Embedding Ultrametrics Into Low-Dimensional Spaces

Anastasios Sidiropoulos

Joint work with: Mihai Badoiu, Julia Chuzhoy, Piotr Indyk

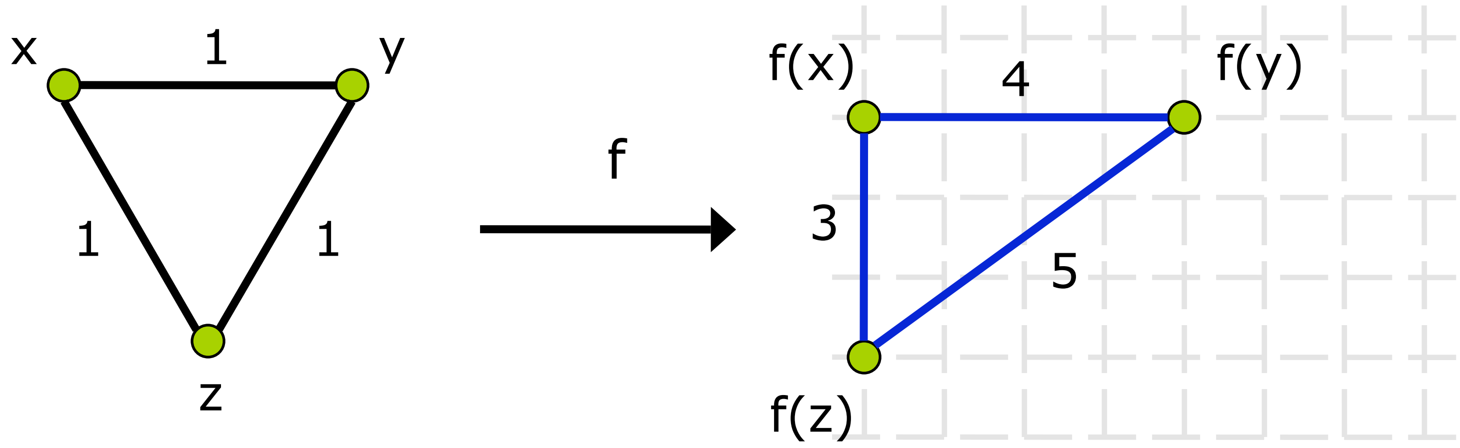
Embeddings of Metric Spaces

- Given finite metric space (X, D)
  - D(p,q)=0 ⇔ p=q
  - D(p,q)=D(q,p)
  - D(p,q)≤D(p,r)+D(r,q)
- Mapping f : X→Y
- Distortion of f is:

$$\max_{x_1,x_2} \frac{D'(f(x_1), f(x_2))}{D(x_1, x_2)} \times \max_{y_1,y_2} \frac{D(y_1, y_2)}{D'(f(y_1), f(y_2))}$$

