Fast Distance Queries

Laurent Viennot

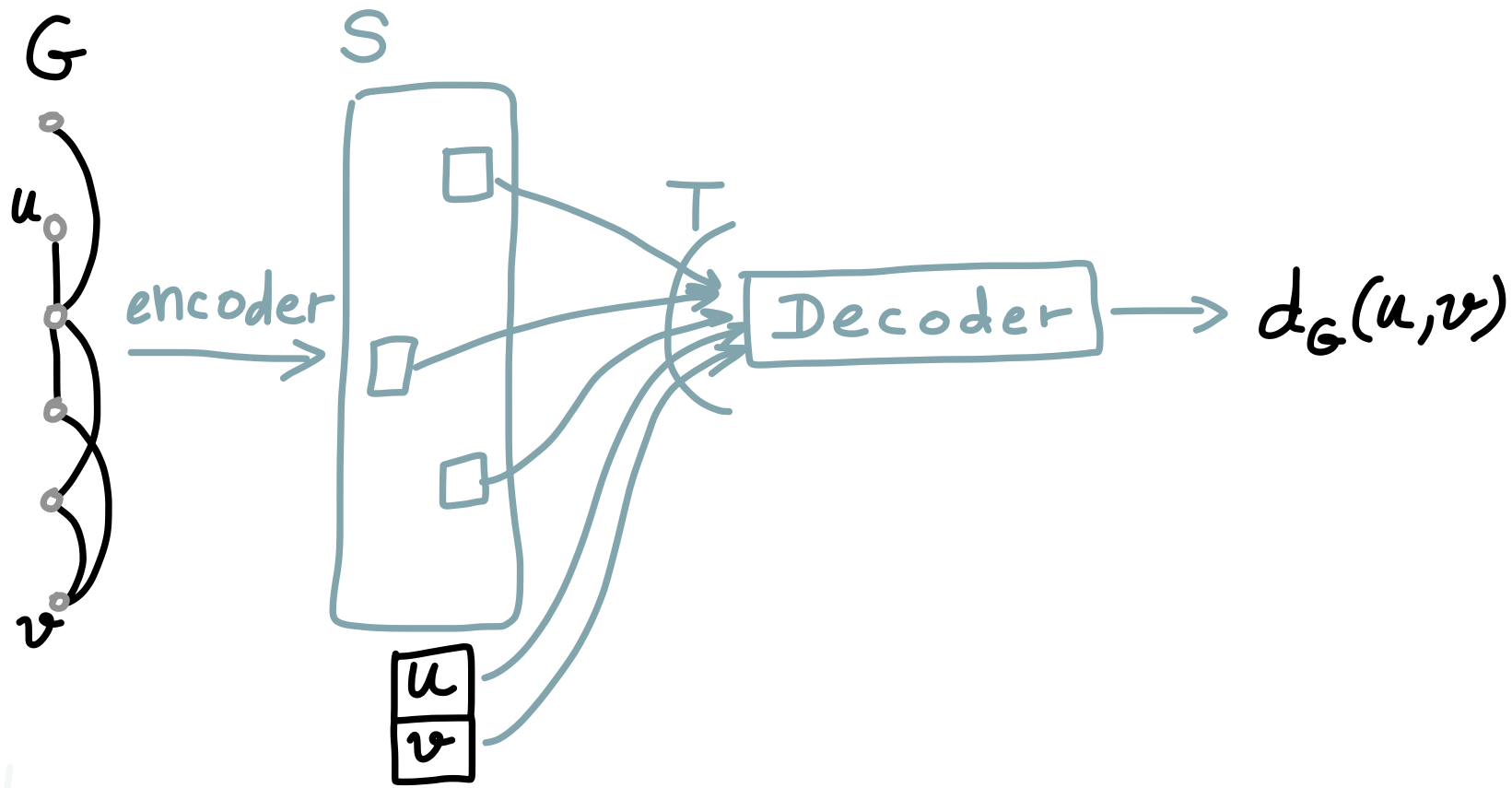
MPRI – Theory of practical graph algorithms

Encoding distances in a graph

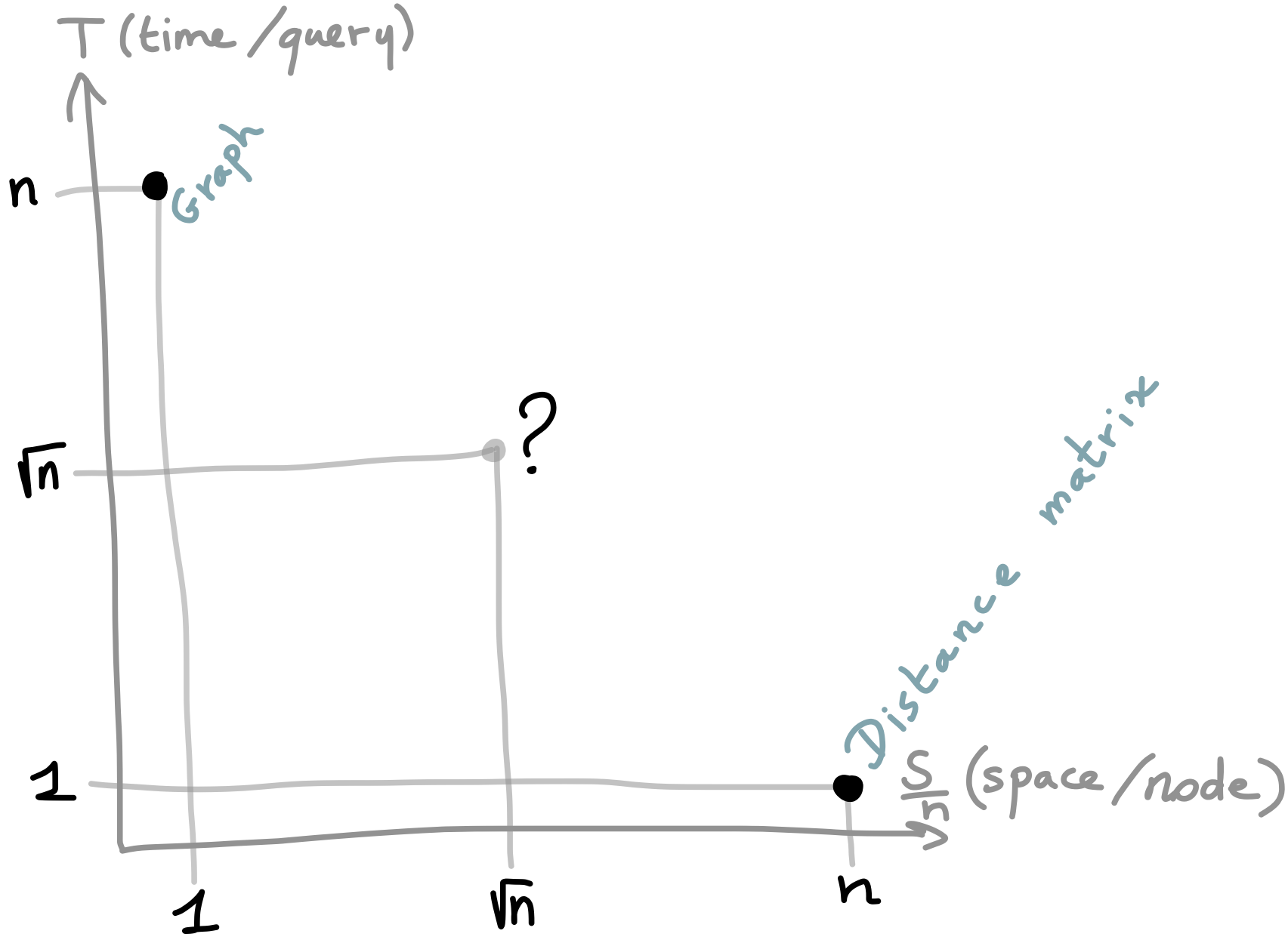
We are given a (weighted) (di-) graph $G = (V, E)$ with n nodes and m edges.

Make any useful pre-computation to answer efficiently online distance queries : what is distance $d(u_1, v_1)$?, $d(u_2, v_2)$?,$d(u_3, v_3)$?,...

