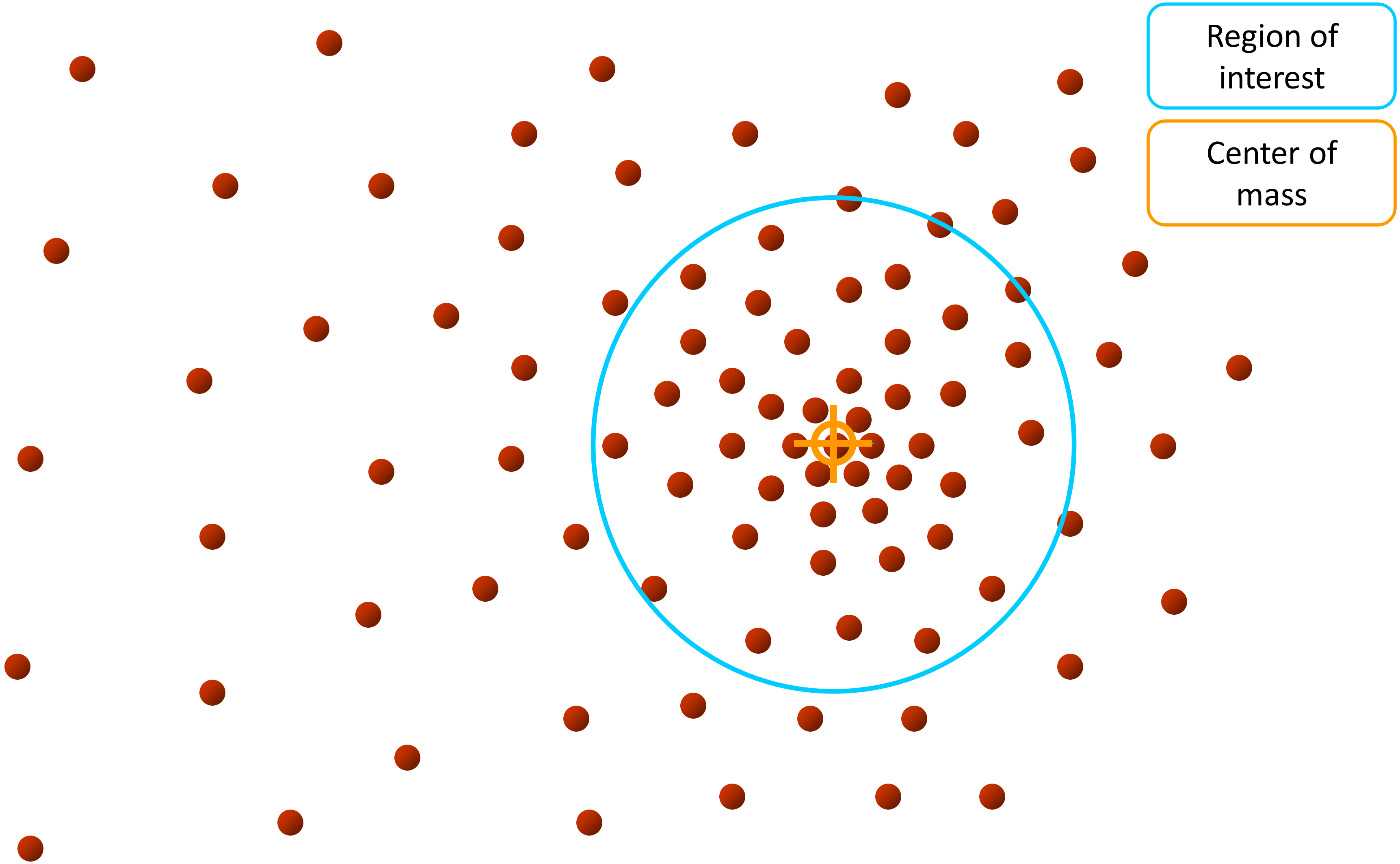
Objective : Find the densest region

Intuitive Description



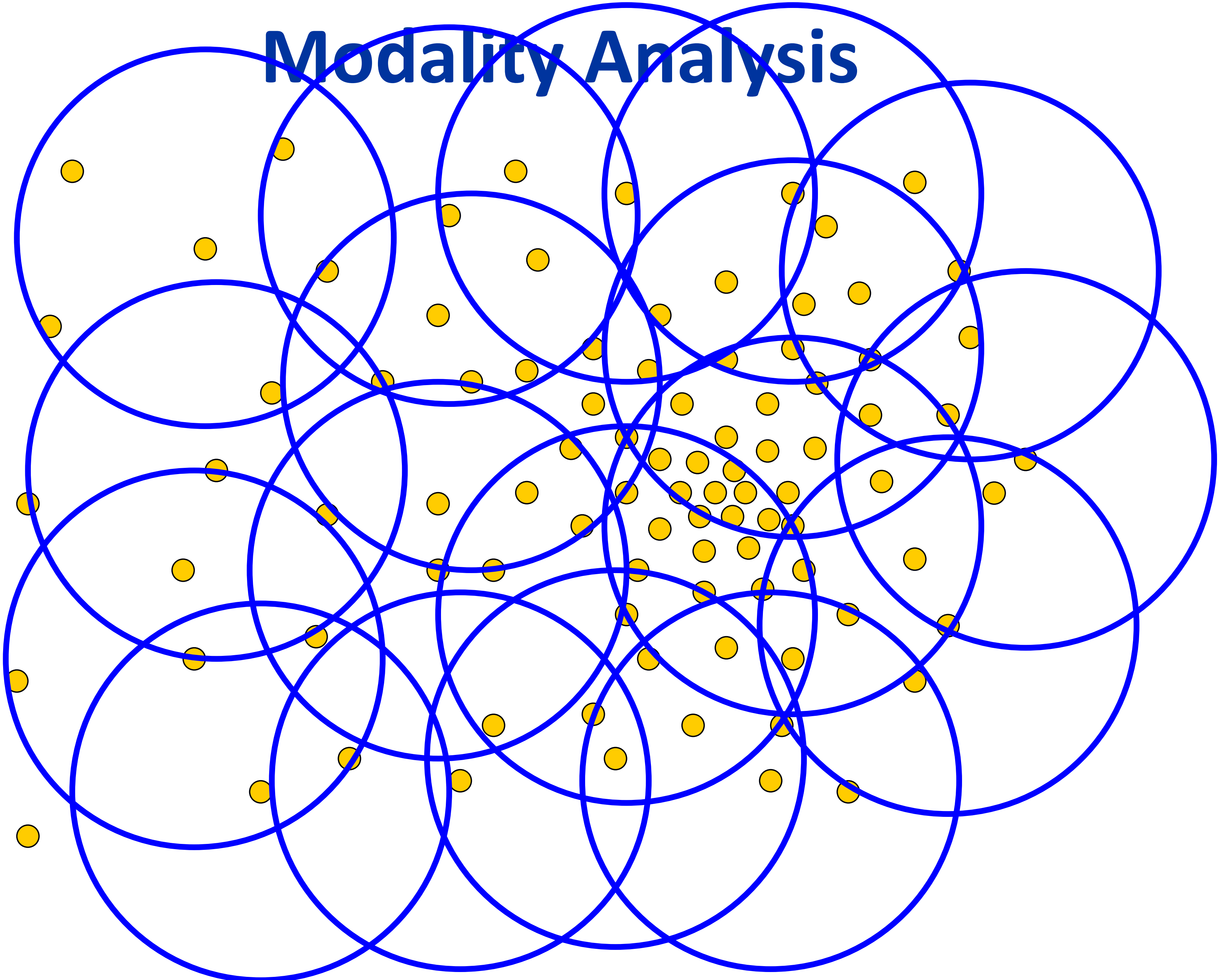
Objective : Find the densest region

Intuitive Description

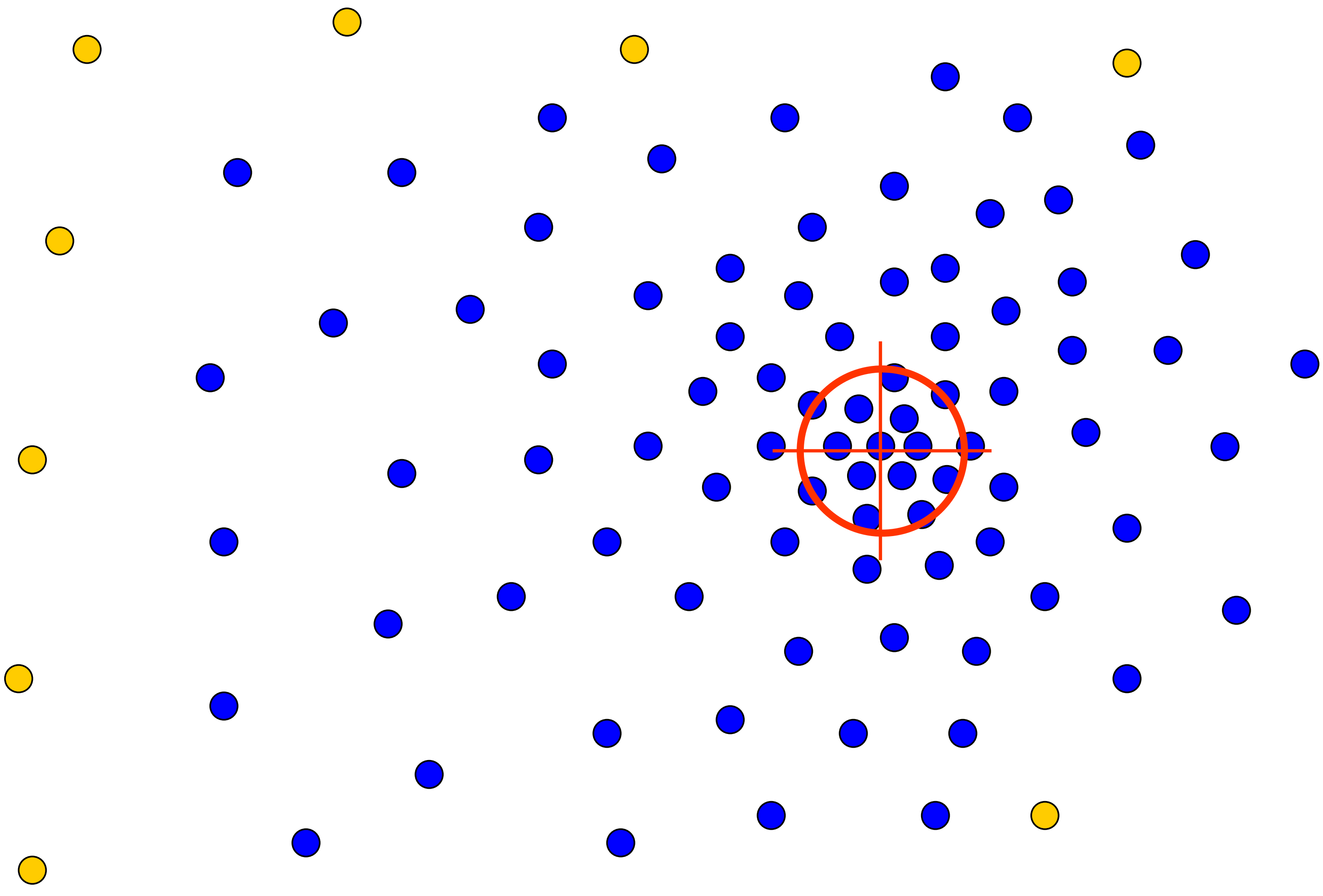


Objective : Find the densest region

Modality Analysis



Modality Analysis