GOAL: Minimize distrotion

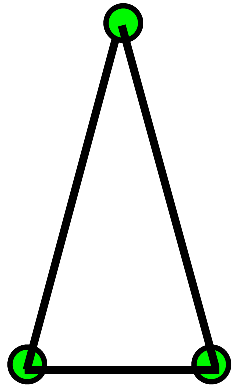
Metric Embedding - Example

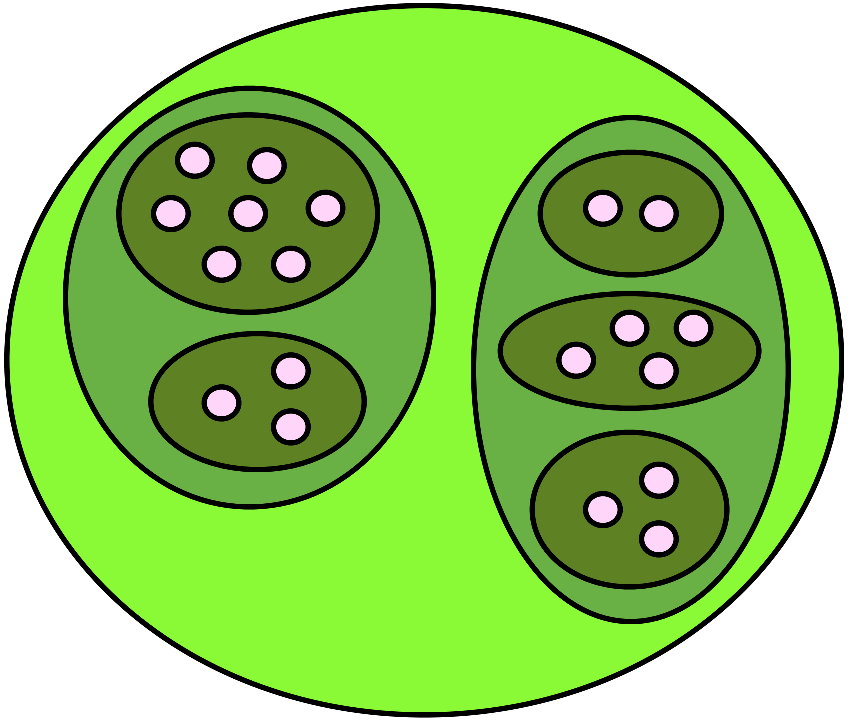


distortion = 5 ·(1/3) = 5/3

Ultrametrics

- An ultrametric is a metric space (X, D)
  - $D(p,q)=0 \Leftrightarrow p=q$
  - $D(p,q)=D(q,p)$
  - $D(p,q)\leq\max\{D(p,r), D(r,q)\}$
- Ideal clustering





Motivation

- Why Embeddings?
  - Compact data representation
  - Embedding into algorithmically nice spaces (e.g.Euclidean spaces, trees)
  - Visualization
  - Characterization of metric properties
- Why Ultrametrics?
  - Computational Biology
  - Clustering

Results on Worst-Case Embeddings

