Ordinal Embeddings of Minimum Relaxation: General Properties, Trees, and Ultrametrics

Noga Alon (Tel Aviv University)
Mihai Bădoiu (MIT)
Erik D. Demaine (MIT)
Martín Farach-Colton (Rutgers University)
MohammadTaghi Hajiaghayi (MIT)
Anastasios Sidiropoulos (MIT)

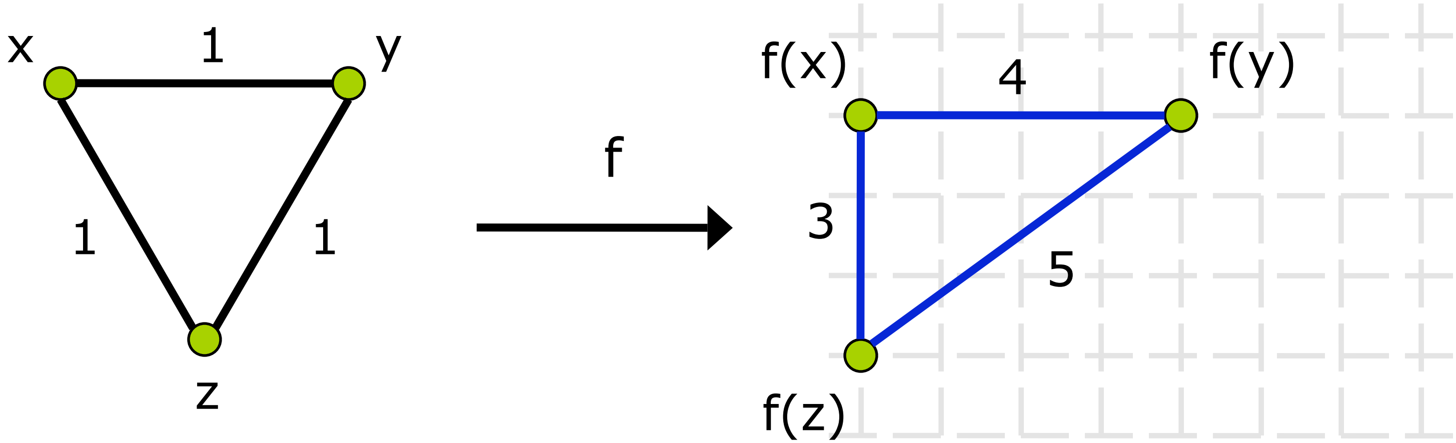
Embeddings of Metric Spaces

- Given a finite metric space (X,D)
  - $D(p,q)=0 \Leftrightarrow p=q$
  - $D(p,q)=D(q,p)$
  - $D(p,q) \leq D(p,r)+D(r,q)$
- Mapping $f:X \rightarrow Y$
- Distortion of f is:

$$\max_{p,q} \frac{D'(f(p),f(q))}{D(p,q)} \times \max_{p,q} \frac{D(p,q)}{D'(f(p),f(q))}$$

Goal: Minimize distortion

Metric Embedding - Example



distortion = 5·(1/3) = 5/3

Motivation

- Compact data representation
- Embedding into algorithmically good spaces (e.g. Euclidean spaces, trees)
- Visualization / Clustering

