Probabilities and Provenance on Trees and Treelike Instances

Antoine Amarilli[1], Pierre Bourhis[2], Pierre Senellart[1,3,4]

September 7th, 2016

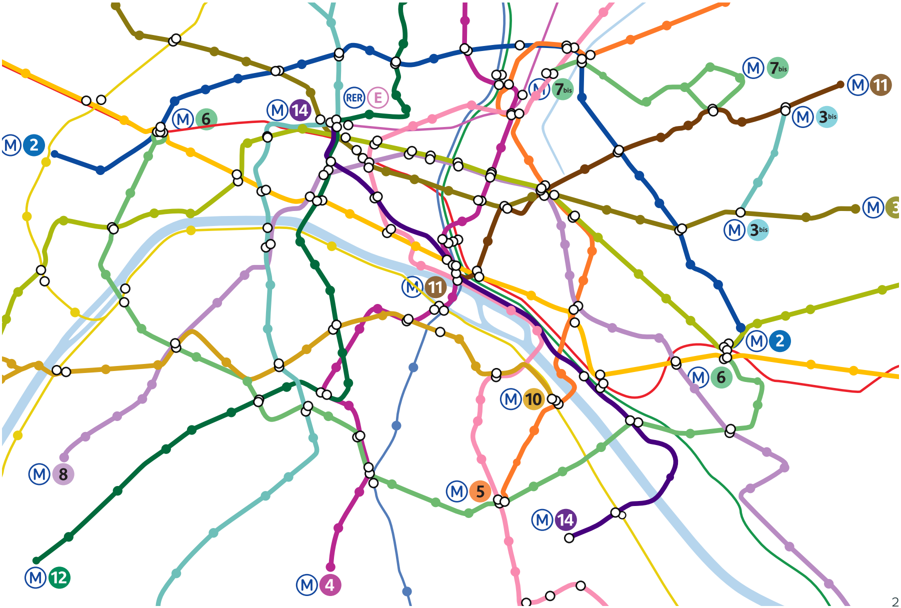
[1]**Télécom ParisTech**

[2]CNRS CRIStAL

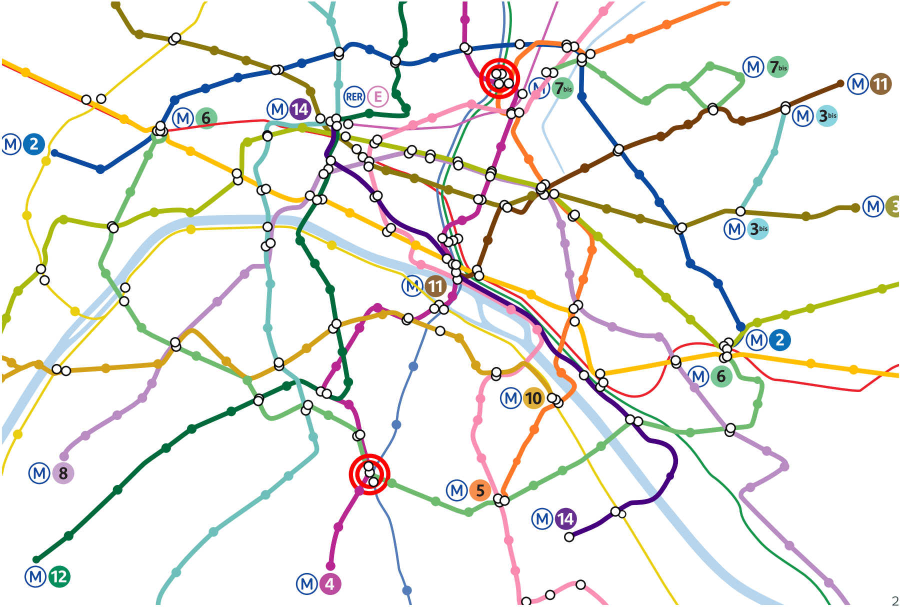
[3]National University of Singapore

[4]École normale supérieure de Paris

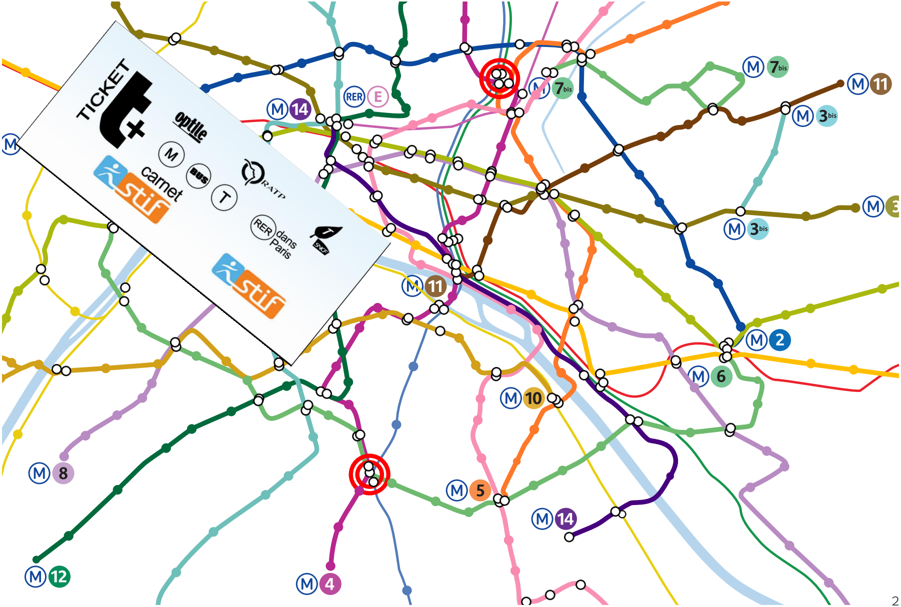
How to travel to Highlights from Paris?