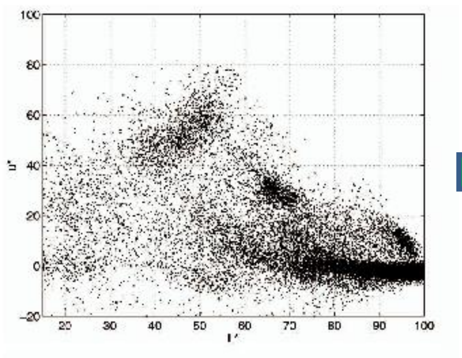


The blue data points were traversed by the windows towards the mode

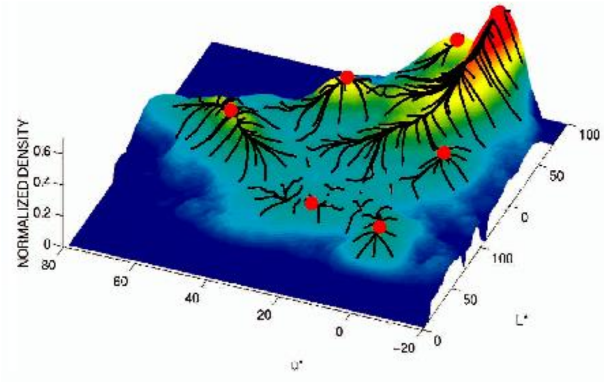
Example excerpted from Yaron Ukrainitz & Bernard Sarel's slides

Conclusion - Mean Shift

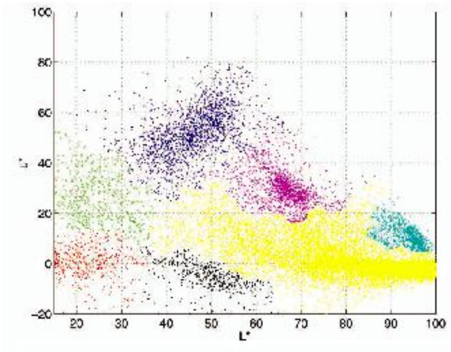
2D (L*u) space representation



Mean shift (mode seeking)