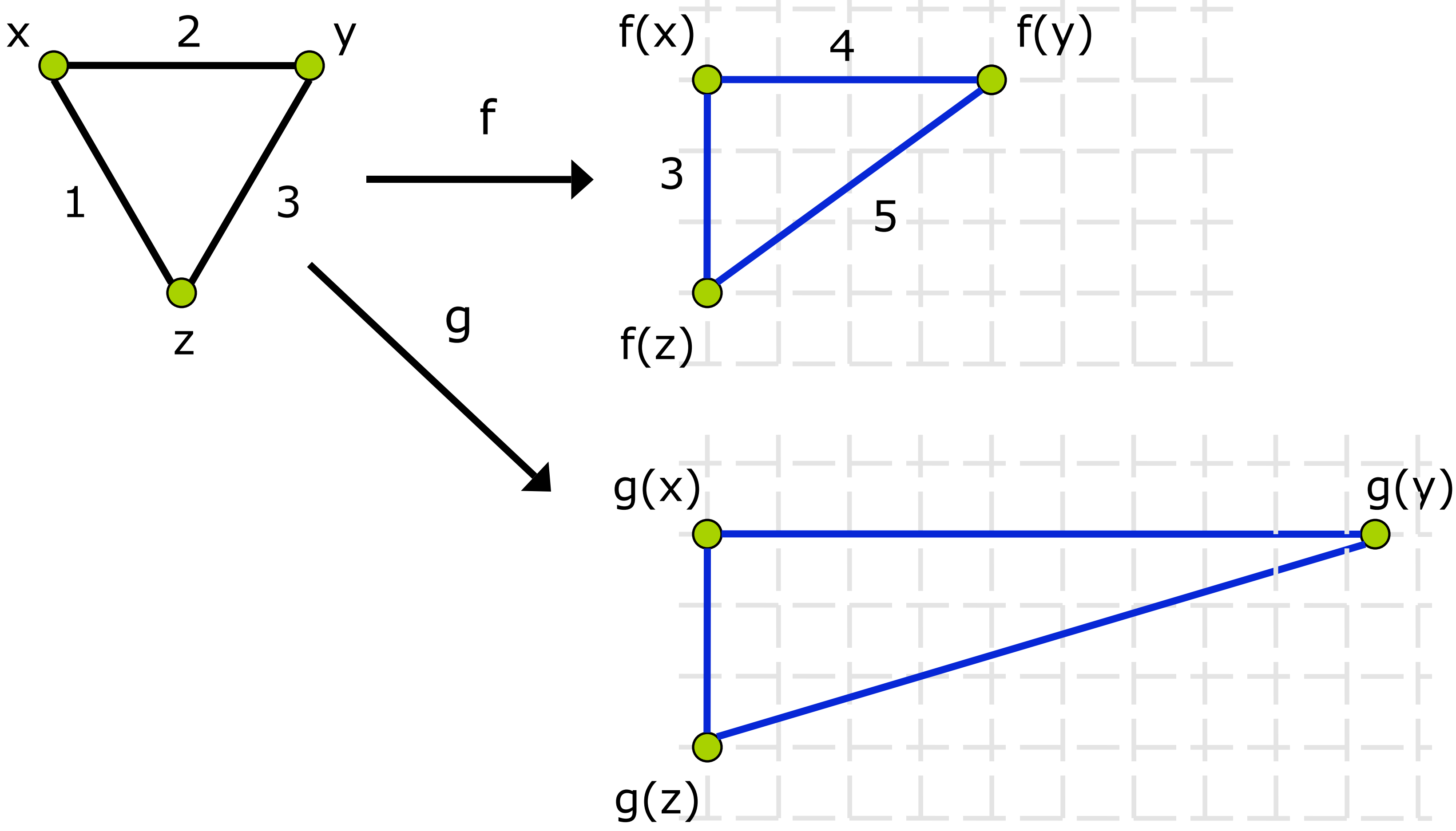
Results on Low-Distortion Embeddings

- Worst-case bounds
  - Any n-point metric into Euclidean space with O (log n) distortion. [Bourgain 1985]
  - Ω(log n) bound. [Linial, London, Rabinovich 1995]
- Approximation algorithms
  - Any n-point metric into ℓ_2 with OPT distortion. [Linial, London, Rabinovich 1995]
  - Unweighted graphs into line, with O(OPT^2), etc. [Bădoiu, Dhamdhere, Gupta, Rabinovich, Raecke, Ravi, S. 2005], also [Bădoiu,Indyk,Rabinovich,S. 2004]
  - General metrics into Trees (additive) [Agarwala, Bafna, Farach, Narayan, Paterson, Thorup 1999]

Ordinal Embeddings

- Relax constraints on embedded lengths:
  - Ignore exact distances
  - Require only the total order on the distances to match between source and target metrics
- Such an embedding called ordinal embedding
- "Normal" embedding called metric embeddings

Ordinal Embeddings - Example



Ordinal Embeddings – Motivation

- Sometimes order is all that matters
- Nearest neighbors
  - Preserved by ordinal embedding
- Visualization
  - Distinguish large from small distances.
  - Classical approach in Visualization/MDS in early 60s.

Known Results on Ordinal Embedding

- NP-hard to decide whether a distance matrix can be ordinally embedded into a tree metric [Shah & Farach-Colton 2004]
- A metric is an ultrametric iff it requires n-1 dimensions [Holman 1972]
- Every distance matrix on n points can be ordinally embedded into $(n-1)$-dimensional Euclidean space, and almost every distance matrix requires $\Omega(n)$ dimensions [Bilu & Linial 2004]

Relaxing ordinal embeddings

- Instead preserving the total order, preserve a partial order.

Question:

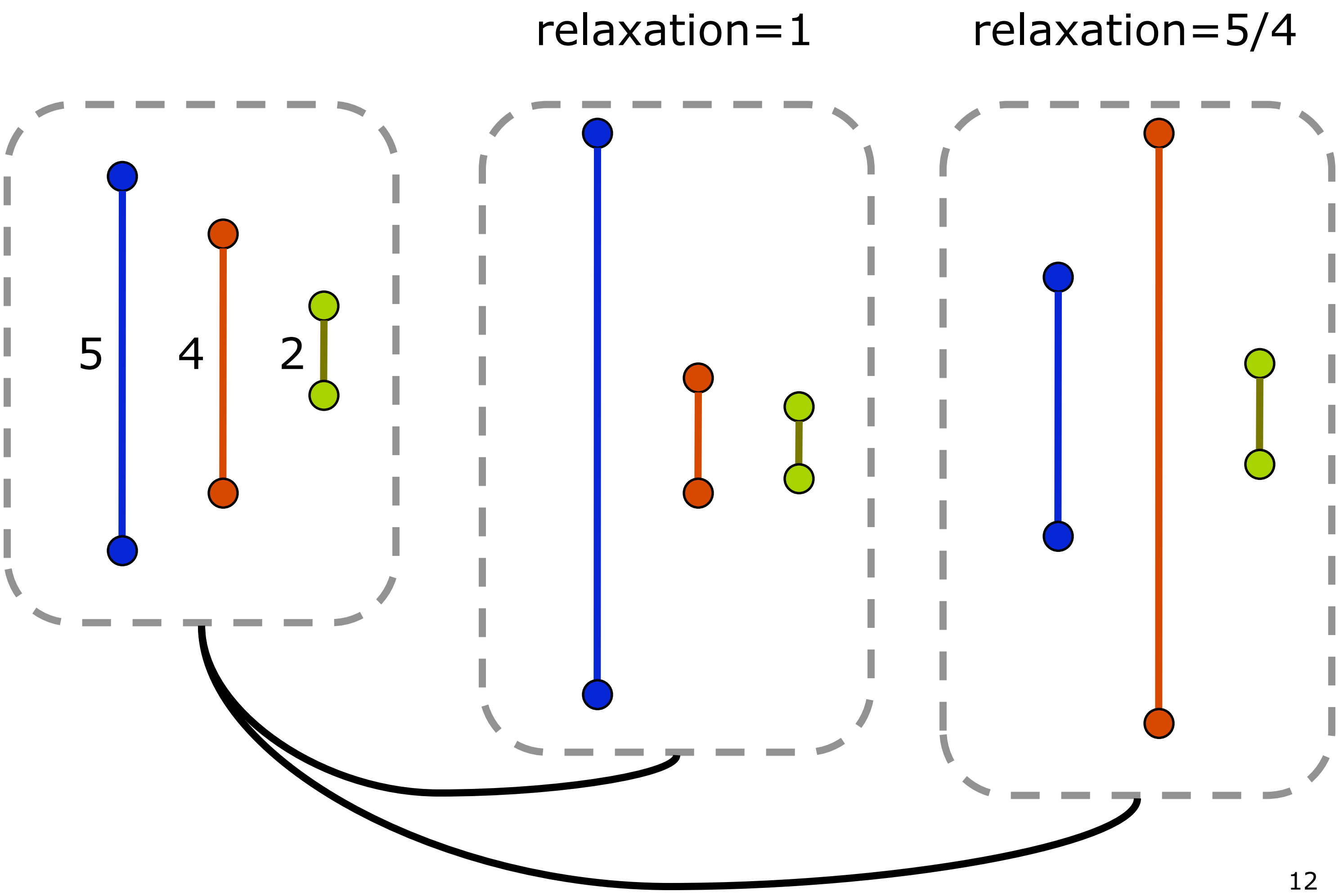
Which orders should we preserve?

Ordinal Relaxation

- Analog to metric distortion
- Embedding f has relaxation $\alpha \geq 1$ if
  - $\alpha \cdot D(x,y) < D(z,w) \Rightarrow D'(f(x),f(y)) < D'(f(z),f(w))$
- I.e., must preserve the order between distances that are different by a factor of more than α
- Note: $\alpha \leq c$

