How to complete a doubling metric

Anupam Gupta
Carnegie Mellon University

work with Kunal Talwar (Microsoft)

appeared at LATIN 2008

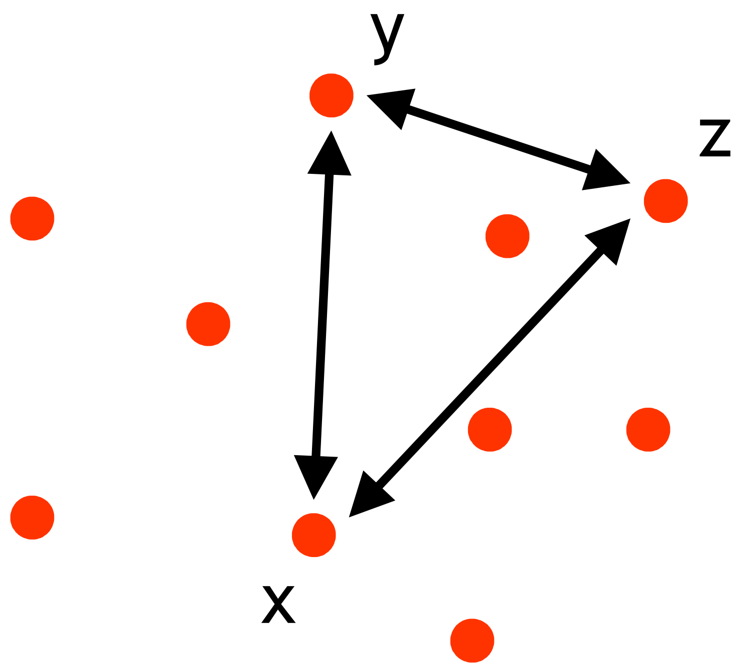
Metric space $M = (V, d)$

set V of points

symmetric distances $d(x,y)$

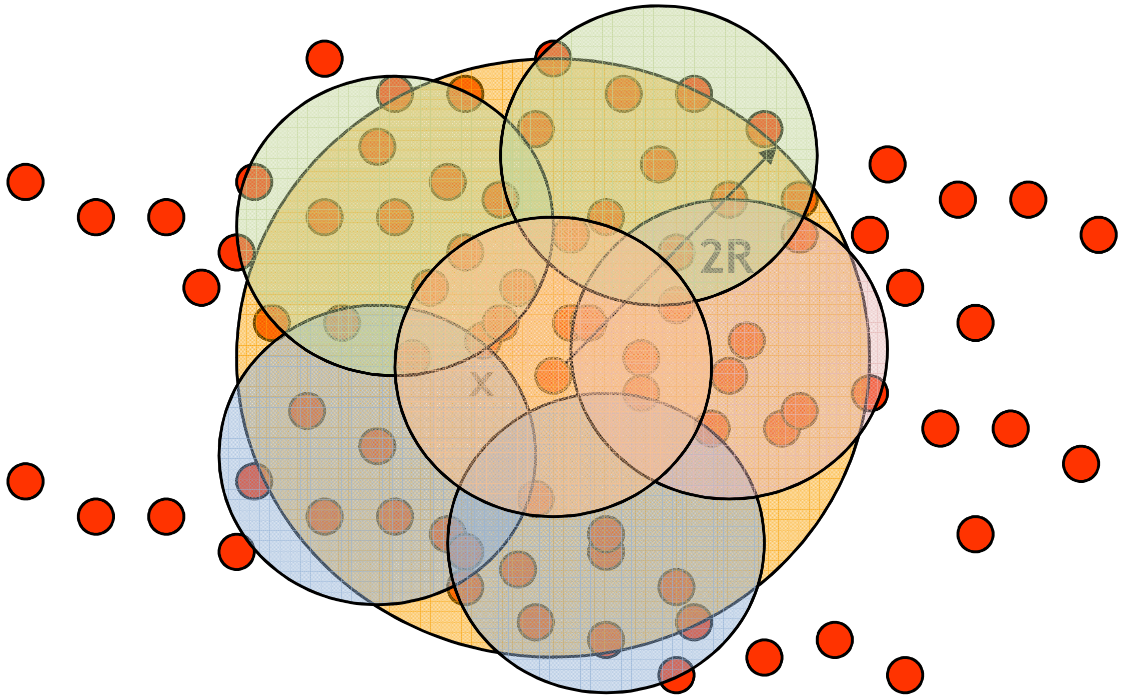
triangle inequality

$$d(x,y) \leq d(x,z) + d(z,y)$$



doubling dimension of a metric space

Dimension $\dim_D(M)$ is the smallest k such that
every ball $B(x, 2R)$ with x in V
can be covered by 2^k balls $B(y, R)$ for y in V.



facts about doubling

Euclidean space $(\Re^k, |\cdot|_p)$ has doubling dimension $\approx$ k

The notion of doubling dimension behaves smoothly
under metric distortion

definition (almost) closed under taking submetrics

Turns out to be super-useful as a notion of metric "complexity"

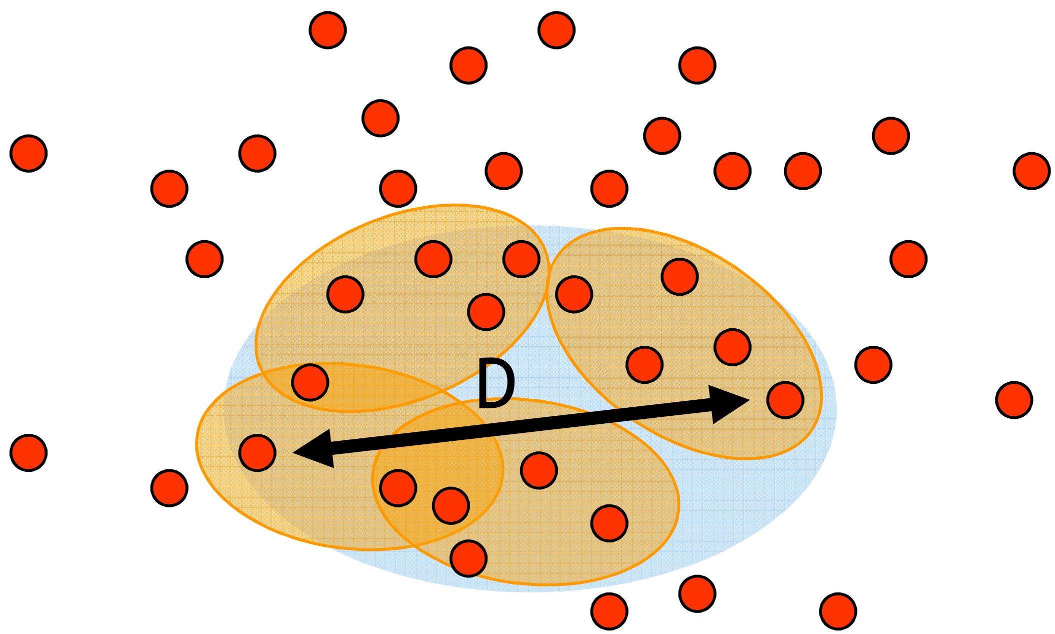
jargon: "doubling" = family of metrics with doubling dimension
bounded by some absolute constant c independent of n.

the doubling dimension

Dimension $\dim_D(M)$ is the smallest k such that

every set S with diameter D_S

can be covered by 2^k sets of diameter $\frac{1}{2}D_S$



a property of doubling

Suppose a metric (X,d) has doubling dimension k.

