Circular Partitions with Applications to Visualization and Embeddings

Anastasios Sidiropoulos (MIT)

Joint work with Krzysztof Onak (MIT)

Metric spaces

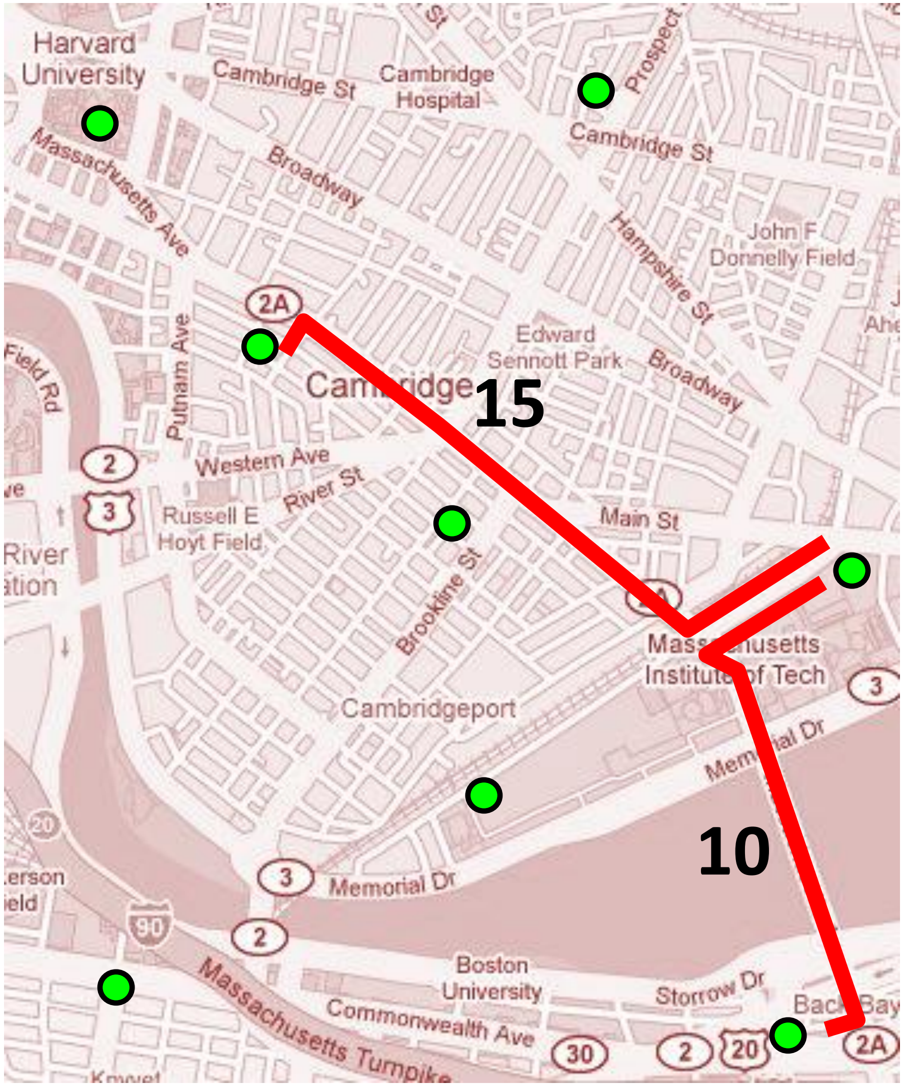
Metric space M=(X,D)