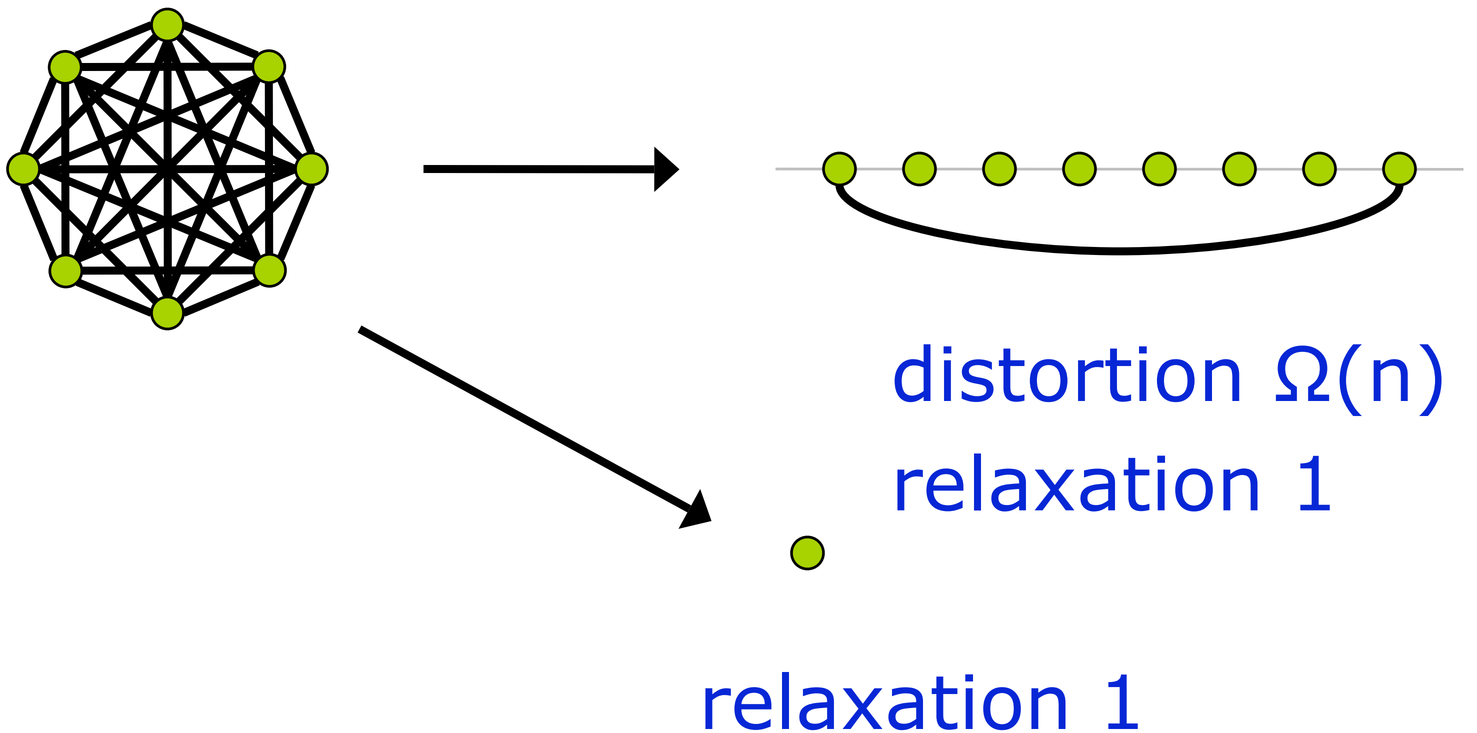
Goal: Minimize relaxation

Ordinal relaxation - Example



Tie breaking

- Uniform metric into the line:



Our Results

- When is it relaxation = distortion?
- Worst-case bounds of unweighted trees into d-dimensional Euclidean space
- O(1)-approximation algorithm for embedding unweighted trees into the line
- Ultrametrics into the line with relaxation 1
- OPT for embedding into ultrametrics

Our Results (cont.)

- Worst case relaxation for embedding into d-dimensional Euclidean space is at least

  $\log n/(\log d + \log\log n + O(1))$

- For d-dimensional ℓ_p space, for every even integer p

  $\log n/(\log d + \log(\log n + \log p) + O(1))$

- For d-dimensional ℓ_p space, for every odd integer p

  $\log n/(\log 2d^2 + 3d \log n + d \log p + O(d))$

- For d-dimensional ℓ_∞ space

  $\log n/(\log d + \log\log n + O(1))$

Lower bound for ℓ_2^d

- Let $P_1, \ldots, P_m$ be m polynomials of degree at most k, on t real variables. If $2m \geq t$, then the number of *sign-patterns* of $(P_1, \ldots, P_m)$ is at most $(8ekm/t)^t$. [Alon 1995]

- For every $g \geq 3$, $n \geq 3$, there are n-vertex graphs with at least $n^{1+1/g}/4$ edges, and girth at least g. [Erdős, Sachs 1963]

Lower bound for ℓ_2^d (cont.)

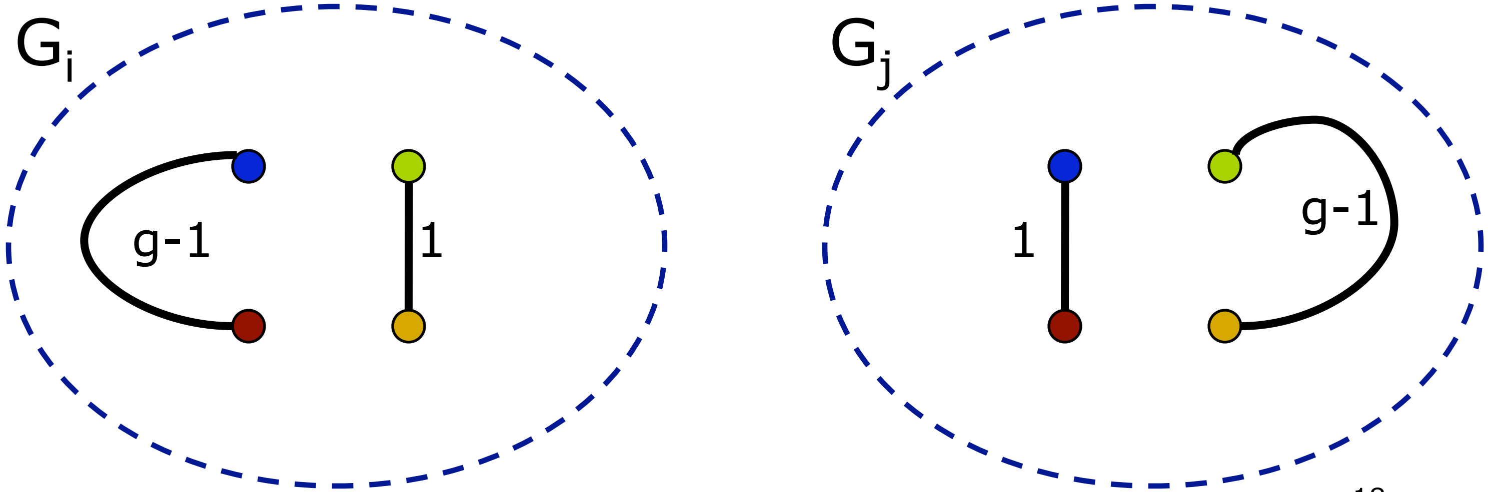
- In Euclidean embedding:
  - Each edge-edge order is specified by a quadratic equation.
  - There are $n^4/4$ such order polynomials on nd variables.
  - Therefore there are few possible orderings in our target space.

Lower bound for ℓ_2^d (cont.)

- Since there exists a dense high-girth graph, it has many subgraphs with of m/2 edges,

  $G_1, G_2, \ldots$

- By PHP, two such graphs must end up with same ordering.