Final clusters

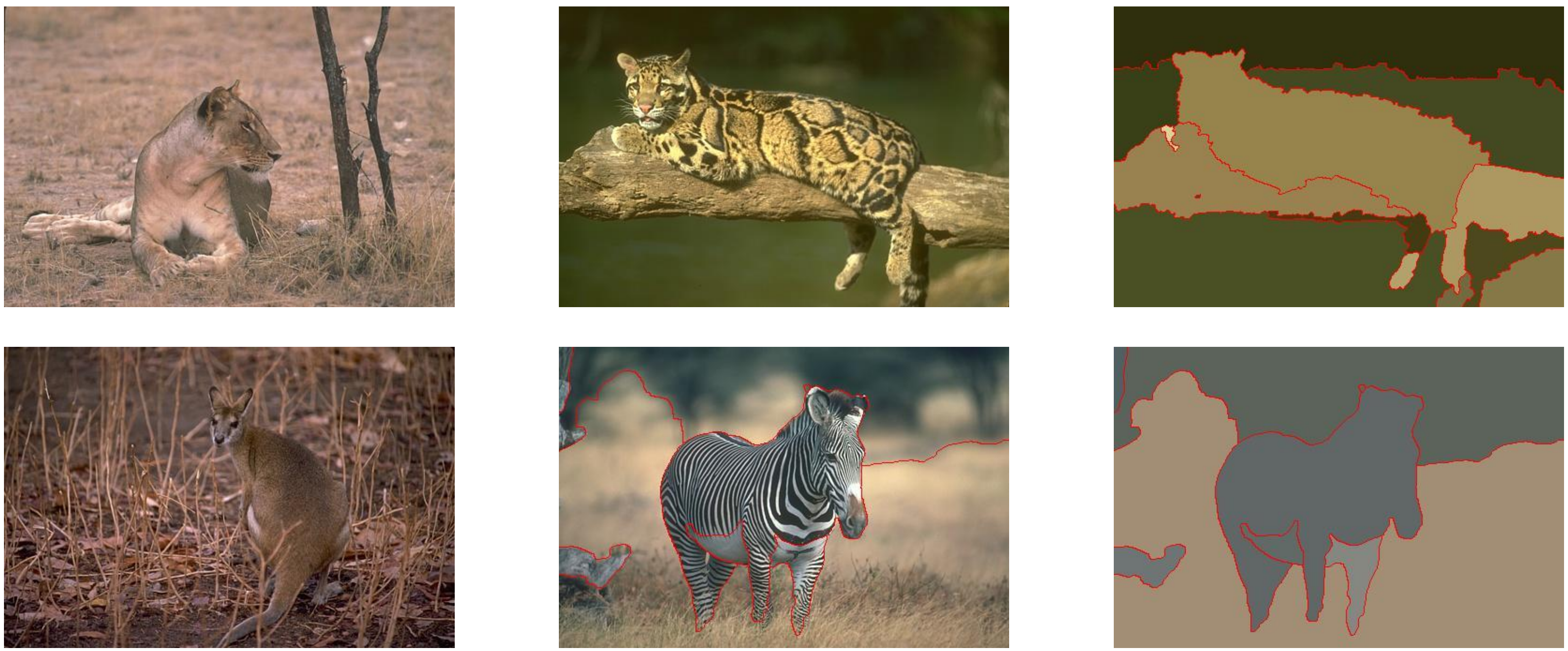


	Performance	Efficiency	Complexity (# parameters)	Theoretical support
Mean shift	Reasonable (↑ for oversegments)	Reasonable	Simple	Good

It is a good choice for general purpose applications if you don't mind the following drawback: Slow in high resolution images.

Common limitation of above methods

- Basic assumption:
  - color in different segments should be different



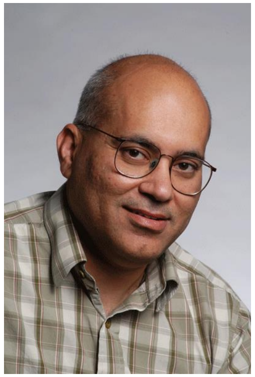
- How to integrate more features (e.g. texture)?
- Can we learn to distinct segments using training data?

Methods

Ncut (Normalized Cut)
FH (Efficient graph-based image segmentation)
Mean Shift
gPb-OWT-UCM
ISCRA (Image Segmentation by Cascaded Region Agglomeration)

Citation: 408 + 245
(since year 2011,2009)



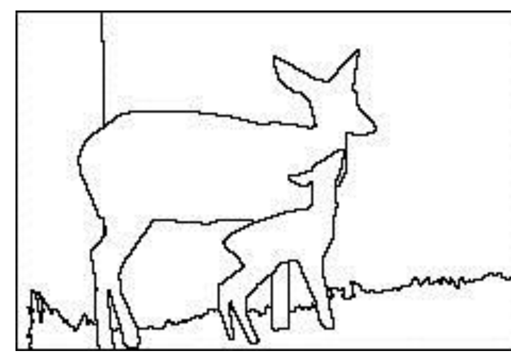



gPb-OWT-UCM

Contour Detection and Hierarchical Image Segmentation, PAMI 2011

- Instead of clustering pixels, it solves the segmentation problem by contour (object boundary) detection.
  - Classify each pixel into "boundary" or "not boundary"
- It combines several different techniques
  - Supervised boundary classifier → NCut globalization → removing noise by watershed → minimal spanning tree merging

Clustering pixels
search space $\sim\mathbf{13}^{\mathbf{150000}}$

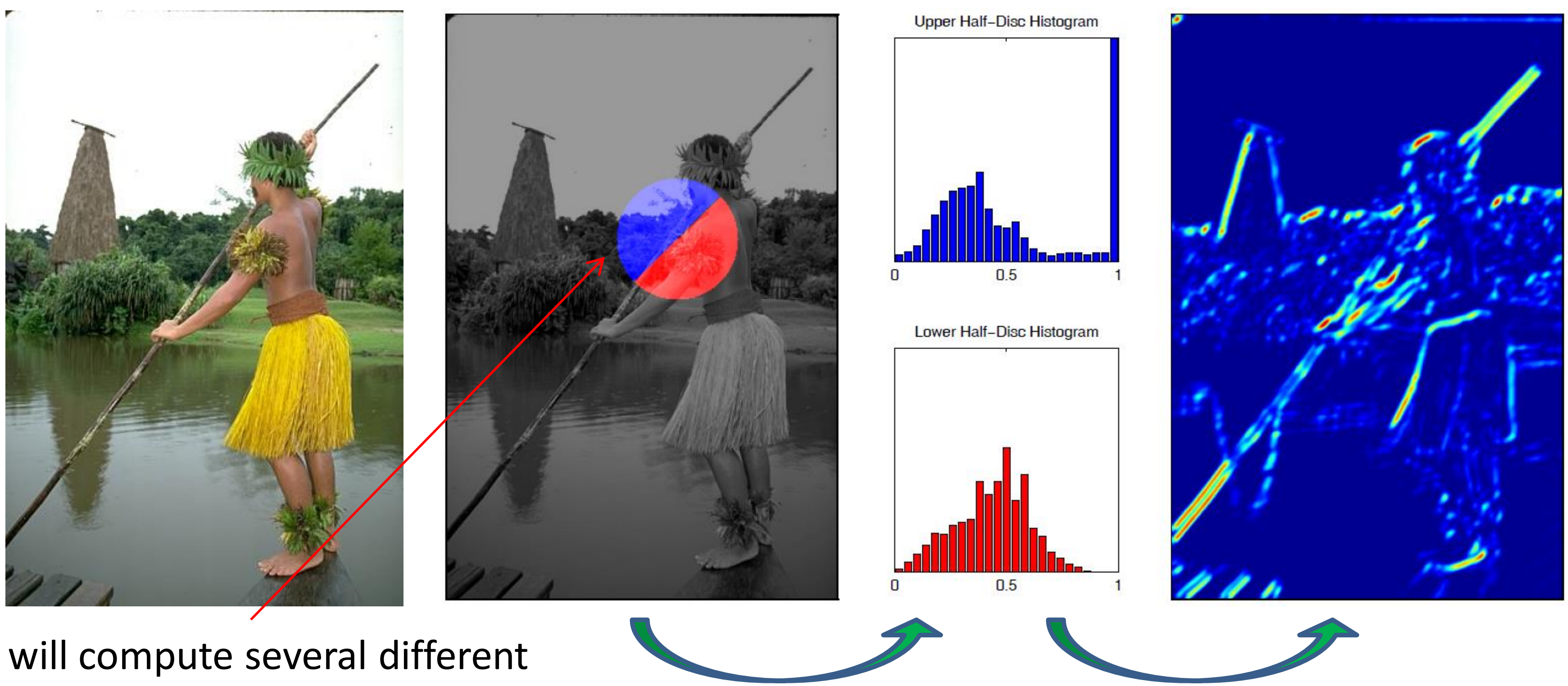


Contour detection
search space $\sim\mathbf{2}^{\mathbf{150000}}$

mPb – Feature extraction for the classifier

- Feature: Histogram difference between two half disks
  - By assuming that the object boundaries are long enough straight lines