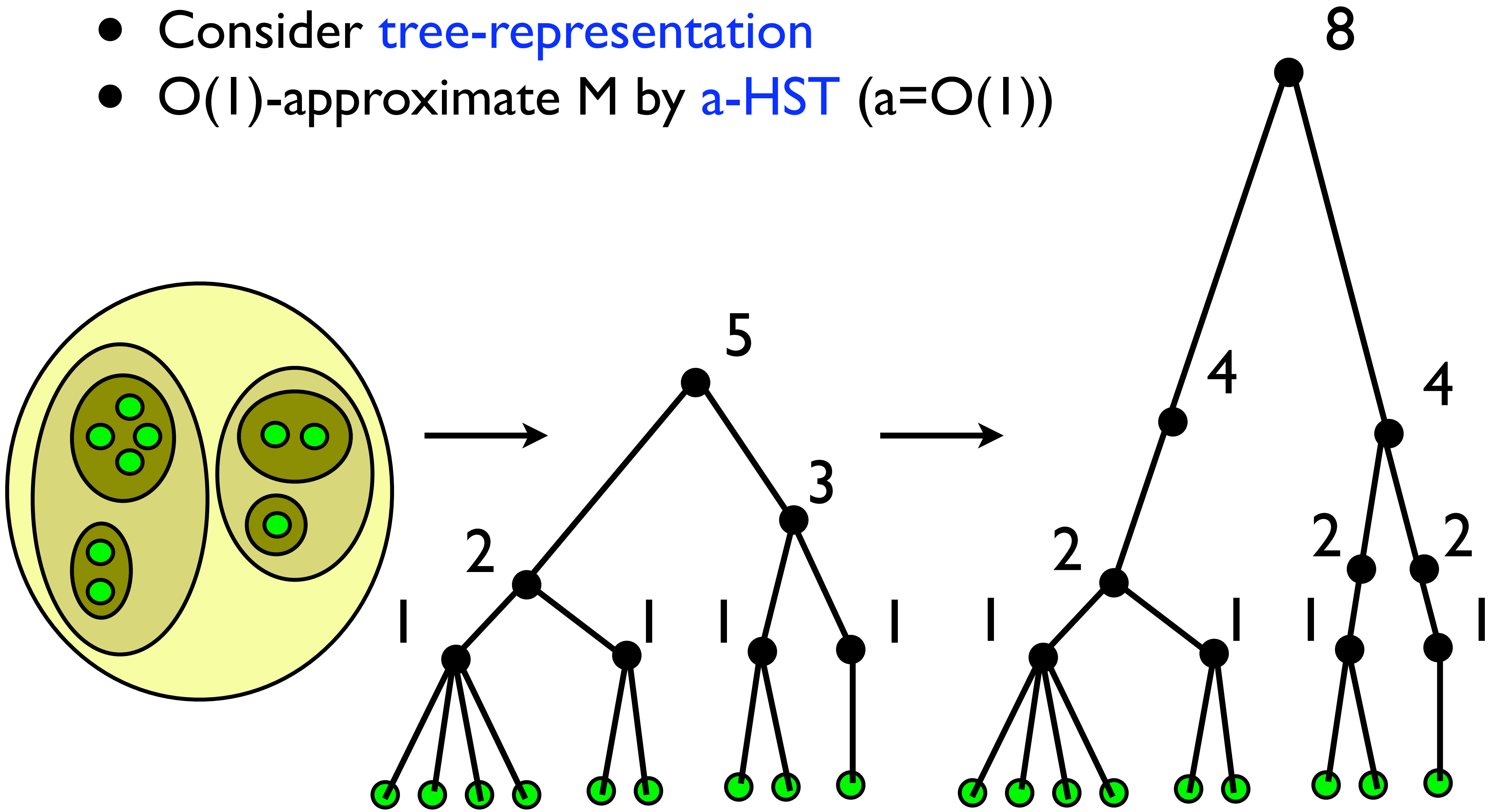
From	Into	Upper Bound	Lower Bound	ref
general	ℓ_2^d	$\tilde{O}(n^{2/d})$	$\Omega(n^{1/\lfloor (d+1)/2 \rfloor})$	[Matousek 90]
trees	ℓ_2^d	$\tilde{O}(n^{1/(d-1)})$	$\Omega(n^{1/d})$	[Gupta 99]
weighted stars	ℓ_2^d	$O(n^{1/d})$	$\Omega(n^{1/d})$	[Gupta 2000]
unweighted trees	plane	$O(n^{1/2})$	$\Omega(n^{1/2})$	[Babilon,Matousek Maxova,Valtr 02]
planar graphs	plane	$O(n)$	$\Omega(n^{2/3})$	[Betani,Demaine,Hajiaghayi,Moharrami 06]
unweighted outerplanar	plane	$O(n^{1/2})$	$\Omega(n^{1/2})$	[Betani,Demaine,Hajiaghayi,Moharrami 06]
ultrametrics	ℓ_2^d	$O(n^{1/d})$	$\Omega(n^{1/d})$	[Badoiu,Chyzhoy, Indyk,S. 06]

