

Compact representation of distances in a graph : a tour around 2-hop labelings

Laurent Viennot (Inria Dyogene/Argo)

Joint work with Siddharth Gupta (Univ. of Warwick),
Adrian Kosowski (NavAlgo)
and Przemysław Uznański (NavAlgo - Univ. of Wrocław)



Encoding distances in a graph

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

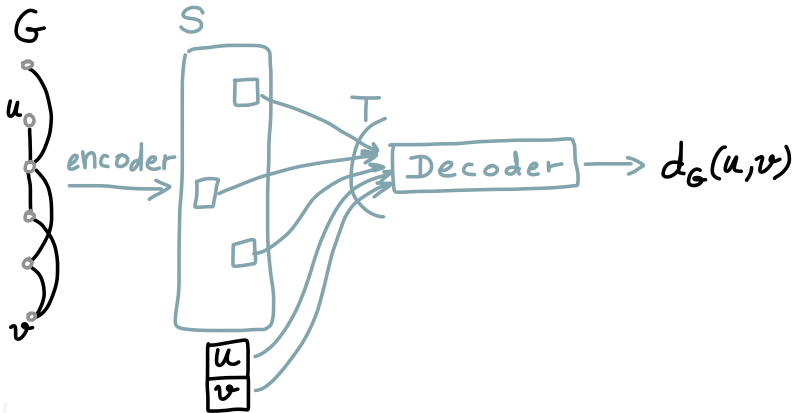
Make any useful pre-computation to answer efficiently online distance queries : what is distance $d(u_1, v_1)$?, $d(u_2, v_2)$?, $d(u_3, v_3)$?, ...

Encoding distances in a graph

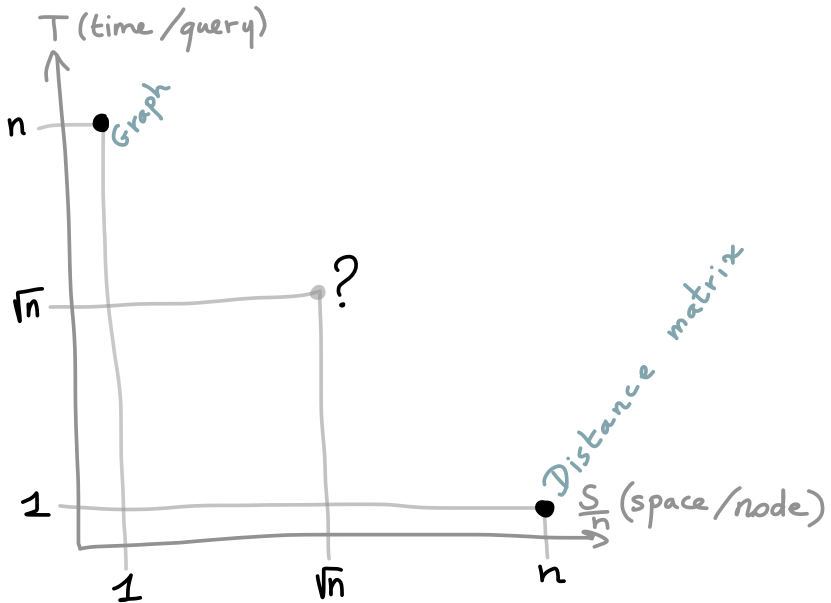
We are given a (weighted) (di-) graph $G = (V, E)$ with n nodes and m edges.

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

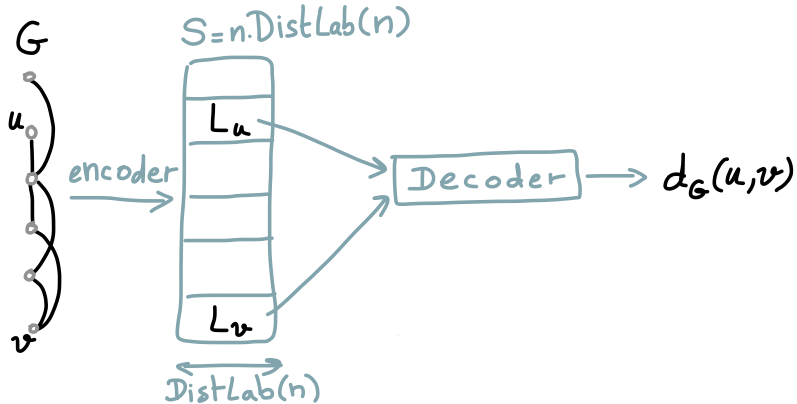
Encoding a graph metric : distance oracles



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



Encoding a graph metric : distance labelings



Encoding a graph metric : 2-hop labelings

A 2-hop labelings is a very simple kind of distance labeling.

The main idea is to associate a set $H_u \subseteq V$ of "hubs" to each node u and to store the distances $d(u, v)$ for all $v \in H_u$.

Also known as hub labeling, or landmark labeling.

One of the two main building blocks of classical distance labelings (the other being tree labelings).

Encoding a graph metric : 2-hop labelings

A 2-hop labelings is a very simple kind of distance labeling.

The main idea is to associate a set $H_u \subseteq V$ of "hubs" to each node u and to store the distances $d(u, v)$ for all $v \in H_u$.

Also known as hub labeling, or landmark labeling.

One of the two main building blocks of classical distance labelings (the other being tree labelings).

Encoding a graph metric : 2-hop labelings

A 2-hop labelings is a very simple kind of distance labeling.

The main idea is to associate a set $H_u \subseteq V$ of "hubs" to each node u and to store the distances $d(u, v)$ for all $v \in H_u$.

Also known as hub labeling, or landmark labeling.

One of the two main building blocks of classical distance labelings (the other being tree labelings).

Encoding a graph metric : 2-hop labelings

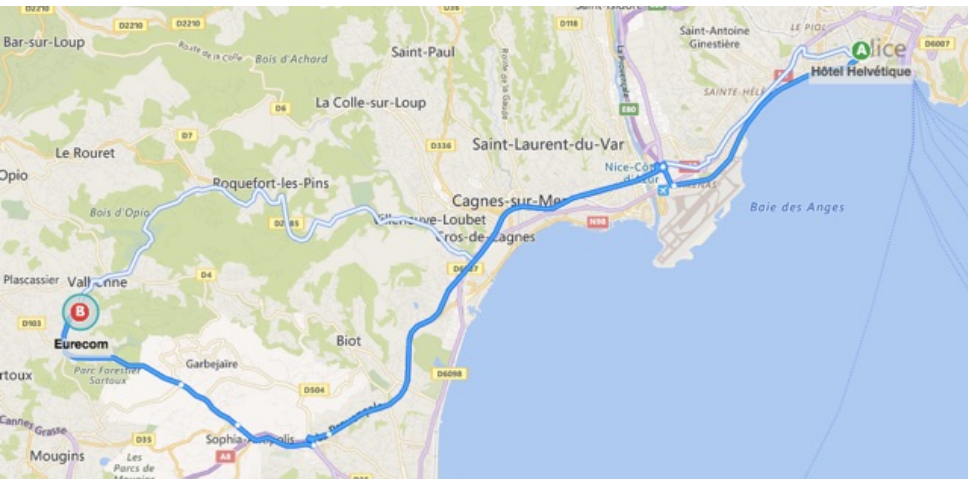
A 2-hop labelings is a very simple kind of distance labeling.

The main idea is to associate a set $H_u \subseteq V$ of "hubs" to each node u and to store the distances $d(u, v)$ for all $v \in H_u$.

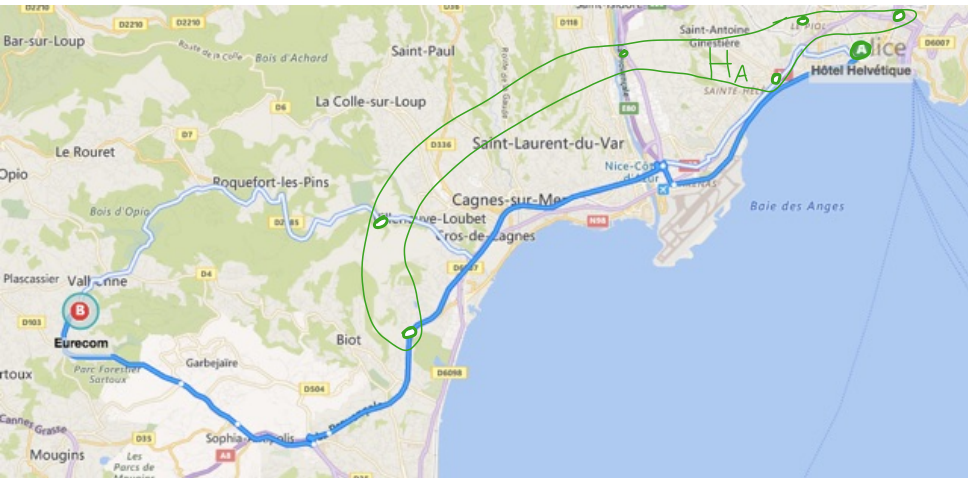
Also known as hub labeling, or landmark labeling.

One of the two main building blocks of classical distance labelings (the other being tree labelings).