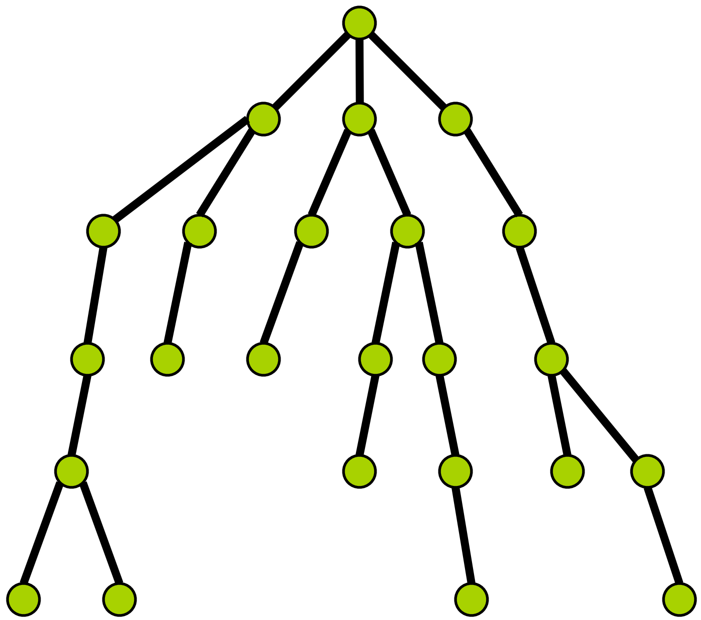
Lower bound for ℓ_2^d (cont.)

- Thus, relaxation >
  g - 1 = log n / (log d + loglog n + 5) - 1

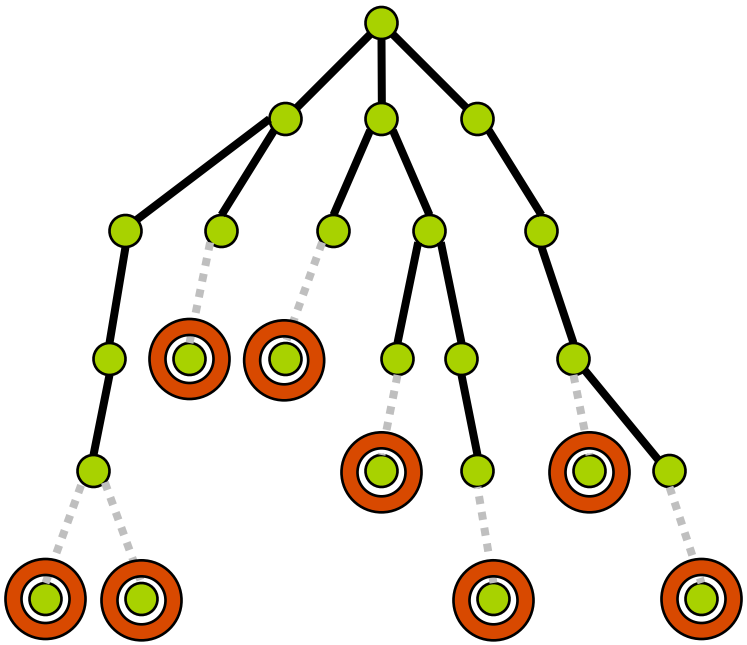
Unweighted Trees into R^d

- **Theorem**: Any unweighted tree can be embedded into d-dimensional Euclidean space with relaxation $\tilde{O}(n^{1/d})$.
- **Theorem**: There is a tree for which every embedding has relaxation $\Omega(n^{1/(d+1)})$.
- Any tree can be embedded into d-dimensional euclidean space with distortion $\tilde{O}(n^{1/(d-1)})$. [Gupta 2000]
- Any embedding of the n-star has distortion $\Omega(n^{1/d})$.