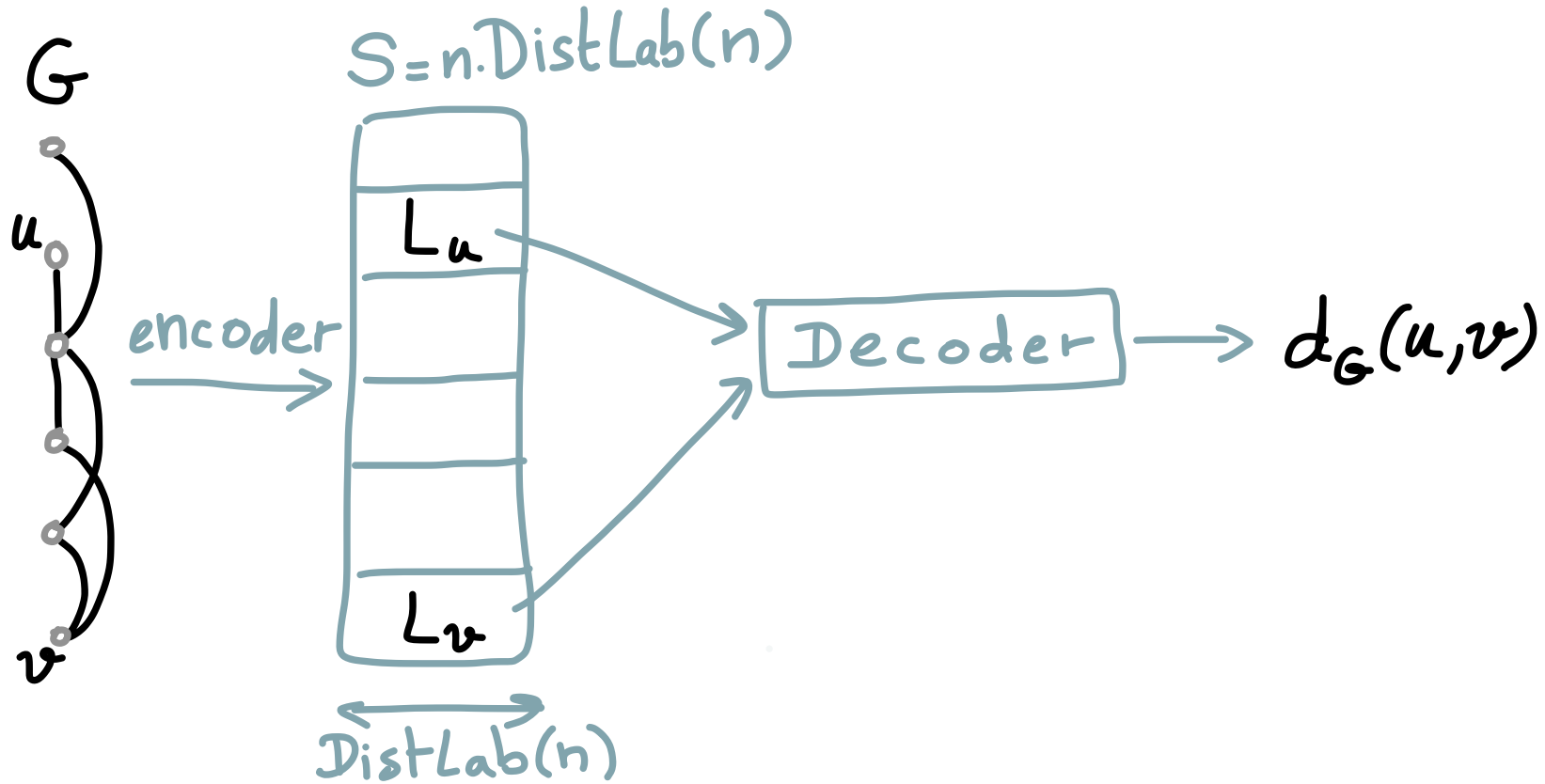
Encoding a graph metric : distance oracles



Size S vs query time T tradeoff (sparse graphs, i.e. $m = O(n)$)



Encoding a graph metric : distance labelings



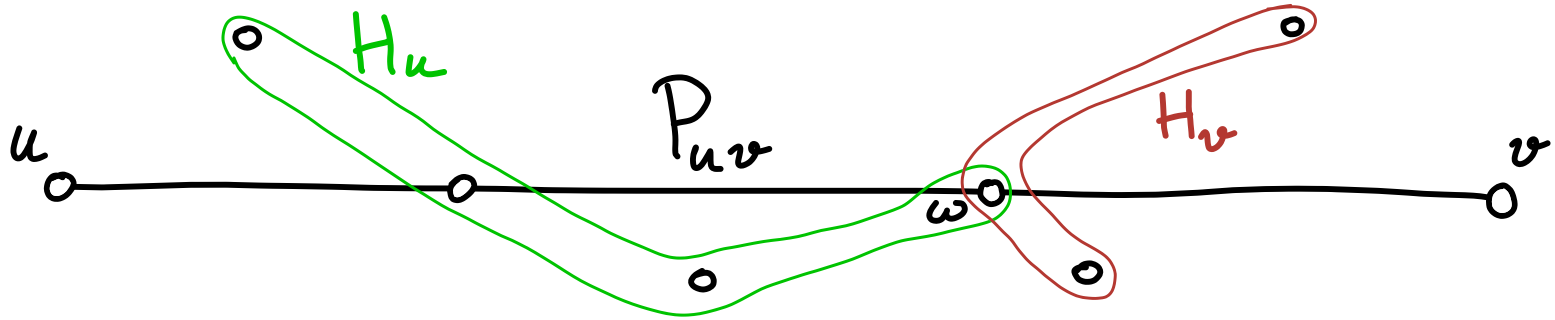
Distance labeling [Gavoille, Peleg, Pérennes, Raz '04]

Problem

Given a graph G assign a label L_u to each node u s.t. for all s, t $d_G(s, t)$ can be computed from L_s and L_t.

Hub sets

Given a graph G, assign a hub set $H_u \subseteq V(G)$ to each node u, s.t. for all u, v there exists $a \in H_u \cap H_v$ with $a \in P_{uv}$.



Distance labels : $L_u = \{(a, d(u, a)) : a \in H_u\}$

Distance query : Dist $(L_s, L_t) = \min_{a \in H_s \cap H_t} d(s, a) + d(a, t)$
in $O(|H_s| + |H_t|)$ time.

Labeling with hub sets : hub labeling a.k.a. 2-hop labeling

Exercise : Hub labeling for a path ?, a tree ? a graph with treewidk k ? a planar graph ?

Open pb : Increase the best known lower bound for unweighted planar graphs ($\Omega(n^{1/3})$ [Gavoille et al. 2004]).

Exercise : Hub labeling for a grid ?

2-hop labeling [Cohen, Halperin, Kaplan, Zwick '03]

Problem : Compute hub sets $(H_u)_{u \in V}$ of minimum size $|H| = \sum_u |H_u|$.

Greedy cover all shortest paths :

- smallest avg. hub size is equivalent to min. cost set cover,
- $O(\log n)$-approximation is possible :
- select a, S s.t.

  rel. cost $\dfrac{|S|}{\text{nb path cov. if a added to all } (H_u)_{u \in S}}$ is min.

Problem : set cover instance with $n \times 2^n$ sets!

Solution : fix a, what is the best S?

G_a graph with edges uv s.t. P_{uv} still uncov. and $a \in P_{uv}$.

Exercise : Propose a greedy algorithm for 2-approximating the best S.

Corollary : Hubsets with smallest average size can be $O(\log n)$-approximated in polynomial time.

Highway dimension [Abraham, Delling, Fiat, Goldberg, Werneck '10-13]

Definition

Highway dimension $h = \max_{u,r} \min_{H \text{ hitting set of } \mathcal{P}_{ur}} |H|$ where $\mathcal{P}_{ur} = \{P \in \mathcal{P}_r \mid \overline{P} \cap B(u,r) \neq \emptyset\}$, $\mathcal{P}_r = \{P \mid \ell(\overline{P}) > \frac{r}{2}\}$, and $\overline{P}$ is any shortest path extending P by 0 or 1 edge at each extremity.

Theorem

Any graph G with highway dimension h and diameter D admits a node ordering π s.t. CH_π^{opt} produces at most $O(nh \log D)$ shortcuts and CH_π + RP bidir. Dij. visits $O(h \log D)$ nodes.

Exercise : use CH to compute a HHL.

Highway dimension [Abraham, Delling, Fiat, Goldberg, Werneck '10-13]

Hierarchical Hub Labeling (HHL) [BGKSW'15]

A hub labeling is hierarchical if it respects an order π such that hubs are more important : $v \in H_u \Rightarrow \pi(u) \leq \pi(v)$
(the graph with edges from nodes to their hubs is a DAG).