- Positive definiteness

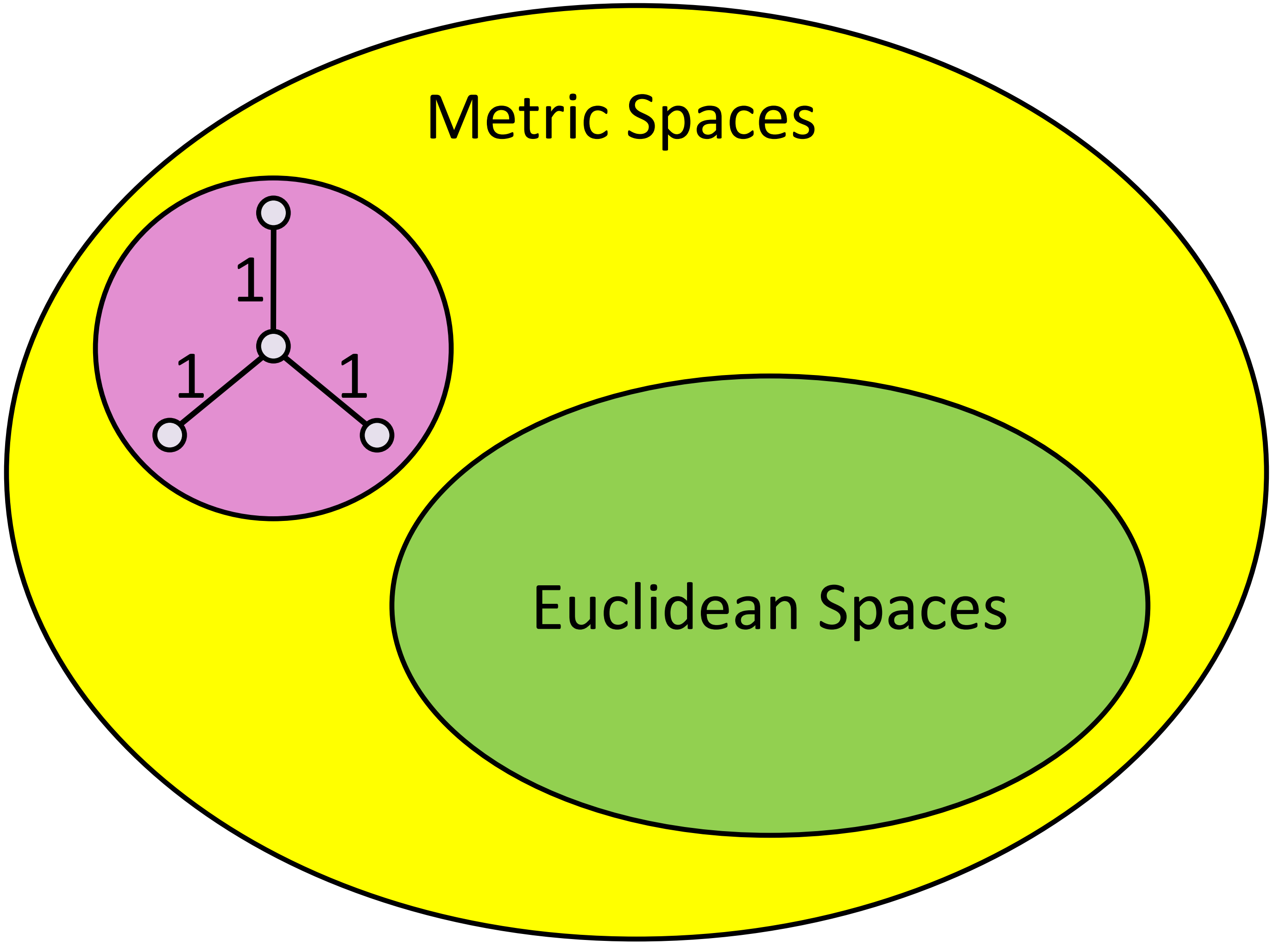
  D(p,q) = 0 iff p = q
- Symmetry

  D(p,q) = D(q,p)
- Triangle inequality

  D(p,q) ≤ D(p,r) + D(r,q)

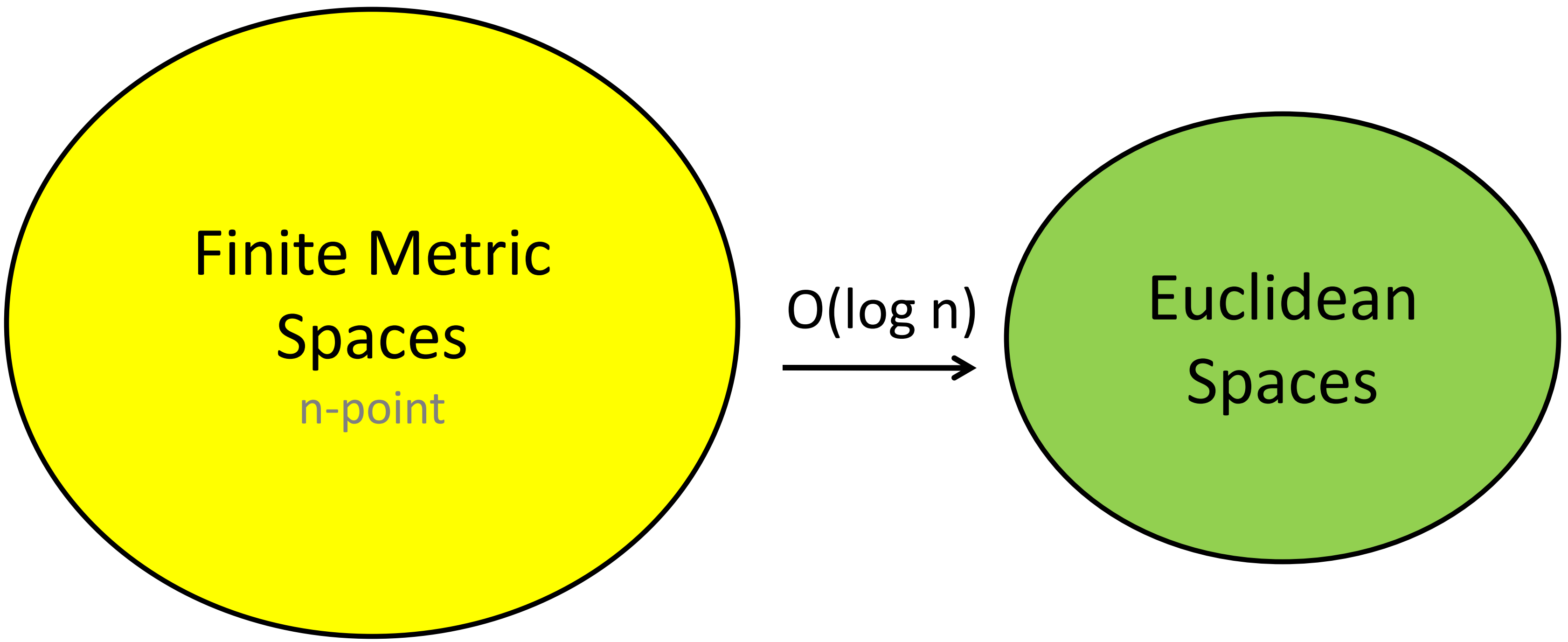


Metric spaces



Metric embeddings

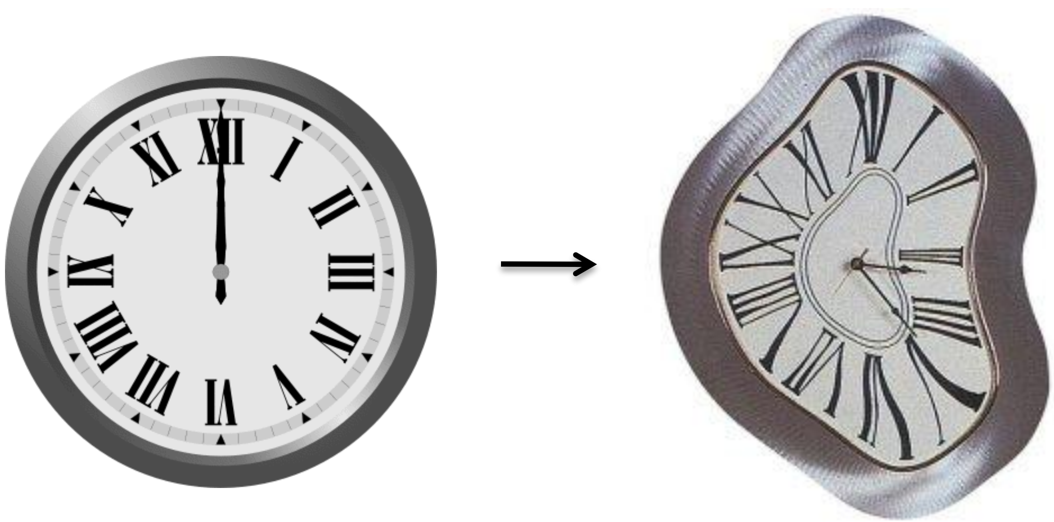
[Bourgain '85]