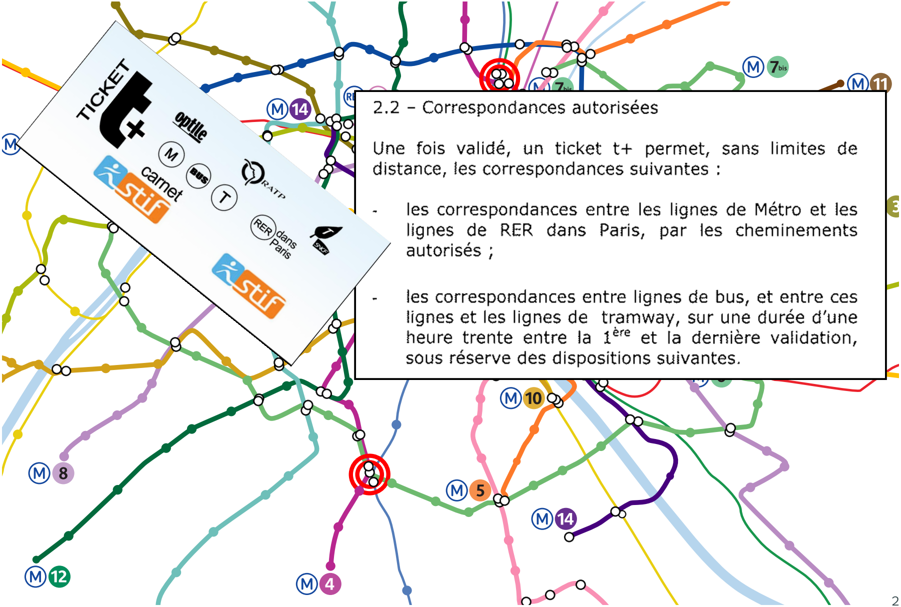
How to travel to Highlights from Paris?



How to travel to Highlights from Paris?



How to travel to Highlights from Paris?



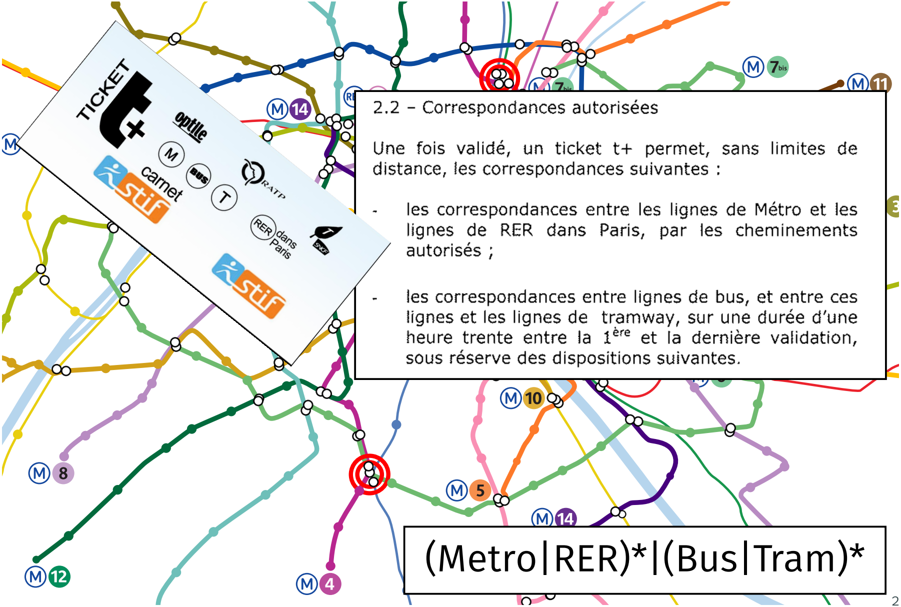
2.2 – Correspondances autorisées

Une fois validé, un ticket t+ permet, sans limites de distance, les correspondances suivantes :

- les correspondances entre les lignes de Métro et les lignes de RER dans Paris, par les cheminements autorisés ;

- les correspondances entre lignes de bus, et entre ces lignes et les lignes de tramway, sur une durée d'une heure trente entre la 1ère et la dernière validation, sous réserve des dispositions suivantes.

How to travel to Highlights from Paris?