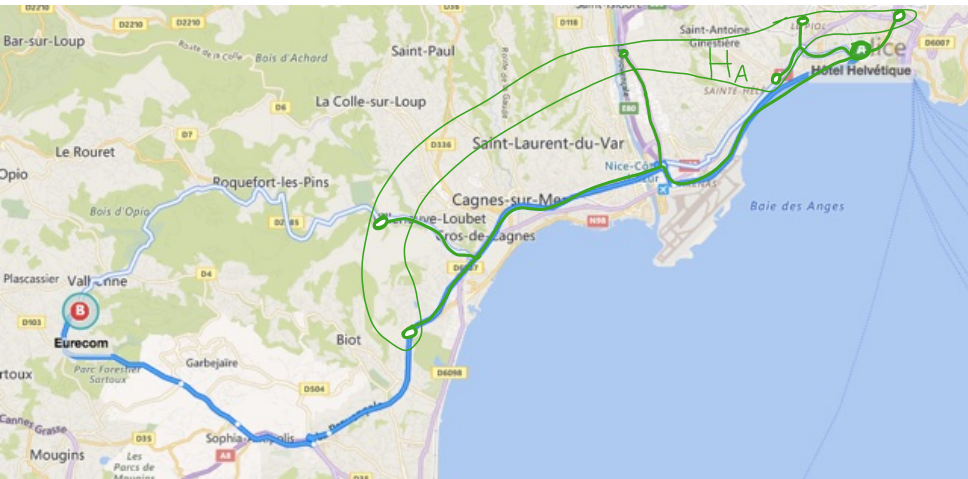
2-hop labeling



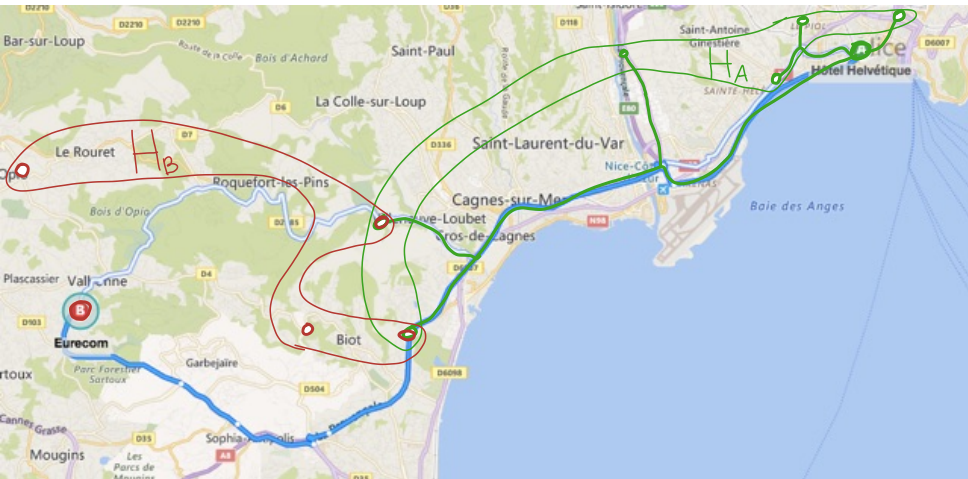
2-hop labeling



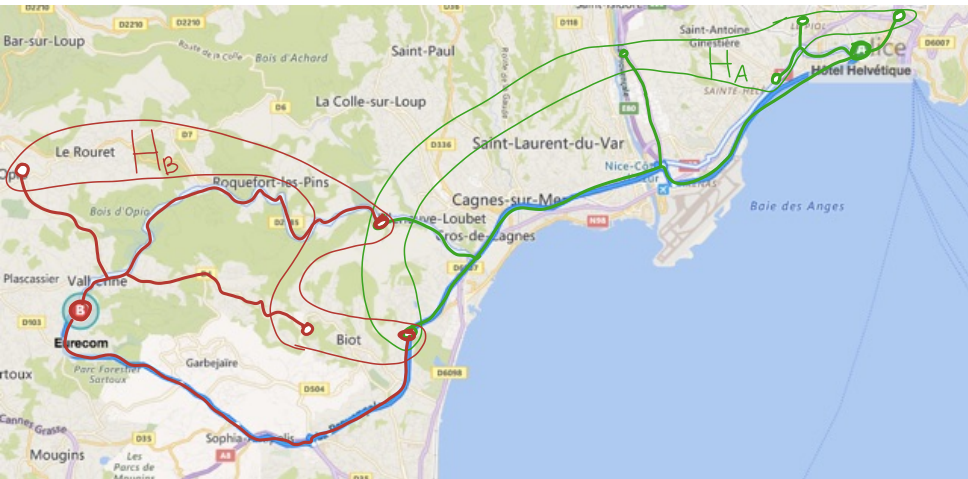
2-hop labeling



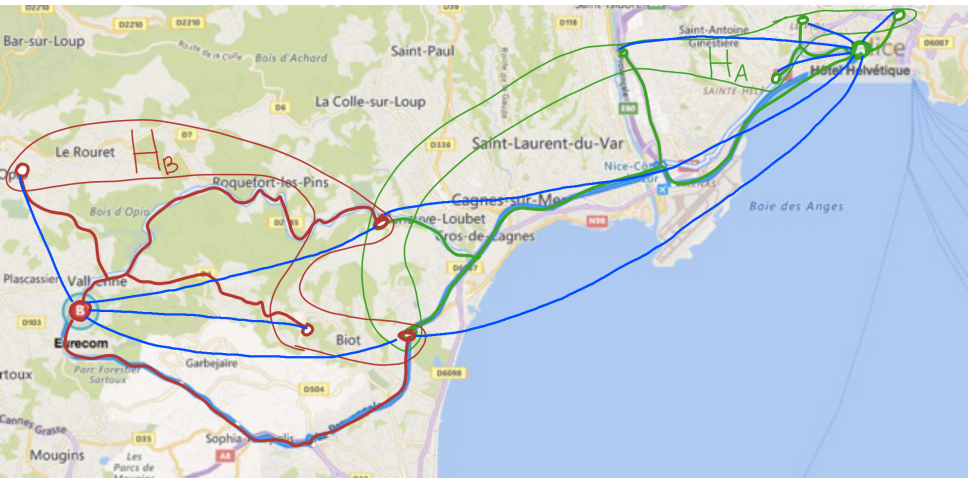
2-hop labeling



2-hop labeling



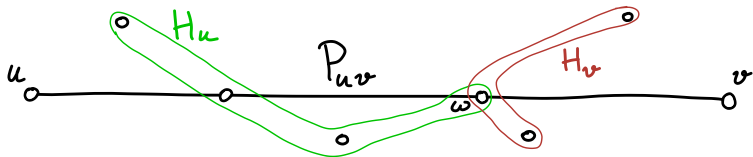
2-hop labeling



Hub sets

Covering property :

A collection of **hub sets** $H_u \subseteq V$ for all $u \in V$ is said to **cover** graph G if for all u, v , there exists $w \in H_u \cap H_v$ with $w \in P_{uv}$, where P_{uv} is a shortest uv -path.



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

Distance query : $\text{Dist}(L_u, L_v) = \min_{w \in H_u \cap H_v} d(u, w) + d(w, v)$

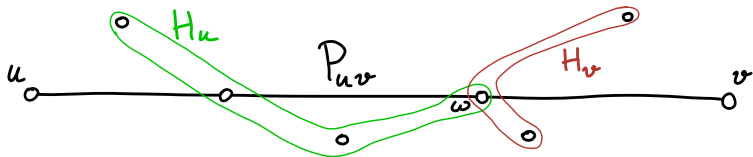
Introduced by [Gavoille et al. '04; Cohen et al. 2003],
applied to road networks [Abraham et al. 2010-2013],
and other practical networks [Akiba et al. 2013].

Approximability results : [Babenko et al. 2013, Angelidakis et al. 2017].

Hub sets

Covering property :

A collection of **hub sets** $H_u \subseteq V$ for all $u \in V$ is said to **cover** graph G if for all u, v , there exists $w \in H_u \cap H_v$ with $w \in P_{uv}$, where P_{uv} is a shortest uv -path.



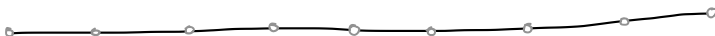
Distance labels : $L_u = \{(w, d(u, w)) : w \in H_u\}$

Distance query : $\text{Dist}(L_u, L_v) = \min_{w \in H_u \cap H_v} d(u, w) + d(w, v)$

Introduced by [Gavoille et al. '04; Cohen et al. 2003],
applied to road networks [Abraham et al. 2010-2013],
and other practical networks [Akiba et al. 2013].

Approximability results : [Babenko et al. 2013, Angelidakis et al. 2017].

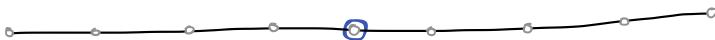
Hub sets covering a path



This results in **covering hub sets** of **size $O(\log n)$** .

A similar construction works for **trees**, **bounded-treewidth graphs** and **planar graphs**.

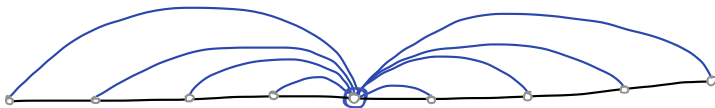
Hub sets covering a path



This results in **covering hub sets** of **size $O(\log n)$** .

A similar construction works for **trees**, **bounded-treewidth graphs** and **planar graphs**.

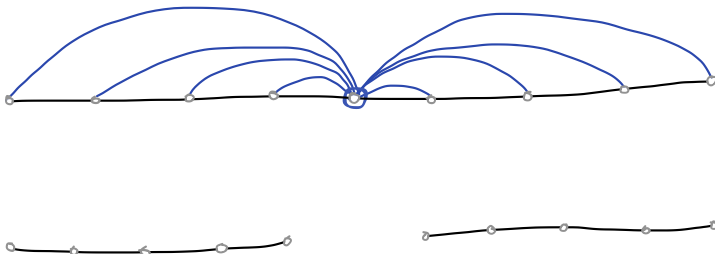
Hub sets covering a path



This results in **covering hub sets** of **size $O(\log n)$** .

A similar construction works for **trees**, **bounded-treewidth graphs** and **planar graphs**.

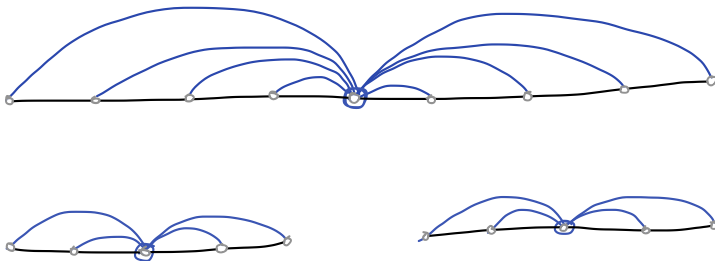
Hub sets covering a path



This results in **covering hub sets** of **size $O(\log n)$** .

A similar construction works for **trees**, **bounded-treewidth graphs** and **planar graphs**.

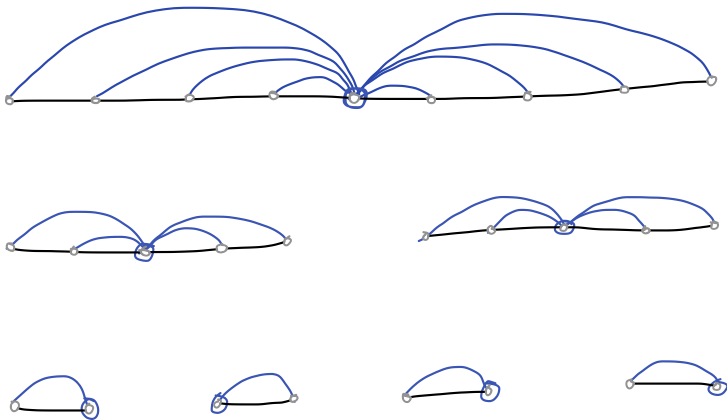
Hub sets covering a path



This results in **covering hub sets** of **size $O(\log n)$** .

A similar construction works for **trees**, **bounded-treewidth graphs** and **planar graphs**.

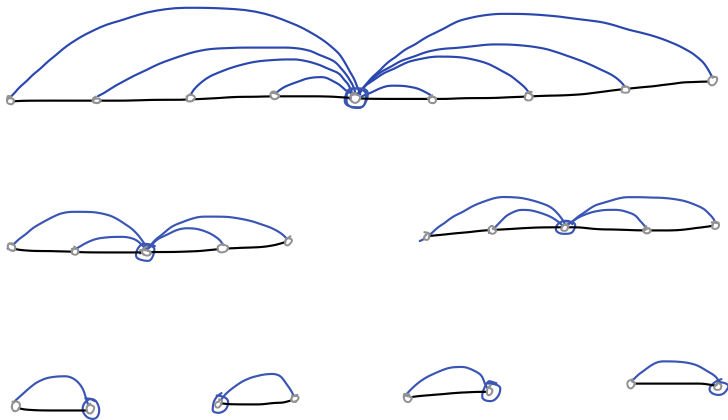
Hub sets covering a path



This results in covering hub sets of size $O(\log n)$.

A similar construction works for trees, bounded-treewidth graphs and planar graphs.

Hub sets covering a path

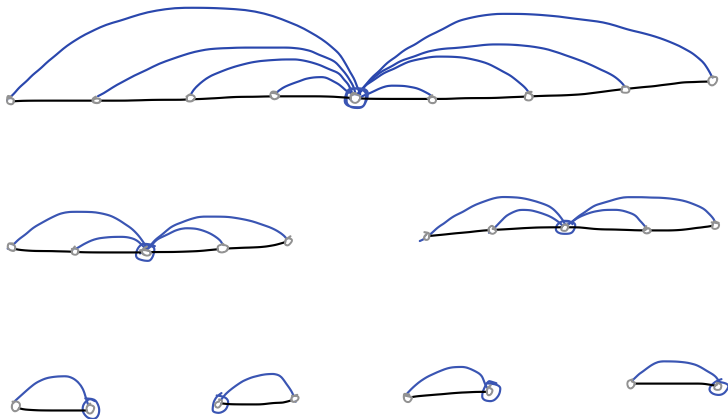


$\log n$
levels

This results in **covering hub sets** of **size $O(\log n)$** .