Canonical HHL

Given an ordering π, for all u, v add $\max_\pi P_{uv}$ to H_u and H_v.

Proposition

Canonical HHL for π is the minimum HHL that respects π.

Exercise : show that any minimal HHL is canonical.

Pruned Labeling [Akiba, Iwata, Yoshida '13]

Procedure PrunedLab (G, π)
  Distance labels $L_u := \emptyset$ for all u.
  For each $a \in V(G)$ in decreasing order of π **do**
    PrunedDijkstra (G, a, L)
    Add $(a, d(a, u))$ to L_u for each visited node u.

Procedure PrunedDijkstra (G, a, L)
  Starting from $u = a$, visit u if $d(u) < \text{Dist}(L_a, L_u)$.

Theorem
PL computes the canonical HHL associated to π in $O(nL \log n + mL^2)$ time where L is maximum label size.

Open pb : charac. classes of graphs with $|HHL| = O(|HL|)$.

HL on massive networks [Delling, Goldberg, Pajor, Werneck '14]

HHL using random sampling to approximate greedy cover (for π) in combination with pruned labeling (for hub sets).

$O(\log n)$ approximation in theory, smallest hub labelings in practice (and fastest distance oracle).

Skeleton dimension [Kosowski et al. '17]

Graph property ensuring small hub sets.

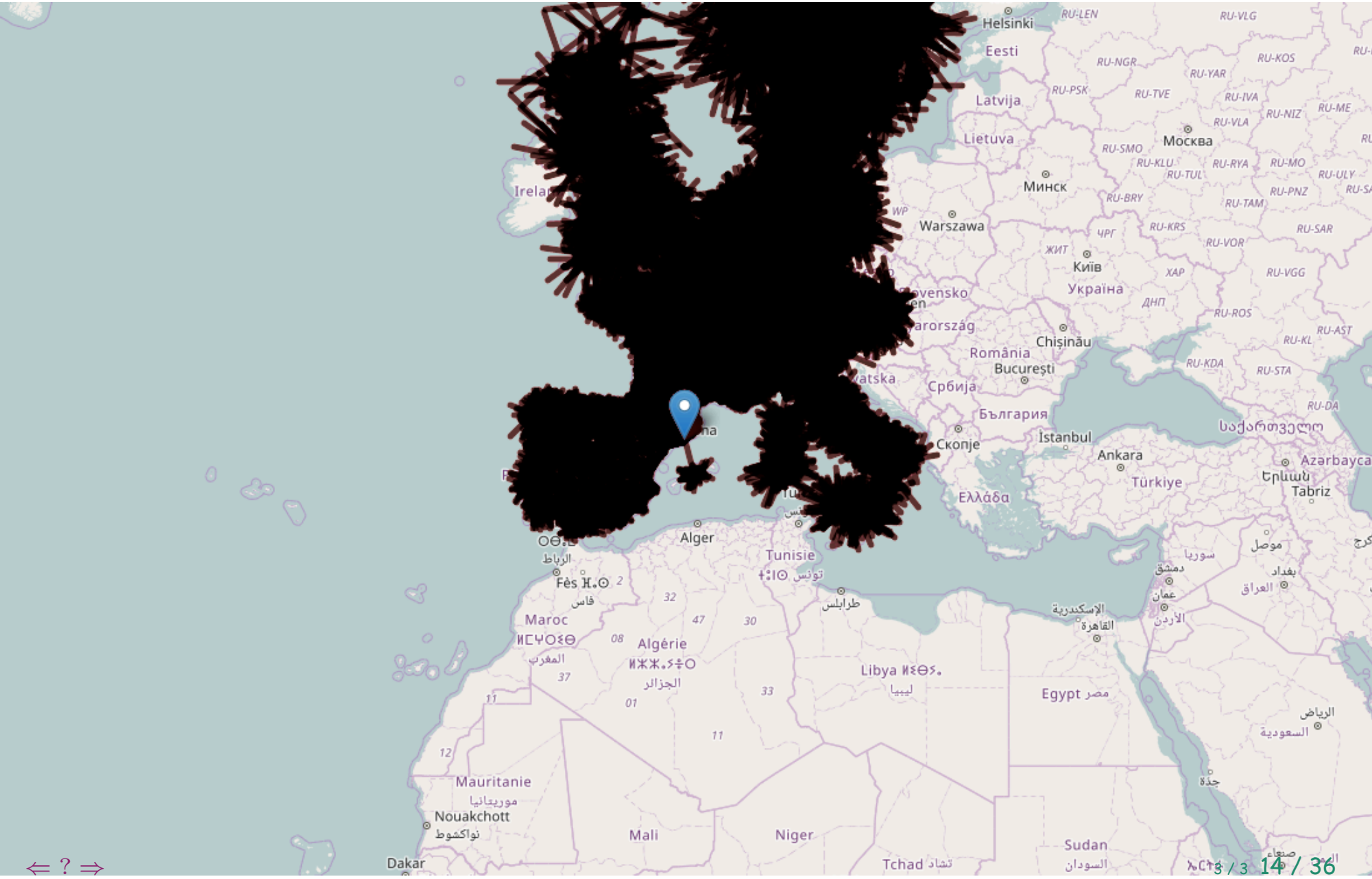
The skeleton dimension k of G is the maximum "width" of a "pruned" shortest path tree (see pres.).

Theorem
Any graph G with skeleton dimension k and diameter D has hub sets of size $O(k \log\log k \log D)$ (polyn. time constr. w.h.p.).

Open pb : what additional property ensures efficient Reach/CH?

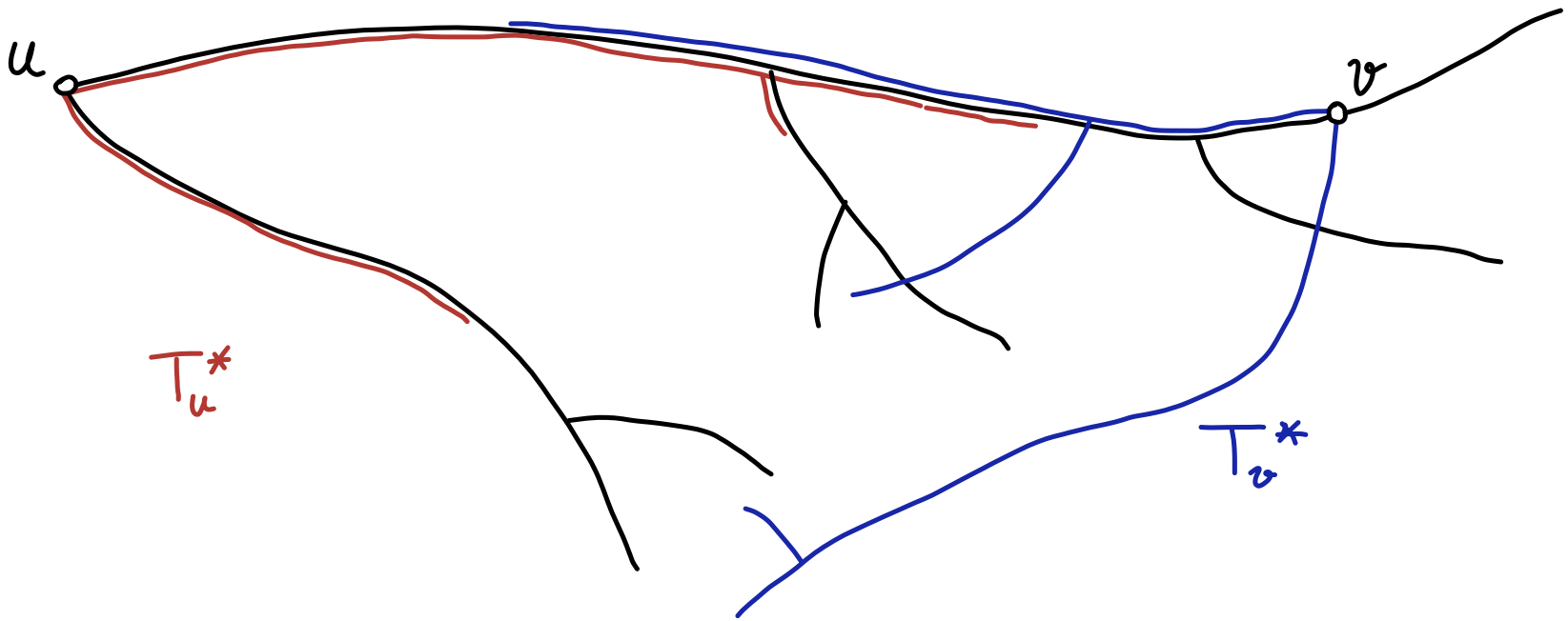
Open pb : tight bounds on HL/HHL in grids?