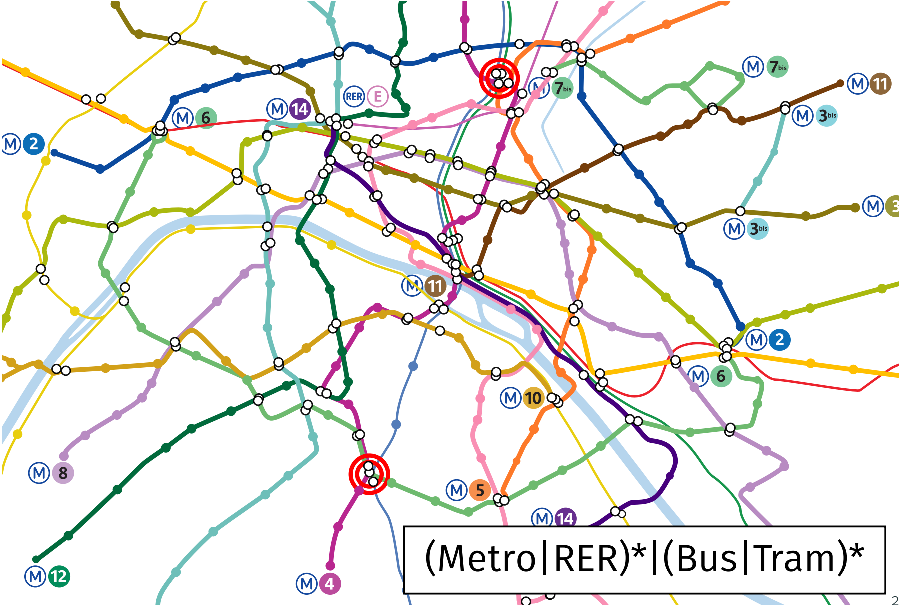
2.2 – Correspondances autorisées

Une fois validé, un ticket t+ permet, sans limites de distance, les correspondances suivantes :

- les correspondances entre les lignes de Métro et les lignes de RER dans Paris, par les cheminements autorisés ;
- les correspondances entre lignes de bus, et entre ces lignes et les lignes de tramway, sur une durée d'une heure trente entre la 1ère et la dernière validation, sous réserve des dispositions suivantes.

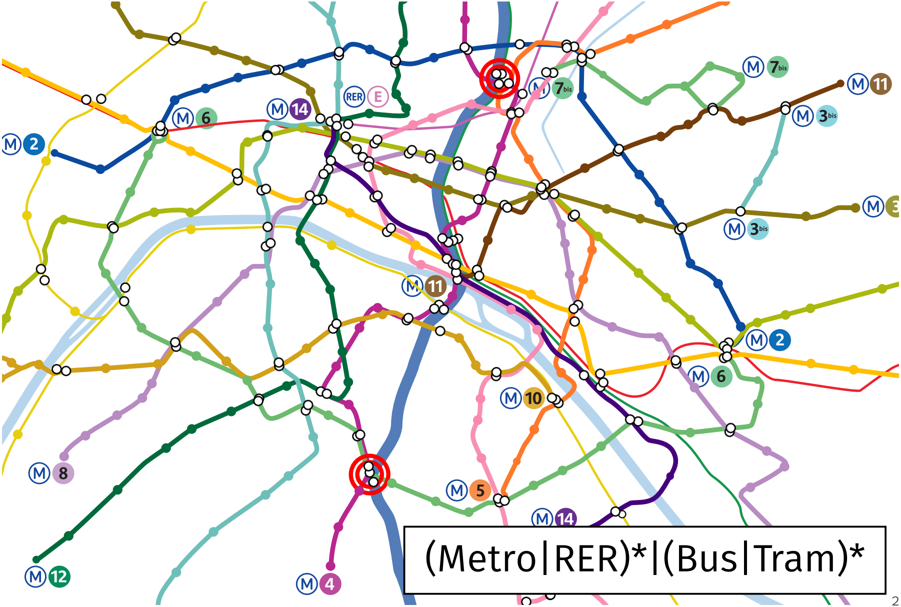
(Metro|RER)*|(Bus|Tram)*

How to travel to Highlights from Paris?