A similar construction works for **trees**, **bounded-treewidth graphs** and **planar graphs**.

Hub sets covering a path



This results in **covering hub sets** of **size $O(\log n)$** .

A similar construction works for **trees**, **bounded-treewidth graphs** and **planar graphs**.

This talk is about

What **graphs** do have **small hubsets**?

No hope for **dense graphs** :

- **average hub-set size** is at least $\frac{m}{n}$ as :
- for each **edge** $uv \in E$, we must have $u \in H_v$ or $v \in H_u$.

Planar graphs have covering hub sets of **size** $O(\sqrt{n})$, with a best known lower bound of $\Omega(n^{1/3})$ (unweighted). [Gavoille, Peleg, Pérennes, Raz '04].

This talk is about

What **graphs** do have **small hubsets**?

No hope for **dense graphs** :

- **average hub-set size** is at least $\frac{m}{n}$ as :
- for each **edge** $uv \in E$, we must have $u \in H_v$ or $v \in H_u$.

Planar graphs have covering hub sets of **size** $O(\sqrt{n})$, with a best known lower bound of $\Omega(n^{1/3})$ (unweighted). [Gavoille, Peleg, Pérennes, Raz '04].

This talk is about

What **graphs** do have **small hubsets**?

No hope for **dense graphs** :

- **average hub-set size** is at least $\frac{m}{n}$ as :
- for each **edge** $uv \in E$, we must have $u \in H_v$ or $v \in H_u$.

Planar graphs have covering hub sets of **size** $O(\sqrt{n})$, with a best known lower bound of $\Omega(n^{1/3})$ (unweighted). [Gavoille, Peleg, Pérennes, Raz '04].

Part I : Do practical graphs have small covering hub sets?

Yes! practical graphs tend to have small covering hub sets.
[Akiba et al. '13] [Delling et al. 14]

What kind of property they have enables that?

Small highway dimension. [Abraham, Fiat, Goldberg, Werneck '10-13]

More generally, small skeleton dimension. [Kosowski, V., SODA'17]

Yes! practical graphs tend to have small covering hub sets.
[Akiba et al. '13] [Delling et al. 14]

What kind of property they have enables that?

Small highway dimension. [Abraham, Fiat, Goldberg, Werneck '10-13]

More generally, small skeleton dimension. [Kosowski, V., SODA'17]

Yes! practical graphs tend to have small covering hub sets.
[Akiba et al. '13] [Delling et al. 14]

What kind of property they have enables that?

Small highway dimension. [Abraham, Fiat, Goldberg, Werneck
'10-13]

More generally, small skeleton dimension. [Kosowski, V.,
SODA'17]

Yes! practical graphs tend to have small covering hub sets.
[Akiba et al. '13] [Delling et al. 14]

What kind of property they have enables that?

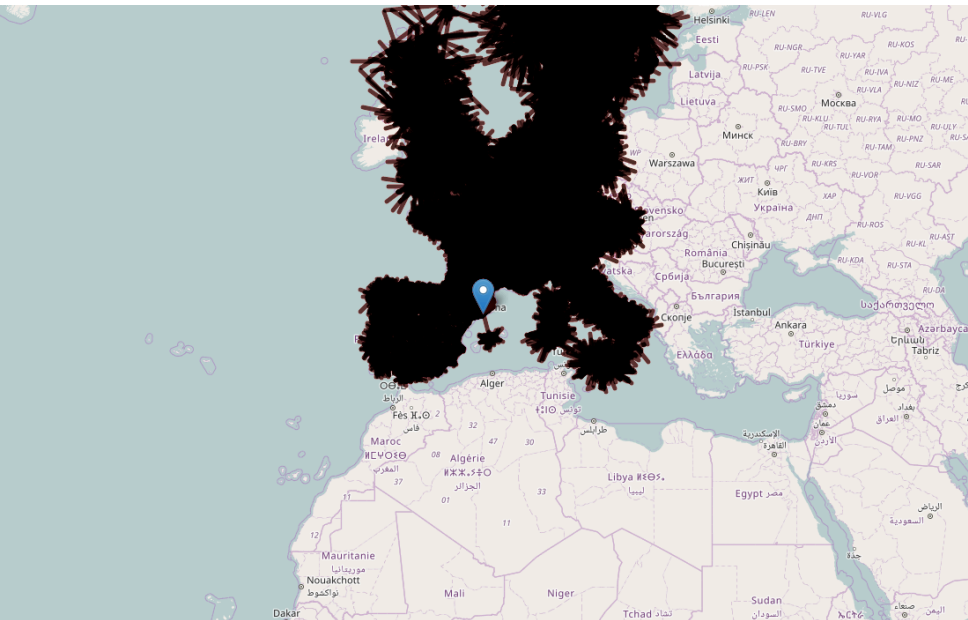
Small highway dimension. [Abraham, Fiat, Goldberg, Werneck '10-13]

More generally, small skeleton dimension. [Kosowski, V., SODA'17]

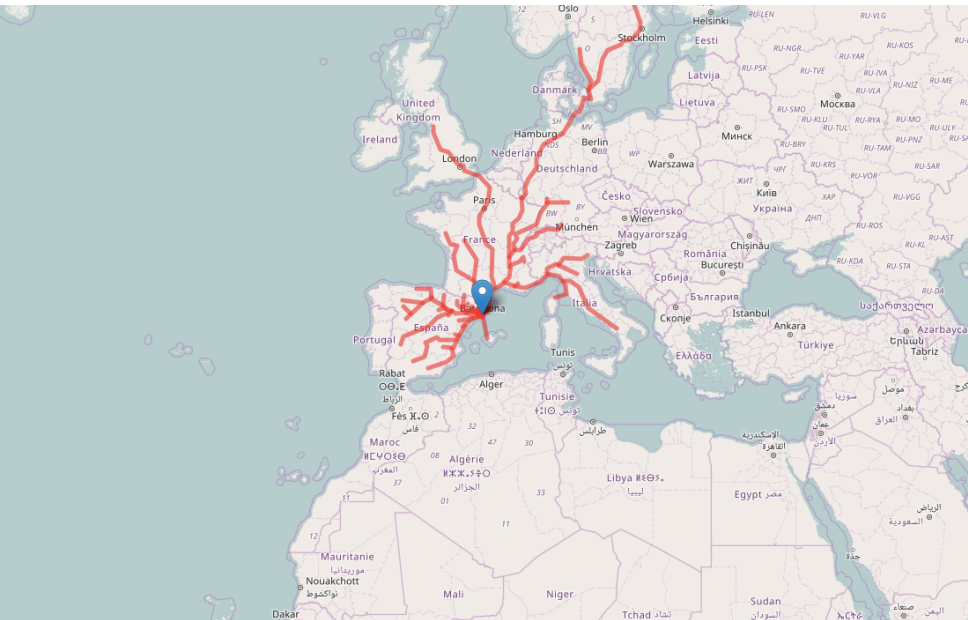
Skeleton dimension

The **skeleton dimension k** of G is the maximum “width” of a “pruned” shortest path tree.

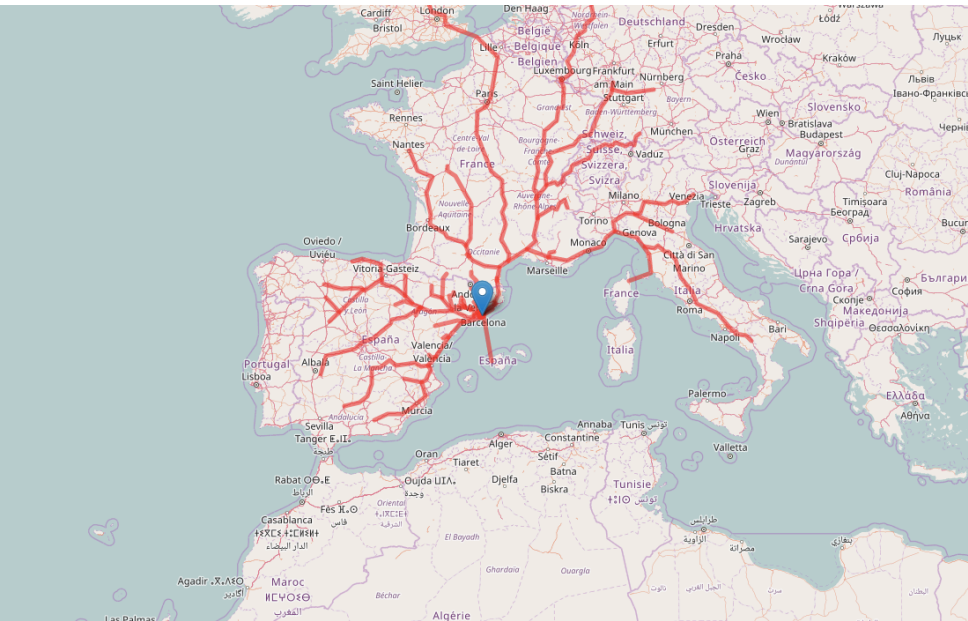
Barcelona shortest path tree



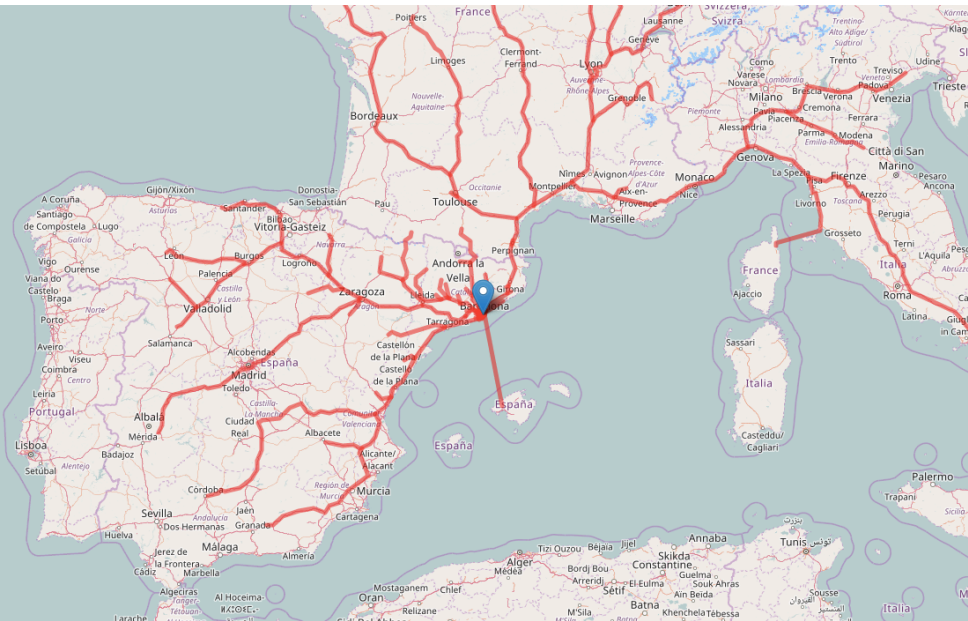
Barcelona tree skeleton : prune last third