Barcelona shortest path tree skeleton : prune last third



Tree skeleton

Hub set selection



Hub set selection (proof for unweighted graphs)

Draw $\rho(w) \in [0,1]$ u.a.r. for all $w \in V(G)$.

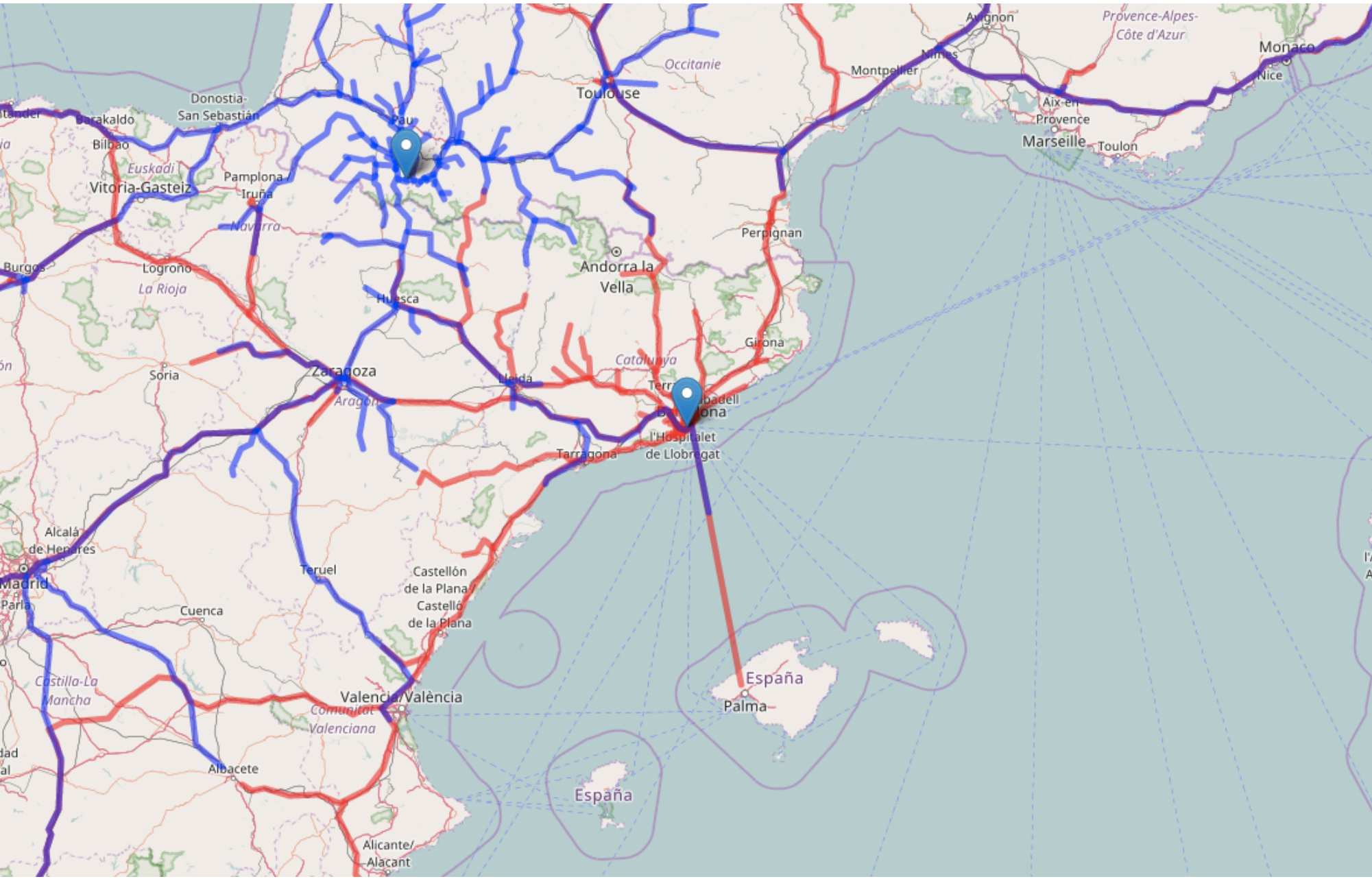
$H_u = \{w \mid \rho(w) \text{ min. in } P_{xwy}\} \cup \{x_y \mid \rho(x_y) \text{ min. in } P_{xyy}\}$

(Can be computed in $\widetilde{O}(n+m)$ separately for each node with shared randomness.)

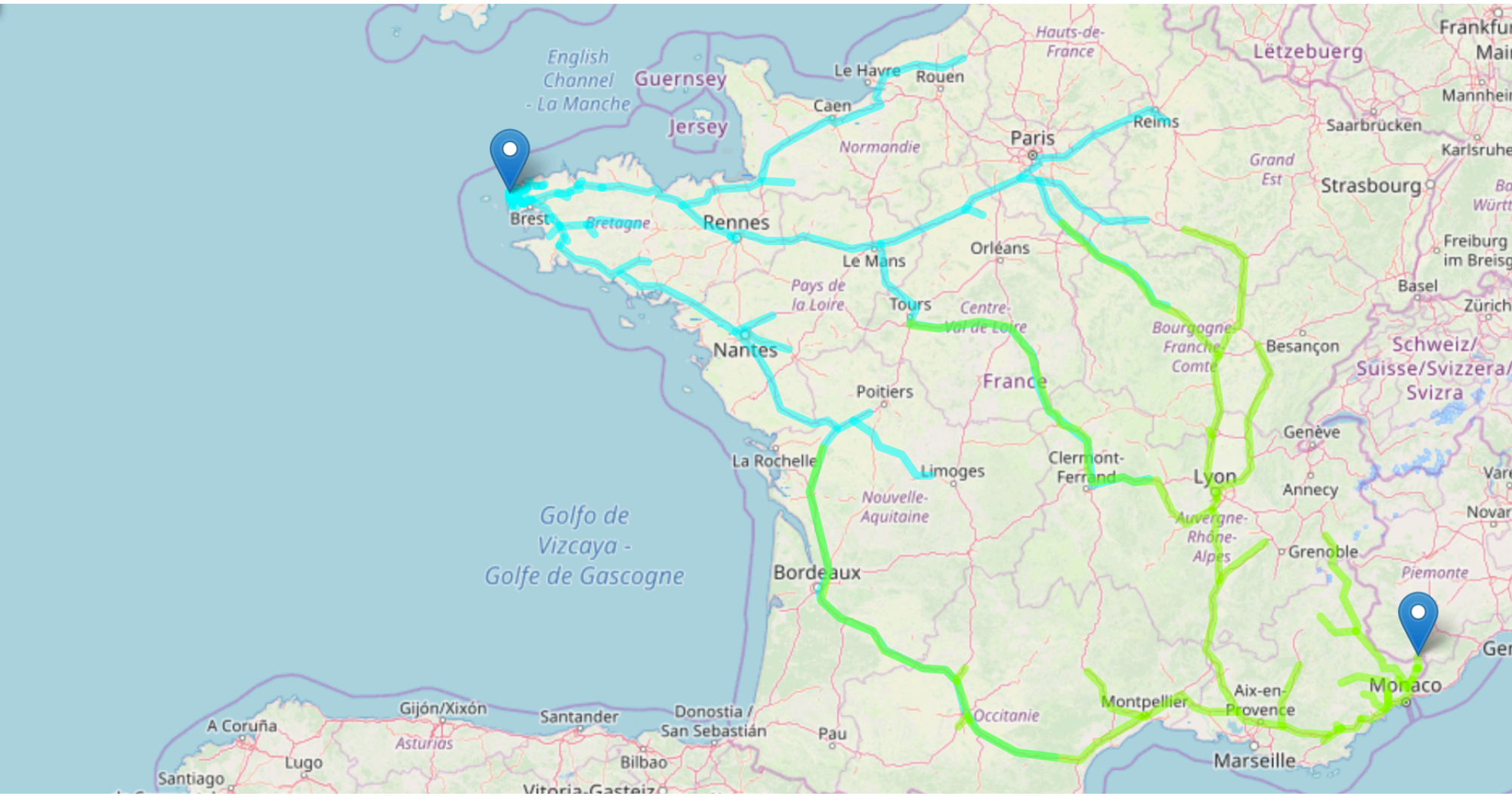
A sub-path P_{xyy} has length $\frac{d(u,y)}{6}$ and generates a hub in H_u with probability at most $\frac{12}{d(u,y)}$.