Results on Approximate Embeddings

From	Into	Upper Bound	Lower Bound	ref
general	ℓ_2	c		[Linial,London, Rabinovich 94]
unweighted graphs	subtrees	$O(c \log n)$	$\Omega(c)$	[Emek, Peleg 04]
general	trees	$(c \log n)^{O(\sqrt{\log \Delta})}$		[Badoiu, Indyk, S]
unweighted graphs	line	$O(c^2)$	$1.01c$	[BDGRRRS 05]
unweighted trees	line	$O(c^{3/2})$		[BDGRRRS 05]
general metrics	line	$O(\Delta^{3/4} c^{11/4})$		[Badoiu,Chyzhoy, Indyk, S. 05]
trees	line	$c^{O(1)}$	$\Omega(n^{1/12} c)$	[Badoiu,Chyzhoy, Indyk, S. 05]
ultrametrics	ℓ_2^d	$c^{O(d)}$	NP-complete	[Badoiu,Chyzhoy, Indyk, S. 06]

Ultrametrics Into Plane: Worst Case

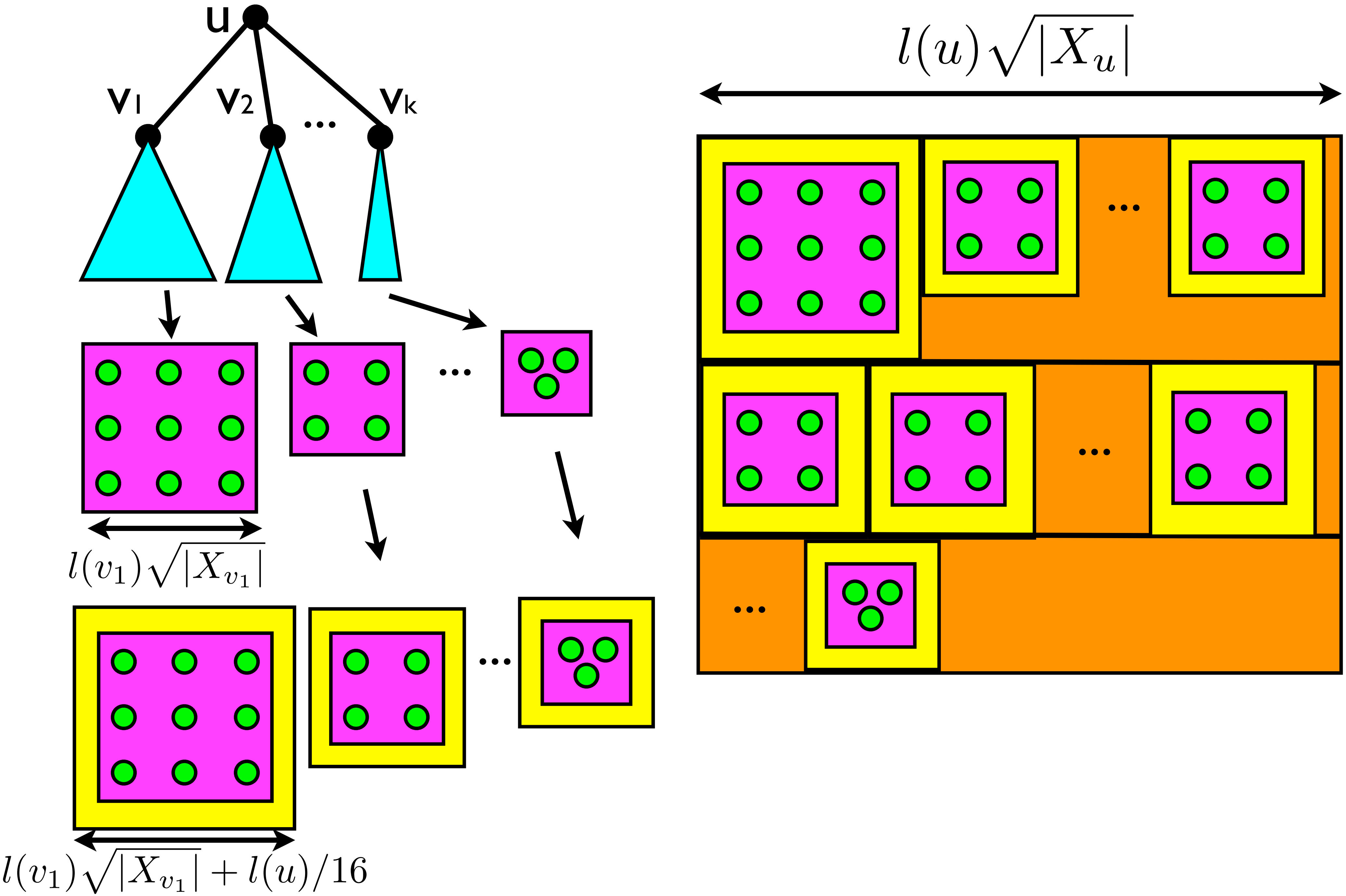
- Given ultrametric M=(X,D)
- Consider tree-representation
- O(1)-approximate M by a-HST (a=O(1))