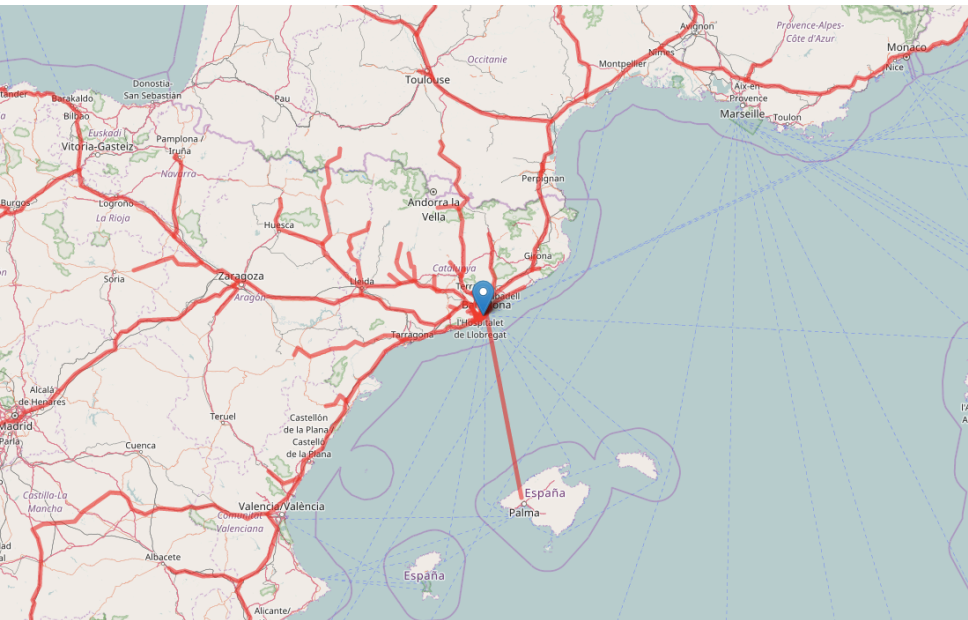
Barcelona tree skeleton : prune last third



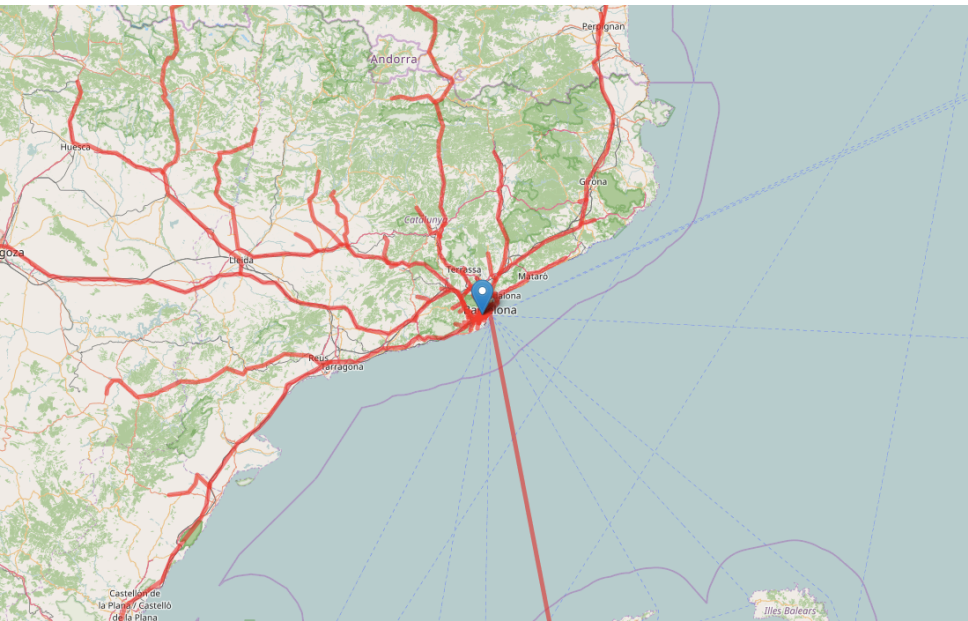
Barcelona tree skeleton : prune last third



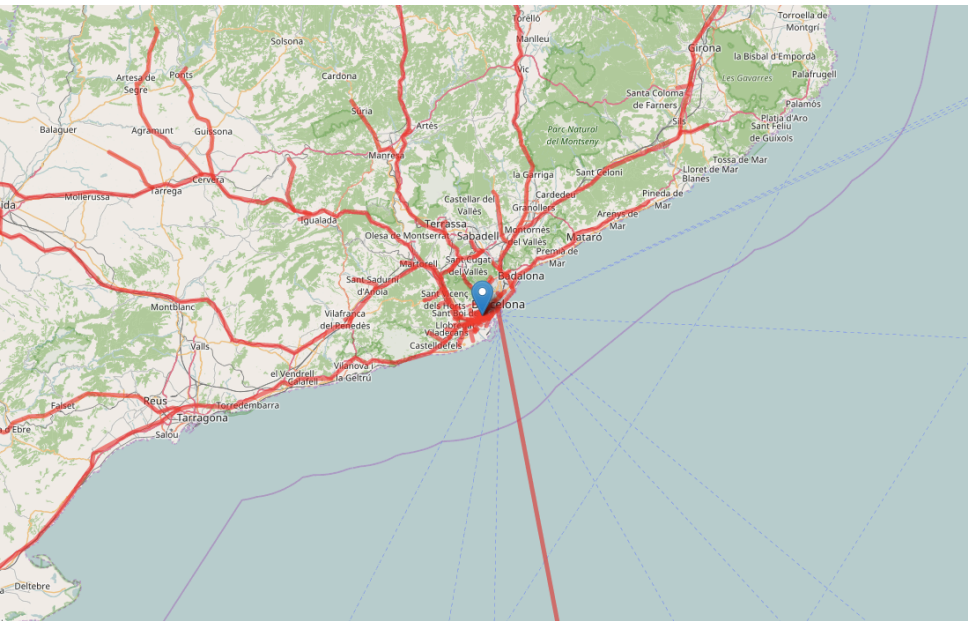
Barcelona tree skeleton : prune last third



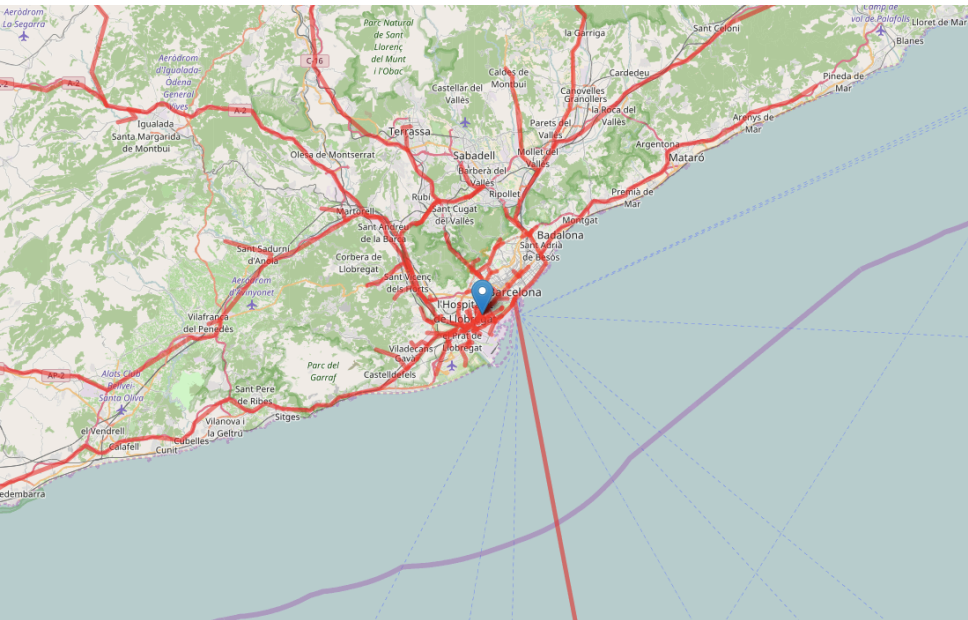
Barcelona tree skeleton : prune last third



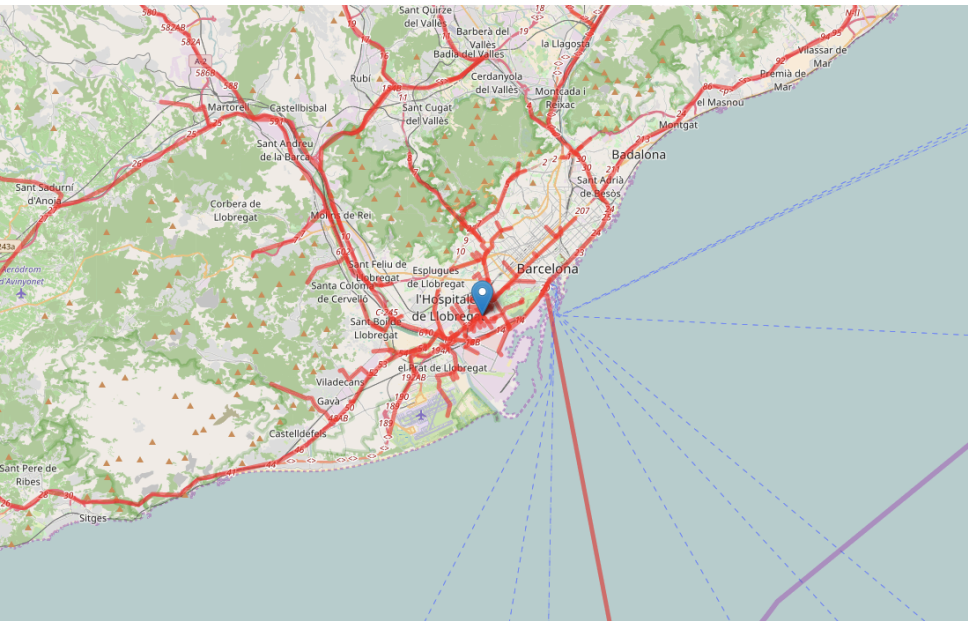
Barcelona tree skeleton : prune last third



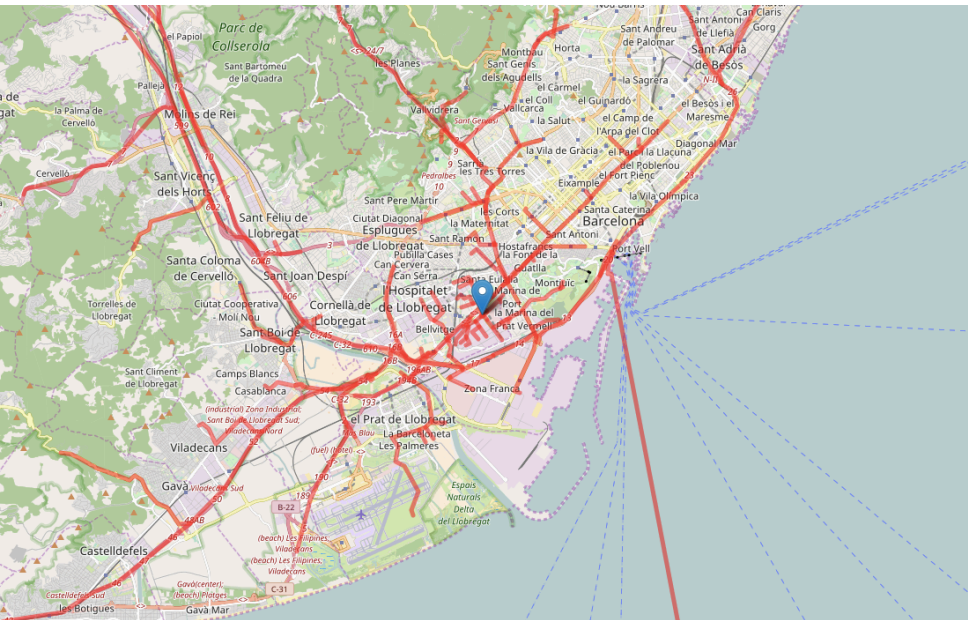
Barcelona tree skeleton : prune last third