Input image | One feature channel | Feature on each pixel

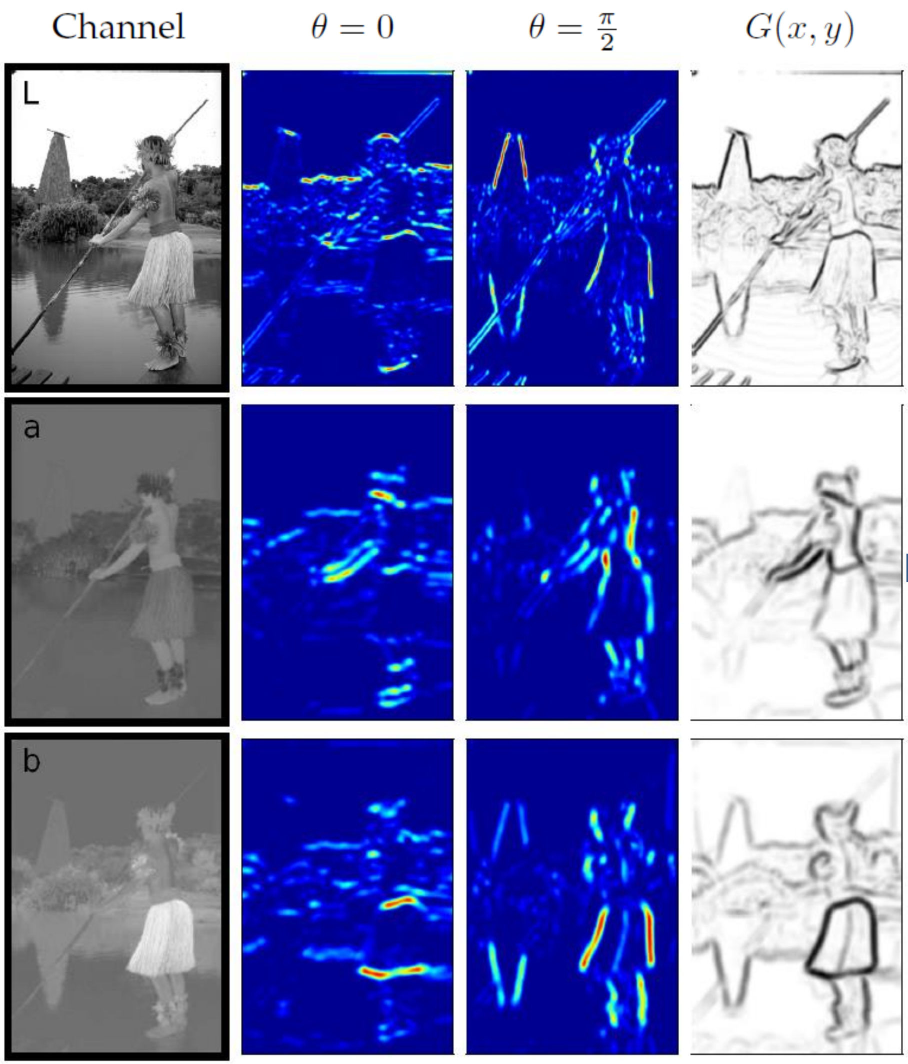


It will compute several different angles (θ) and disk sizes (scales).

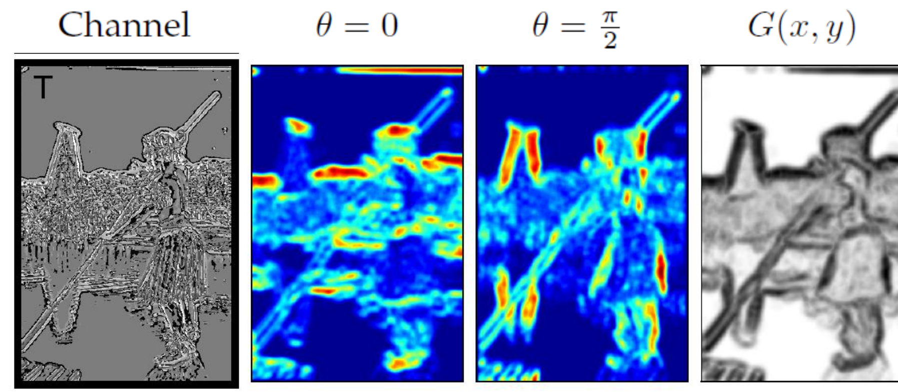
Extract feature histograms within the disk

Compute the distance between histograms

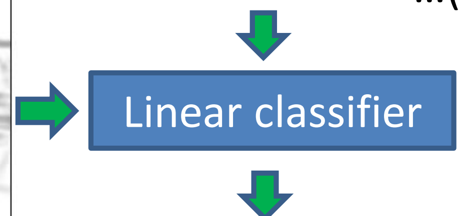
mPb – Classification using multiple features



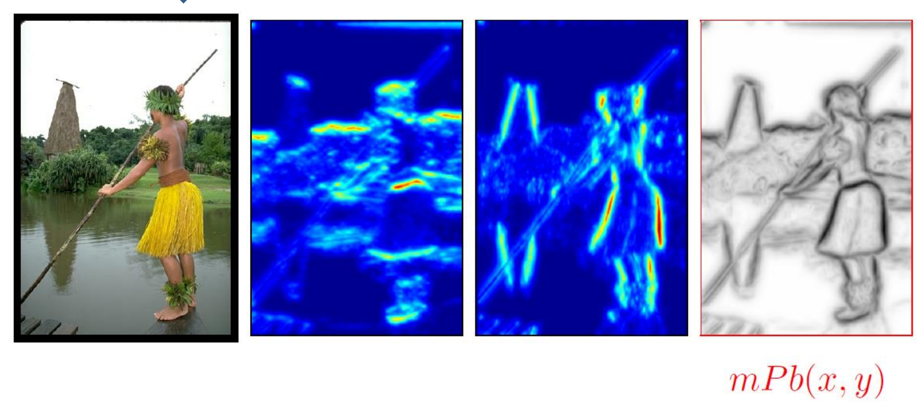
...(other scales)



...(other scales)