If any subset $S \subseteq X$ of points has
all inter-point distances lying between δ and Δ

then $|S| \leq (2\Delta/\delta)^k$

Proof: recursively apply the definition…

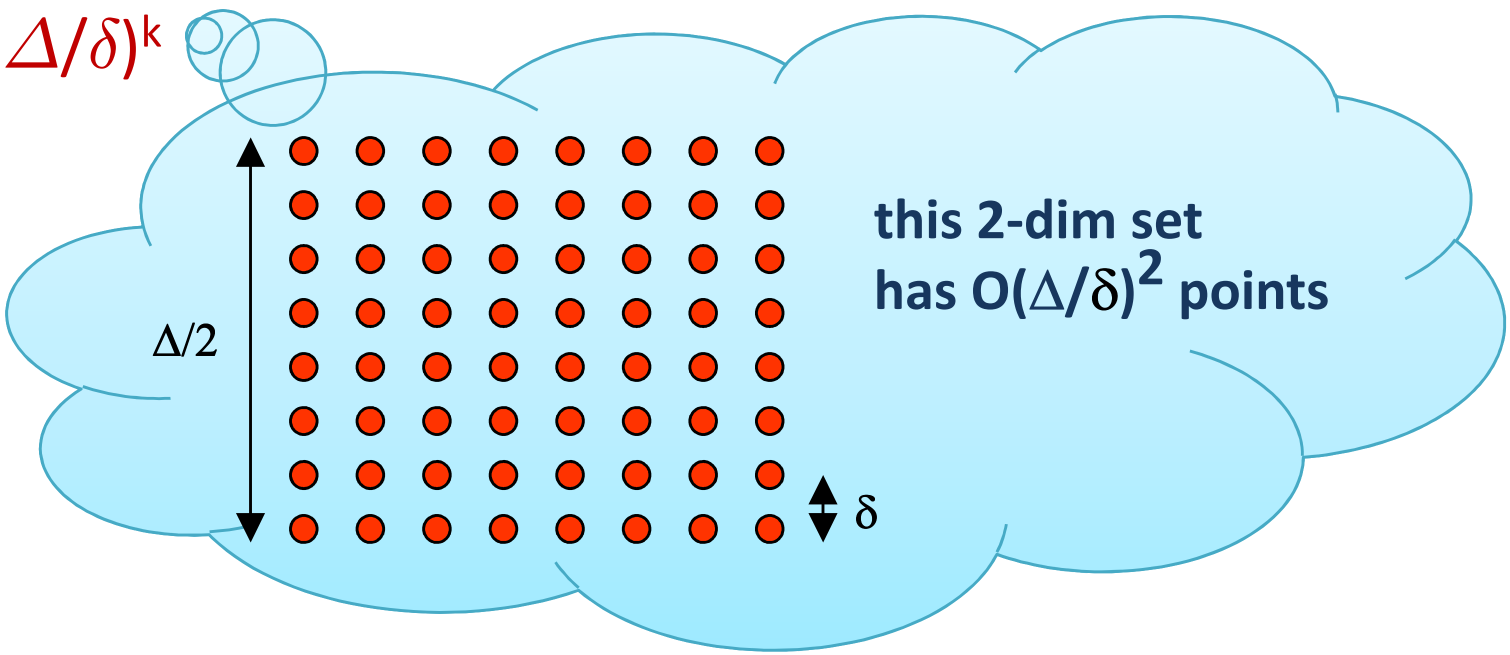
a property of doubling

Suppose a metric (X,d) has doubling dimension k.

If any subset $S \subseteq X$ of points has

all inter-point distances lying between δ and Δ

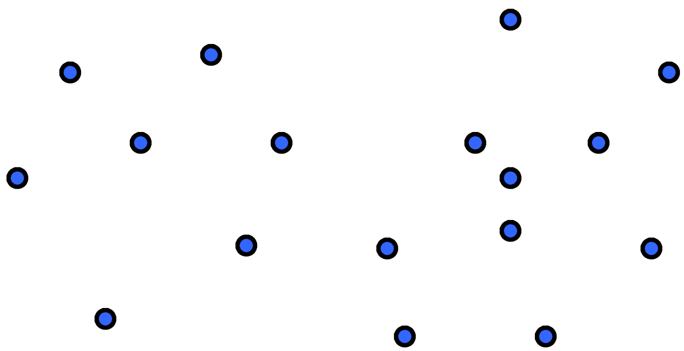
then $|S| \leq (2\Delta/\delta)^k$



alternate characterization

Uniform metric: All non-zero distances equal to R

2-uniform metric: All non-zero distances in [R,2R]

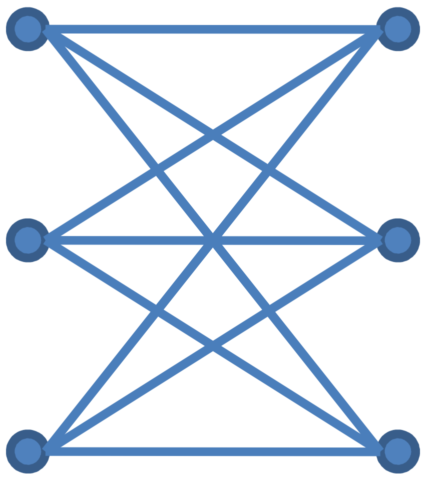


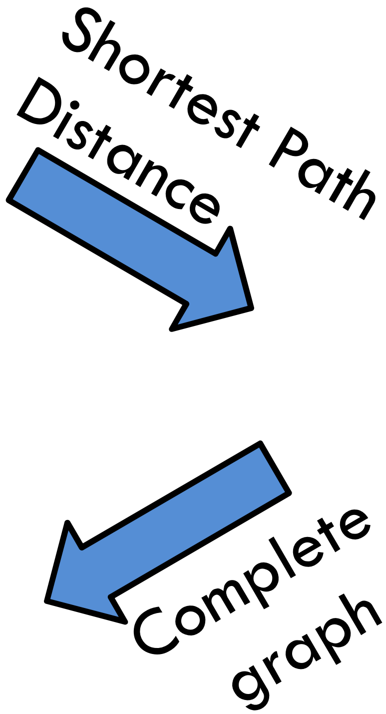
Doubling Dimension k

iff

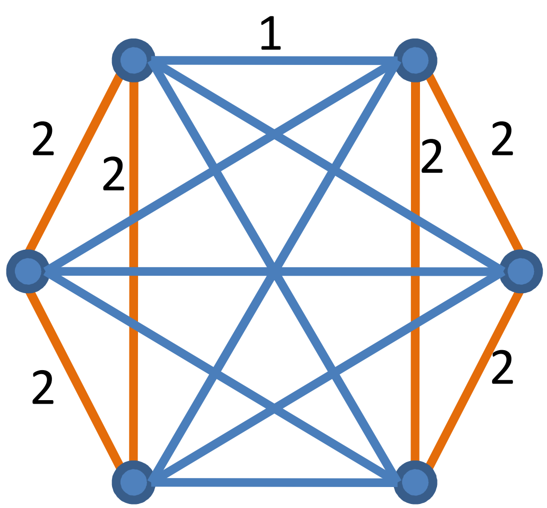
largest 2-uniform submetric has $\approx 2^{O(k)}$ points

graphs and metrics





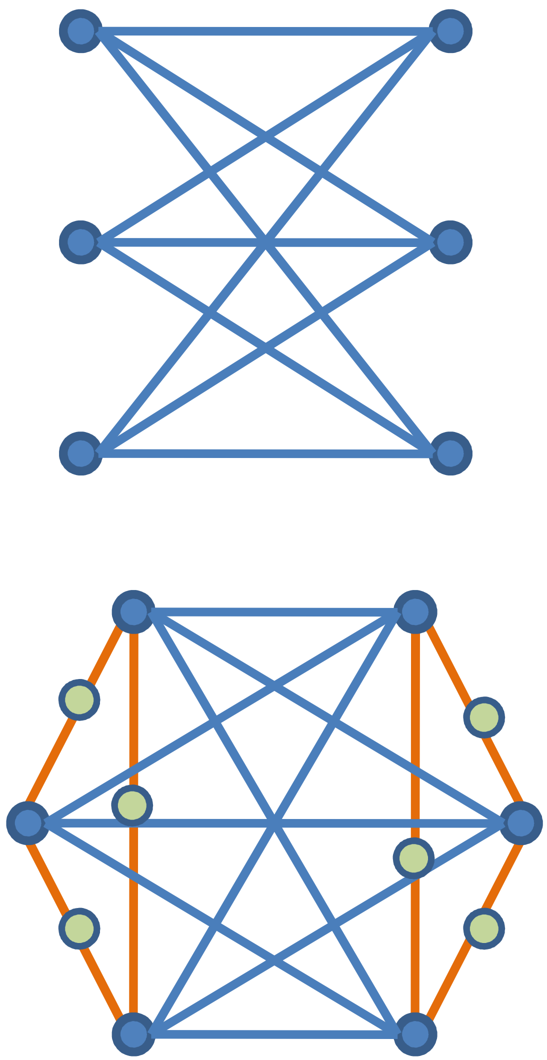
0	2	2	1	1	1
	0	2	1	1	1
		0	1	1	1
			0	2	2
				0	2
					0