Metric embeddings

- Given spaces $M=(X,D)$, $M'=(X',D')$
- Mapping $f:X \longrightarrow X'$
- Distortion c if:

$$D(x_1,x_2) \leq D'(f(x_1),f(x_2)) \leq c \cdot D(x_1,x_2)$$



Motivation

- Geometric interpretation
- Succinct data representation
  - Embedding into low-dimensional spaces
- Visualization
  - Embedding into the plane
  - Multi-dimensional scaling
- Optimization
  - Embedding into “easy” spaces
- Phylogenetic reconstruction
  - Embedding into trees

Known results

Host space	**Distortion**	**Citation**
O(log n) –dimensional L_2 (also true for L_p)	O(log n)	[Bourgain ’85], [Johnson-Lindenstrauss], [Alon], [Linial, London, Rabinovich ’94], [Abraham, Bartal, Neiman ‘06]
d-dimensional L_2	$\tilde{O}(n^{2/d})$	[Matousek ’90]

Absolute vs. relative embeddings

- Small dimension $\longrightarrow$ high distortion $(n^{\Omega(1/d)})$
  - E.g. embedding a cycle into the line
- What if a particular metric embeds with small distortion?
- Computational problem:

Approximate best possible distortion

Relative embeddings into R^d

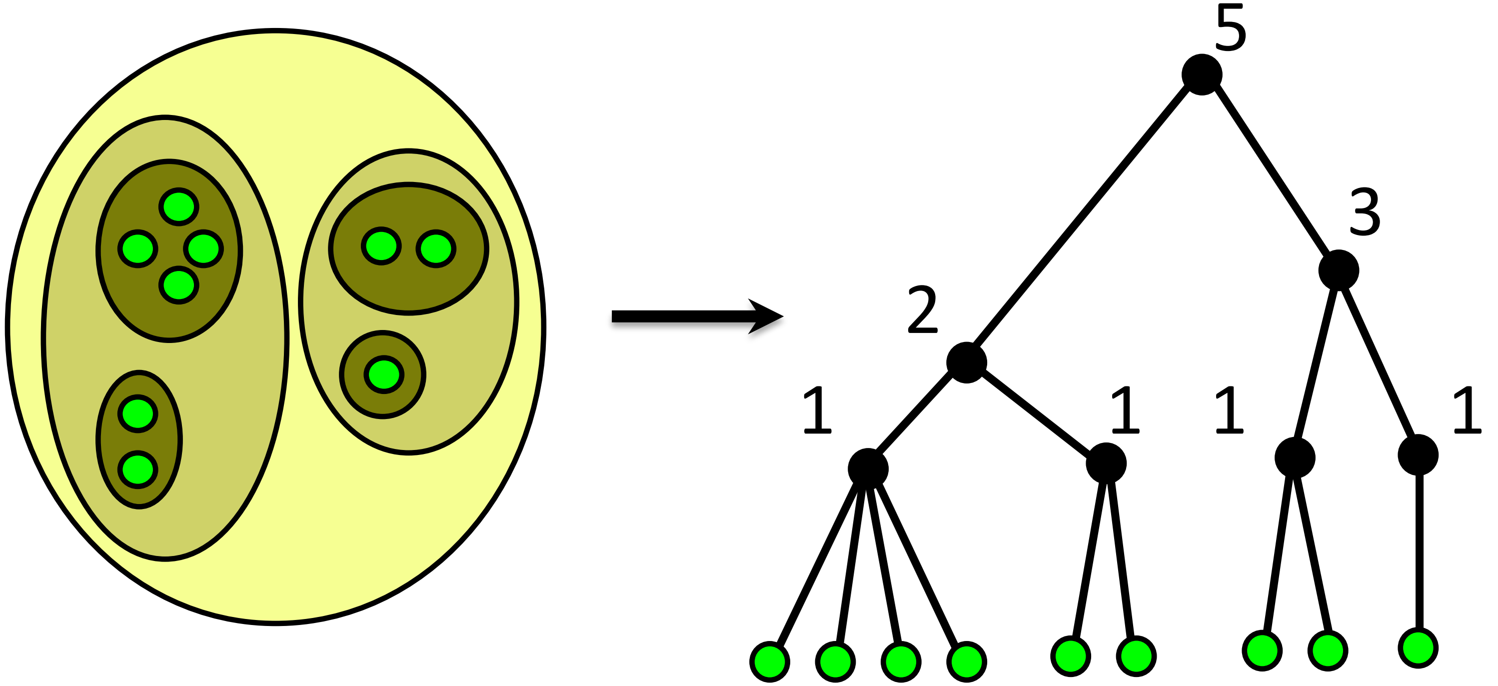
Input	Host	Distortion	Citation
ultrametrics	R^d	$OPT^{O(d)}$	[Badoiu, Chuzhoy, Indyk, S ’06]
ultrametrics	R^d	$OPT \cdot \log^{O(d)} \Delta$	[Onak, S ‘08]

Input	Host	Hardness	Citation
ultrametrics	R^d	NP-hard	[Badoiu, Chuzhoy, Indyk, S ‘06]
general	R^d	$\Omega(n^{1/17d}) \cdot OPT$	[Matousek, S ‘08]