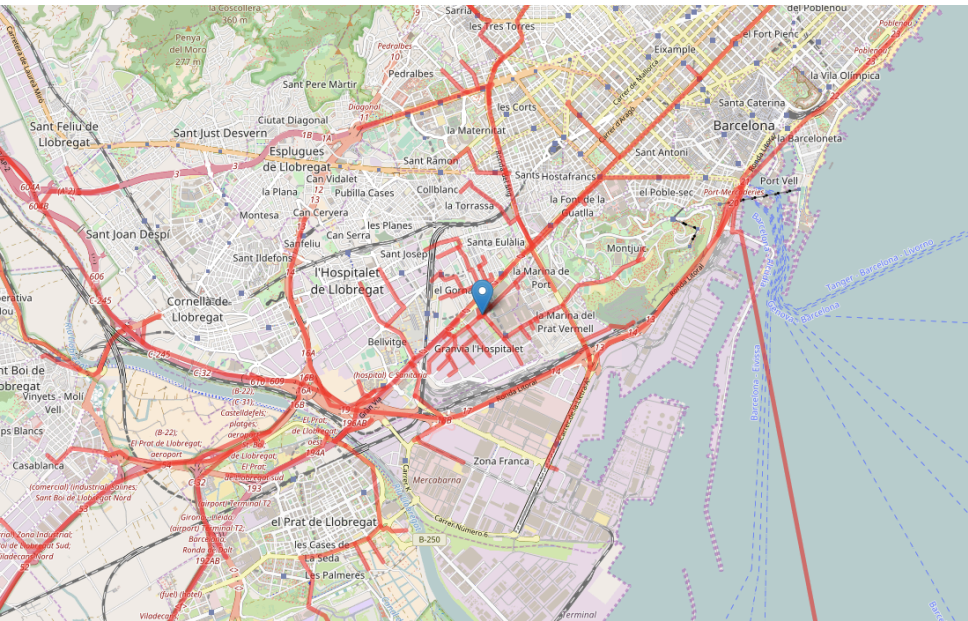
Barcelona tree skeleton : prune last third



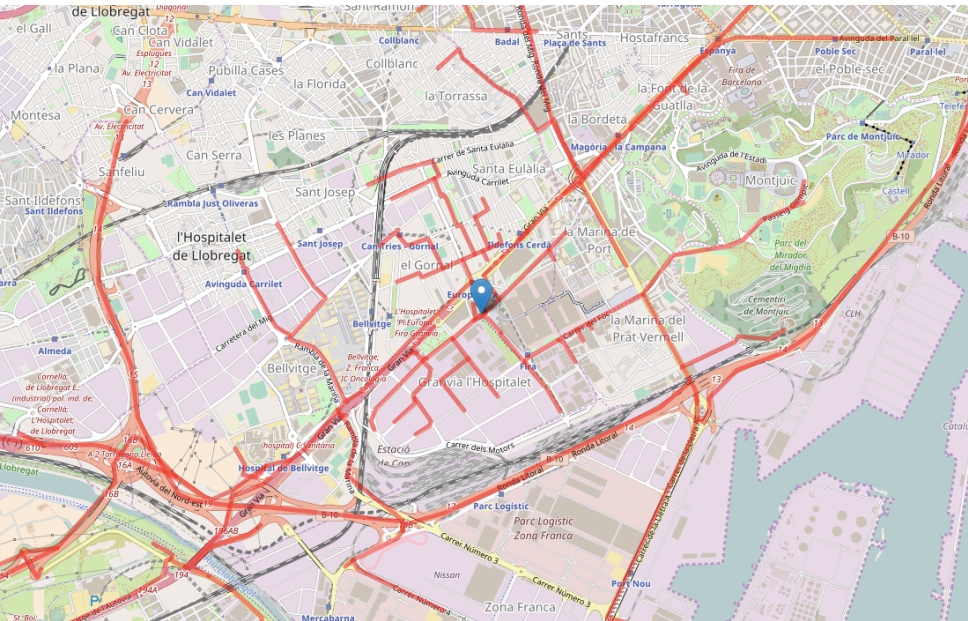
Barcelona tree skeleton : prune last third



Barcelona tree skeleton : prune last third



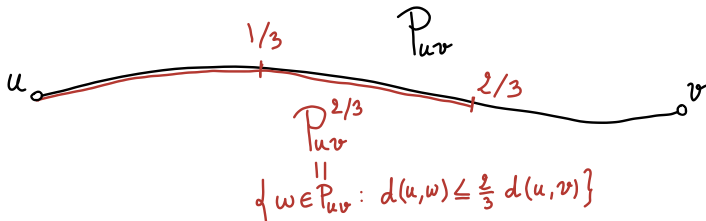
Barcelona tree skeleton : prune last third



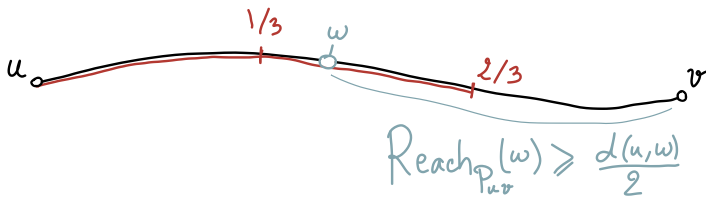
Tree skeleton



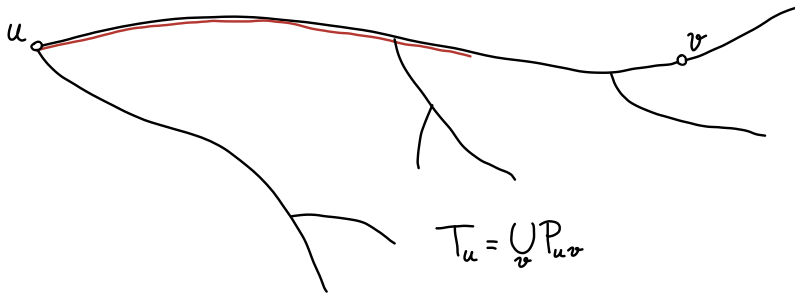
Tree skeleton



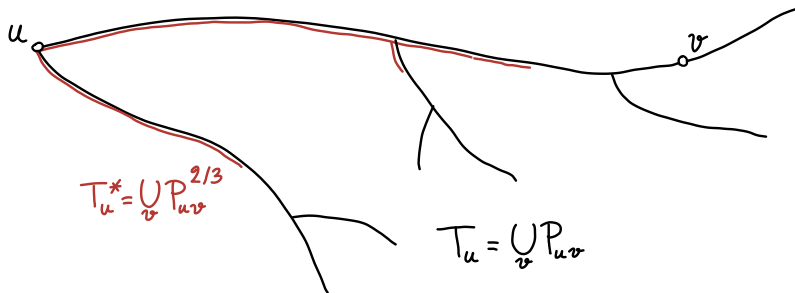
Tree skeleton



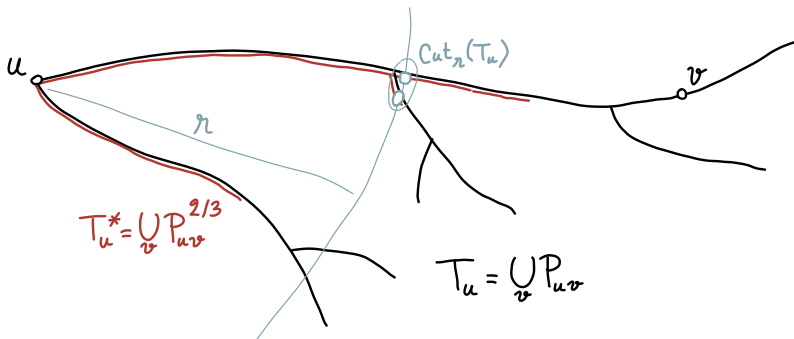
Tree skeleton



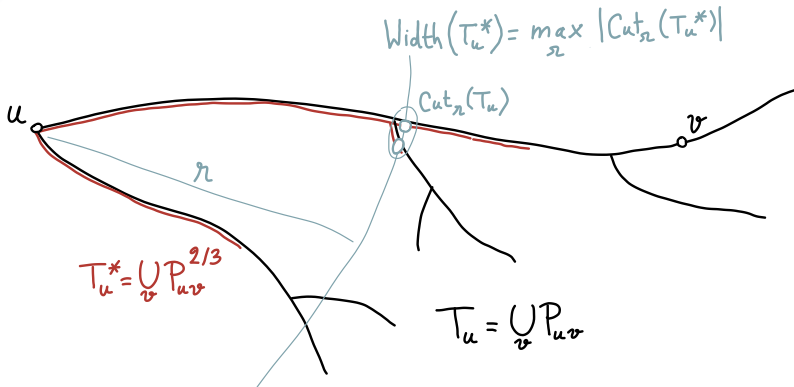
Tree skeleton



Tree skeleton

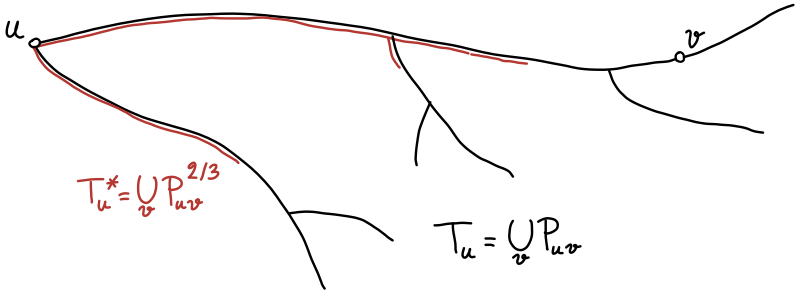


Tree skeleton



Tree skeleton

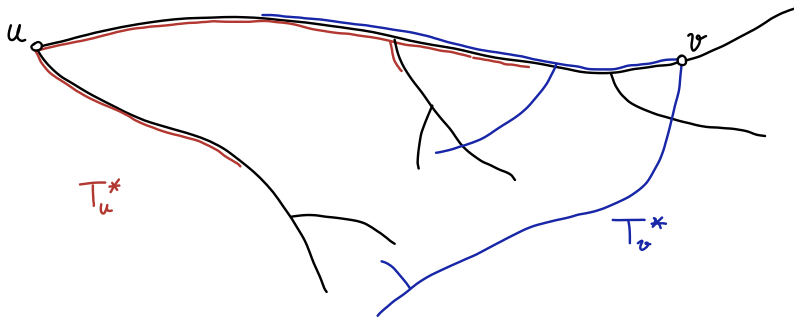
$$\text{skel. dim. } k = \max_u \text{Width}(T_u^*)$$



Theorem (Kosowski, V., SODA'17)

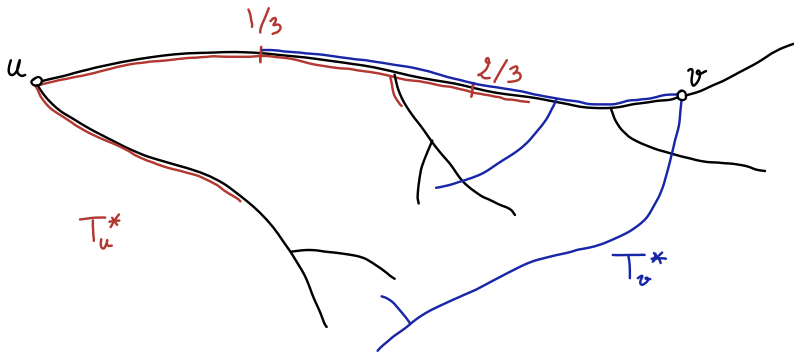
Given a graph G with **skeleton dimension k** and diameter D , a simple random sampling technique allows to find in polynomial time **hub sets** with size $O(k \log D)$ on average and maximum size $O(k \log \log k \log D)$ with high probability.