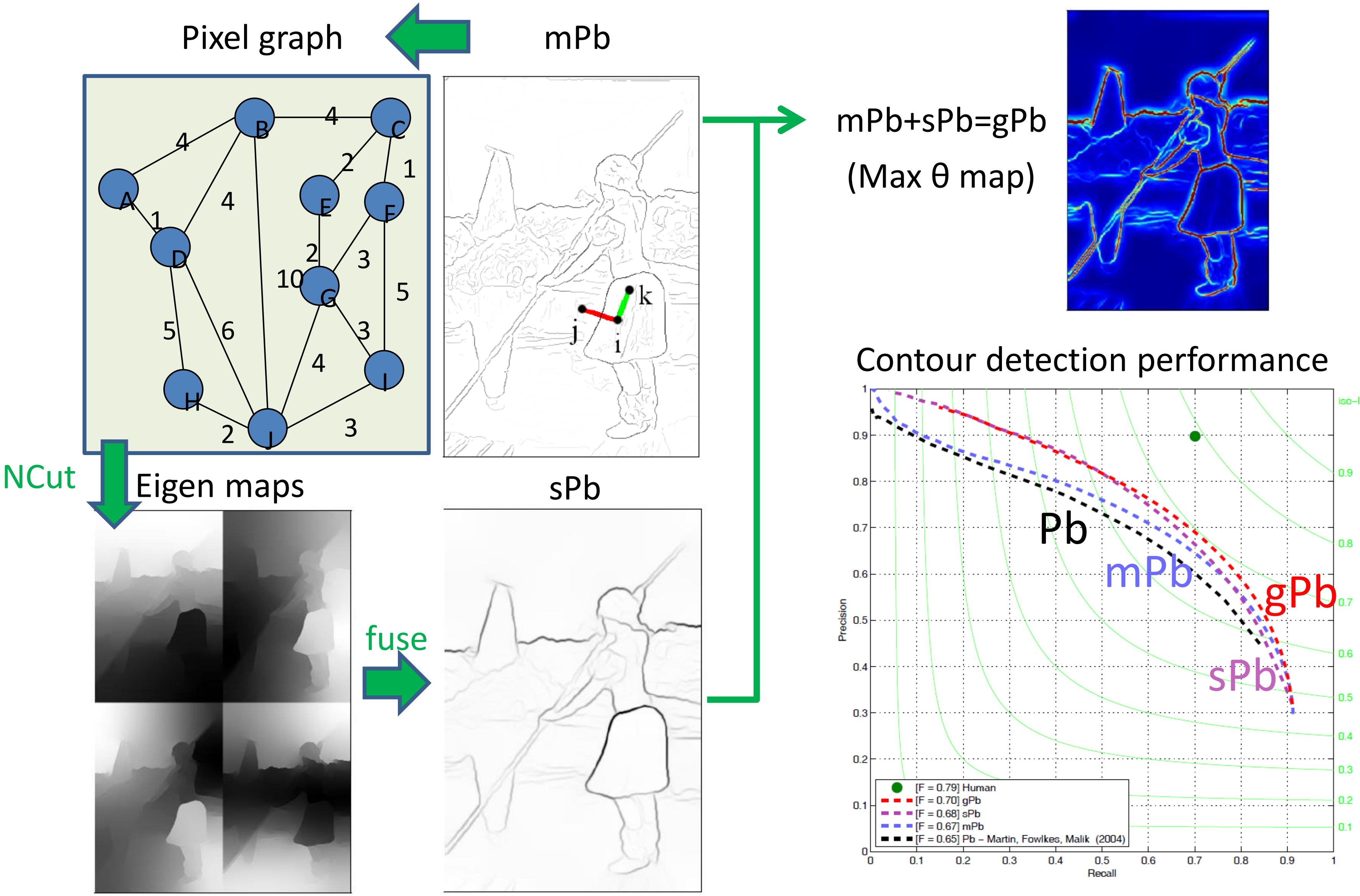


Other researchers use dictionary learning [1] or random forest [2], etc...

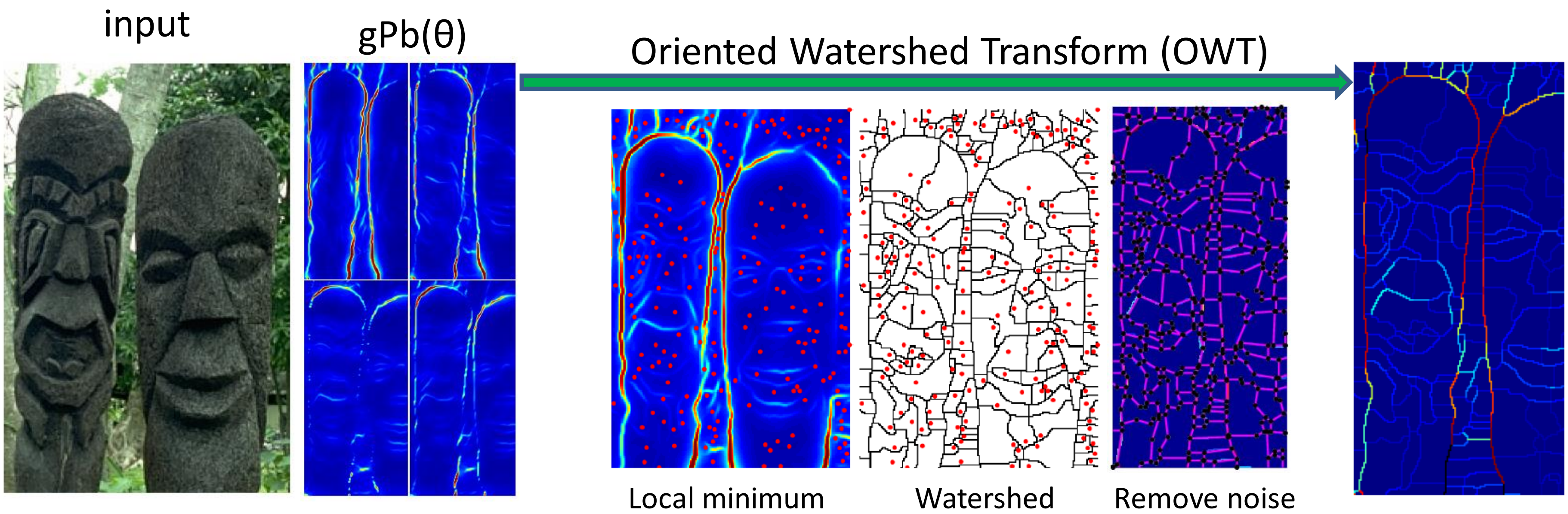


[1] Discriminatively Trained Sparse Code Gradients for Contour Detection, NIPS 2012
[2] Structure Forests for Fast Edge Detection, ICCV 2013

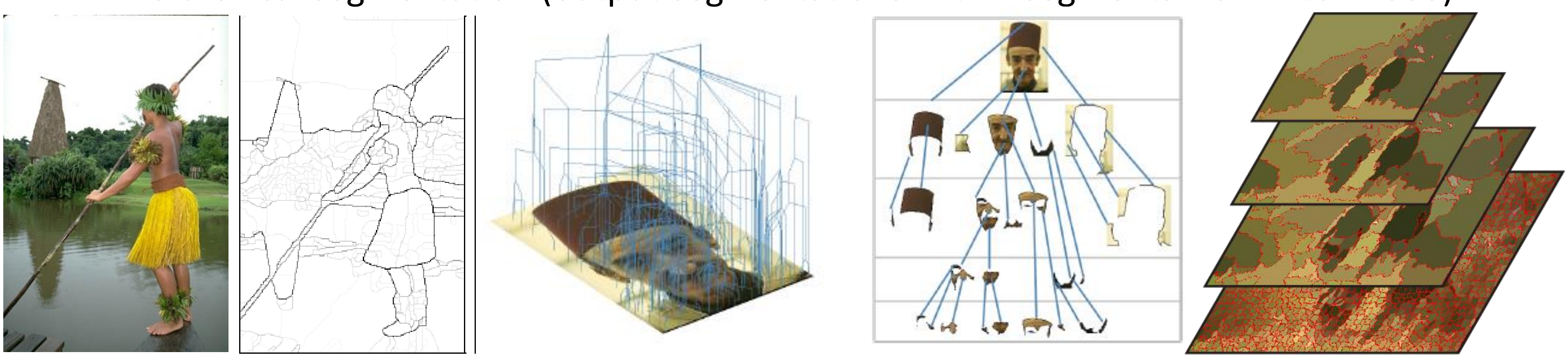
gPb (global Probability boundary)