Embedding ultrametrics into R^d

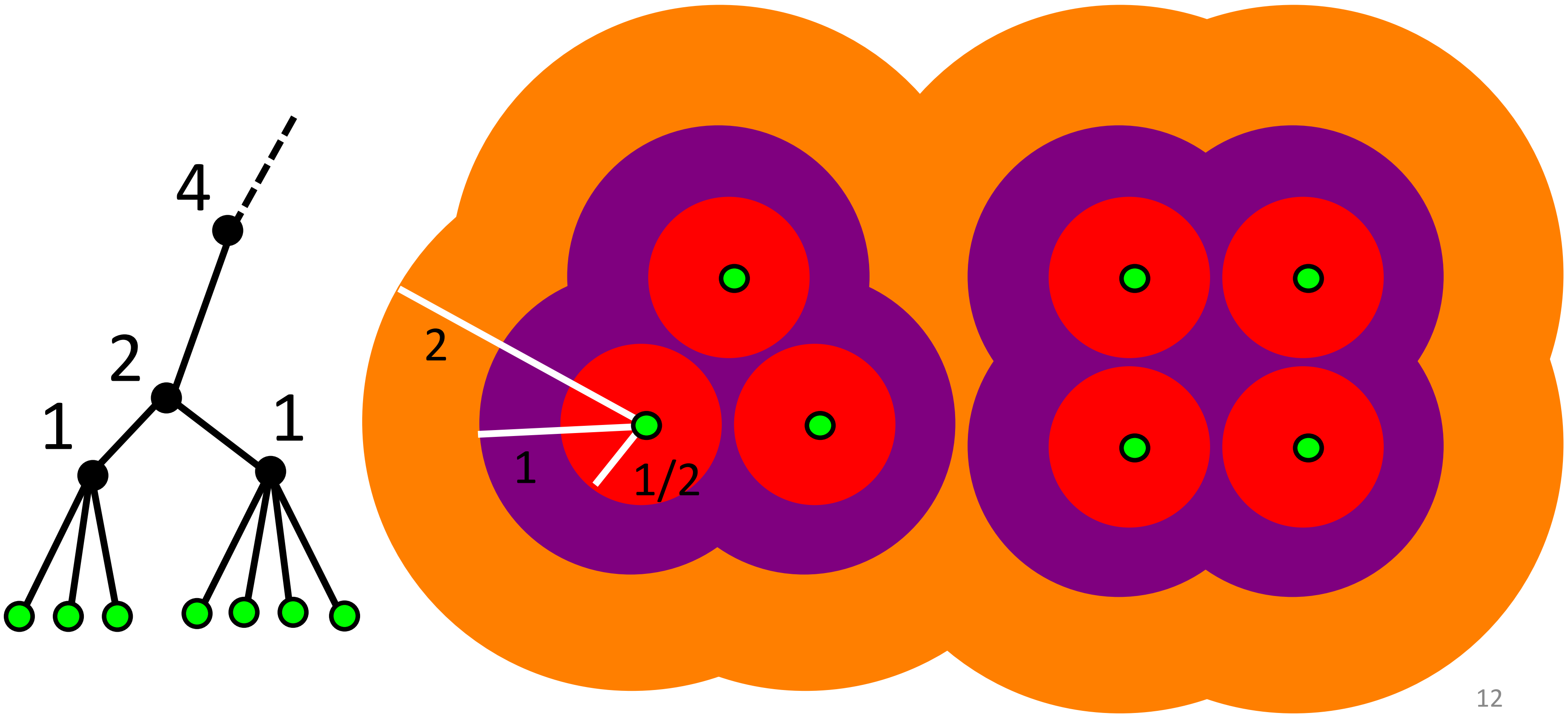
Ultrametrics

- l(u): label of u
- D(u,v)=l(nca(u,v))

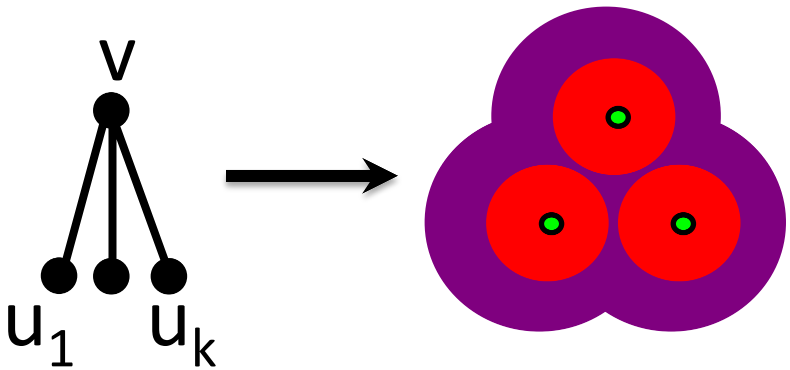


A lower bound on the optimal

- Many disjoint "areas"
- Small distortion $\longrightarrow$ small total "area"



Lower bound: Main idea



[Badoiu,Chuzhoy,Indyk,S'07]