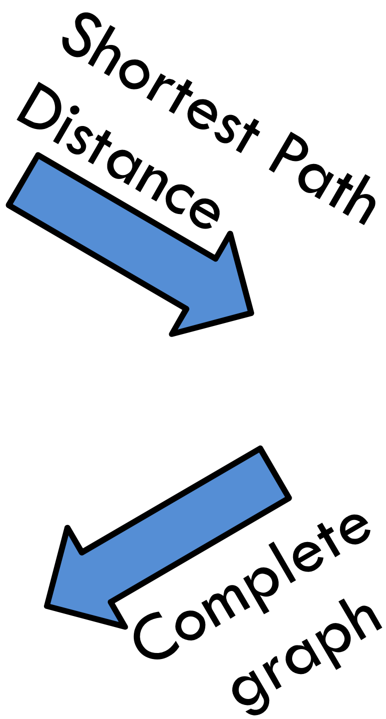


Graphs

Distance functions

graphs and metrics





$$\begin{bmatrix} 0 & 2 & 2 & 1 & 1 & 1 \\ & 0 & 2 & 1 & 1 & 1 \\ & & 0 & 1 & 1 & 1 \\ & & & 0 & 2 & 2 \\ & & & & 0 & 2 \\ & & & & & 0 \end{bmatrix}$$

Graphs

Distance functions

multiple representations…

two kinds of metric properties

intrinsic or “geometric” or simply “metric”

properties that depend merely on interpoint distances

e.g.:

the metric has doubling dimension at most 50.

or the metric embeds isometrically into normed space N

or the metric satisfies the 4-point condition

for any four points $i, j, k, l \in V$, $d_{ij} + d_{kl} \leq \max\{d_{ik} + d_{jl}, d_{il} + d_{jk}\}$

two kinds of metric properties

intrinsic or “geometric” or simply “metric”

properties that depend merely on interpoint distances

representational or “graph theoretic” or “topological”

properties related to the graphs that generate the metric

e.g., the metric can be generated by a tree

or generated by a planar graph…

their interplay

sometimes things work out perfectly:

a metric satisfies the 4-point condition

for any four points $i, j, k, l \in V$, $d_{ij} + d_{kl} \leq \max\{d_{ik} + d_{jl}, d_{il} + d_{jk}\}$

iff

it is representable by a (graph-theoretic) tree

their interplay

sometimes things work out perfectly:

any metric that is generated by an outerplanar graph

embeds into ℓ_1 isometrically

their interplay

sometimes the problems are harder:

can we find geometric properties that characterize representability by planar graphs?

or:

given a planar graph, how well can it embed into ℓ_1?

the high-level question

What are connections between graph structure and the properties of metrics generated by these graphs?

a more specific question…

given a doubling metric,

can it be represented as a graph

duh…

that is unweighted

of course…

such that the doubling dimension
of the resulting graph metric is also small?

hmm, let me think…

the q., rephrased

Given a metric (V,d) with doubling dimension k

Is there an ***unweighted*** graph (V',E') with V ⊆ V' such that

a) its shortest-path metric d'(.,.) agrees with d(.,.) when restricted to V × V

   i.e., d'(x,y) = d(x,y) for all x,y in V × V

b) the doubling dimension of d' is close to k

why do we care?

Unweighted graphs are simpler to argue about.

E.g., doubling tree metrics embed into constant-dimensional Euclidean space with constant distortion. [GKL ’03, LNP ‘06]

Proof for unweighted trees:	2 pages
Proof for weighted trees:	20 pages
Having this “reduction” theorem:	priceless

the q., rephrased

Given a metric (V,d) with doubling dimension k

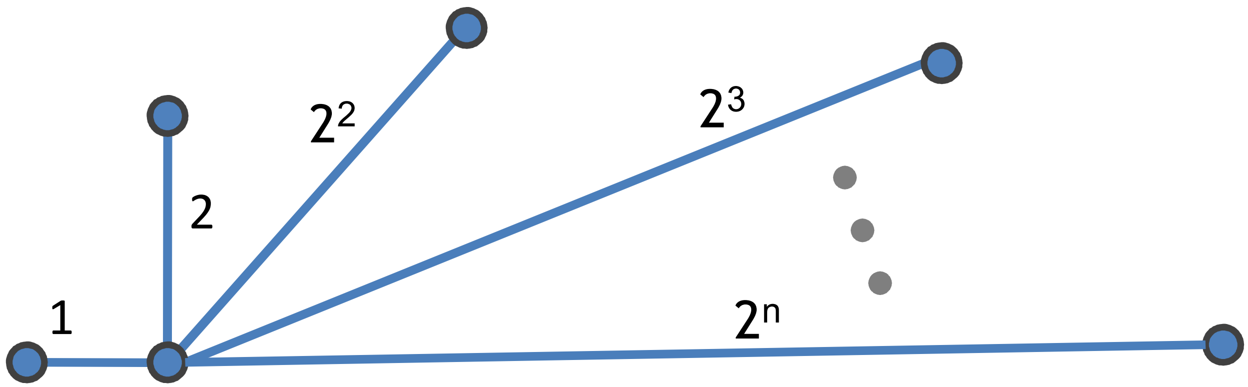
Is there an ***unweighted*** graph (V',E') with $V \subseteq V'$ such that

a) its shortest-path metric d'(.,.) agrees with d(.,.) when restricted to $V \times V$

i.e., $d'(x,y) = d(x,y)$ for all x,y in $V \times V$

b) the doubling dimension of d' is close to k

simple answer: **no**



how about a little distortion?

Given a metric (V,d) with doubling dimension k

Is there an ***unweighted*** graph (V',E') with $V \subseteq V'$ such that

a) its shortest-path metric $d'(.,.)$ *almost* agrees with $d(.,.)$ when restricted to $V \times V$

i.e., $d'(x,y) \approx d(x,y)$ for all x,y in $V \times V$

b) the doubling dimension of d' is close to k

our results$_{(1)}$

Given a metric (V,d) with doubling dimension k

there is an unweighted graph (V',E') with $V \subseteq V'$ such that

- Distances in (V',E') are within $(1+\epsilon)$ of d
- Doubling dimension of (V',E') is $O(k \log \epsilon^{-1})$

“completing” a metric

Given a graph, view each edge of length ℓ_e as a continuous segment of length ℓ_e

We find a graph G = (V’,d’) representing metric (V,d) such that even when we complete it to get Conv(G), we still have the bound on the dimension.

our results$_{(1)}$

Theorem 1:

Given a metric (V,d) with doubling dimension k

there is a weighted graph $G' = (V',E')$ with $V \subseteq V'$ such that

- Distances in G' are within $(1+\epsilon)$ of d
- Doubling dimension of $\mathrm{Conv}(G')$ is $O(k \log \epsilon^{-1})$

our results$_{(2)}$

Theorem 2:

Given a tree metric $T = (V,d)$ with doubling dimension k

there is a weighted tree $T' = (V',E')$ with $V \subseteq V'$ such that

– Distances in T' are within $(1+\epsilon)$ of d

– Doubling dimension of $Conv(T')$ is $O(k + \log\log \epsilon^{-1})$

our results$_{(3)}$

Lower bounds:

Dimension increase of loglog ϵ^{-1} for trees is tight.

Dimension blowup of log ϵ^{-1} for general metrics (sort of) tight.*

q: what is the correct bound for general metrics?