Ultrametrics Into Plane: Worst Case

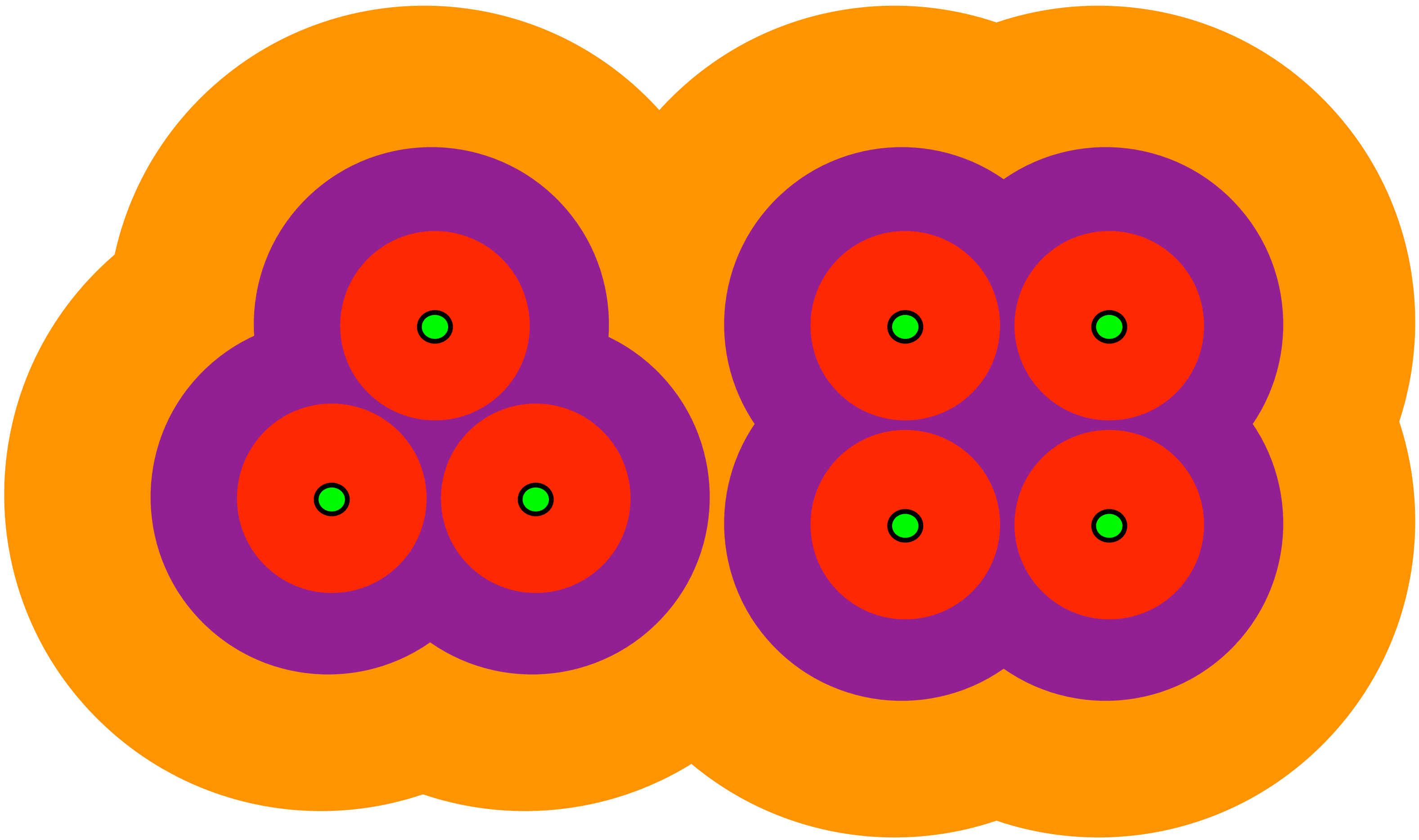
- X_u: Set of points with ancestor u
- l(u): label of u
- Compute f inductively on the HST
- For each node u of the HST:
  - f has contraction O(1)
  - $f(X_u)$ is contained in a square of side $l(u)\sqrt{|X_u|}$

Ultrametrics Into Plane: Worst Case



Approximation: A lower bound

- Observation: In a non-contracting embedding, there are many disjoint "areas"
- If the distortion is small, the total "area" should be small



Approximation: A lower bound

- [Brunn-Minkowski inequality] For any $A, B \subset \mathbf{R}^2$

$$\sqrt{Vol(A \oplus B)} \geq \sqrt{Vol(A)} + \sqrt{Vol(B)}$$

Let $V(r) = \pi r^2$, and $\rho(\alpha) = V^{-1}(\alpha)$

For each leaf u, define C(u) = V(1/2)