Hub set selection : random sampling



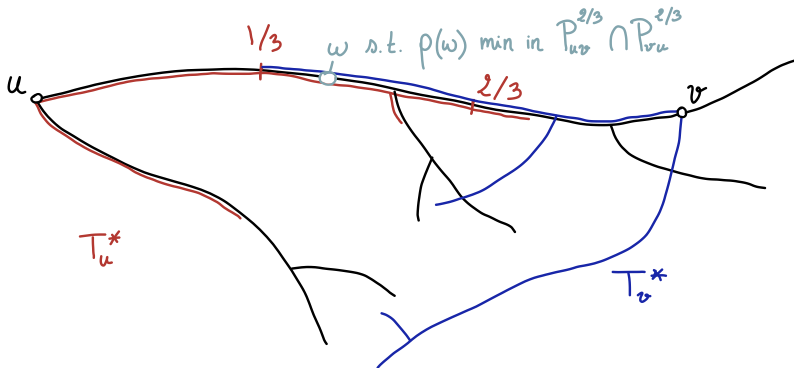
The probability to select a node x is $\propto \frac{1}{d(u,x)}$.

Hub set selection : random sampling



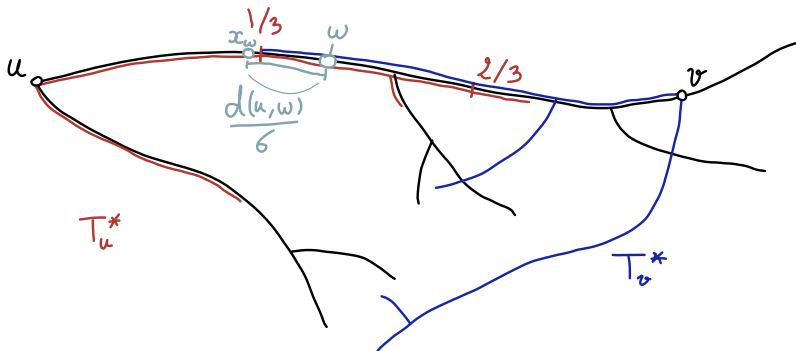
The probability to select a node x is $\propto \frac{1}{d(u,x)}$.

Hub set selection : random sampling



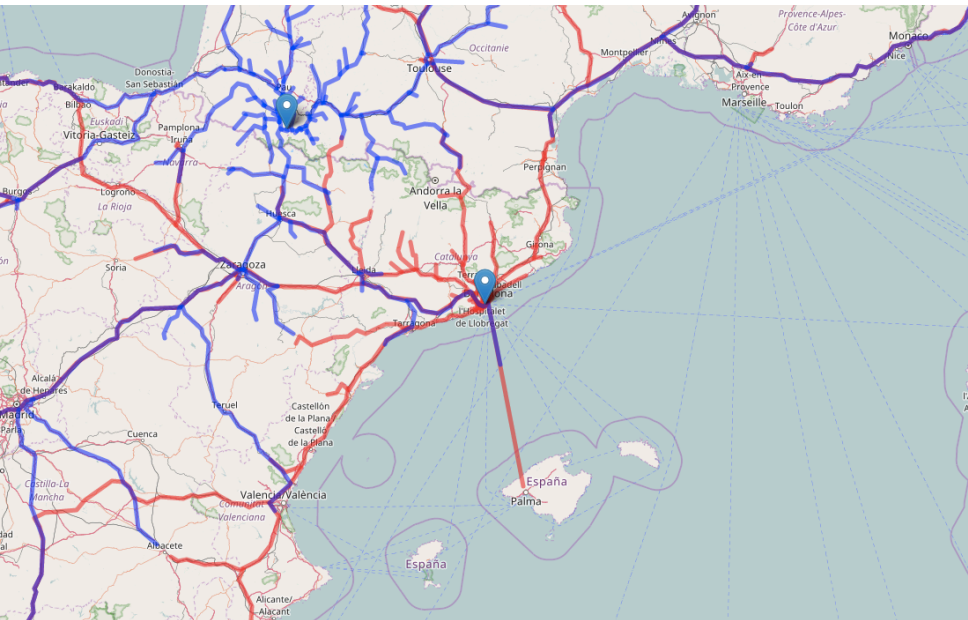
The probability to select a node x is $\propto \frac{1}{d(u,x)}$.

Hub set selection : random sampling

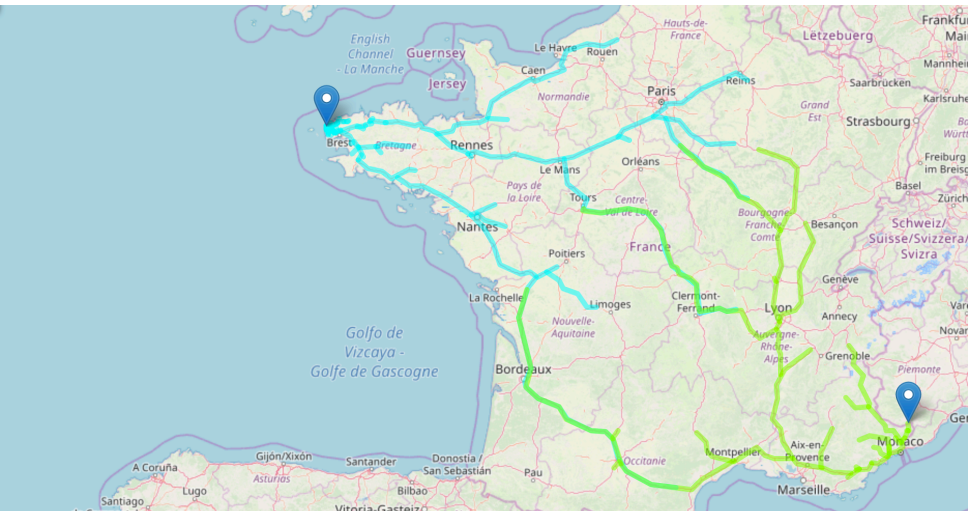


The probability to select a node x is $\propto \frac{1}{d(u, x)}$.

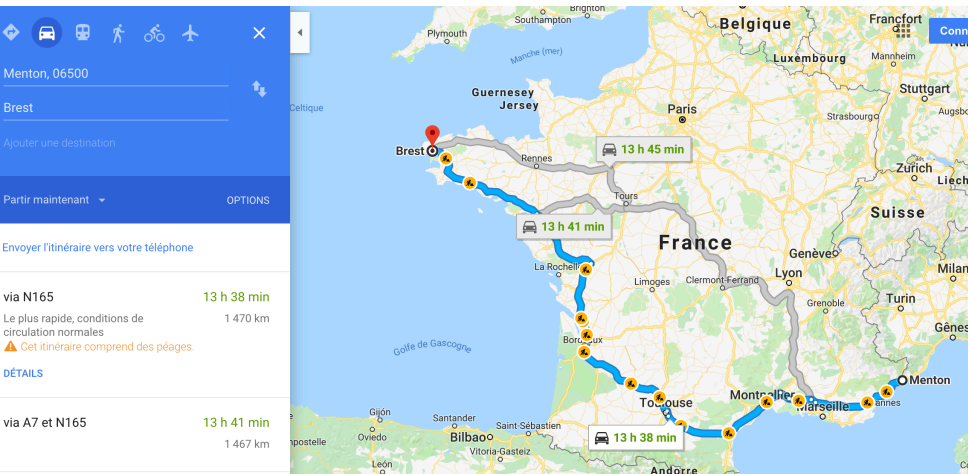
Road networks : two tree skeletons



What ...maps do?



What ...maps do?



What ...maps do?

De Menton à Brest

Options ▼

Menton

Brest

Ajouter une destination

Partir maintenant ▼

OK

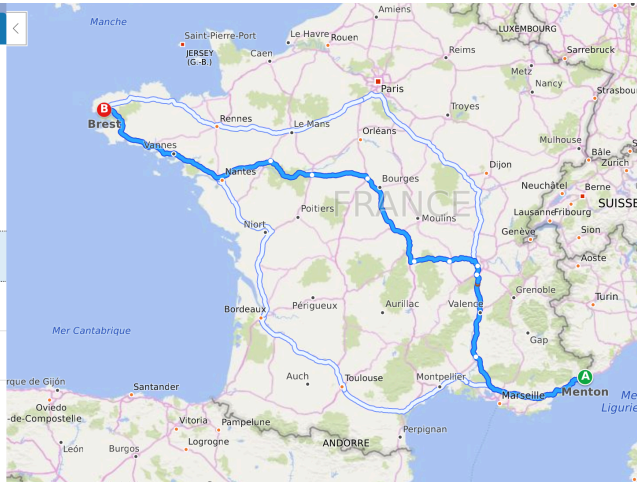
16:08 h min	Trafic modéré Par E80, E60	Retard de 2 h 44 min <i>Itinéraire avec péage</i>	1468 km
16:12 h min	Trafic modéré Par E80, E72	Retard de 2 h 43 min <i>Itinéraire avec péage</i>	1471 km
16:26 h min	Trafic modéré Par E15, E50	Retard de 2 h 39 min <i>Itinéraire avec péage</i>	1509 km