* if we restrict ourselves to metrics that don't have extra "Steiner" points

in the rest of the talk…

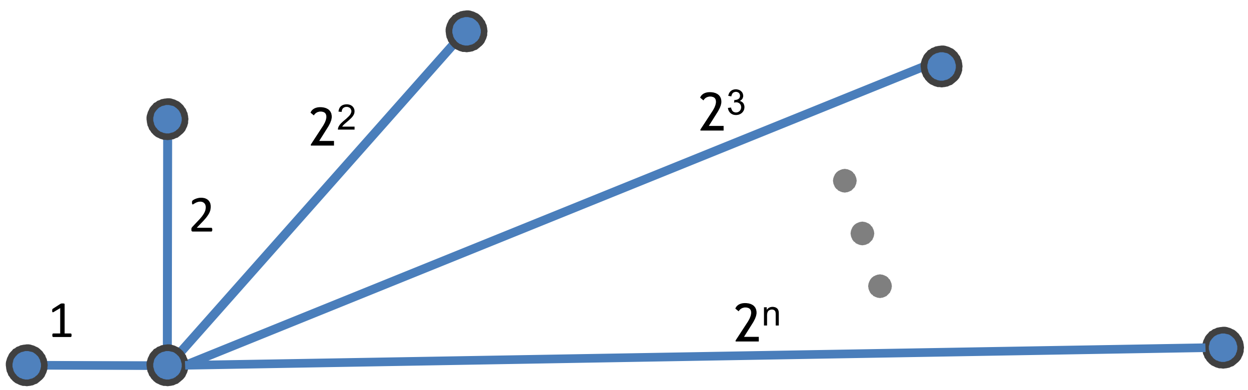
The lower bound for trees.

A structure theorem.

The upper bound for trees.

Outline of the general upper bound.

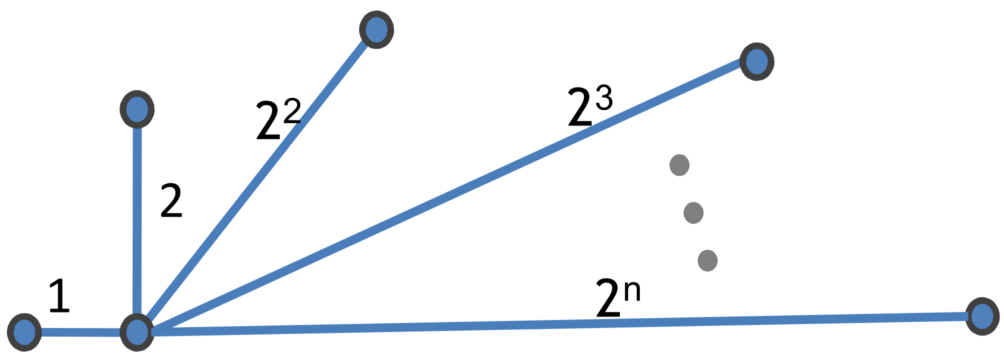
the lower bound example



This metric has doubling dimension O(1).

Any unweighted graph representing this metric to within distortion $(1+\epsilon)$ has doubling dimension $\Omega(\log\log \epsilon^{-1})$

the lower bound example



in the rest of the talk…

The lower bound for trees.

A structure theorem.

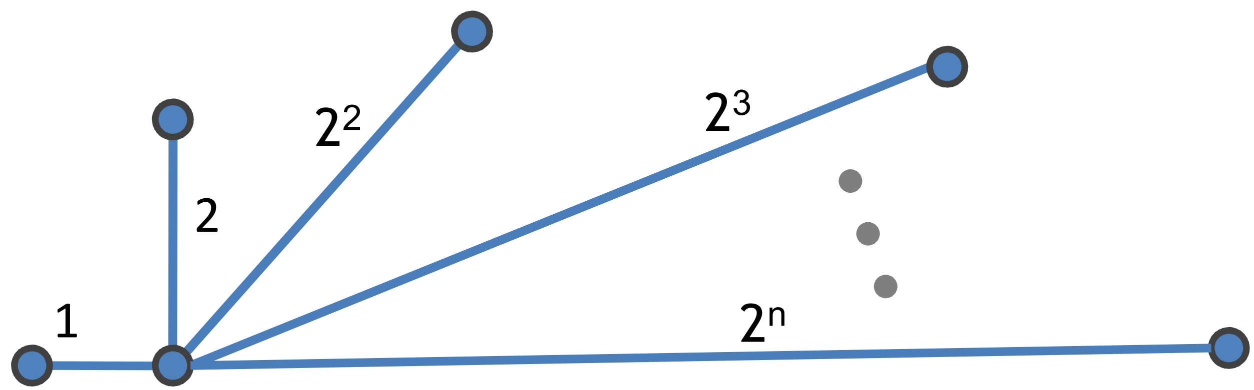
The upper bound for trees.

Outline of the general upper bound.

a structure theorem

Long edge: Given a weighted graph $H=(V_H,E_H)$, a vertex $x \in V_H$ and a radius $R>0$, an edge $\{v,w\}$ is *long* w.r.t. x,R if

- $d(x,v) \leq R$
- $\ell(e) > R$



Structure Theorem:

the convex completion conv(H) has doubling dimension $\Theta(k)$ iff at most 2^k long edges for any x,R. (Assume $\dim(H) \leq k$.)

the proof (sketch)

the convex completion conv(H) has doubling dimension $\Theta(k)$ iff at most 2^k long edges for any x,R.   (Assume dim(H) ≤ k.)

the proof (sketch)

the convex completion conv(H) has doubling dimension $\Theta(k)$ iff at most 2^k long edges for any x,R. (Assume dim(H) ≤ k.)

so, can now redefine goal

Goal: Find a weighted graph $H = (V_H, E_H)$ with $V \subseteq V_H$ such that

- Shortest path metric of H within $(1+\epsilon)$ of d
- H has only a few long edges for any node x, radius R

in the rest of the talk…

The lower bound for trees.

A structure theorem.

The upper bound for trees.

Outline of the general upper bound.

recall the result…

Theorem 2:

Given a tree metric $T = (V,d)$ with doubling dimension k

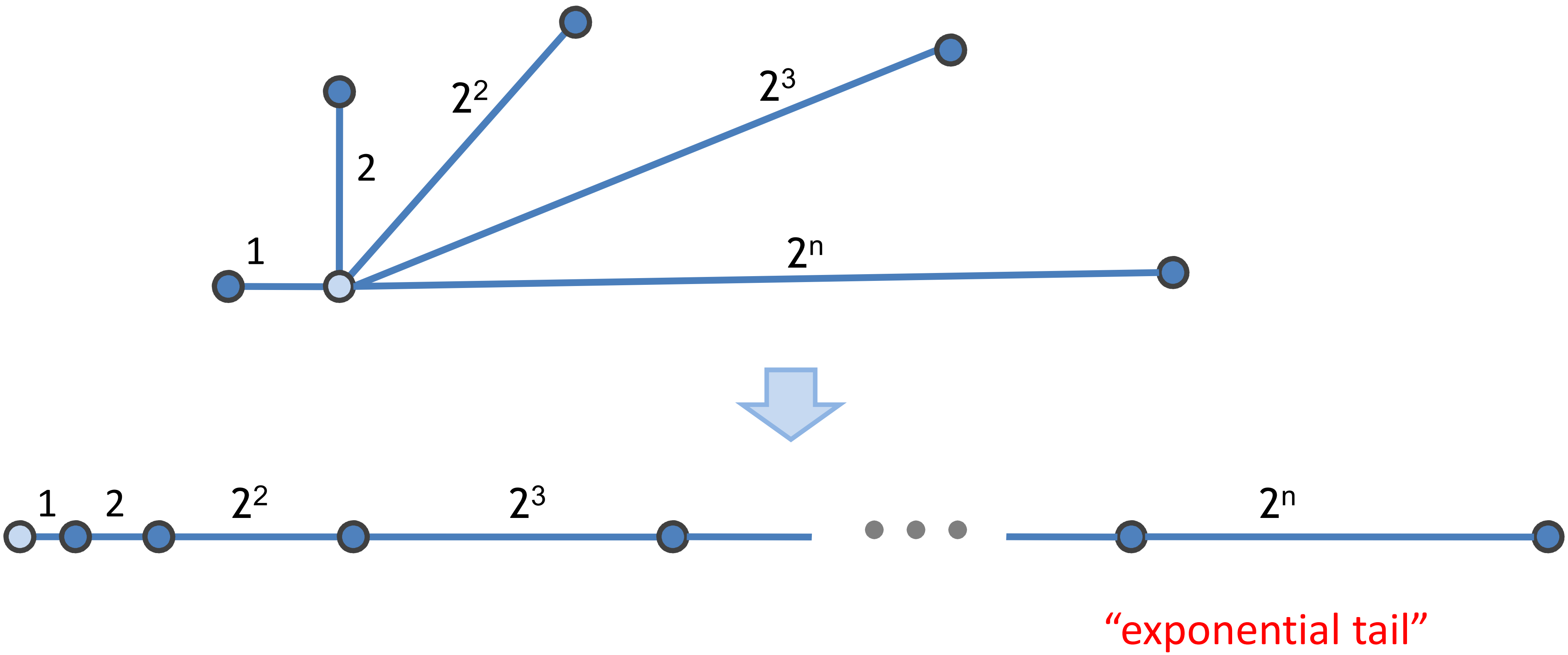
there is an weighted tree $T' = (V',E')$ with $V \subseteq V'$ such that