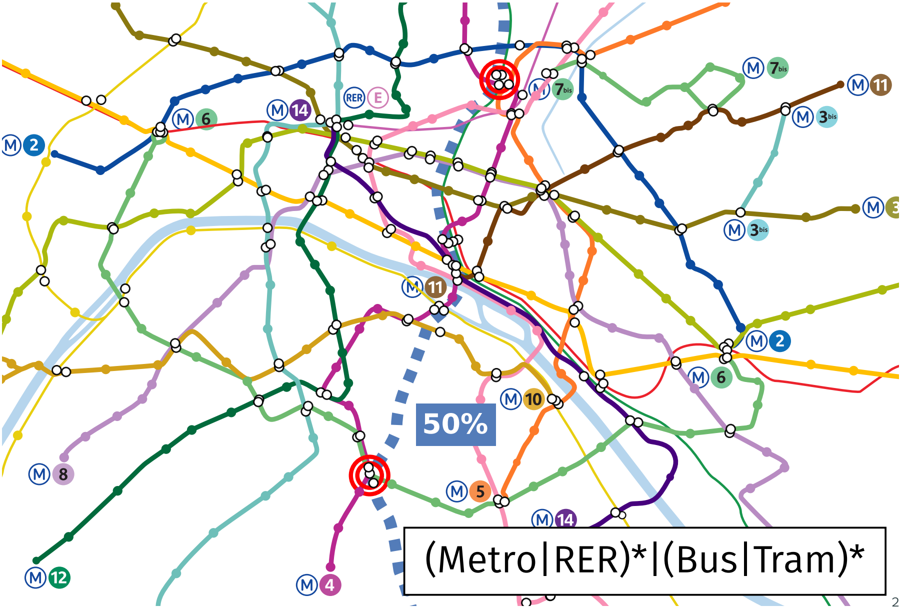
(Metro|RER)*|(Bus|Tram)*

How to travel to Highlights from Paris?

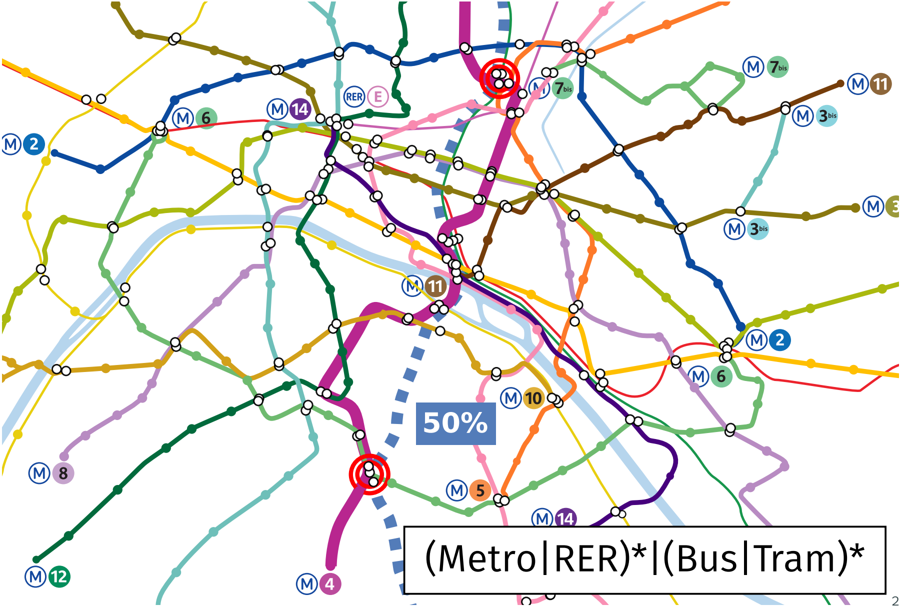


(Metro|RER)*|(Bus|Tram)*

How to travel to Highlights from Paris?

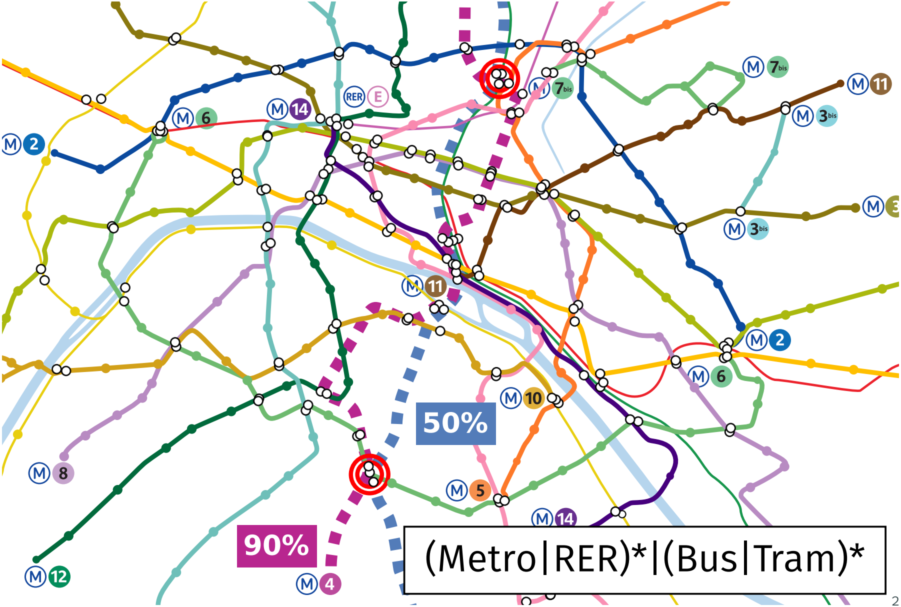


How to travel to Highlights from Paris?