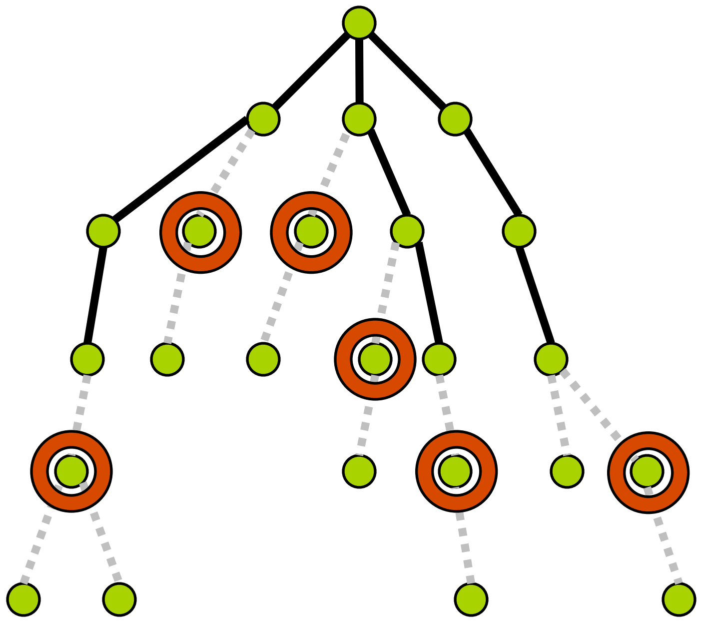
Unweighted Trees into R^d



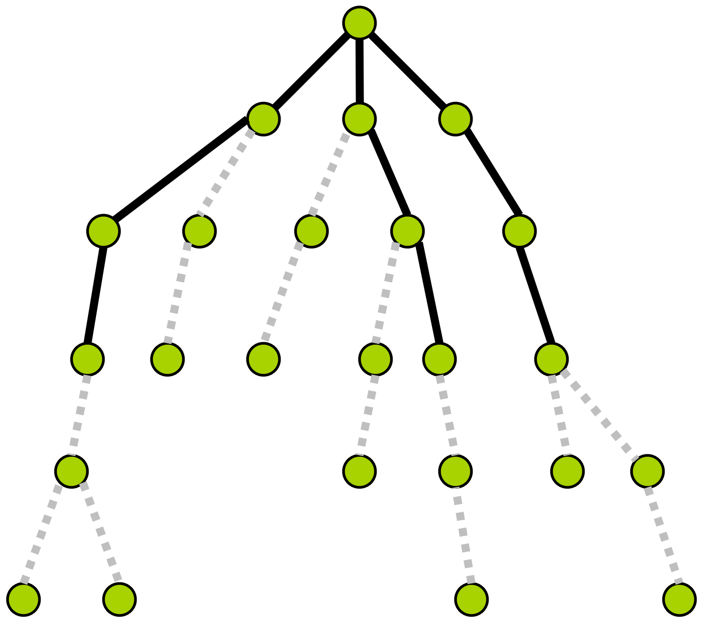
Unweighted Trees into R^d



Unweighted Trees into R^d



Unweighted Trees into R^d



Repeat $n^{1/d}$ times.

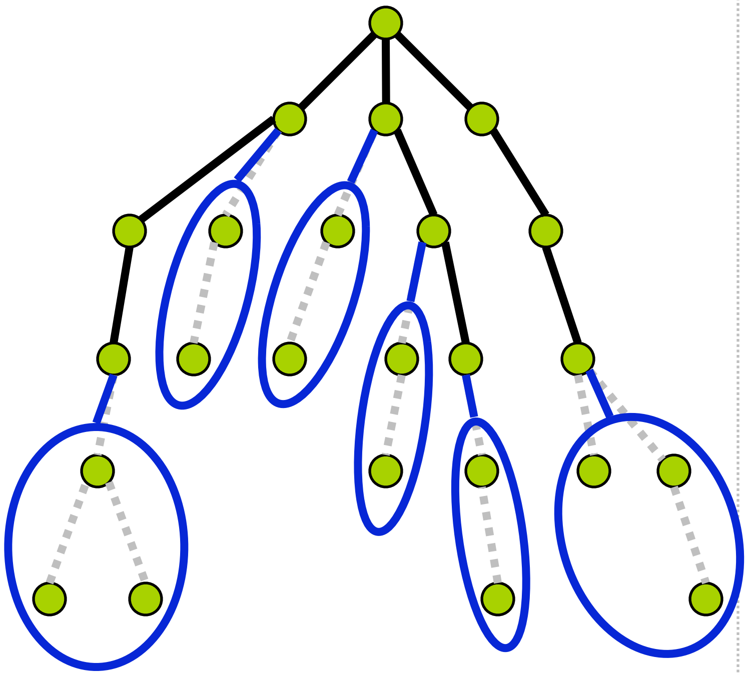
$\Rightarrow$

$O(n^{(d-1)/d})$ leaves.

Using [Gupta 2000]

$\tilde{O}(n^{1/d})$ distortion.

Unweighted Trees into R^d



Map every subtree into its root

$\Rightarrow$

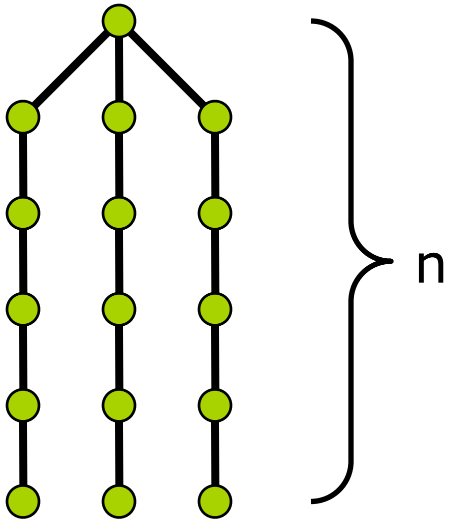
$\tilde{O}(n^{1/d})$ relaxation

Approximation Algorithm for Unweighted Trees into the Line

- **Theorem:** There is a 3-approximation poly-time algorithm for minimizing relaxation of ordinal embedding of an unweighted tree into line.
- In contrast, best approximation algorithm for minimum-distortion embedding is $\tilde{O}(n^{O(1)})$-approximation. [Bădoiu,Dhamdhere, Gupta,Rabinovich,Raecke,Ravi,S. 2005], also [Bădoiu,Indyk,Rabinovich,S. 2004]