$E[|H_u|] \leq \sum_{y \in V(T_u^*)} \frac{12}{d(u,y)} \leq \sum_r |Cut_r(T_u^*)| \frac{12}{r} = O(k \log D)$

Road networks : two tree skeletons



What ...maps do?


English Channel - La Manche
Guernsey
Jersey
Le Havre
Rouen
Caen
Normandie
Paris
Reims
Hauts-de-France
Lëtzebuerg
Saarbrücken
Grand Est
Strasbourg
Brest
Bretagne
Rennes
Le Mans
Orléans
Pays de la Loire
Tours
Centre-Val de Loire
Nantes
Bourgogne-Franche-Comté
Besançon
Basel
Zürich
Schweiz/Suisse/Svizzera/Svizra
France
Poitiers
La Rochelle
Limoges
Clermont-Ferrand
Lyon
Genève
Annecy
Nouvelle-Aquitaine
Auvergne-Rhône-Alpes
Grenoble
Golfo de Vizcaya - Golfe de Gascogne
Bordeaux
Piemonte
Monaco
Montpellier
Aix-en-Provence
Occitanie
Marseille
Pau
Donostia / San Sebastián
Santander
Gijón/Xixón
Asturias
A Coruña
Lugo
Santiago
Bilbao

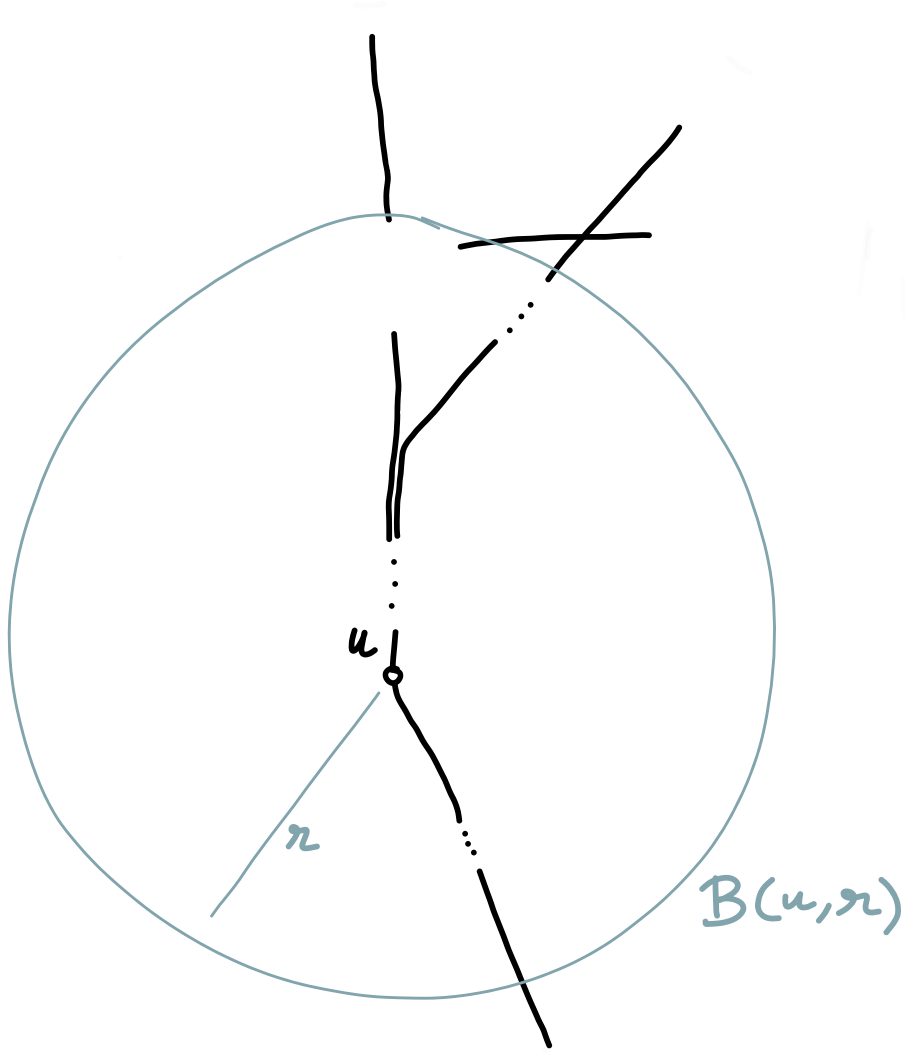
Highway dimension $\geq$ skeleton dimension

$\mathcal{P}_{ur} = \left\{ P : \ell(\overline{P}) > \frac{r}{2} \text{ and } \overline{P} \cap B(u, r) \neq \emptyset \right\}$

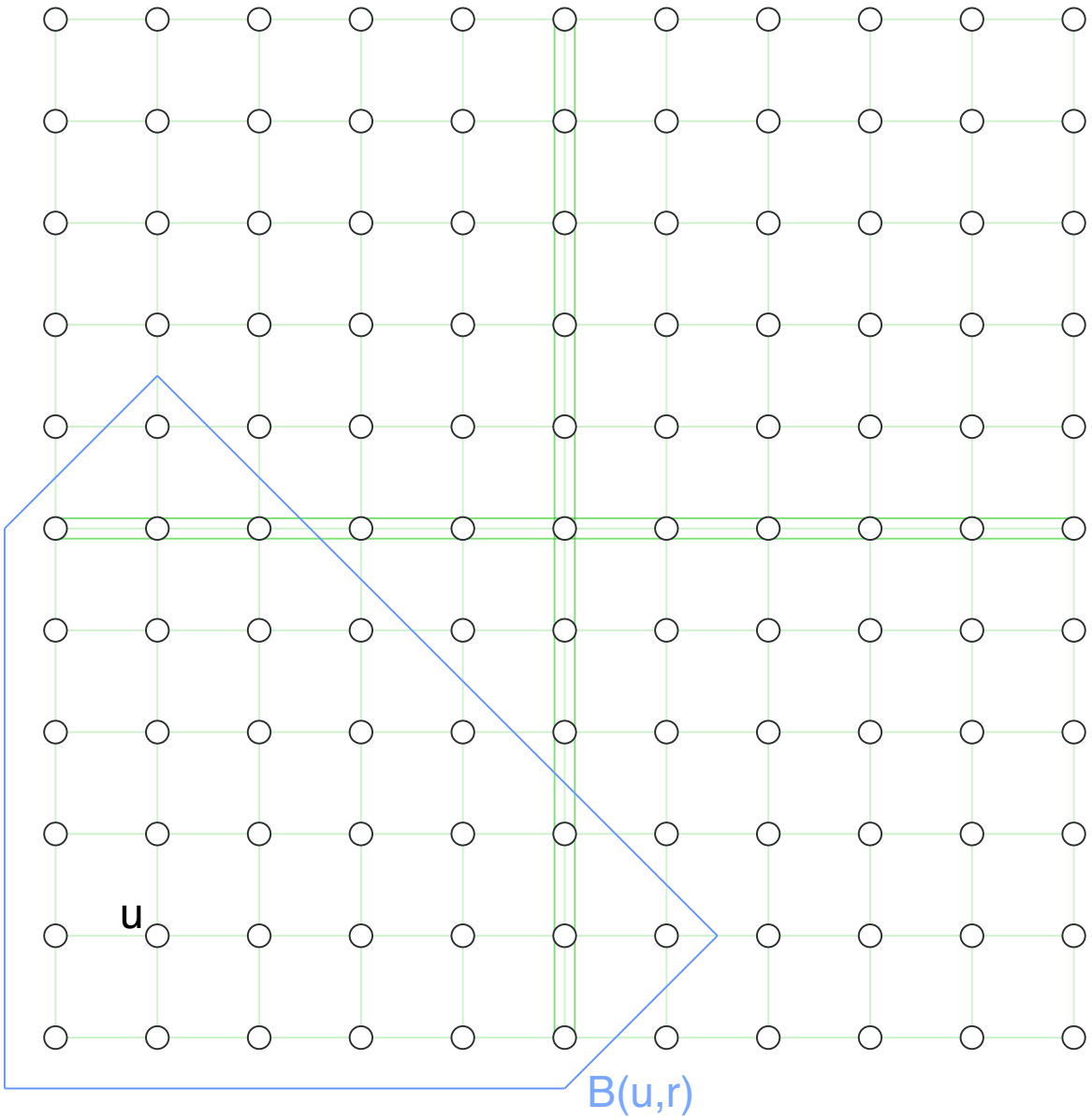
H hits $\mathcal{P}_{ur}$ if $H \cap P \neq \emptyset$ for all $P \in \mathcal{P}_{ur}$