Imprimer

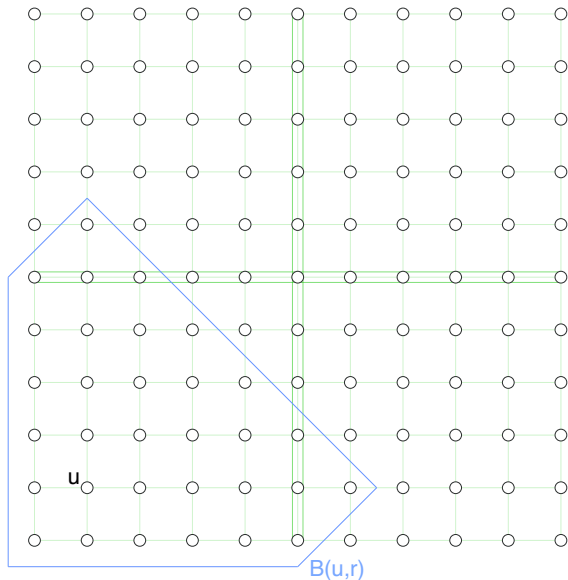
Menton

Quitter Promenade du Soleil / D6007
en direction de Traversée Saint-Michel

0,6 km

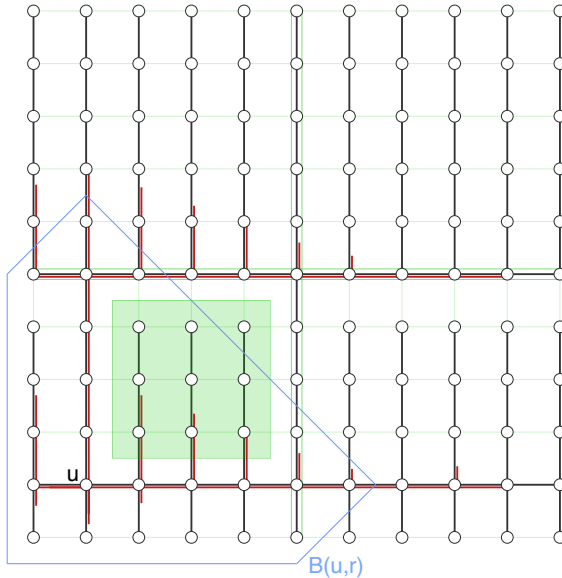


Skeleton dimension of grids



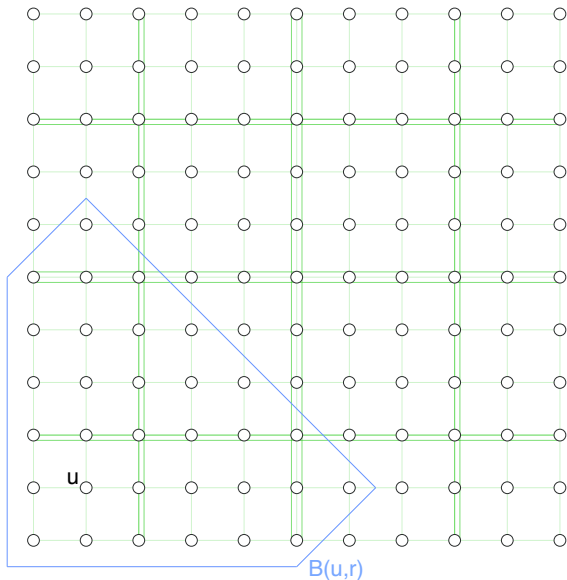
$$k = \Theta(\log n)$$

Skeleton dimension of grids



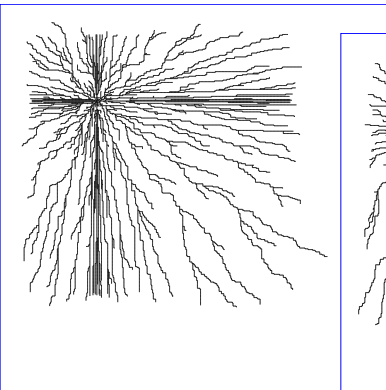
$$k = \Theta(\log n)$$

Skeleton dimension of grids

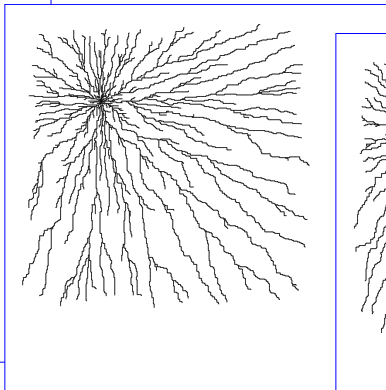


$$k = \Theta(\log n)$$

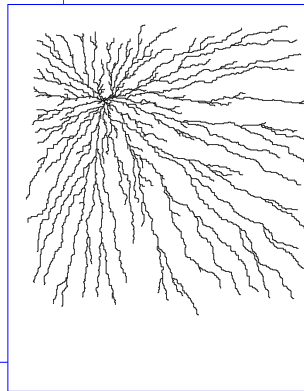
Open : random grid (here 500×500)



$k = 70$



$k = 49$ (fpp $[1, 4]$)



$k = 49$ (prob $2/3$)

Related to **first-passage percolation** [Licea, Newman, Piza '96]
[Aldous '14].

Part II : Do sparse graphs have covering hub sets with $o(n)$ size?

Can we have **sublinear size** for **sparse graphs** ($m = O(n)$)?

Or even **constant degree** graphs?

Best known **upper bound** is $O(\frac{n}{\log n})$. [Alstrup, Dahlgaard, Beck, Knudsen, Porah '16] [Gawrychowski, Kosowski, Uznanski '16]

Best known **lower bound** for distance labels is $\Omega(\sqrt{n})$.
[Gavoille, Peleg, Pérennes, Raz '04]

Can we have **sublinear size** for **sparse graphs** ($m = O(n)$)?

Or even **constant degree** graphs?

Best known **upper bound** is $O(\frac{n}{\log n})$. [Alstrup, Dahlgaard, Beck, Knudsen, Porah '16] [Gawrychowski, Kosowski, Uznanski '16]

Best known **lower bound** for distance labels is $\Omega(\sqrt{n})$.
[Gavoille, Peleg, Pérennes, Raz '04]

Can we have **sublinear size** for **sparse graphs** ($m = O(n)$)?

Or even **constant degree** graphs?

Best known **upper bound** is $O(\frac{n}{\log n})$. [Alstrup, Dahlgaard, Beck, Knudsen, Porah '16] [Gawrychowski, Kosowski, Uznanski '16]

Best known **lower bound** for distance labels is $\Omega(\sqrt{n})$.
[Gavoille, Peleg, Pérennes, Raz '04]

Can we have **sublinear size** for **sparse graphs** ($m = O(n)$)?

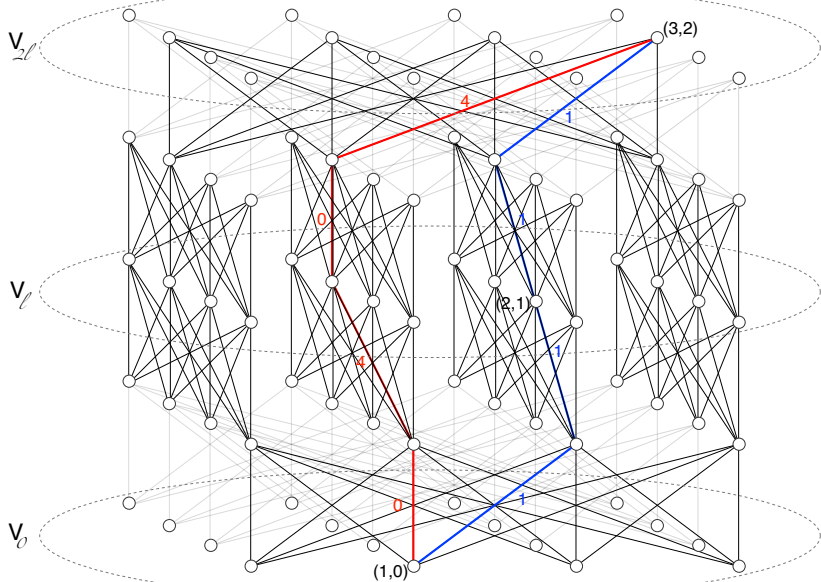
Or even **constant degree** graphs?

Best known **upper bound** is $O(\frac{n}{\log n})$. [Alstrup, Dahlgaard, Beck, Knudsen, Porah '16] [Gawrychowski, Kosowski, Uznanski '16]

Best known **lower bound** for distance labels is $\Omega(\sqrt{n})$.
[Gavoille, Peleg, Pérennes, Raz '04]

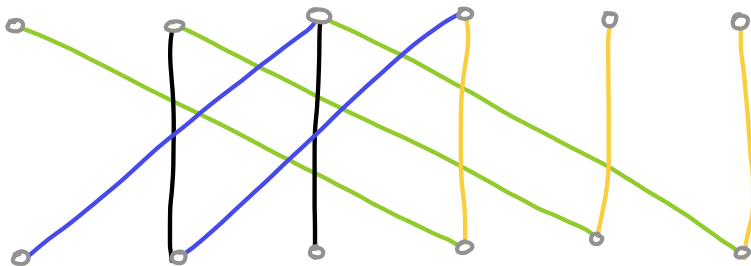
Theorem (Kosowski, Uznański, V., PODC'19)

- (1) There exists graphs of degree at most 3 where any collection of covering hub sets has average size $\frac{n}{2^{O(\sqrt{\log n})}}$.
- (2) Any graph has a collection of hub sets of $O(\frac{n}{RS(n)^{1/7}})$ size where $2^{\Omega(\log^* n)} \leq RS(n) \leq 2^{O(\sqrt{\log n})}$ is a number related to Ruzsa-Szemerédi graphs.



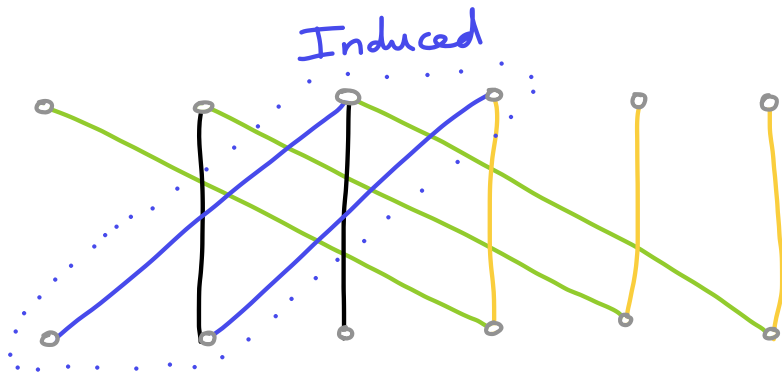
Each V_i is a regular $2^\ell \times \dots \times 2^\ell$ **lattice** of dim. $\ell \approx \sqrt{\log n}$ (here $\ell = 2$).
Edges from V_{i-1} to V_i connect nodes differing on **i th coordinate**.

Ruzsa-Szemerédi



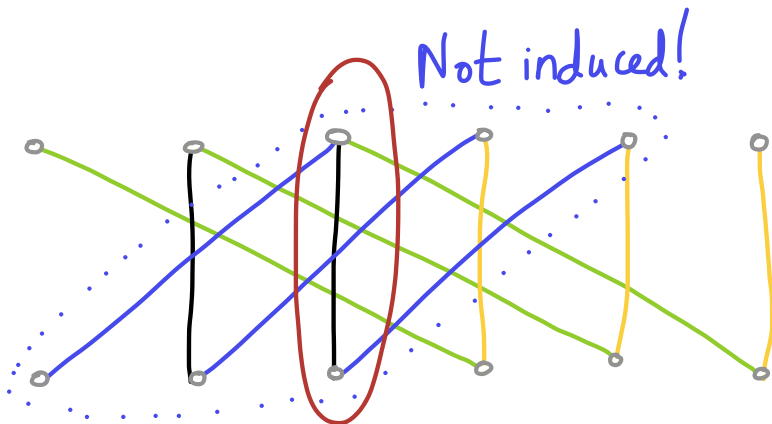
A graph is an **RS-graph** if it can be decomposed into n induced matchings.

Ruzsa-Szemerédi



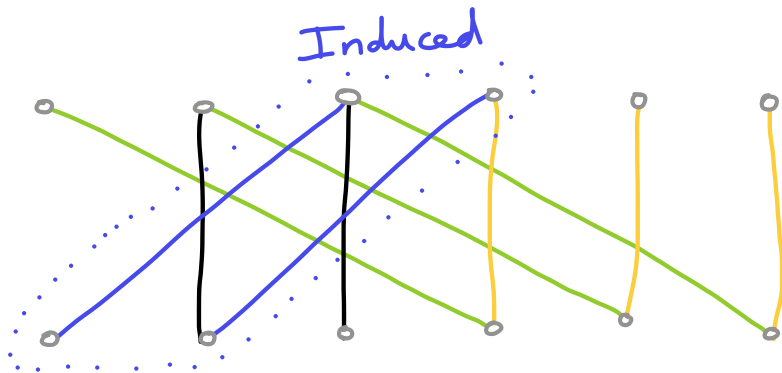
A graph is an **RS-graph** if it can be decomposed into n induced matchings.

Ruzsa-Szemerédi



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

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 $\frac{n^2}{\text{RS}(n)}$ edges.

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

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

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 $\frac{n^2}{\text{RS}(n)}$ edges.

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

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

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 $\frac{n^2}{\text{RS}(n)}$ edges.

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

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

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 $\frac{n^2}{\text{RS}(n)}$ edges.

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

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

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

What are the **densest RS-graphs**?