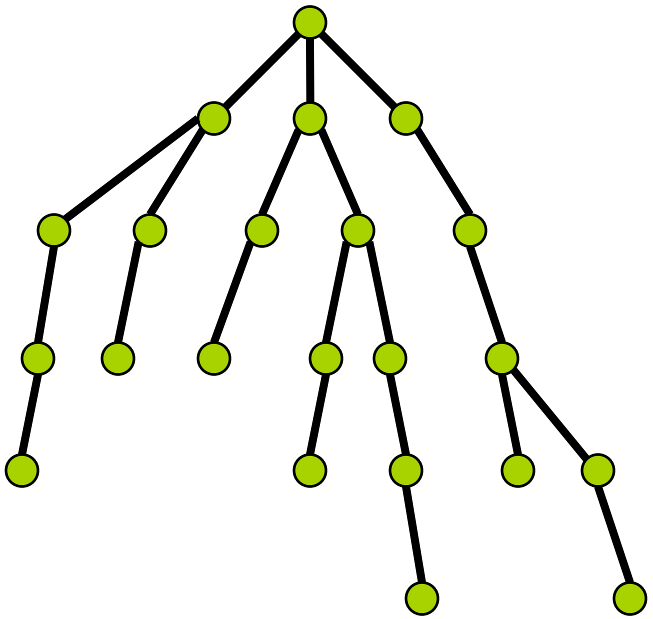
Approximation Algorithm for Unweighted Trees into the Line

- Lower bound: 3-spider



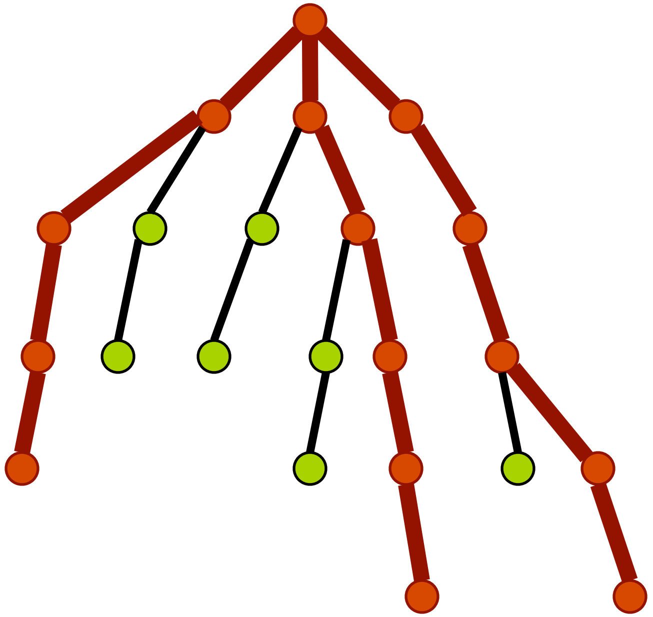
$\Rightarrow$ relaxation $\Omega(n)$

Approximation Algorithm for Unweighted Trees into the Line



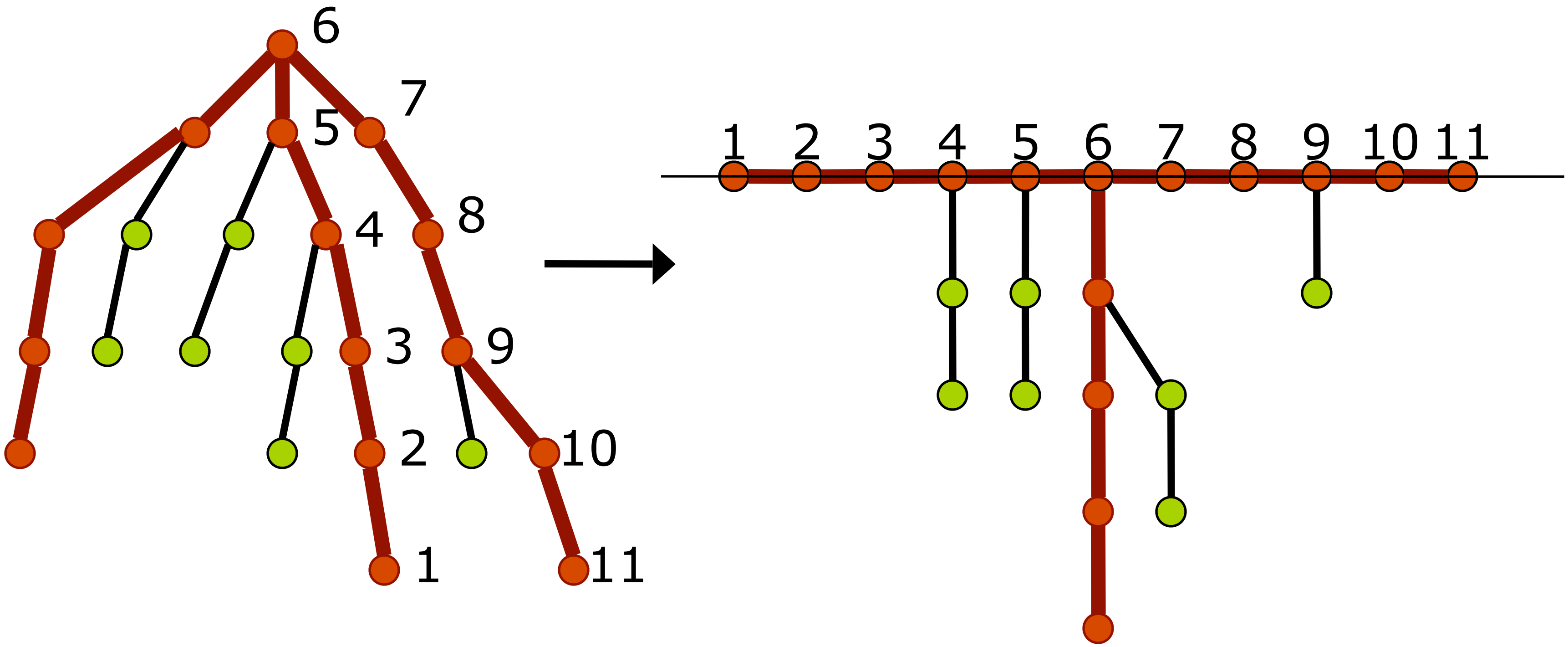
Approximation Algorithm for Unweighted Trees into the Line