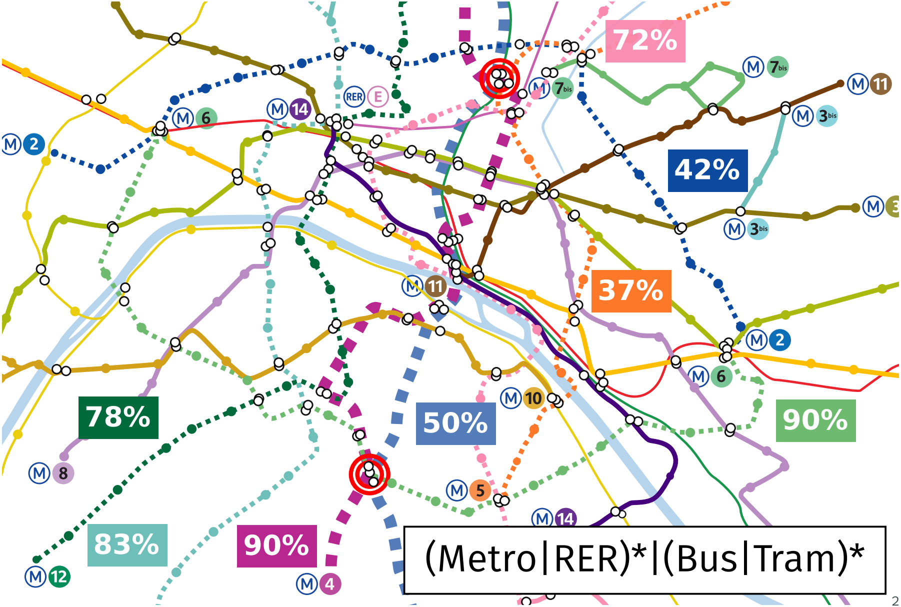
(Metro|RER)*|(Bus|Tram)*

How to travel to Highlights from Paris?



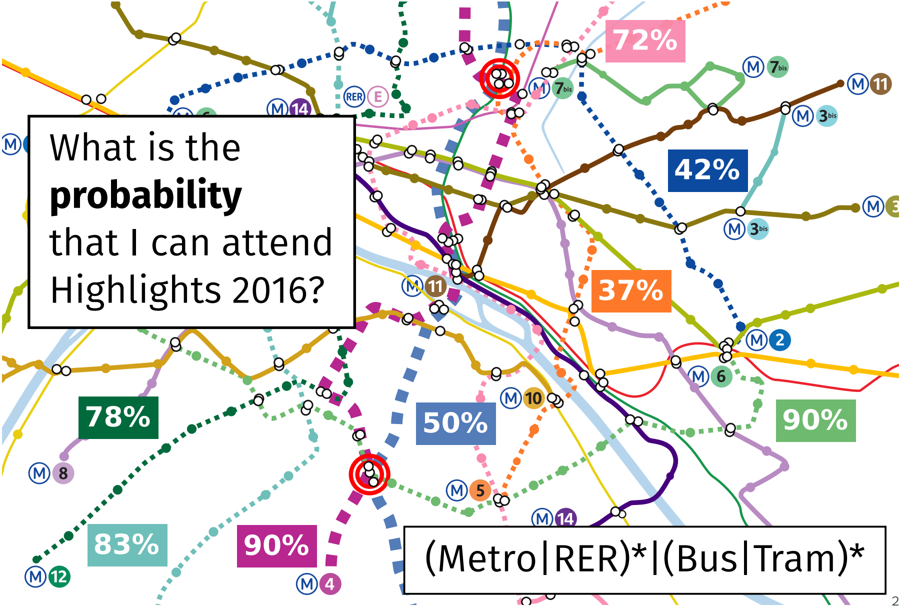
(Metro|RER)*|(Bus|Tram)*

How to travel to Highlights from Paris?



(Metro|RER)*|(Bus|Tram)*

How to travel to Highlights from Paris?



What is the **probability** that I can attend Highlights 2016?

(Metro|RER)*|(Bus|Tram)*

Problem statement

Input:

- Query Q (Metro|RER)*|(Bus|Tram)*

Problem statement

Input:

- Query Q — (Metro|RER)*|(Bus|Tram)*
- Database D or graph

Problem statement

Input:

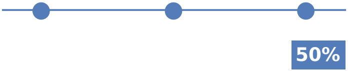
- Query *Q* — (Metro|RER)*|(Bus|Tram)*
- Database *D* or graph

- Probabilities on facts or edges — 50%

Problem statement

Input:

- Query Q — (Metro|RER)*|(Bus|Tram)*
- Database D or graph
- Probabilities on facts or edges

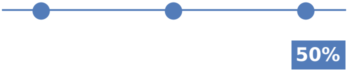


Output: the probability that the query is true under the distribution (assuming independence of all probabilistic events)

Problem statement

Input:

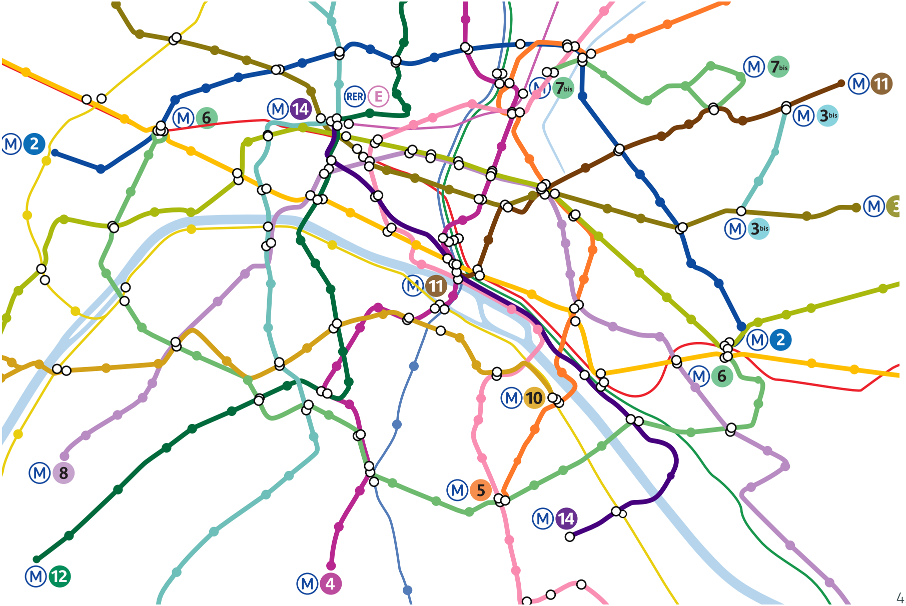
- Query Q — (Metro|RER)*|(Bus|Tram)*
- Database D or graph
- Probabilities on facts or edges



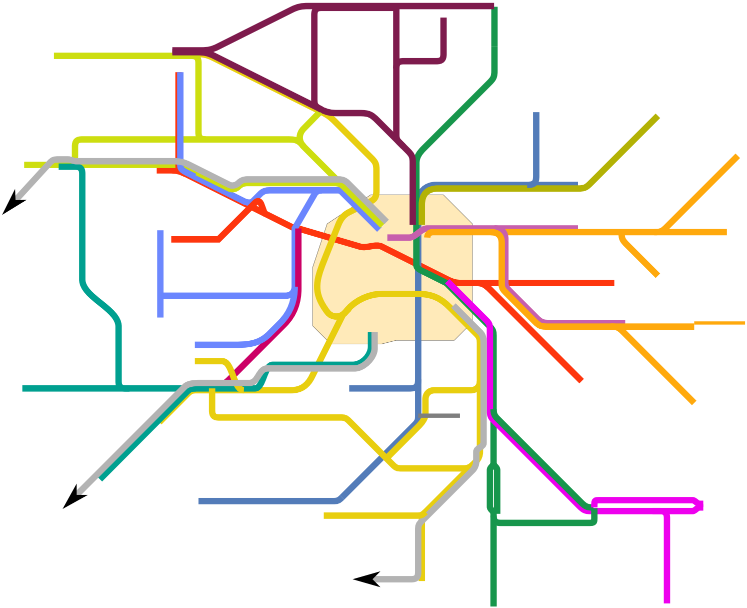
Output: the probability that the query is true under the distribution (assuming independence of all probabilistic events)

Complexity: already #P-hard in the input database!
(from #MONOTONE-SAT)

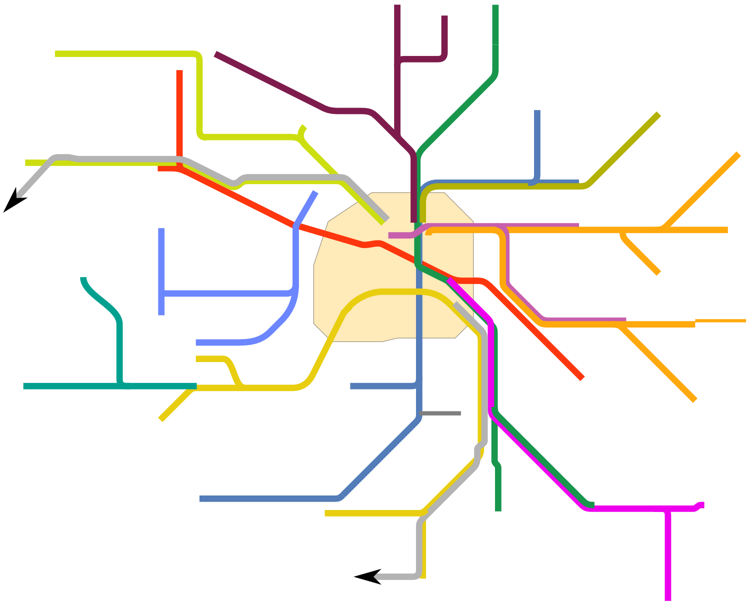
Using treewidth to make the problem tractable



Using treewidth to make the problem tractable

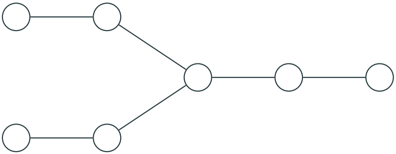


Using treewidth to make the problem tractable



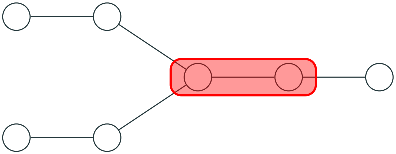
Using treewidth to make the problem tractable