For each non-leaf u, define:

$$C(u) = \sum_{i=1}^{k} \left(\sqrt{C(u_i)} + \sqrt{V(l(u)/4)} \right)^2$$

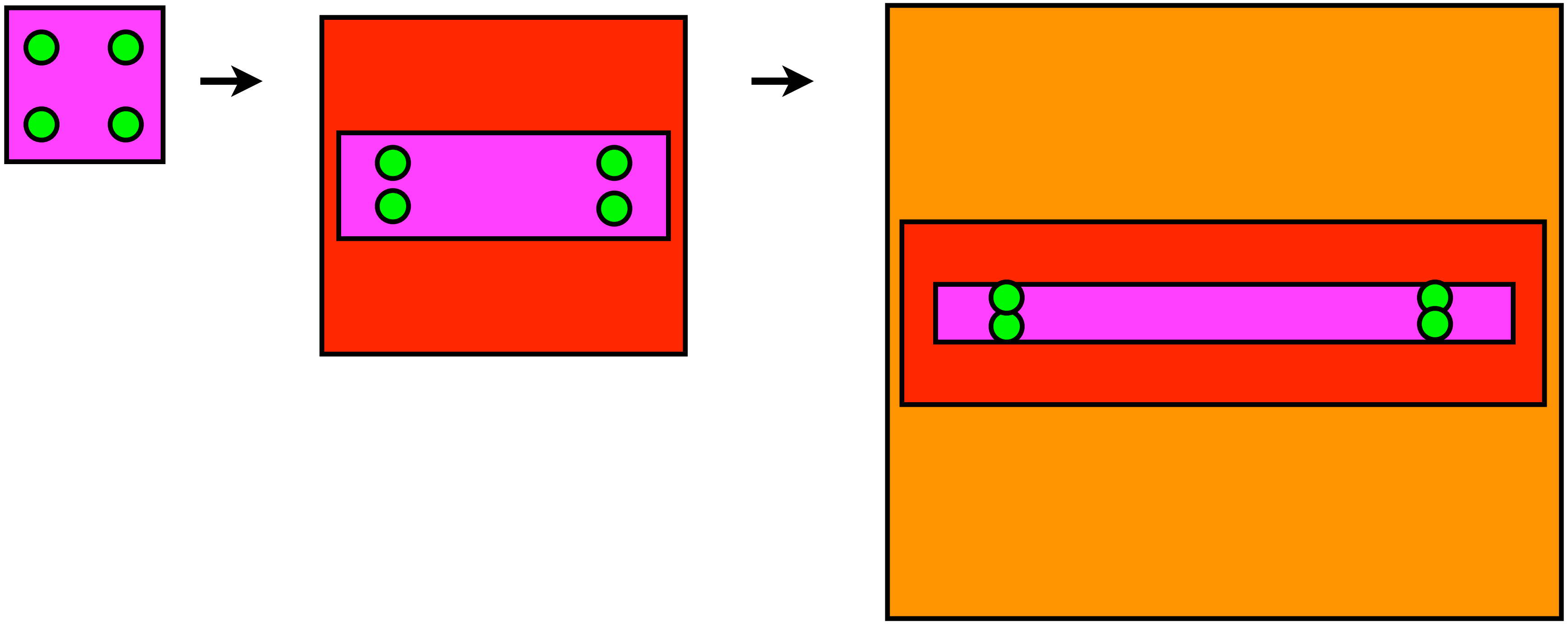
Lemma: For each u,

$$c \geq \rho(C(v))/l(v) - 1$$

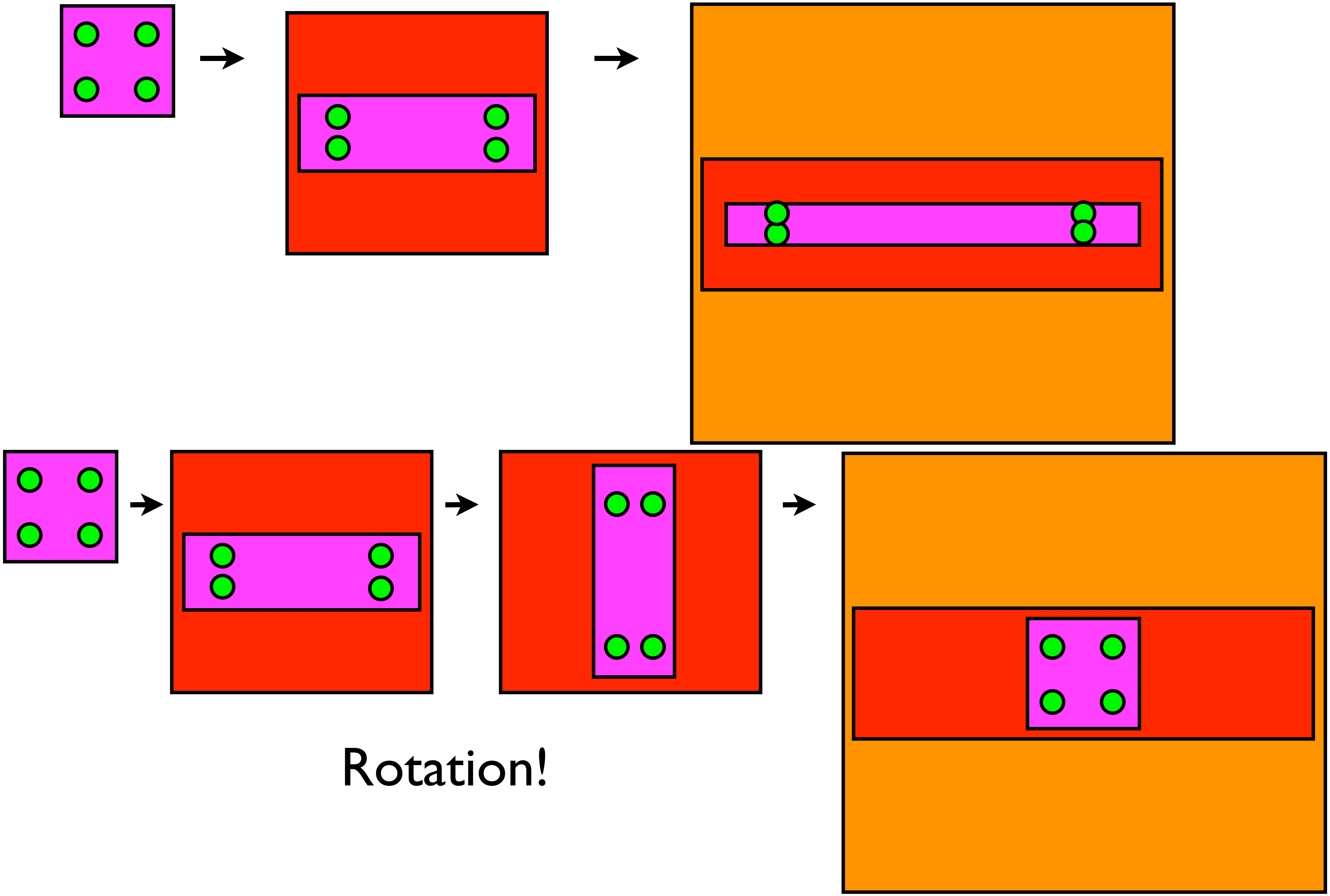
Approximation: The algorithm

- We will try to match the lower bound
- It suffices to embed each X_u inside a square with side length $\sqrt{C(u)}$
- We can do that if we stretch each square by a factor of c
- PROBLEM: The streching accumulates!