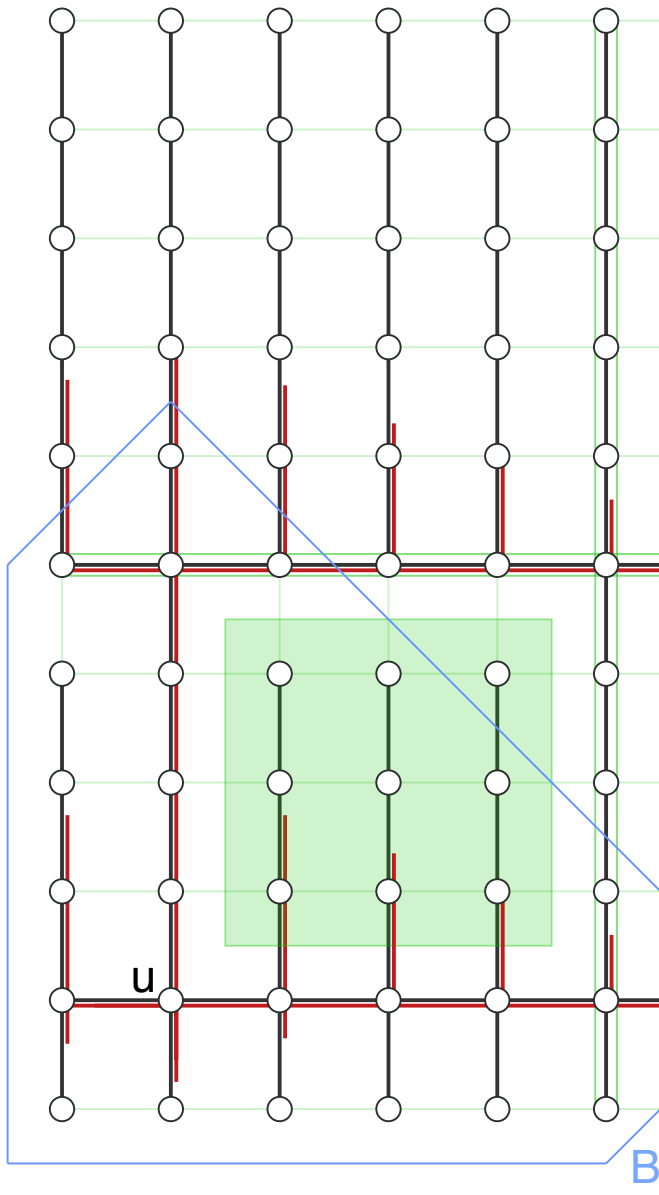
Highway dim. $h = \max_{ur} \min_{H \text{ hits } \mathcal{P}_{ur}} |H|$

$k \leq h$: $Cut_r(T^*_u)$ induces a packing in $\mathcal{P}_{ur}$, and $|Cut_r(T^*_u)| \leq |H|$.



Dimension of grids



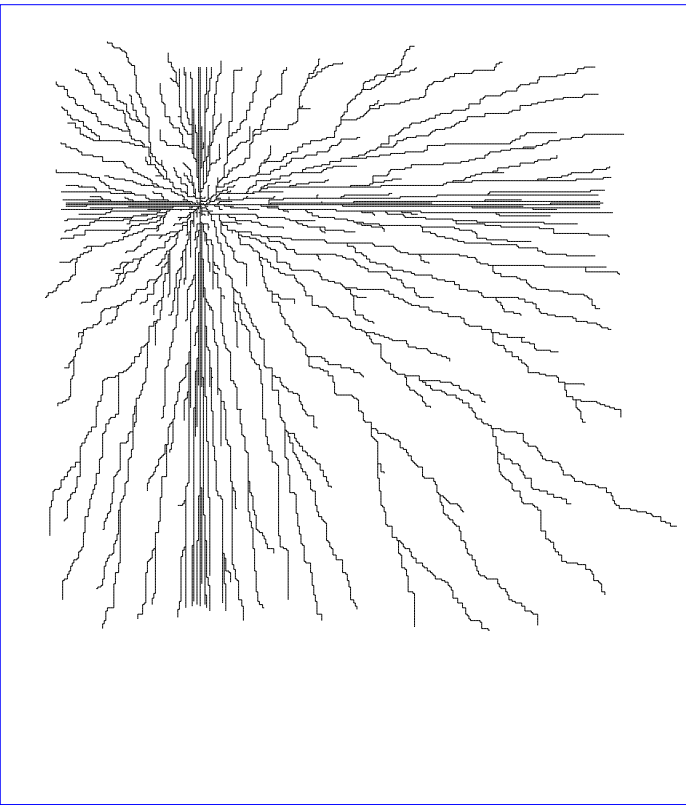


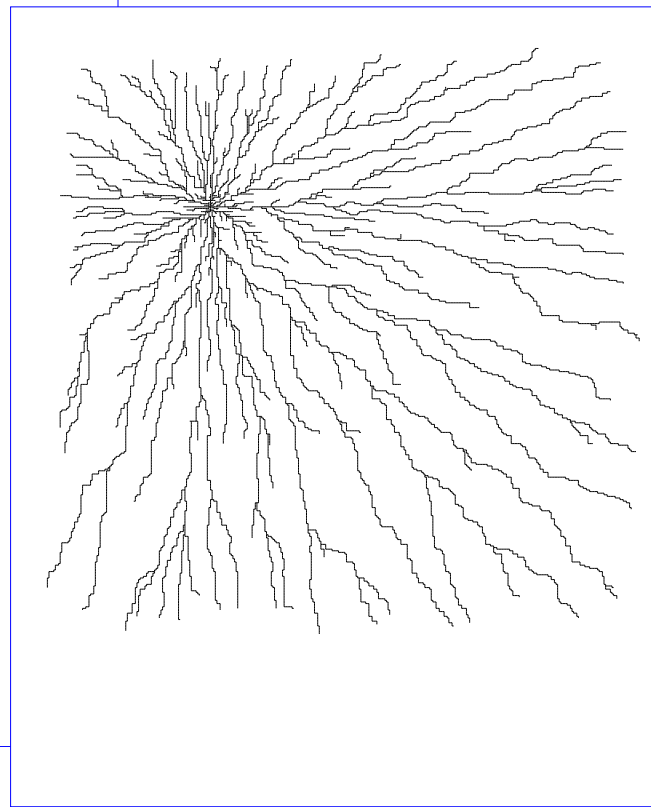
Highway vs skeleton in Brooklyn

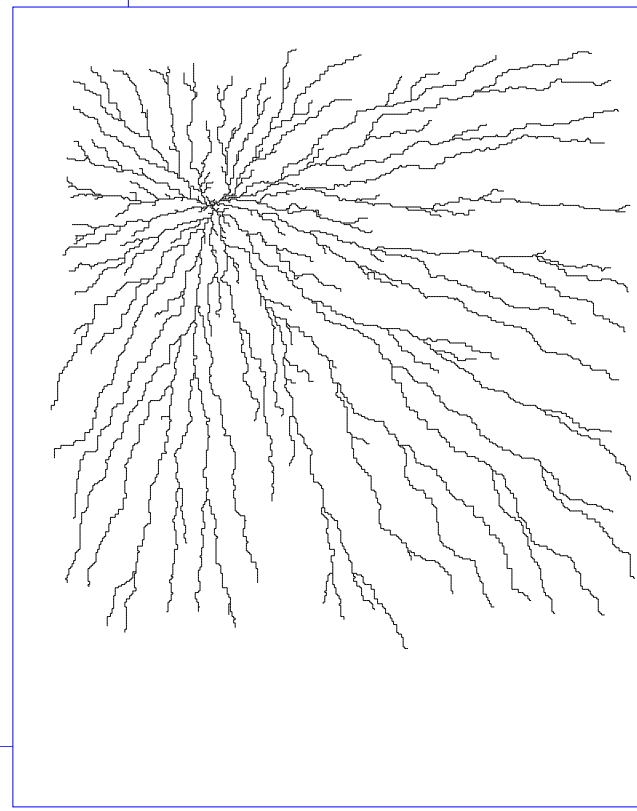
Packing of 172 paths

Skeleton width 48

Open : random grid (here 500×500)







$k = 70$ $k = 49$ (fpp $[1, 4)$) $k = 49$ (prob 2/3)

What about general graphs?