Estimate volumes via Brunn-Minkowski inequality

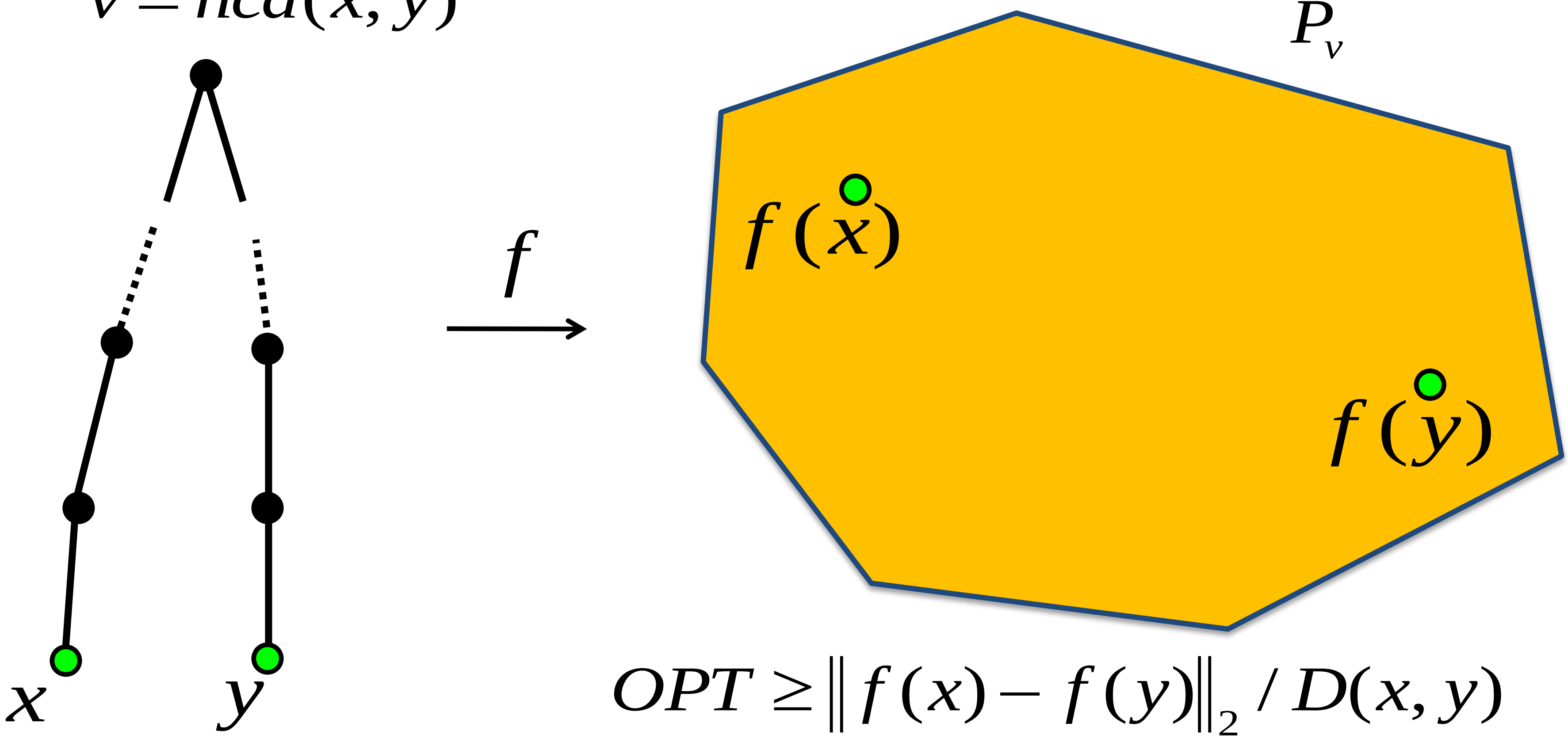
$$Vol(leaf) = O(1)$$

$$Vol(v) = \sum_{i=1}^{k} \left(\sqrt{Vol(u_i)} + l(v) \right)^2$$

Lemma: $OPT = \Omega\left(\sqrt{Vol(v)} \,/\, l(v) \right)$

Why does this work?

$$v = nca(x, y)$$

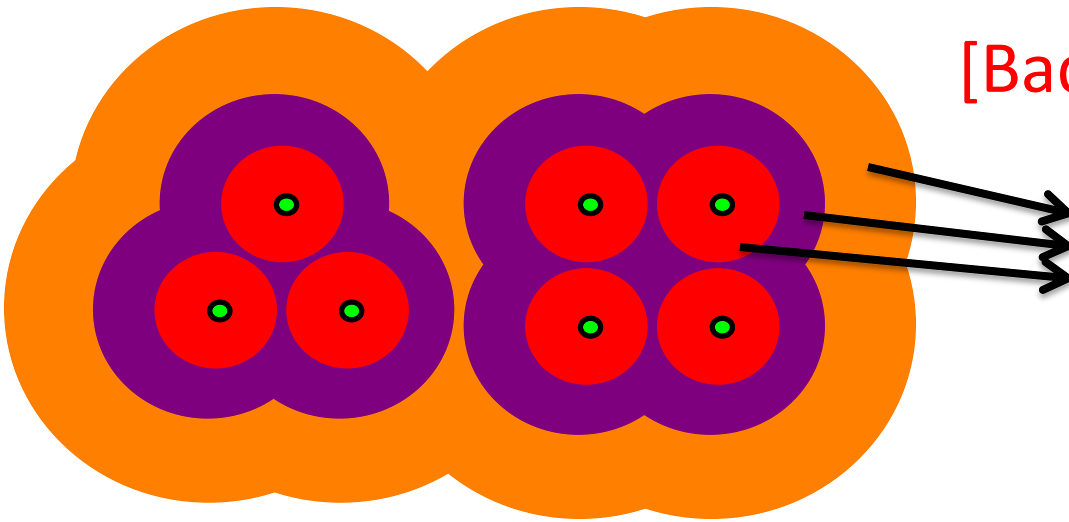


$$D(x, y) = l(u)$$

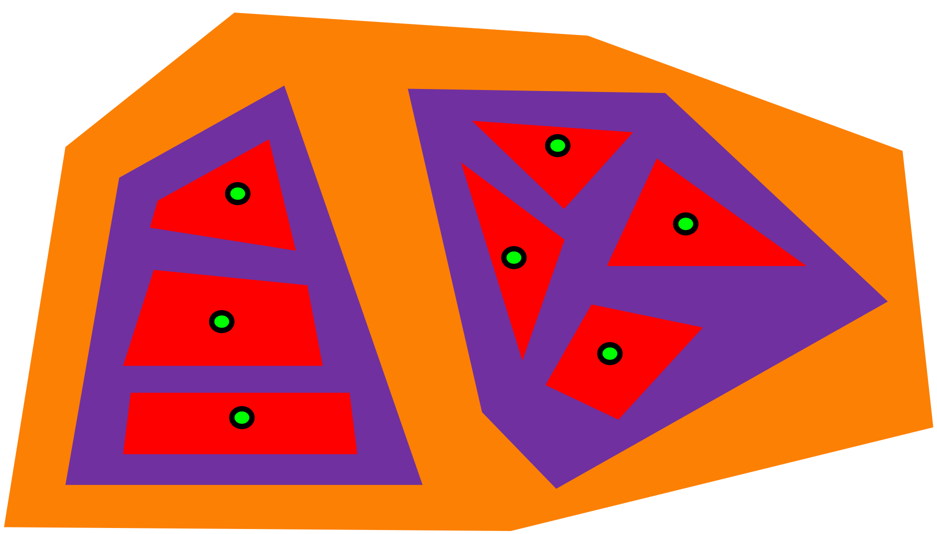
$$OPT \geq \| f(x) - f(y) \|_2 / D(x, y)$$
$$= \Omega(diam(P_v) / l(v))$$
$$= \Omega\left(\sqrt{Vol(v)} / l(v)\right)$$

Approximation: Main idea

[Badoiu,Chuzhoy,Indyk,S'07]