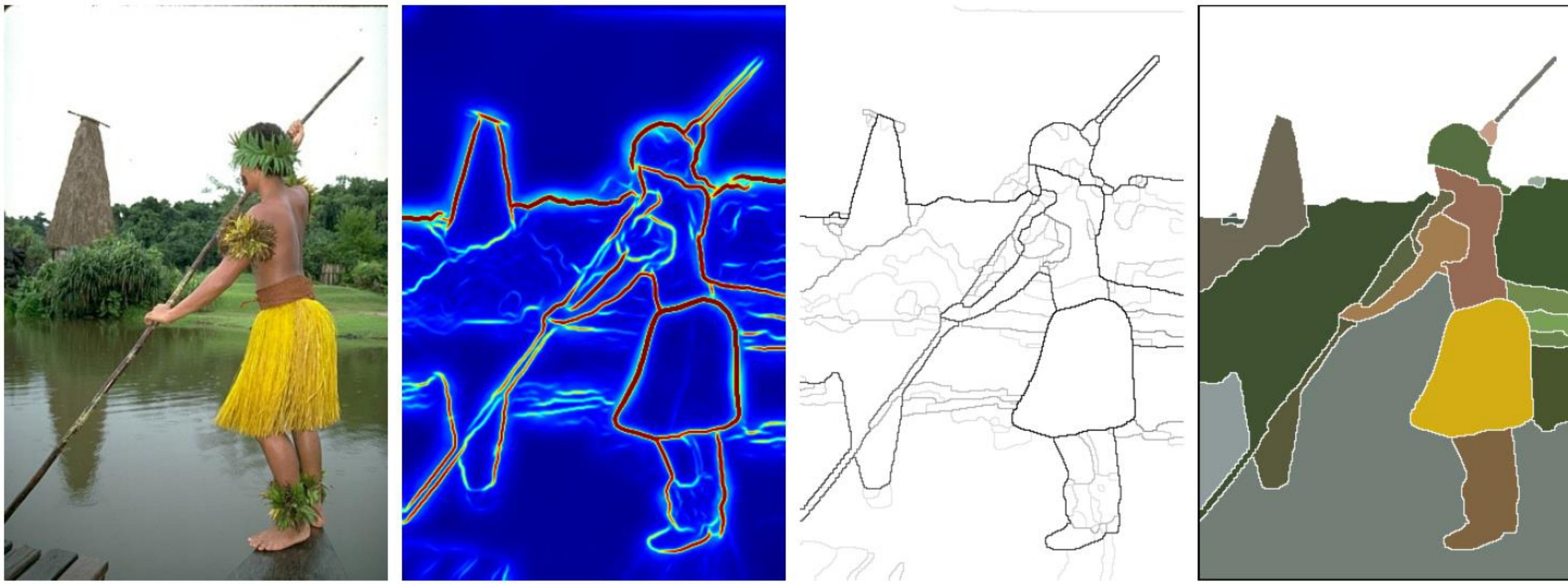
OWT-UCM



UCM (greedy merge like FH segmentation) →
Hierarchical segmentation (output segmentations with # segments from 1 to ~1000)



Conclusion - gPb-OWT-UCM



input | global Probability boundary (gPb) | Oriented watershed transform (OWT) | Ultrametric Contour Map (UCM)

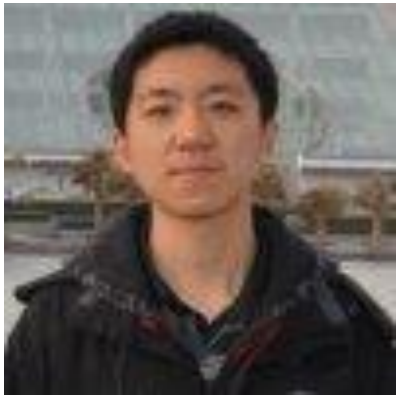
	Performance	Efficiency	Complexity (# parameters)	Theoretical support
gPb-OWT-UCM (need training) (2011)	Good	Slow	Complex	Empirical

It outperforms all methods which don't use gPb, and is widely used in recognition applications.
The concept of contour detection could be easily generalized to RGB-D images

Methods

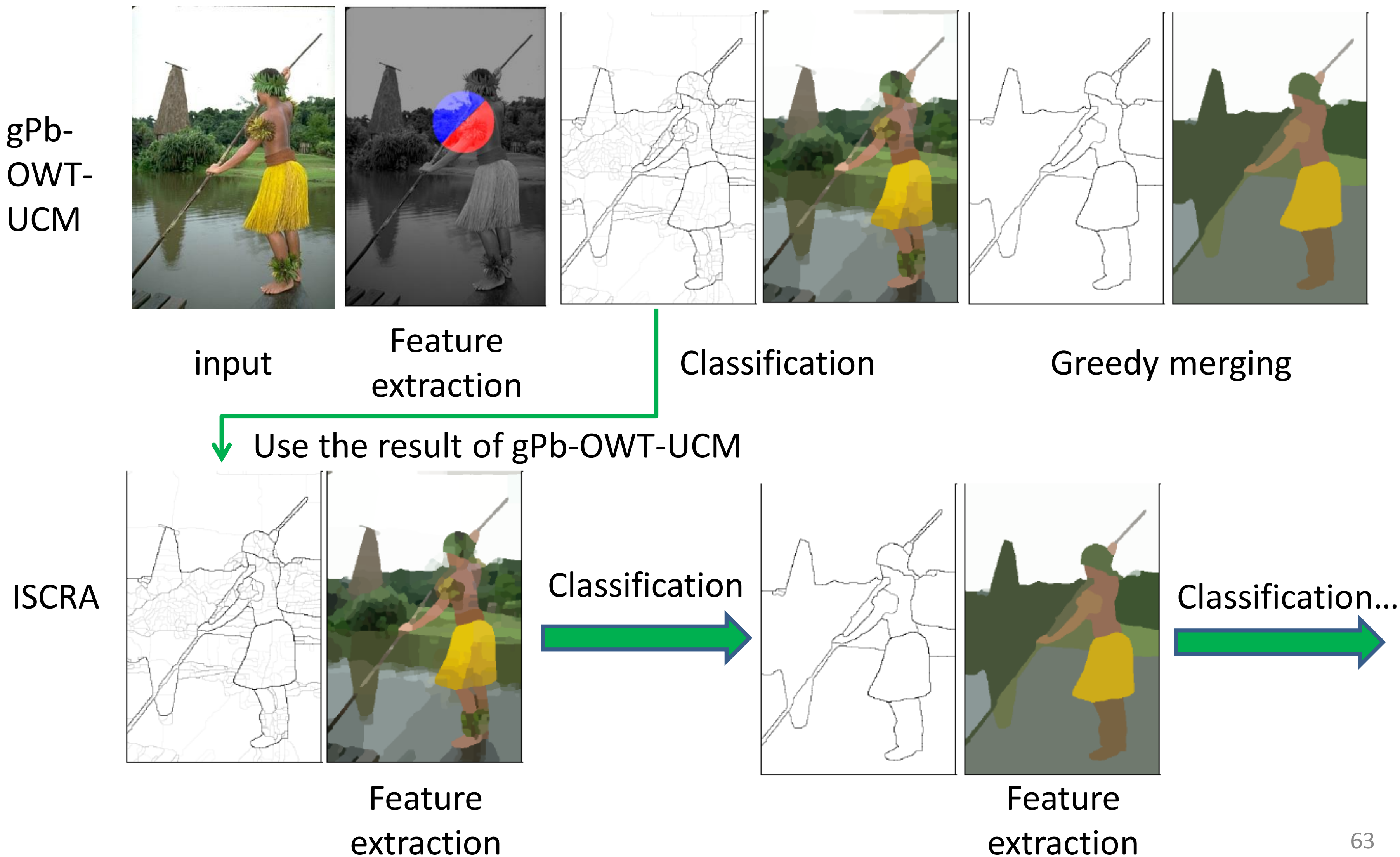
Ncut (Normalized Cut)
FH (Efficient graph-based image segmentation)
Mean Shift
gPb-OWT-UCM
ISCRA (Image Segmentation by Cascaded Region Agglomeration)