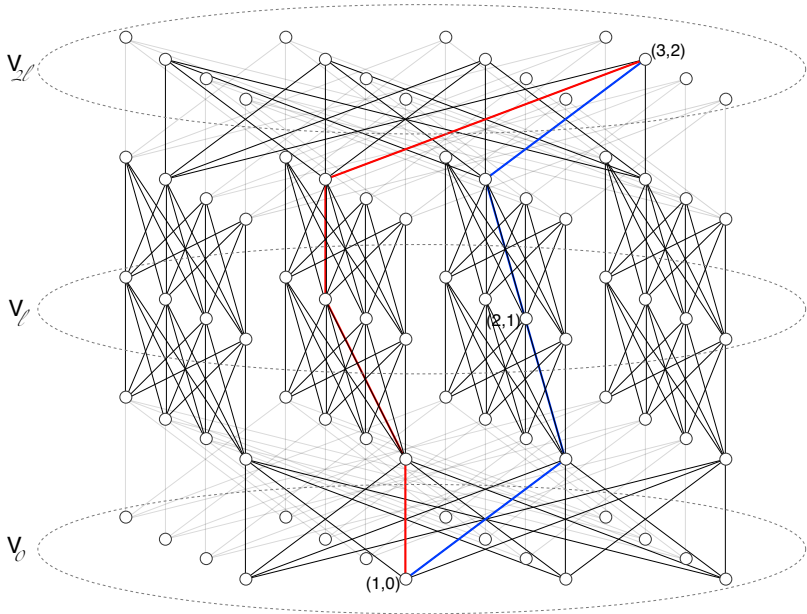
Theorem ([Ruzsa, Szemerédi '78])

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 $\frac{n^2}{\text{RS}(n)}$ edges.

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

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



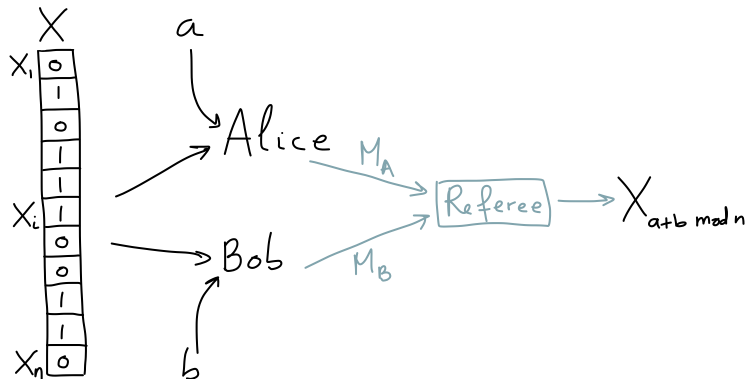
$$G_Y^D = \{x_0 z_{2l} \mid y = \frac{x+z}{2} \text{ and } d_G(x, z) = D\} \quad \exists D \text{ s.t. } |\cup_y G_Y^D| \geq \frac{n^2}{2^{O(\sqrt{\log n})}}$$

Converse

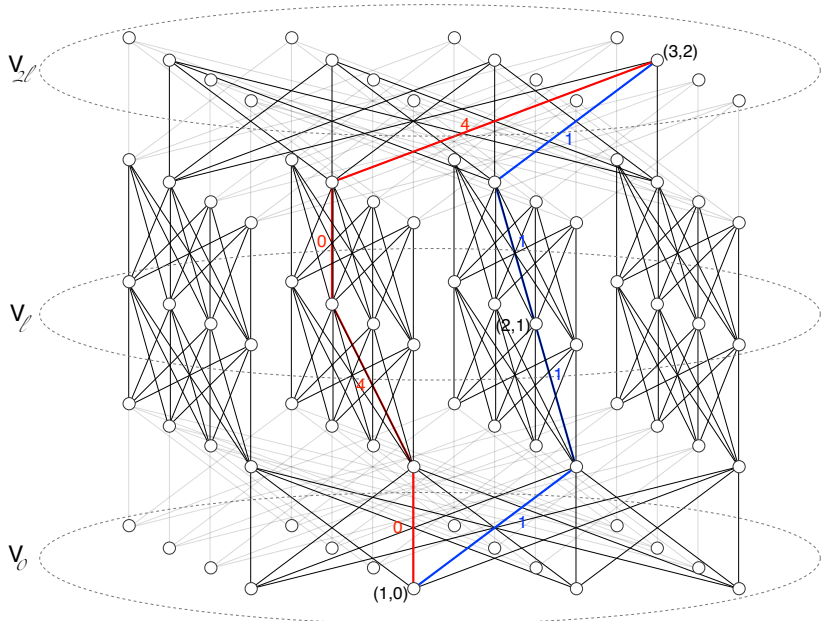
Any cst. deg. graph G has **hub sets** of av. size $O(\frac{n}{RS(n)^{1/7}})$.

Idea : use a **vertex cover** of each ${}^cG_y^D$ ($VC \leq 2MM$).

Connection with SumIndex problem (comm. complexity)



$$\text{SUMINDEX}(n) = \min_{\text{Encoder}} \max_X |M_A| + |M_B|$$
$$\Omega(\sqrt{n}) \leq \text{SUMINDEX}(n) \leq \tilde{O}\left(\frac{n}{2^{\sqrt{\log n}}}\right) \text{ [Pudlak 1994]}$$



$G_X = G \setminus \{y_\ell \mid X_y = 0\}$, send $x = 2a, L_{x_0}, z = 2b, L_{z_{2\ell}}$, check $d(x_0, z_{2\ell})$.

Part III : what about 3 hops?

3-hopset of a path



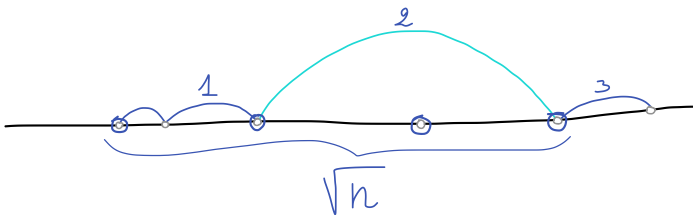
3-hopset of a path



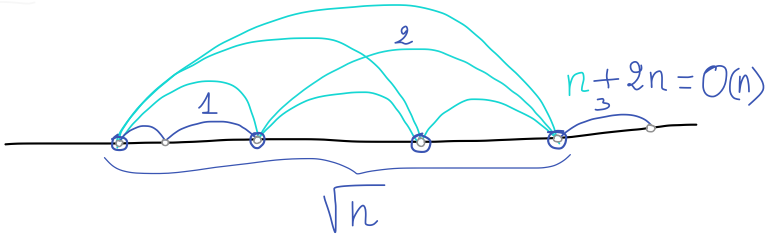
3-hopset of a path



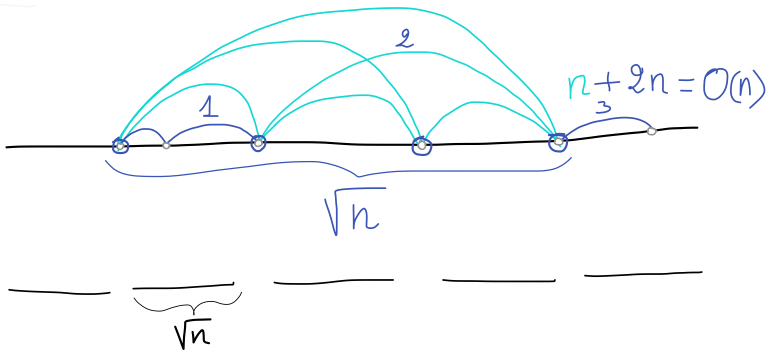
3-hopset of a path



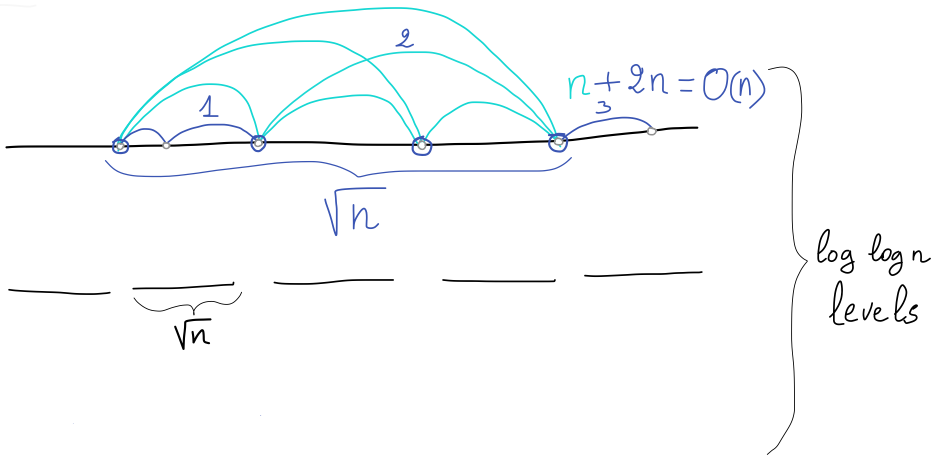
3-hopset of a path



3-hopset of a path



3-hopset of a path



3-hopset distance oracle

Store $x, d_G(u, x)$ for $x \in N_{13}(u)$ ($2 \log \log n$ per node).

Store *middle links* in a hashtable H_2 ($O(n \log \log n)$ size).

Query for $d_G(u, v)$: best 3-hop path length is

$$\min_{x \in N_{13}(u), y \in N_{13}(v), xy \in H_2} d_G(u, x) + d_G(x, y) + d_G(y, v)$$

($O((\log \log n)^2)$ time).

3-hopset distance oracle

Store $x, d_G(u, x)$ for $x \in N_{13}(u)$ ($2 \log \log n$ per node).

Store **middle links** in a hashtable H_2 ($O(n \log \log n)$ size).

Query for $d_G(u, v)$: best 3-hop path length is

$$\min_{x \in N_{13}(u), y \in N_{13}(v), xy \in H_2} d_G(u, x) + d_G(x, y) + d_G(y, v)$$

($O((\log \log n)^2)$ time).

3-hopset distance oracle

Store $x, d_G(u, x)$ for $x \in N_{13}(u)$ ($2 \log \log n$ per node).

Store **middle links** in a hashtable H_2 ($O(n \log \log n)$ size).

Query for $d_G(u, v)$: best 3-hop path length is

$$\min_{x \in N_{13}(u), y \in N_{13}(v), xy \in H_2} d_G(u, x) + d_G(x, y) + d_G(y, v)$$

($O((\log \log n)^2)$ time).

Theorem (Kosowski, Gupta, V., ICALP'19)

For a unique-shortest-path graph with **skeleton dimension** k and average **link length** $L \geq 1$, there exists a randomized construction of a **3-hopset distance oracle** of size $|H| = O(nk \log k (\log \log n + \log L))$, which performs distance queries in expected time $O(k^2 \log^2 k (\log^2 \log n + \log^2 L))$.

Some questions

Improve lower-bounds on **sparse** graphs for **general** distance labelings/oracles.

What is the **skeleton dimension** of a **random grid**?

What graphs have **covering hub sets** of **size $O(1)$** ?

What if the graph evolves with time (**temporal graphs**)?

Some questions

Improve lower-bounds on **sparse** graphs for **general** distance labelings/oracles.

What is the **skeleton dimension** of a **random grid**?

What graphs have **covering hub sets** of size $O(1)$?

What if the graph evolves with time (**temporal graphs**)?

Some questions

Improve lower-bounds on **sparse** graphs for **general** distance labelings/oracles.

What is the **skeleton dimension** of a **random grid**?

What graphs have **covering hub sets** of **size** $O(1)$?

What if the graph evolves with time (**temporal graphs**)?

Some questions

Improve lower-bounds on **sparse** graphs for **general** distance labelings/oracles.

What is the **skeleton dimension** of a **random grid**?

What graphs have **covering hub sets** of **size** $O(1)$?

What if the graph **evolves** with time (**temporal graphs**)?

Thanks.