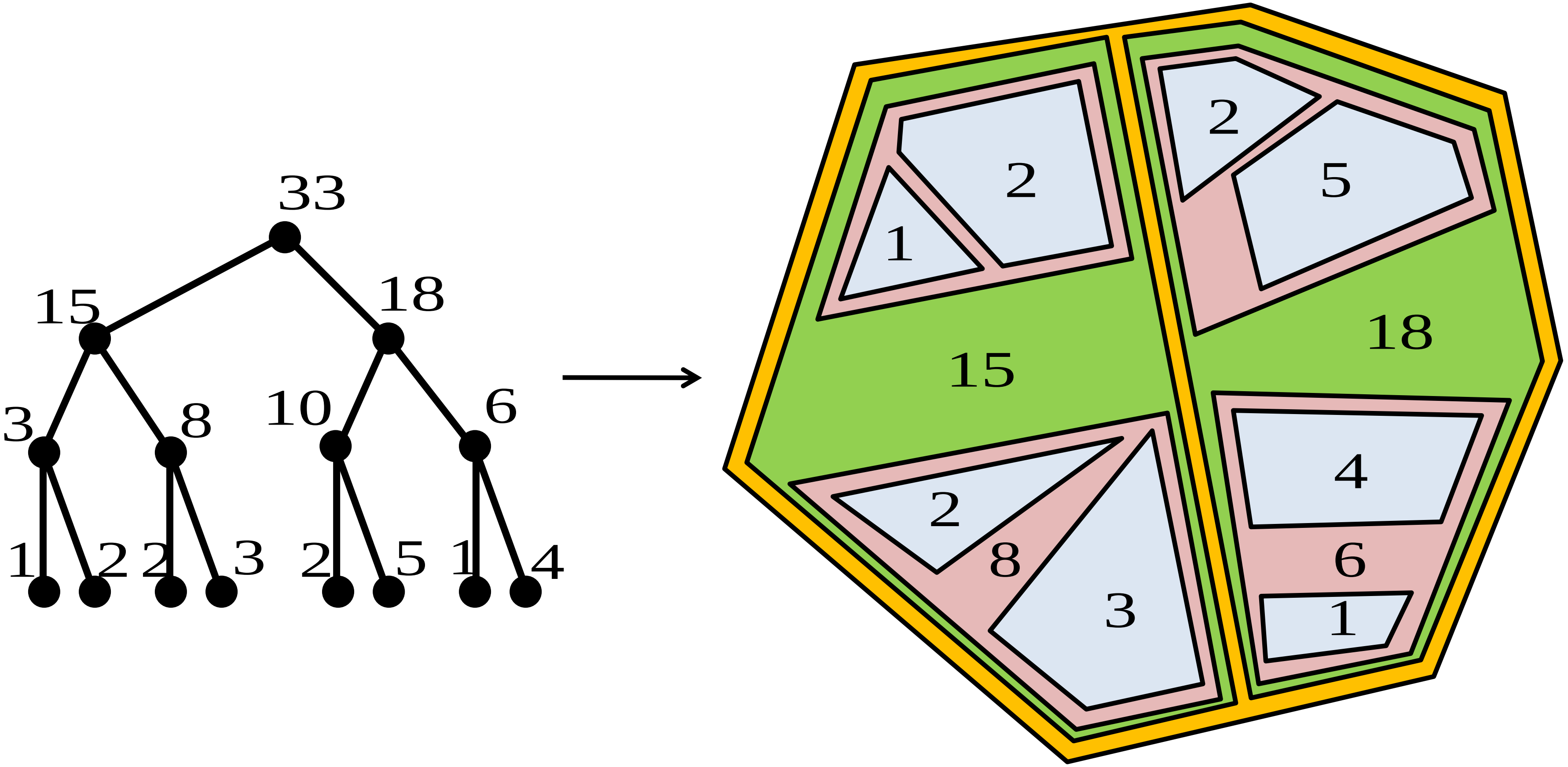
Estimate volumes via Brunn-Minkowski



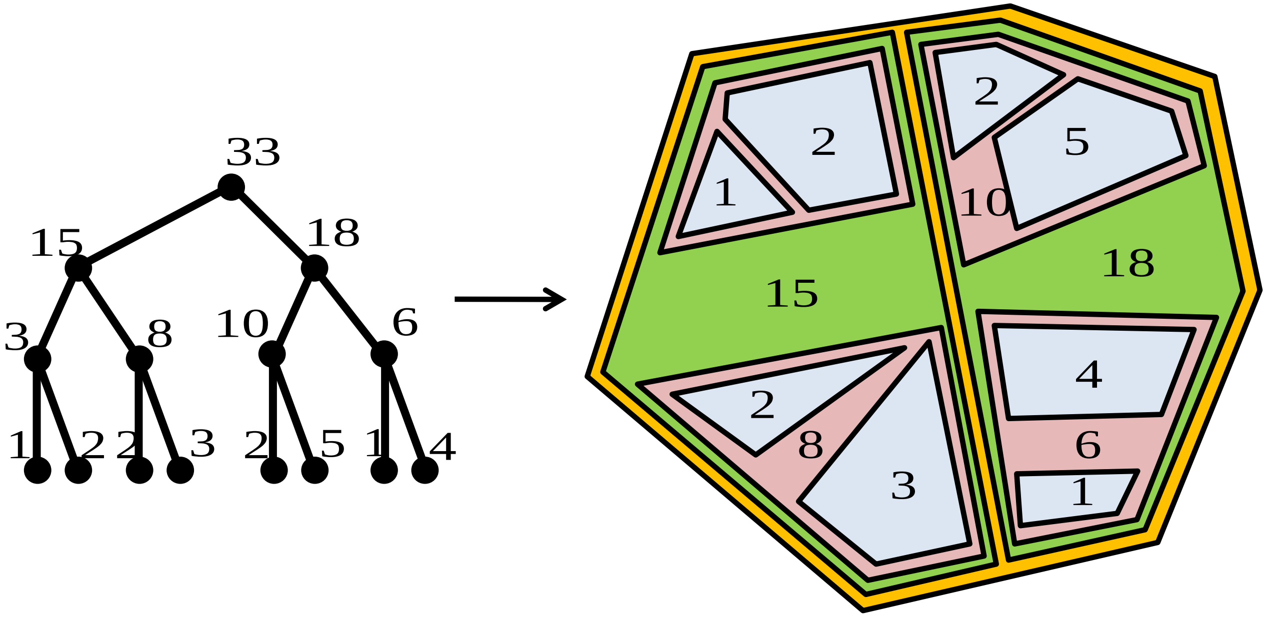
Match volumes
Minimize aspect ratio of polygons:

$$\lambda(A) = \frac{diam(A)}{\sqrt{Vol(A)}}$$

Hierarchical partitions



Hierarchical partitions



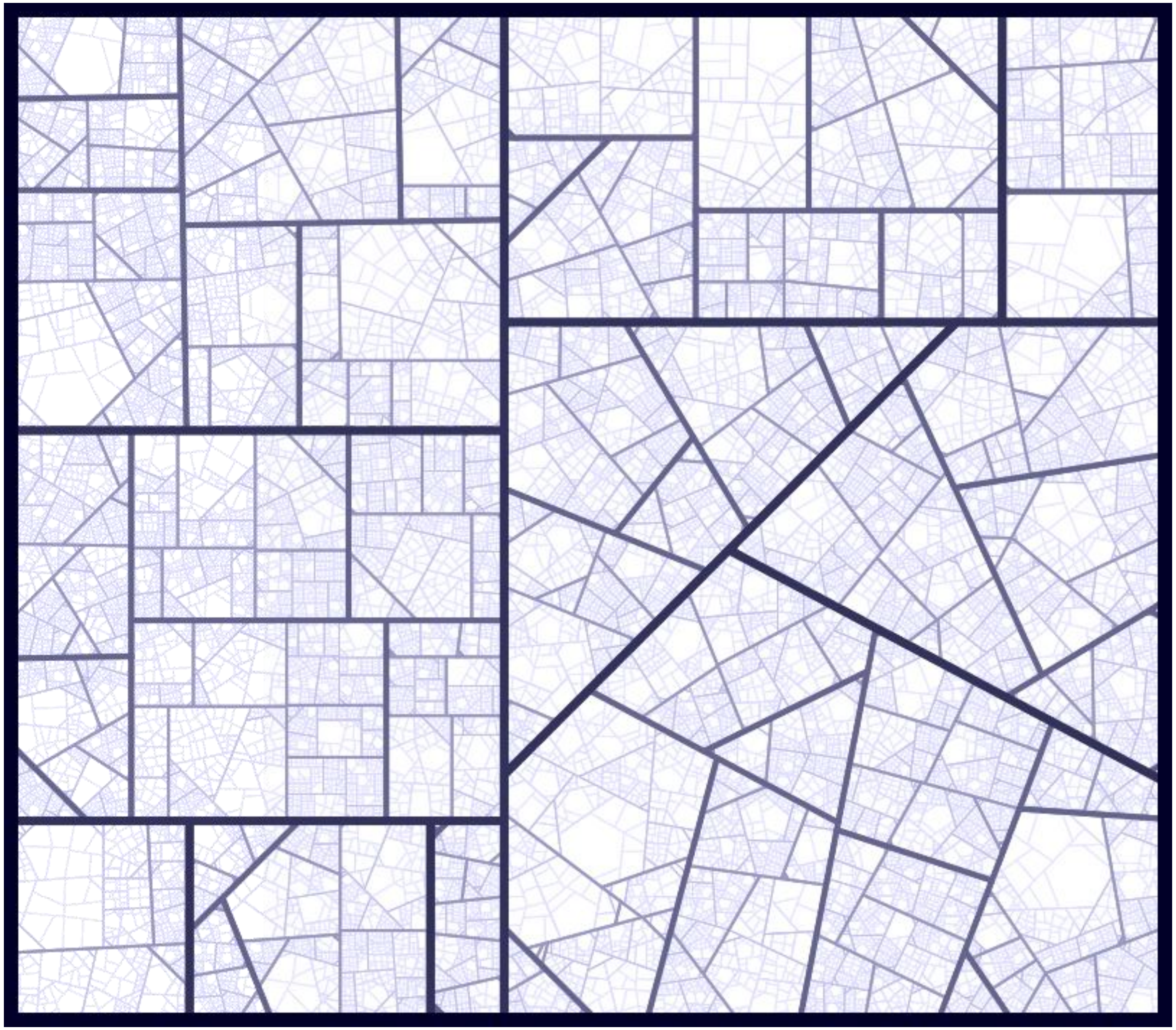
Theorem: [Onak, S '08]
There exist hierarchical partitions with poly-logarithmic aspect ratio.

The algorithm: Summary

- Estimate volumes
- Compute Hierarchical Partition
- Place each point in the center of its polygon