Approximation: The algorithm

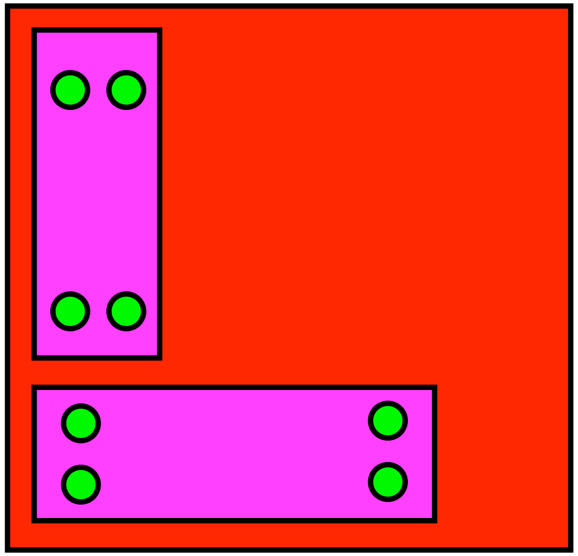


Approximation: The algorithm



Approximation: The algorithm

- We need to “synchronize” the rotations:



In this case rotation does NOT help!

Solution: Top-down preprocessing for calculation of the rotations

Approximation: The algorithm

- The final algorithm:
- Bottom-up: Compute the "volumes" C(u)
- Top-down: Compute the "rotations"
- Bottom-up: Compute the embedding
- Resulting distortion $O(c^3)$

THE END

Questions?