- Distances in T' are within $(1+\epsilon)$ of d

- Doubling dimension of $Conv(T')$ is $O(k + \log \log \epsilon^{-1})$

equivalently, number of long edges in T' is $2^{O(k)} \times O(\log \epsilon^{-1})$

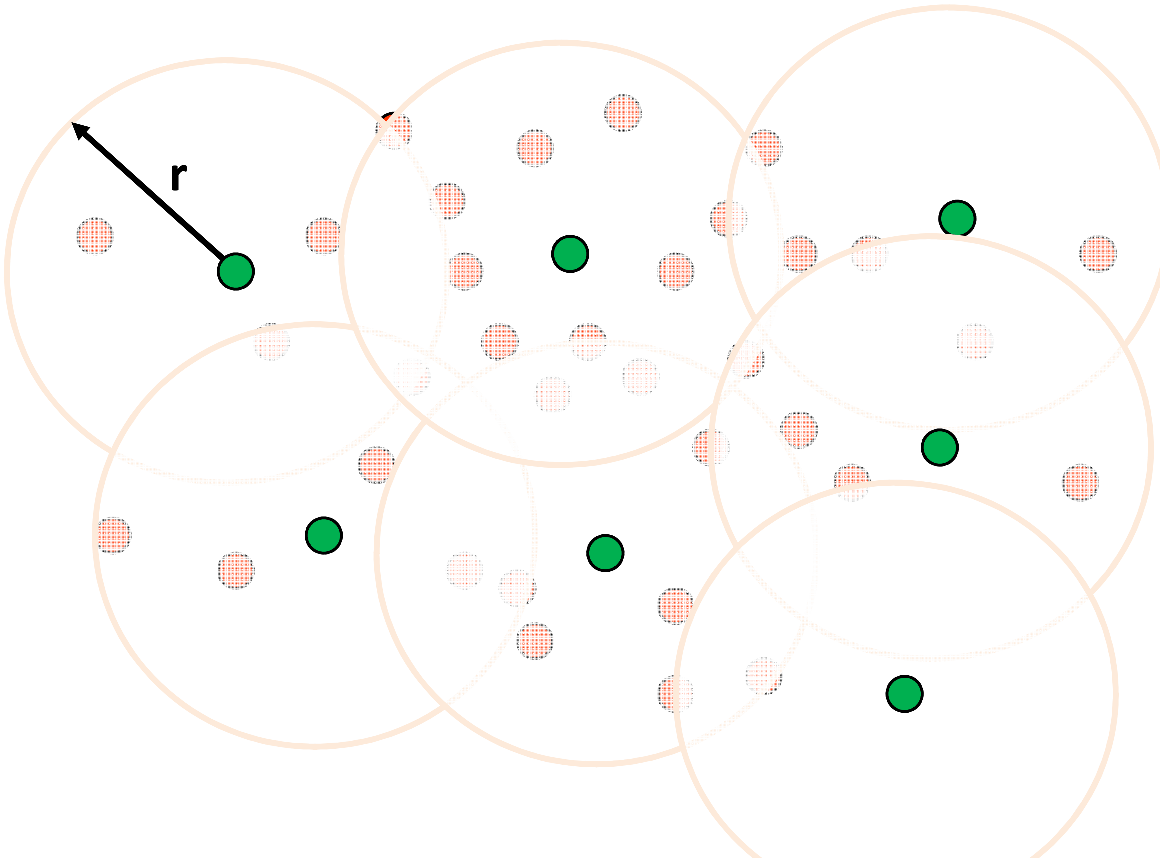
e.g.: exponential weighted star



standard tool: nets

Nets: A set of points N is an r-net of V if

- $d(u,v) \geq r$ for any $u,v \in N$
- For every $w \in V \setminus N$, there is a $u \in N$ with $d(u,w) < r$



standard tool: nets

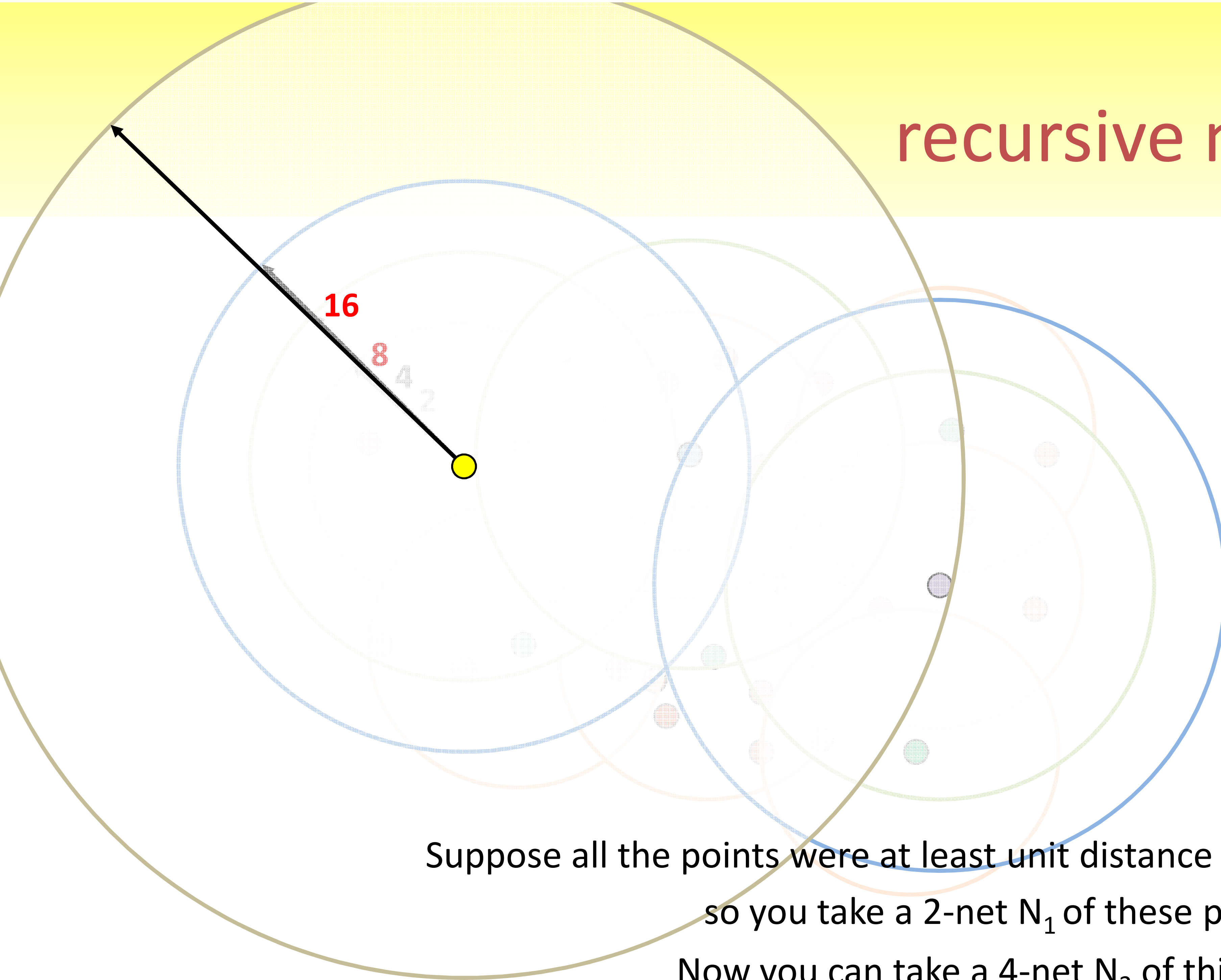
Nets: A set of points N is an r-net of V if