Pre-hub labeling [Angelidakis, Makarychev, Oparin '17]

Hub sets $(H_u)_{u \in V(G)}$ for a graph G form a pre-hub labeling if for all u, v pairs, hubs cross on P_{uv} : $\exists u' \in P_{uv} \cap H_u$ and $\exists v' \in P_{uv} \cap H_v$ with $u' \in P_{v'v}$ and $v' \in P_{uu'}$.

Theorem

If shortest paths are unique,

- PHL 2-approximate HL (and pol. time constr.),
- PHL can be converted to HL with $O(\log D)$ factor.

Theorem

If shortest paths are not unique,

- best polyn. time approx. is $\Omega(\log n)$ (even if $D = O(1)$).

Theorem

In trees, HHL 2-approximate HL (and pol. time constr.).

Exercise : prove it.

Hub labeling of general graphs

Any graph has $O(\frac{n}{\log n})$ hubsets for $m = O(n)$ (combining [Kosowski et al. '17] and [Alstrup et al. '16]).

Idea (for $\Delta = O(1)$) :

- for $r \geq \delta$, the width of a r-cut of a skeleton tree is $k = O(n/\delta)$ (we can get $O(n/\delta)$ hubsets for distances $\geq \delta$),
- select as hubs all nodes at distance less than $\delta = \frac{\log n}{2 \log \Delta}$

(at most $\sqrt{n}$ nodes).

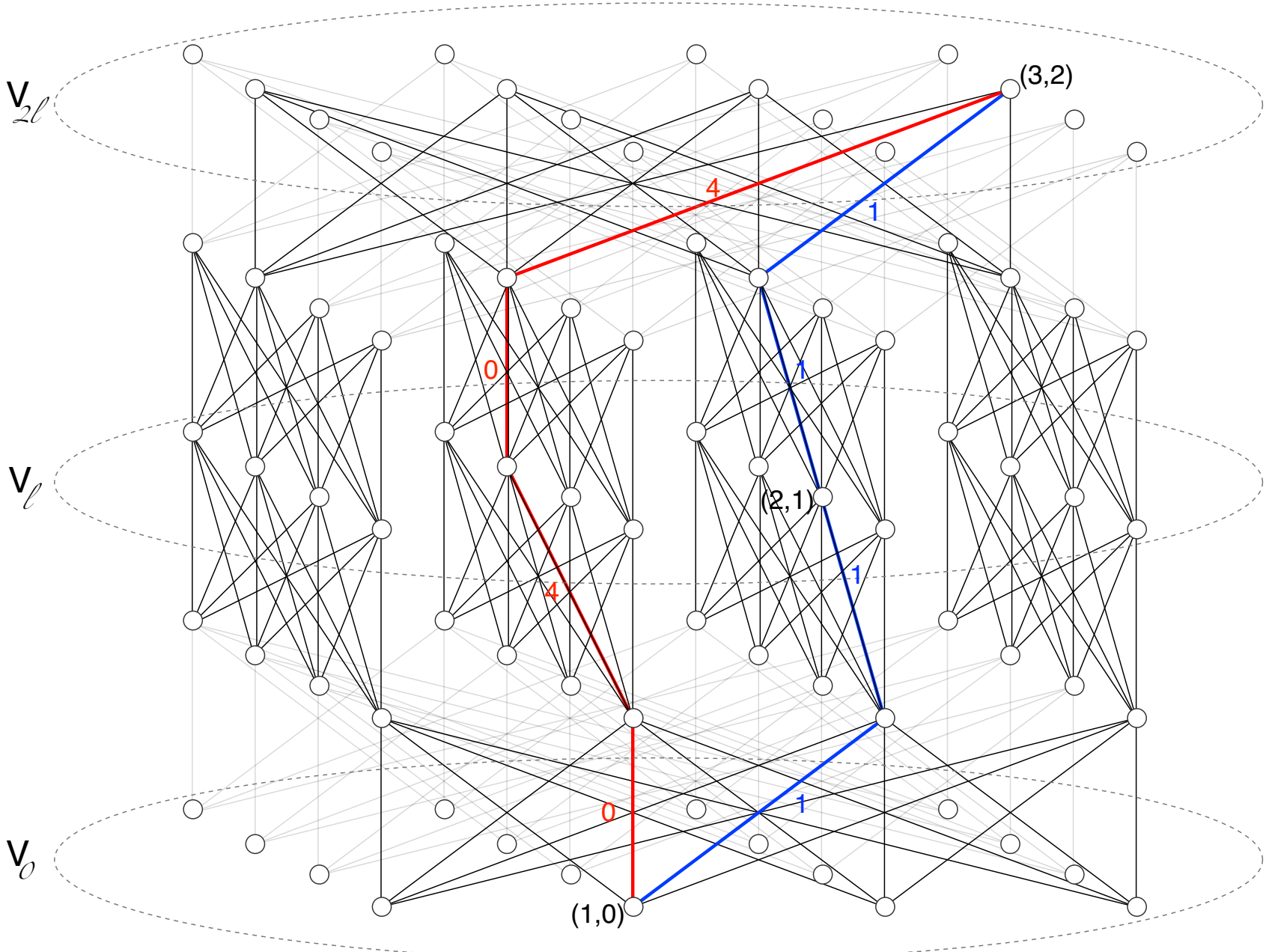
Lower bound on hub labeling of general graphs

Theorem [Kosowski et al. '19] : There exists graphs with max-degree 3 such that any hubsets have average size $\Omega(\frac{n}{2^{O(\sqrt{\log n})}})$.

Linked to Ruzsa-Szemerédi function bounding the number of edges in a graph decomposoble into n induced matchings.

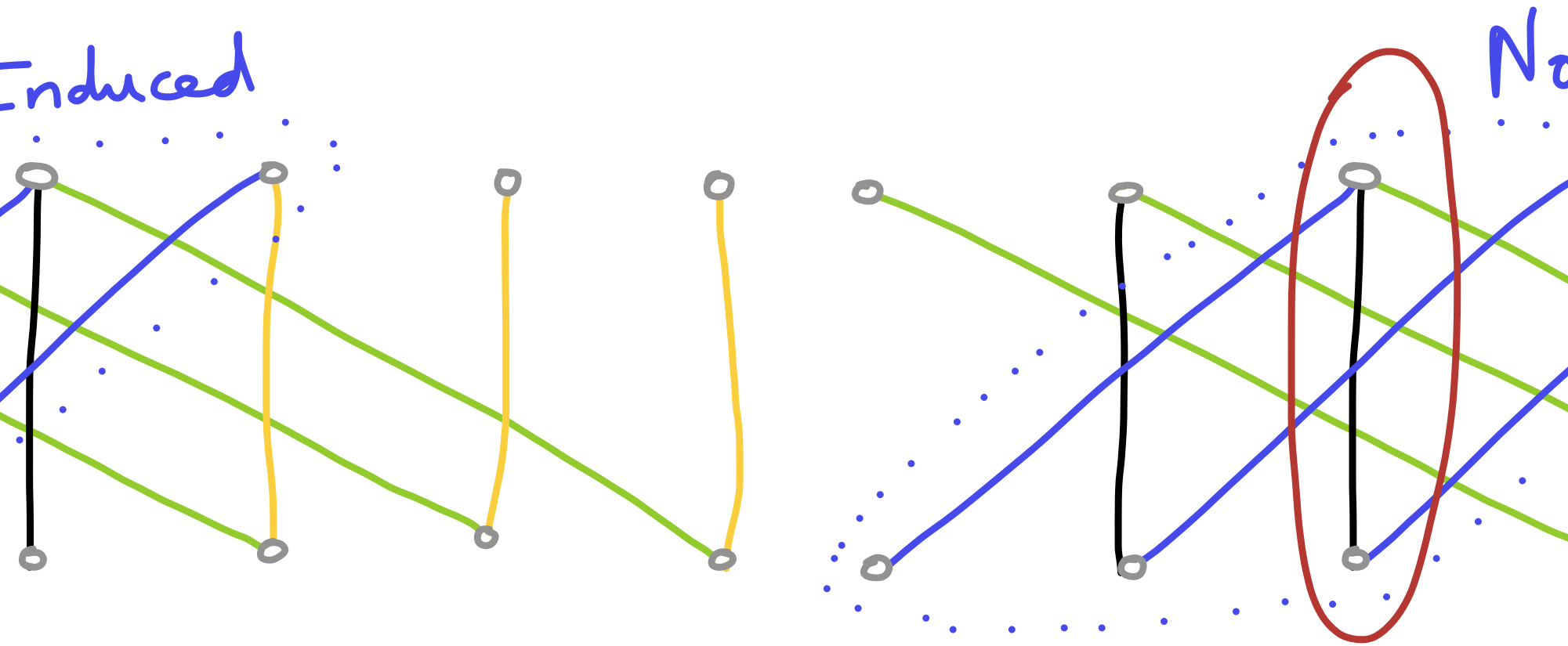
There exists graphs with max-degree 3 such that any distance labels must have average size $\geq \frac{\text{SUMINDEX}(n)}{2^{O(\sqrt{\log n})}}$.