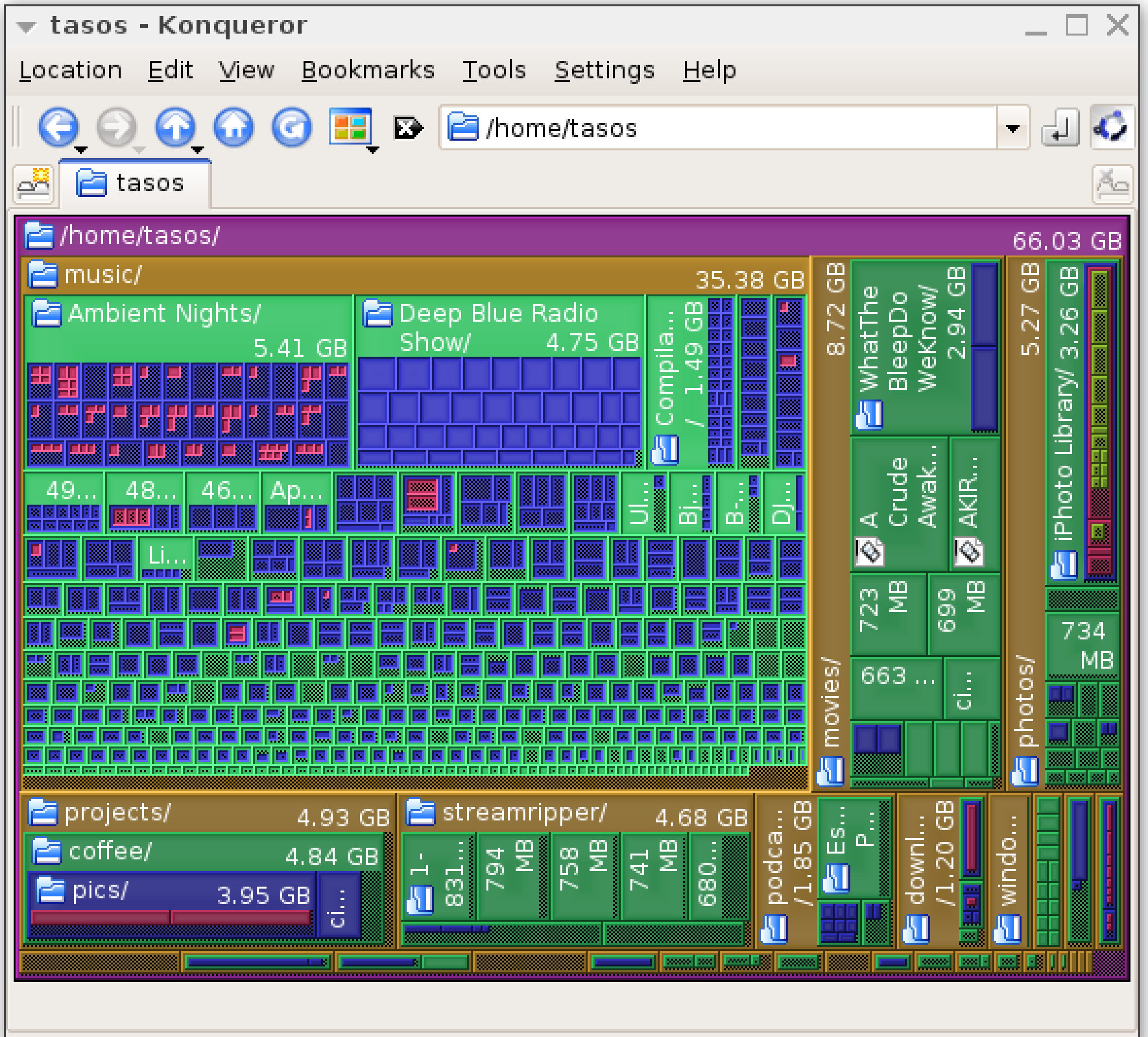
Our distortion = $\mathrm{OPT} \cdot \log^{O(1)} \Delta$

Hierarchical partitions and Treemap



[Onak, S ‘08]

Hierarchical partitions and Treemap



[Shneiderman '91]

Further directions

- Is this modified Treemap useful in practice?
- For $d \geq 3$, NP-hard to distinguish between distortion 100 and $\Omega(n^{0.01/d})$ [Matousek, S '08]
- Intriguing open problem:

  Embedding into R^d, $d \leq 2$.

  Is there an algorithm achieving distortion $OPT^{O(1)}$?

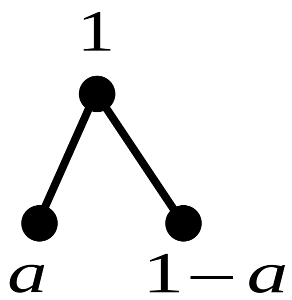
Questions?

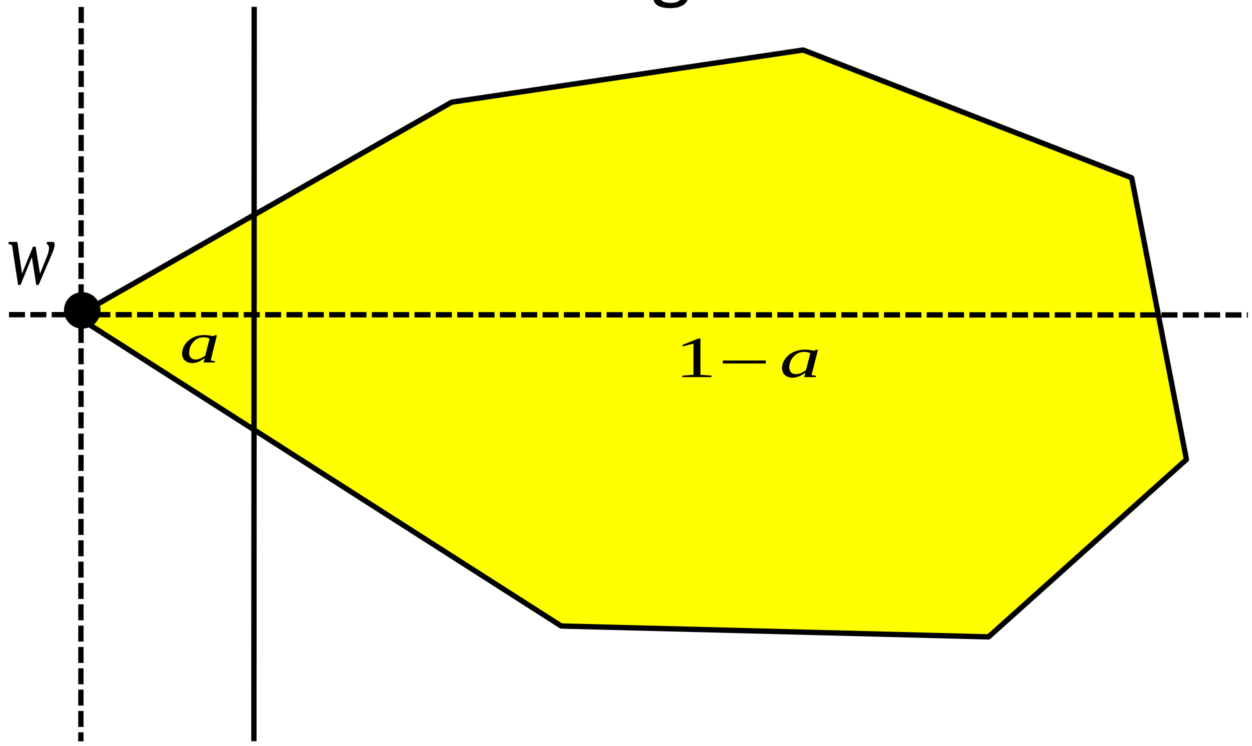
Computing circular partitions

It suffices to show how to perform one split:

Case 1: a is small

Pick smallest angle w





Computing circular partitions

It suffices to show how to perform one split:

Case 2: a is large

Cut along the diameter