- $d(u,v) \geq r$ for any $u,v \in N$
- For every $w \in V \setminus N$, there is a $u \in N$ with $d(u,w) < r$

Fact: If a metric has doubling dim k and N is an r-net

$\Rightarrow B(x,2r) \cap N \leq O(1)^k$

recursive nets



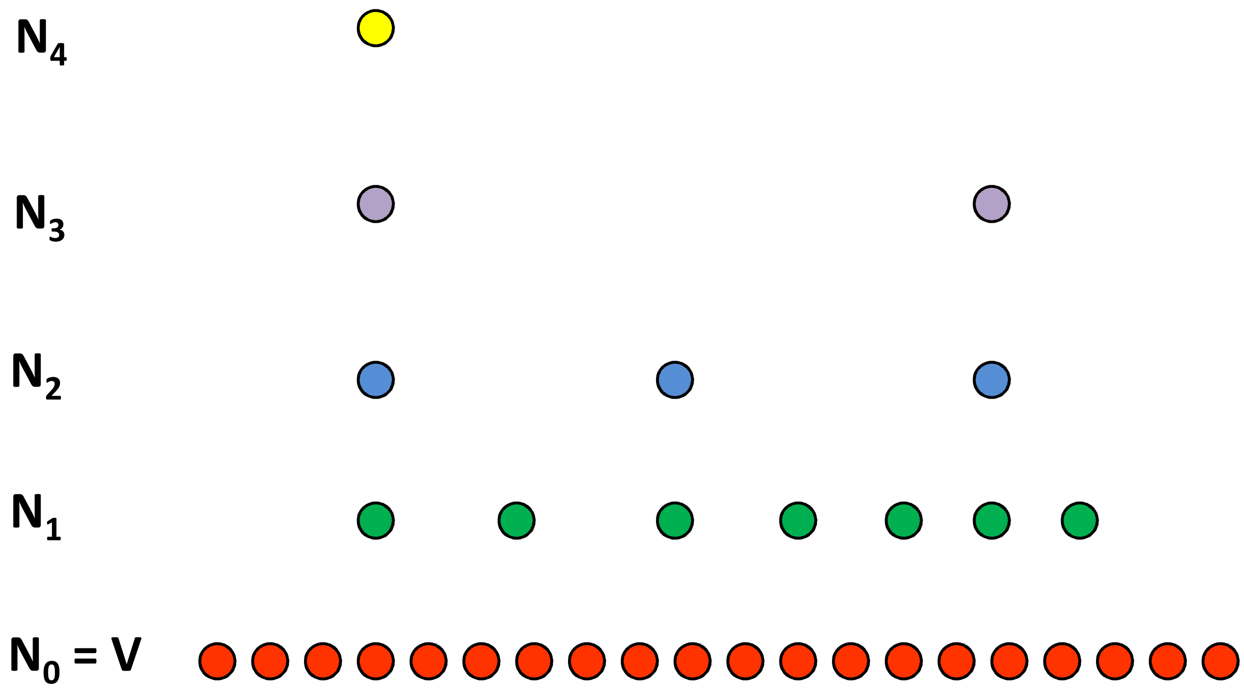
Suppose all the points were at least unit distance apart

so you take a 2-net N_1 of these points

Now you can take a 4-net N_2 of this net

And so on…

recursive nets



N_t is a 2^t-net of the set N_{t-1}

$\Rightarrow N_t$ is a 2^{t+1}-net of the set V (almost)

Algorithm for trees

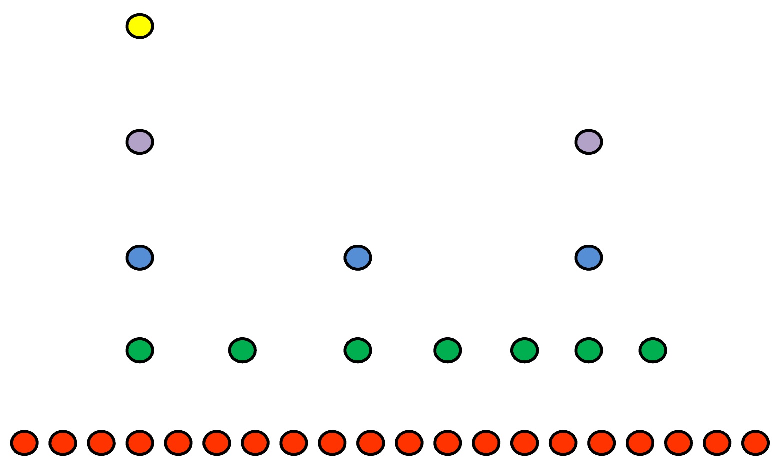
Algorithm for Trees

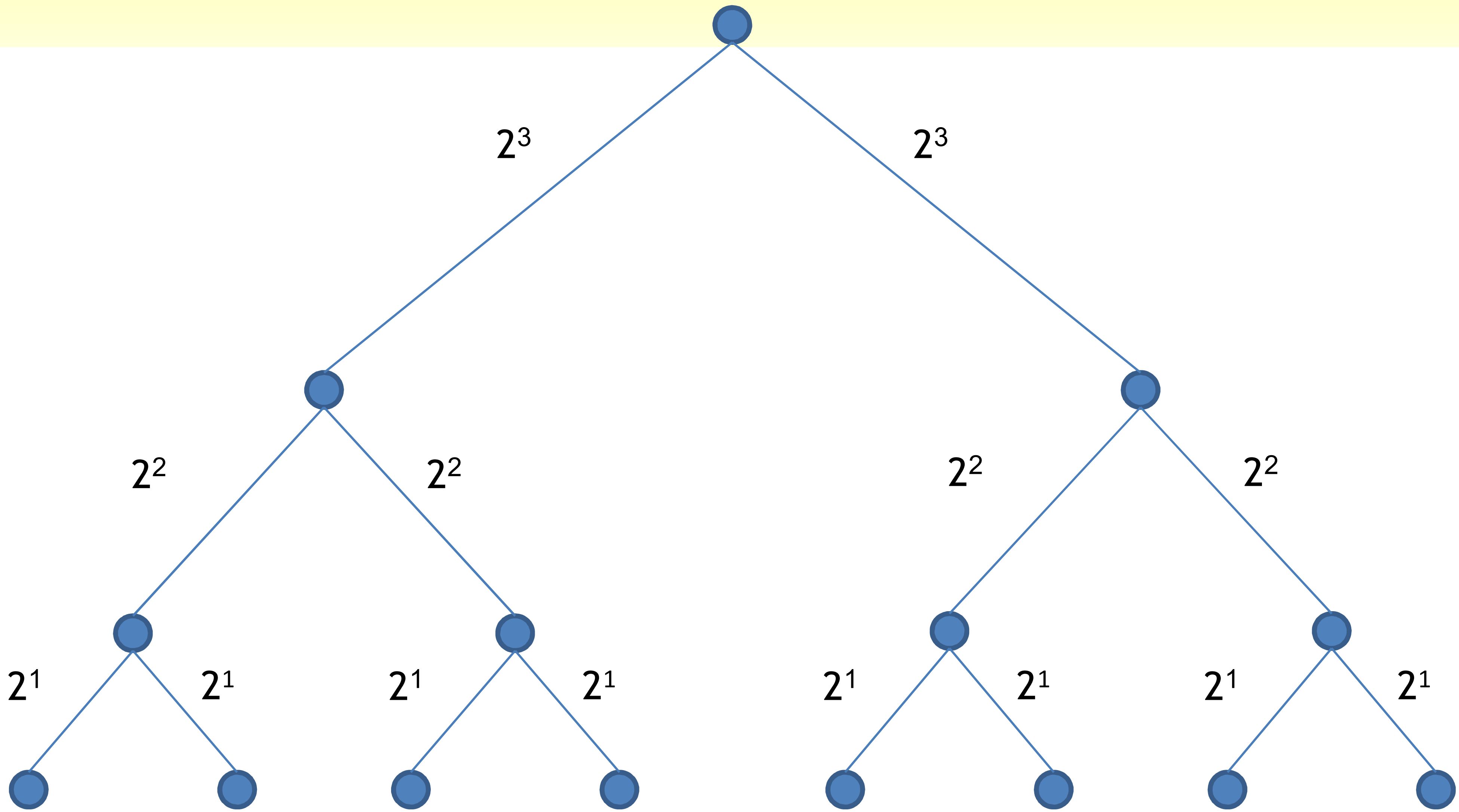
Construct the recursive nets $\{N_t\}$ for the tree metric.