- Find longest 3-spider



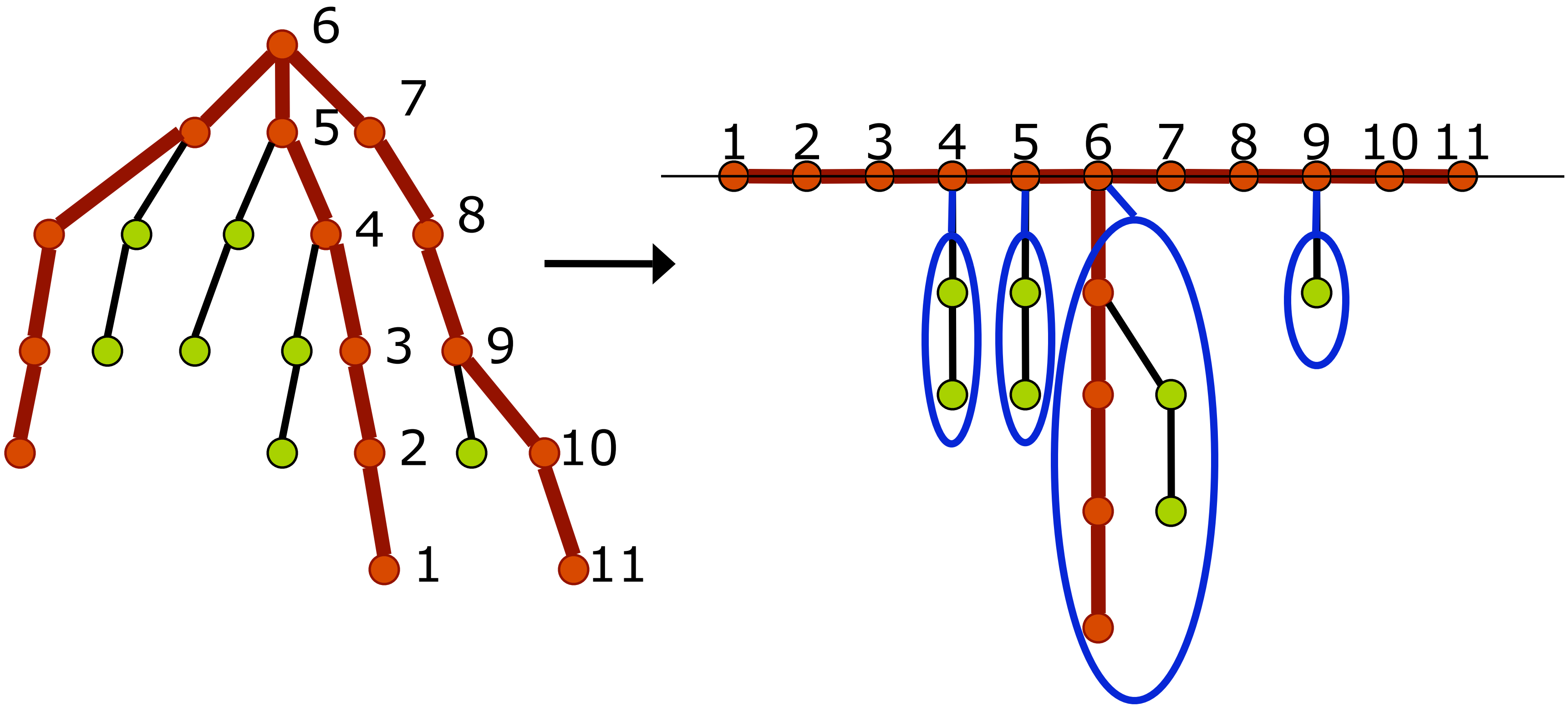
Approximation Algorithm for Unweighted Trees into the Line

- Find longest 3-spider, embed longest hair



Approximation Algorithm for Unweighted Trees into the Line

- Find longest 3-spider, embed longest legs



Map remaining subtrees into their roots

Conclusions – Open problems

- Worst case relaxation for embedding into O(log n)-dimensional Euclidean space is

  $\Omega(\log n / \log\log n)$, and $O(\log n)$.
- Dimensionality reduction in ℓ_1?
- Approximation algorithms