Citation: 1
(since year 2013)



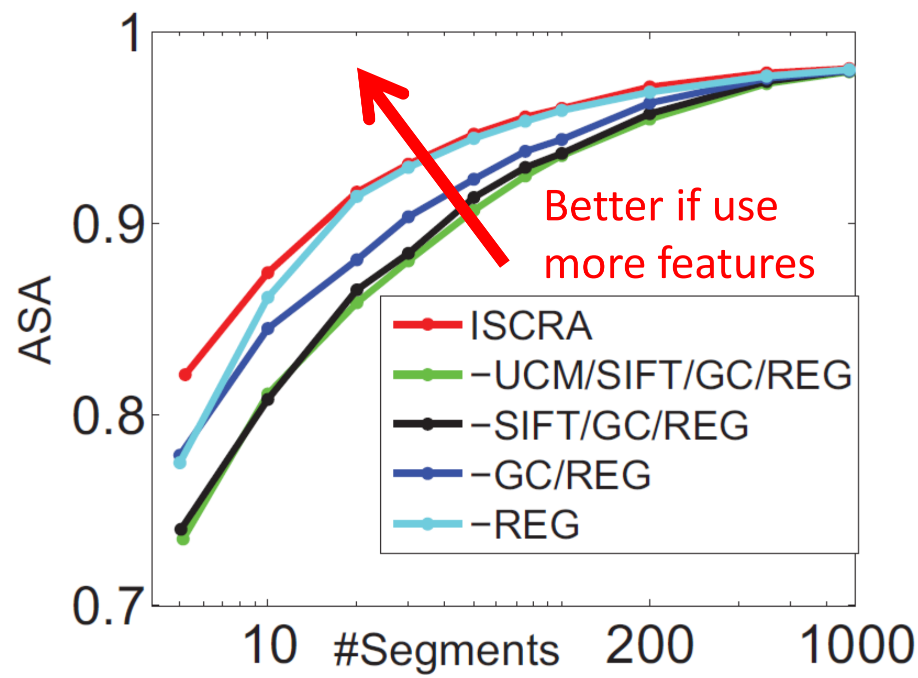
Cascaded Region Agglomeration

Image Segmentation by Cascaded Region Agglomeration, CVPR 2013



Conclusion - ISCRA

- The performance is better than gPb-OWT-UCM because ISCRA
  - uses more features.
  - uses the different classifiers for different scales (layer of hierarchical).
    - could see more global features
  - does not assume that object boundaries are straight lines in the feature extraction.



	Performance	Efficiency	Complexity (# parameters)	Theoretical support
ISCRA (2013+) (code not released)	Best (state-of-the-art)	Slowest (need to use gPb)	Complex	Empirical

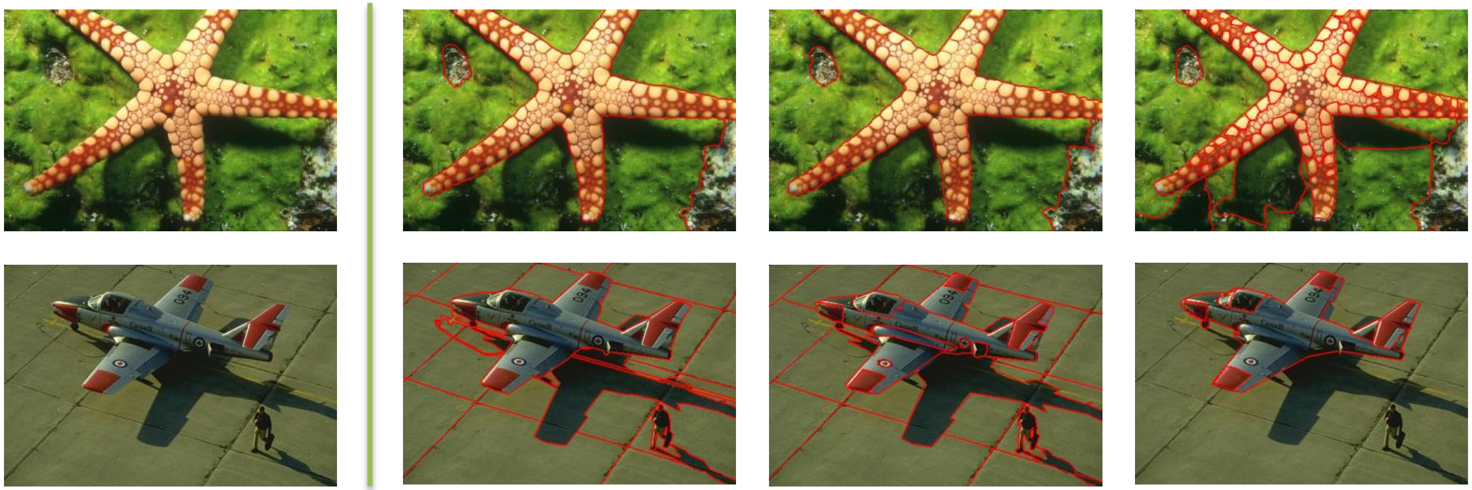
Experimental Results

Berkeley Segmentation Dataset (BSDS)
Evaluation - Quantitative Results
How much room left for improvement?