For each value of t

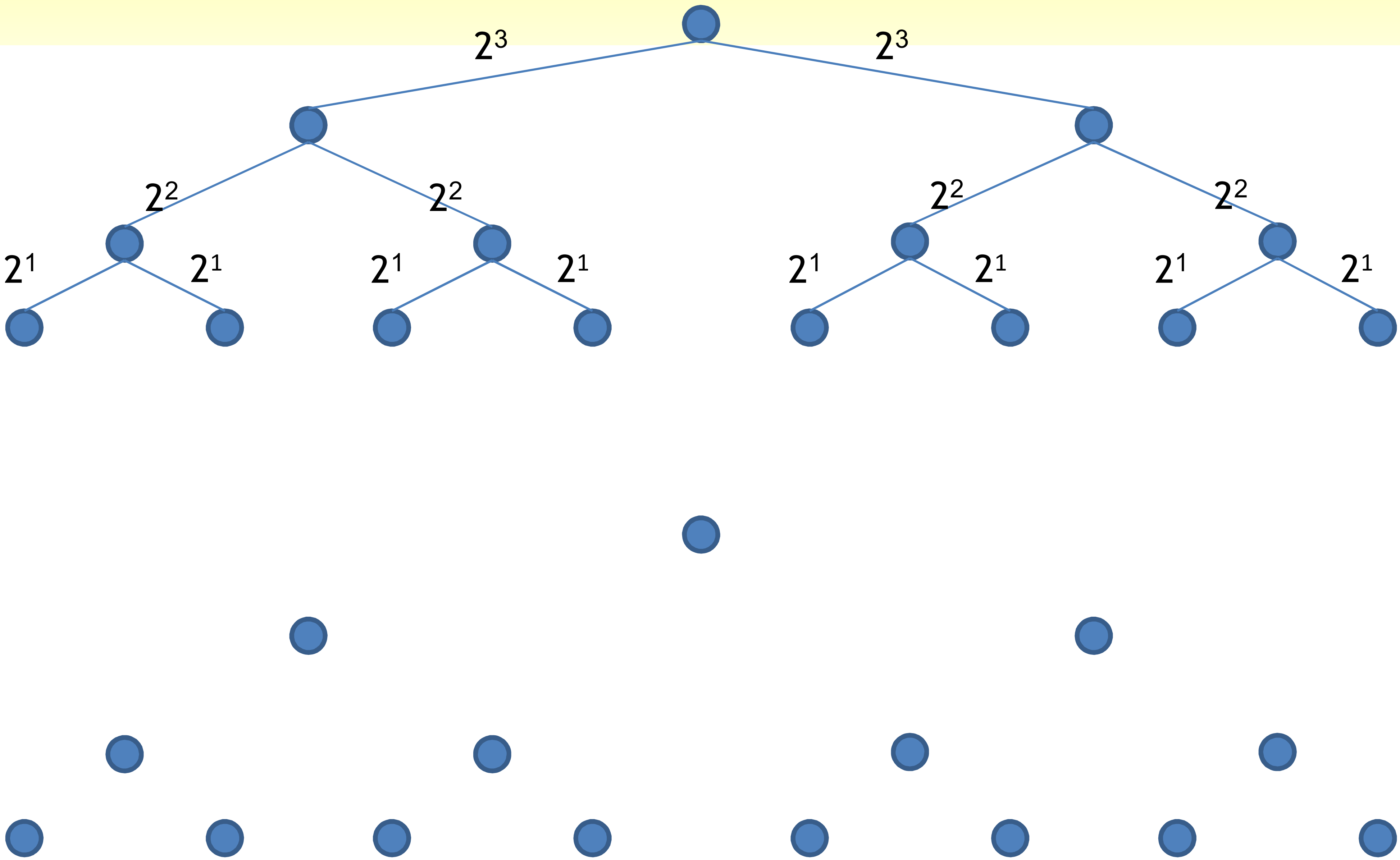
Attach an exponential tail with t nodes to vertices in $N_t \setminus N_{t+1}$