Open problem : does any sparse graph have a (centralized) distance oracle of size $O(n^{1.5})$ and query time $O(n^{0.5})$?



Each V_i is a regular $2^\ell \times \cdots \times 2^\ell$ lattice of dim. $\ell \approx \sqrt{\log n}$ (here $\ell = 2$).
Edges from V_{i-1} to V_i connect nodes differing on ith coordinate.

Ruzsa-Szemerédi



A graph is an RS-graph if it can be decomposed into n induced matchings.

A graph is an RS-graph if it can be decomposed into n induced matchings.

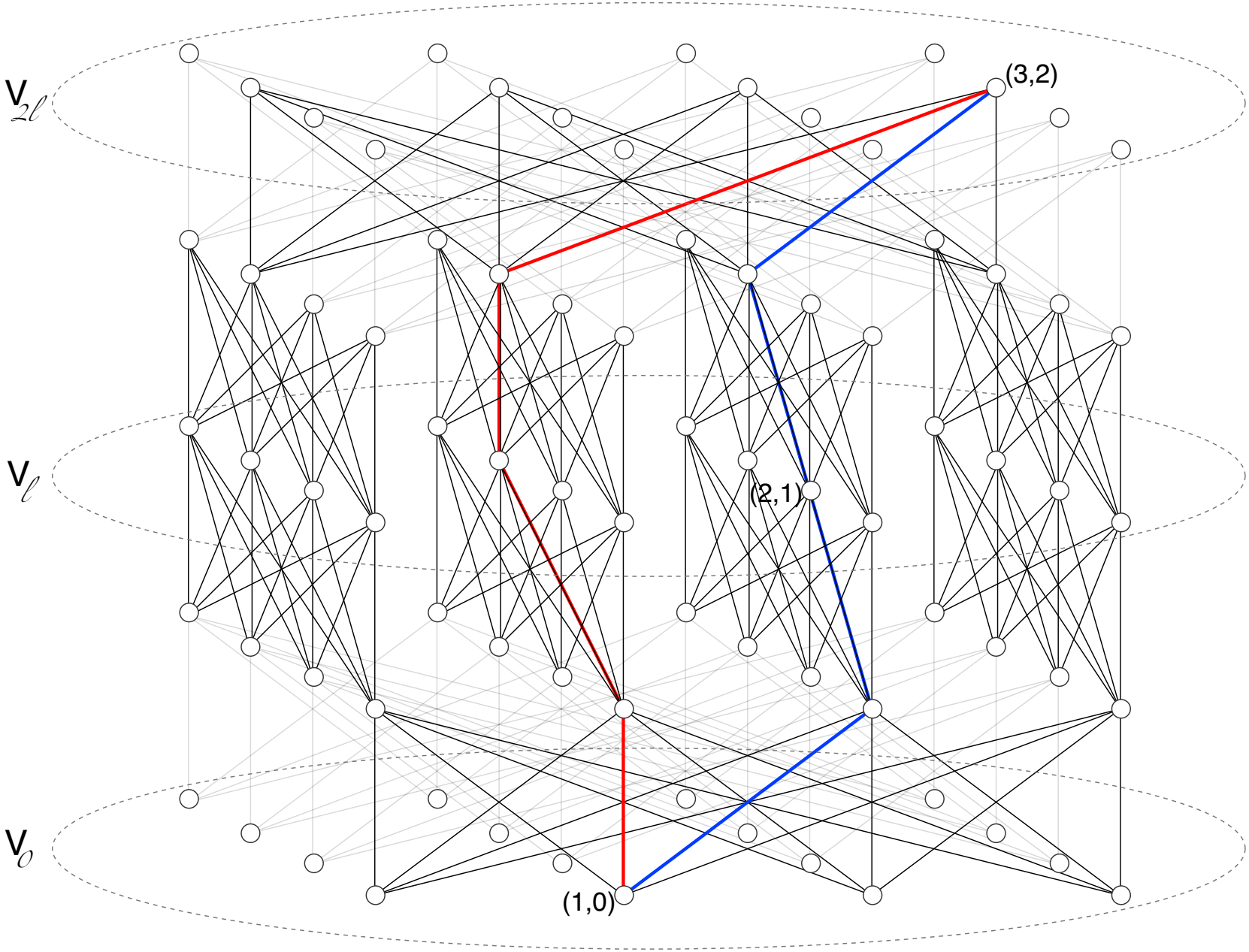
What are the densest RS-graphs?

Theorem ([Ruzsa, Szemerédi '78])
Any RS-graph has at most $\frac{n^2}{2^{O(\log^* n)}}$ edges.

Define $RS(n)$ as the smallest integer such that there exists an RS-graph with n nodes and $\frac{n^2}{RS(n)}$ edges.

$$2^{\Omega(\log^* n)} \leq RS(n) \leq 2^{O(\sqrt{\log n})}$$

[Ruzsa, Szemerédi '78] [Elkin '10] [Fox '11]



$$G_y^D = \{x_0 z_{2\ell} \mid y = \tfrac{x+z}{2} \text{ and } d_G(x,z) = D\} \quad \exists D \text{ s.t. } |\cup_y G_y^D| \geq \frac{n^2}{2^{O(\sqrt{\log n})}}$$

What about more hops?

h-hop distance

$$d_G^h(u,v) = \min_{P\ uv\text{-path of } \leq h \text{ edges}} \ell(P)$$

Usual distance : $d_G(u,v) = d_G^{n-1}(u,v)$

Exercise

We define a h-hopset of G as a set H of edges such that $d^h_{G\cup H}(u,v) = d_G(u,v)$ where each edge uv of H is considered to have length $d_G(u,v)$.

(a) What is the minimum number of edges in $G \cup H$ when H is a 1-hopset of G?

(b) What notion seen in course is tightly related to the notion of 2-hopset?

(c) Suppose that G is a path of length n, propose a 3-hopset of G with as few edges as you can (we do not care about multiplicative constants).

(d) Same question for a 4-hopset.

(e) Consider a 3-hopset H of a graph G. Propose a distance oracle based on distinguishing middle edges of 3-hop shortest paths from the two others. What query time do you obtain when G is a path of length n?

3-Hopsets in Graphs with Bounded Skeleton Dimension

Theorem (Gupta et al. 2019)

For a unique shortest path graph with skeleton dimension k and polylog average link length, there exists a randomized construction of a 3-hopset distance oracle of size $|H| = O(n\,k\log k\log\log n)$, which for an arbitrary queried node pair performs distance queries in expected time $O(k^2\log^2 k\log^2\log n)$ (where the expectation is taken over the randomized construction of the oracle).

Open pb : Does there exists $\varepsilon, \varepsilon' > 0$ and distance oracles for constant degree graphs with size $O(n^{2-\varepsilon})$ and query time $O(n^{1-\varepsilon'})$?

Open pb : Could it be a 3-hopset distance oracle?

Further reading

[Angelidakis, Makarychev, Oparin 2017]
Algorithmic and hardness results for the hub labeling problem.
SODA 2017.
https://arxiv.org/abs/1611.06605

[Kyng, Meierhans, Gutenberg 2022]
Incremental SSSP for sparse digraphs beyond the hopset barrier.
SODA 2022.
https://arxiv.org/abs/2110.11712