Berkeley Segmentation Dataset (BSDS)

Citation: 1884 (since year 2001)

- 500 natural images
- Image size: 321×481 (or 481×321)
    - \# ground truth segmentations per image ≥ 4
        - Manually segmented by different human subjects
        - The quality of ground truth is pretty high.
- Examples



Image

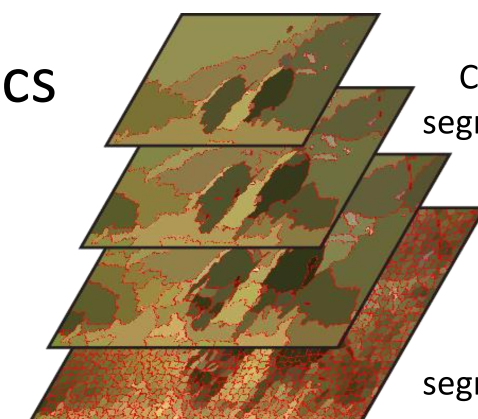
Ground truths

Quantitative Evaluation Criteria

	BSDS 300 (100 test images)			BSDS 500 (200 test images)		
	SC (↑)	PRI (↑)	VoI (↓)	SC (↑)	PRI (↑)	VoI (↓)
NCut	0.44	0.75	2.18	0.45	0.78	1.89
FH	0.51	0.77	2.15	0.52	0.80	2.21
Mean shift	0.54	0.78	1.83	0.54	0.79	1.85
gPb-OWT-UCM	0.59	**0.81**	1.65	**0.59**	**0.83**	1.69
ISCRA	**0.60**	**0.81**	**1.61**	**0.59**	0.82	**1.60**

Outperforms gPb-OWT-UCM more significantly on other datasets such as MSRC or SBD

- Criteria characteristics



Coarse segmention

↑

Fine segmention

VoI

↑

SC

↑

PRI

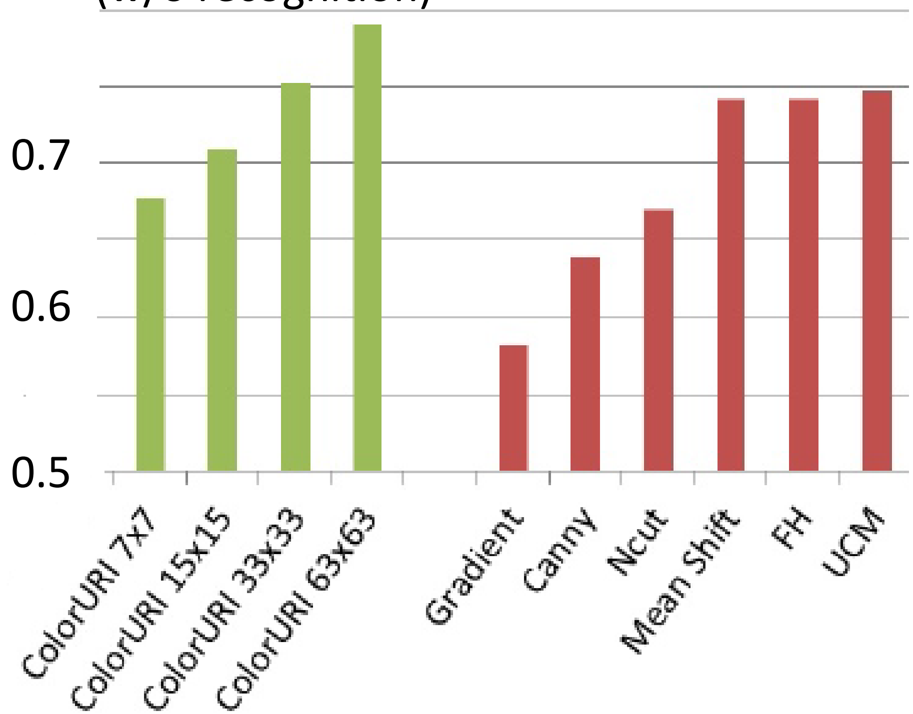
(details)

How much room left for improvement?

- SUN dataset
  - Subset: 185 images
- Interesting facts
  - Learning based method might overfit training data
    - gPb-OWT-UCM is not significantly better
  - Local information has been exhausted
  - There is not much room for improvement

Contour detection accuracy

Human (w/o recognition) Machine



The Role of Image Understanding in Contour Detection, CVPR2012

Conclusion

- After years of research efforts, researchers find that totally unsupervised segmentation is an ill-posed problem, so the methods or problems required learning are actually more well-defined and practical
  - I believe that it is very hard to improve performance in totally unsupervised segmentation problem without overfitting datasets
  - I believe that it is easier and more meaningful to improve performances in supervised setting or other desired characteristics such as preventing overfitting, efficiency of training and testing, complexity and theoretical support …

	Performance	Efficiency	Complexity (# parameters)	Theoretical support
NCut (2000)	Bad (original version)	Reasonable	Simple	Best
FH (2004)	Reasonable (↑ for superpixels)	Fastest	Simple	Empirical
Mean shift (2002)	Reasonable (↑ for oversegments)	Reasonable	Simple	Good
gPb-OWT-UCM (need training) (2011)	Good	Slow	Complex	Empirical
ISCRA (2013+) (code not released)	Best (state of the art)	Slowest (need to use gPb)	Complex	Empirical

Q & A

References

- J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.
- D. Martin, C. Fowlkes, D. Tal, and J. Malik, "*A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*," IEEE International Conference on Computer Vision, 2001.
- D. Comaniciu, and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 2002.
- P. Felzenszwalb, D. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, 2004.
- P. Arbel´aez, M. Maire, C. Fowlkes, and J. Malik, "*Contour Detection and Hierarchical Image Segmentation*," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011.
- Z Ren, G. Shakhnarovich, "*Image Segmentation by Cascaded Region Agglomeration*," IEEE Computer Vision and Pattern Recognition, 2013.

Appendix – NCut derivation

$$\mathrm{Ncut}(A,B) := \mathrm{cut}(A,B)\left(\frac{1}{\mathrm{vol}(A)} + \frac{1}{\mathrm{vol}(B)}\right)$$

Let f = $[f_1\ f_2\ \ldots\ f_n]^T$ with $f_i = \begin{cases} \frac{1}{\mathrm{vol}(A)} & \text{if } i \in A \\ -\frac{1}{\mathrm{vol}(B)} & \text{if } i \in B \end{cases}$

$$\mathbf{f}^T \mathbf{L} \mathbf{f} = \sum_{ij} w_{ij} (\mathbf{f}_i - \mathbf{f}_j)^2 = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{\mathrm{vol}(A)} + \frac{1}{\mathrm{vol}(B)} \right)^2$$

(L=D-W)

$$\mathbf{f}^T \mathbf{D} \mathbf{f} = \sum_j d_i \mathbf{f}_i^2 = \sum_{i \in A} \frac{d_i}{\mathrm{vol}(A)^2} + \sum_{j \in B} \frac{d_i}{\mathrm{vol}(B)^2} = \frac{1}{\mathrm{vol}(A)} + \frac{1}{\mathrm{vol}(B)}$$

$$\mathrm{Ncut}(A,B) = \frac{\mathbf{f}^T \mathbf{L} \mathbf{f}}{\mathbf{f}^T \mathbf{D} \mathbf{f}} \Rightarrow \boxed{\mathbf{L}\mathbf{f} = \lambda \mathbf{D} \mathbf{f}}$$

back

Appendix – Quantitative Evaluation Criteria

- Probabilistic Rand Index (PRI) ↑ $[0, 1]$
  - For any two pixels, what is the probability of the two pixels belonging to the same cluster (or different clusters) in the ground-truth set?
- Variation of Information (VoI) ↓ $[0, \infty)$
  - How many bits are required to describe the difference between a test result and a ground truth?
- Segmentation Covering (SC) ↑ $[0, 1]$
  - Suppose each segment in a ground truth is to be detected, what is the overall quality of detections in a test result according to IOU-overlap?
- Achievable segmentation accuracy (ASA) ↑ $[0, 1]$
  - ASA is a superpixel performance upper bound measure. It gives the highest accuracy achievable for object segmentation that utilizes superpixels as units.

Criteria / Stress on	SC (↑)	PRI (↑)	VoI (↓)
Hierarchical level	Middle	Low	High
Scale of objects	Middle	Small	Large
# segments	Middle	High	Low

back