Move edges of length $2^t/\epsilon$ to tails of "parent" nodes in N_t

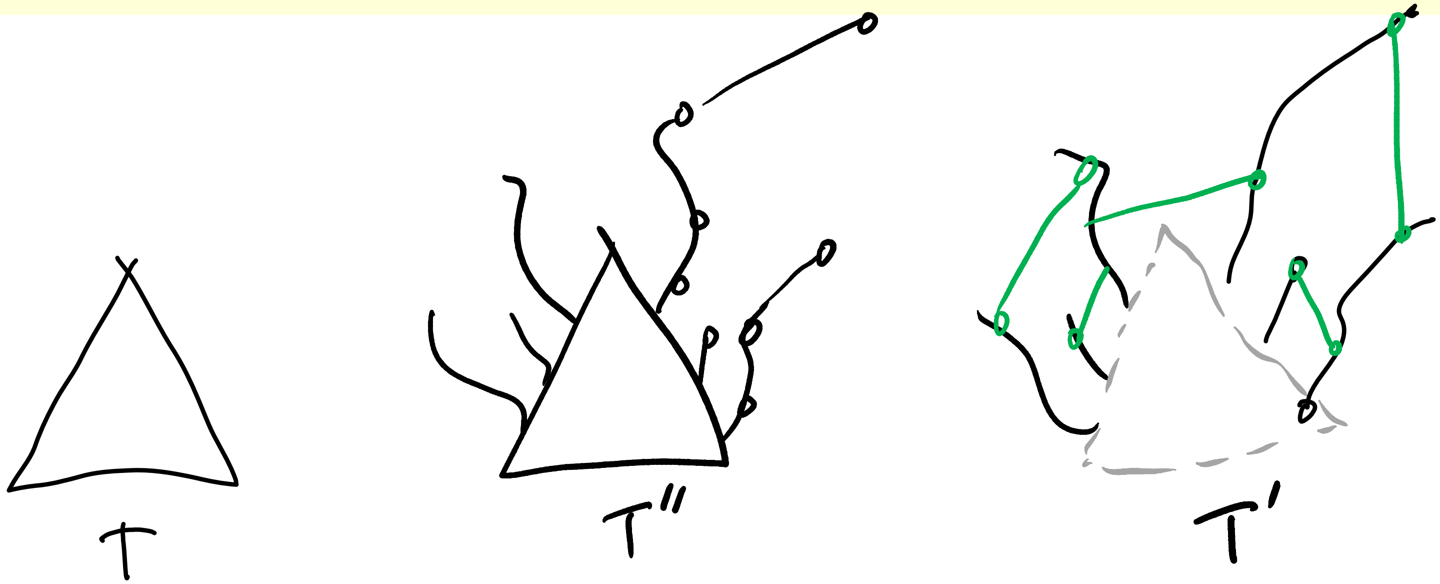




2^3
2^3
2^2
2^2
2^2
2^2
2^1
2^1
2^1
2^1
2^1
2^1
2^1
2^1


2^3
2^3
2^2
2^2
2^2
2^2
2^1
2^1
2^1
2^1
2^1
2^1
2^1
2^1

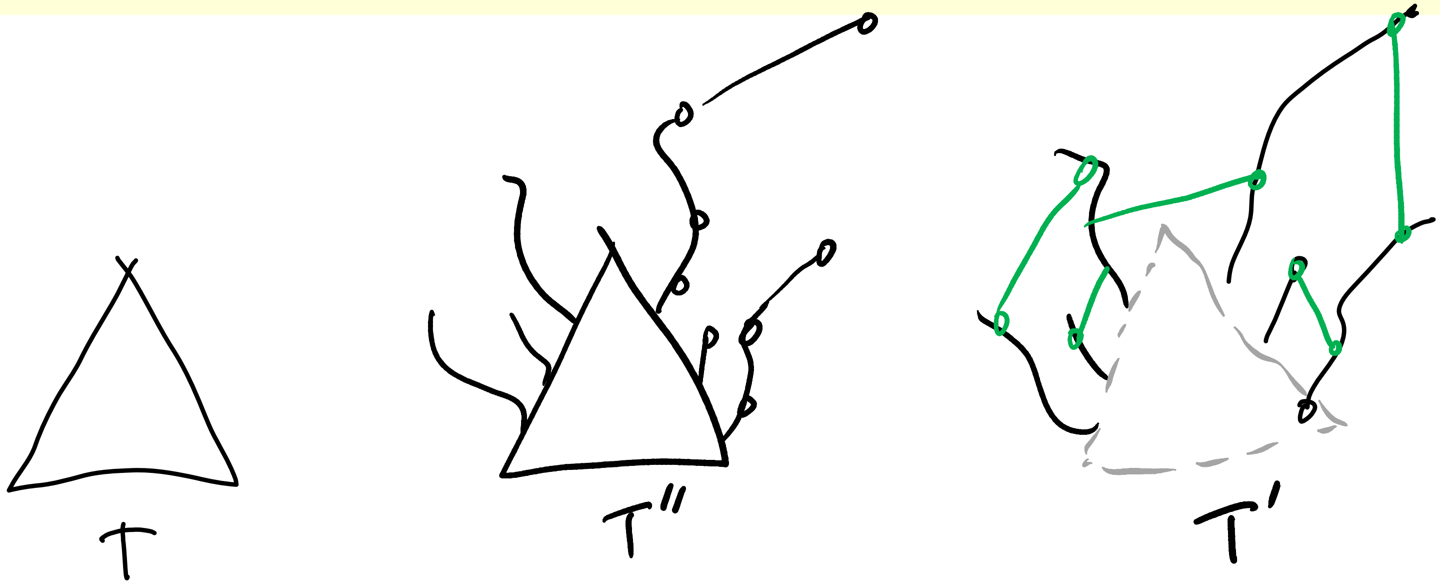
the process in cartoons



Fact 1: no distortion between T and T''

Fact 2: distortion between T'' and T' is at most $(1+O(\epsilon))$

the process in cartoons



Fact 1: no distortion between T and T''

Fact 2: distortion between T'' and T' is at most $(1+O(\epsilon))$

Fact 3: doubling dimension of T'' = O(dim(T))

$\Rightarrow$ dim(T') = O(dim(T)) using fact 2.

the final fact

Lemma: For any node x, radius R,

the number of long edges in T' is $2^{O(k)} \times O(\log \epsilon^{-1})$

> **Recall:** these are the edges that have
> a) one endpoint within B(x,R)
> b) length > R.

Proof sketch. (for interesting case wher

No edge of length > R/ϵ has endpoint in B(x,R) any more.

Hence long edges have length between R and R/ϵ

Need to show at most $2^{O(k)}$ edges of each length scale.

Since $\log 1/\epsilon$ relevant length scales, we'll be done.

so finally… (proof by picture)

so finally… (proof in words)

consider all long edges $\{u_i, v_i\}$ of length $\sim 2^t$ with u_i in $B(x,R)$

all these v_i's contained within $B(x,R+2^{t+1}) \subseteq B(x,2^{t+2})$

the path $x \rightarrow u_p$ has length R, cannot contain long edges

same for $x \rightarrow u_q$

tree path $u_p \rightarrow u_q$ is the symmetric difference of these paths

Hence $v_p \rightarrow v_q$ contains both the long edges, distance at least 2.2^t.

Now all the v_i's are in a ball of radius 2^{t+2}

and are at least 2^{t+1} apart.

So only $2^{O(k)}$ of them.

last part of the talk now

The lower bound for trees.

A structure theorem.

The upper bound for trees.

Outline of the general upper bound.

the theorem for general metrics

Theorem 2:

Given a metric (V,d) with doubling dimension k

there is an weighted graph $G' = (V',E')$ with $V \subseteq V'$ such that

- Distances in G' are within $(1+\epsilon)$ of d

- Doubling dimension of $Conv(G')$ is $O(k \log \epsilon^{-1})$

  equivalently, number of long edges in G' is $2^{O(k \log \epsilon^{-1})}$

General Graphs

[Chan G. Maggs Zhou '05] gave bounded-degree spanners for doubling metrics.

This is a (sparse) graph $G' = (V, E')$ such that

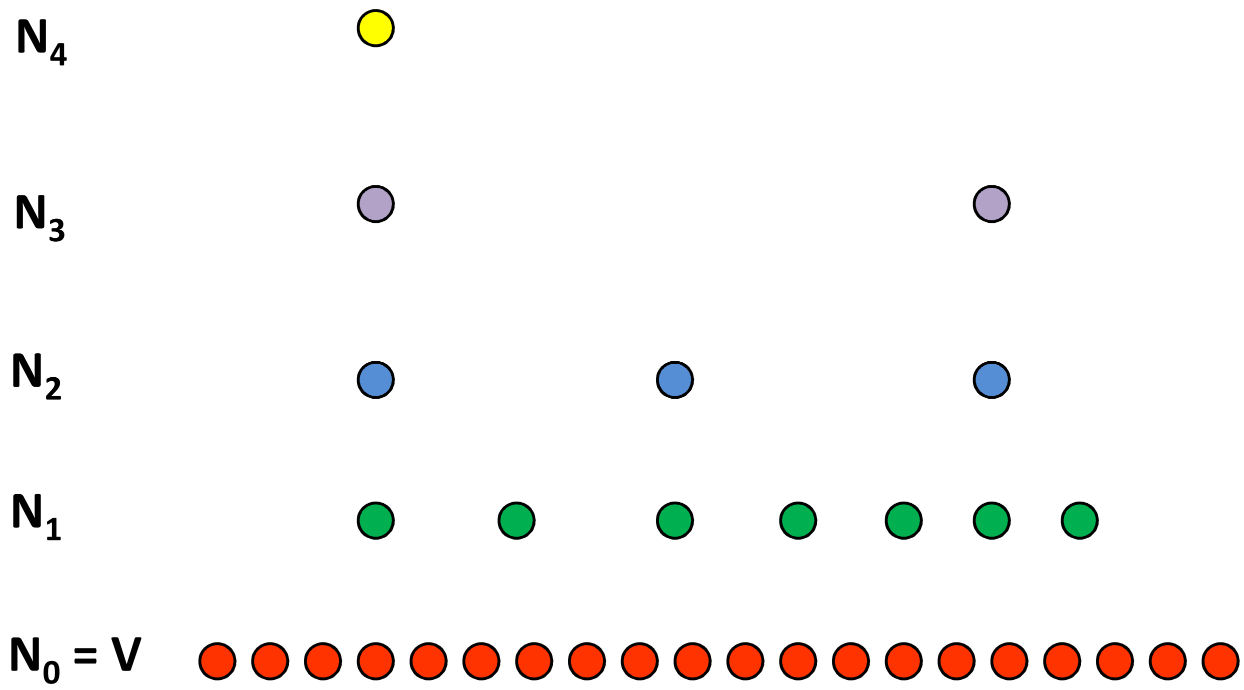
a) degree of each node in G' is $(1/\epsilon)^{O(k)}$

b) distances in (V,d) are maintained to within $(1+\epsilon)$

We can show that this spanner G' has few long edges

Thus taking $conv(G')$ suffices to prove the theorem.

the idea of the construction



Add edges between all pairs in N_t at distance at most $O(2^t/\epsilon)$

This gives a sparse spanner (not constant degree, though)

the idea of the construction$_{(2)}$

summary

Given a doubling metric, we show there is a nearby graph which can be made ‘convex’ without increasing the doubling dimension.

Similar result for doubling tree metrics

Allows us to reason about unweighted graphs/trees.

and two questions

Q. Improve bounds for general graphs?

(need to have Steiner points to get better, remember.)

Q. Greater understanding of the interplay between topology of graphs and properties of the metrics generated on them…

thanks!