Treewidth by example:



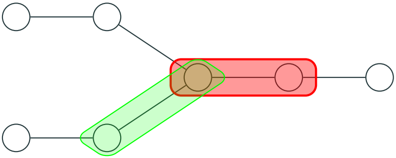
Using treewidth to make the problem tractable

Treewidth by example:



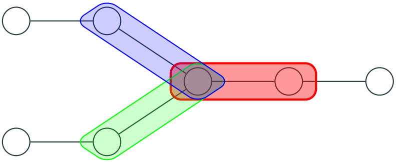
Using treewidth to make the problem tractable

Treewidth by example:



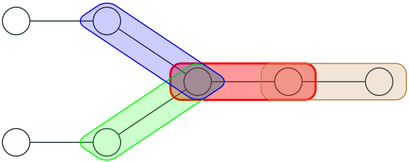
Using treewidth to make the problem tractable

Treewidth by example:



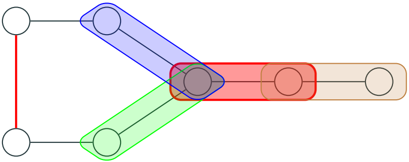
Using treewidth to make the problem tractable

Treewidth by example:



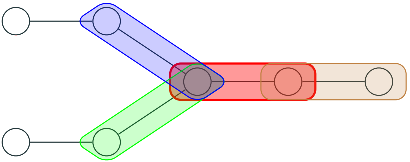
Using treewidth to make the problem tractable

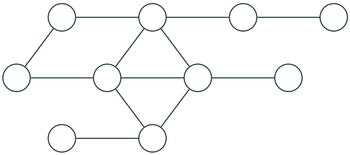
Treewidth by example:



Using treewidth to make the problem tractable

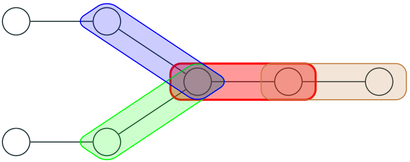
Treewidth by example:

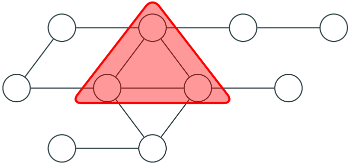




Using treewidth to make the problem tractable

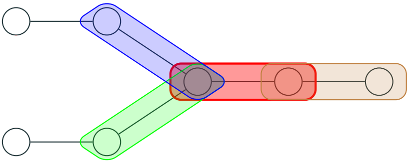
Treewidth by example:

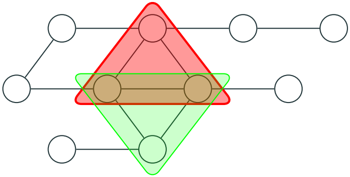




Using treewidth to make the problem tractable

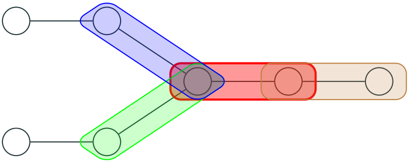
Treewidth by example:

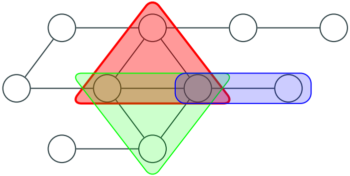




Using treewidth to make the problem tractable

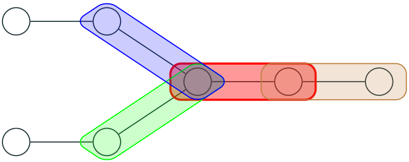
Treewidth by example:

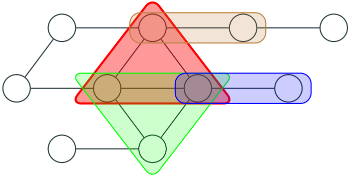




Using treewidth to make the problem tractable

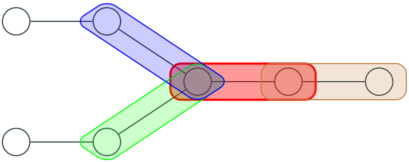
Treewidth by example:

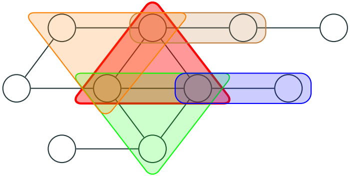




Using treewidth to make the problem tractable

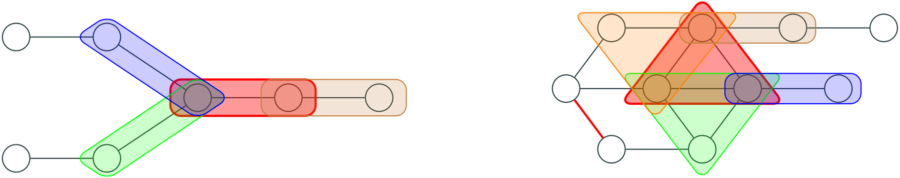
Treewidth by example:



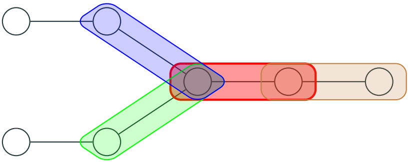


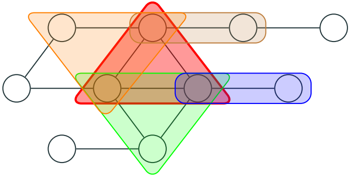
Using treewidth to make the problem tractable

Treewidth by example:



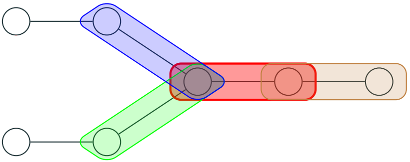
Treewidth by example:

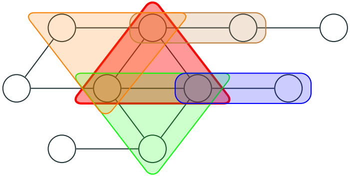




Using treewidth to make the problem tractable

Treewidth by example:

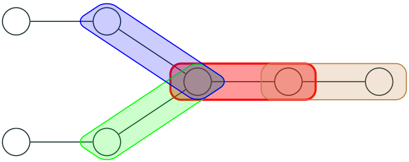


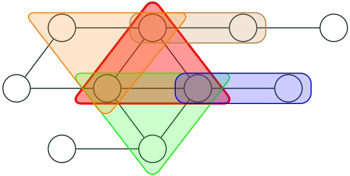


- Trees have treewidth 1
- Cycles have treewidth 2
- k-cliques and $(k-1)$-grids have treewidth $k-1$

Using treewidth to make the problem tractable

Treewidth by example:



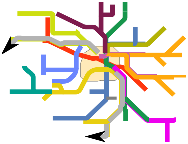


- Trees have treewidth 1
- Cycles have treewidth 2
- k-cliques and $(k-1)$-grids have treewidth $k-1$

$\rightarrow$ Treelike: the treewidth is bounded by a constant

Tractability on treelike instances

Treelike **data**

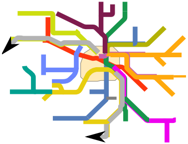


MSO query

(RER|metro)*
|(bus|tram)*

Tractability on treelike instances

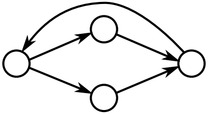
Treelike **data**



MSO query

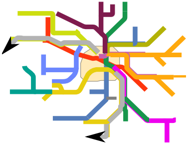
(RER|metro)*
|(bus|tram)*

Tree **automaton**

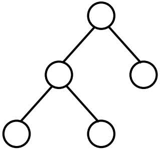


Tractability on treelike instances

Treelike **data**



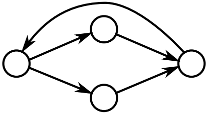
Tree **encoding**



MSO query

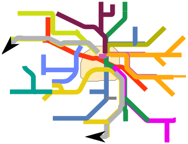
(RER|metro)*
|(bus|tram)*

Tree **automaton**

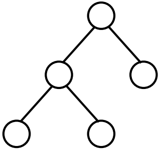


Tractability on treelike instances

Treelike **data**



Tree **encoding**



linear
[Courcelle]

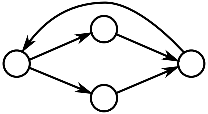
Query **answer**

TRUE

MSO query

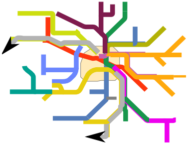
(RER|metro)*
|(bus|tram)*

Tree **automaton**

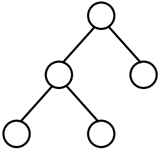


Tractability on treelike instances

Treelike **data**



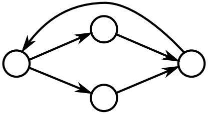
Tree **encoding**



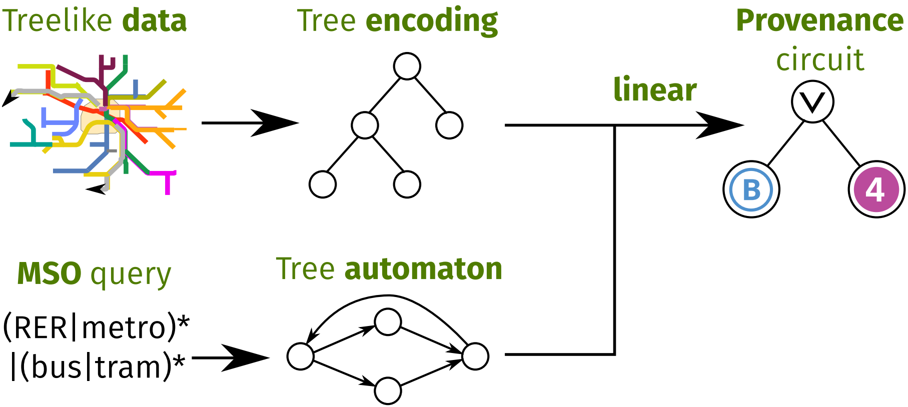
MSO query

(RER|metro)*
|(bus|tram)*

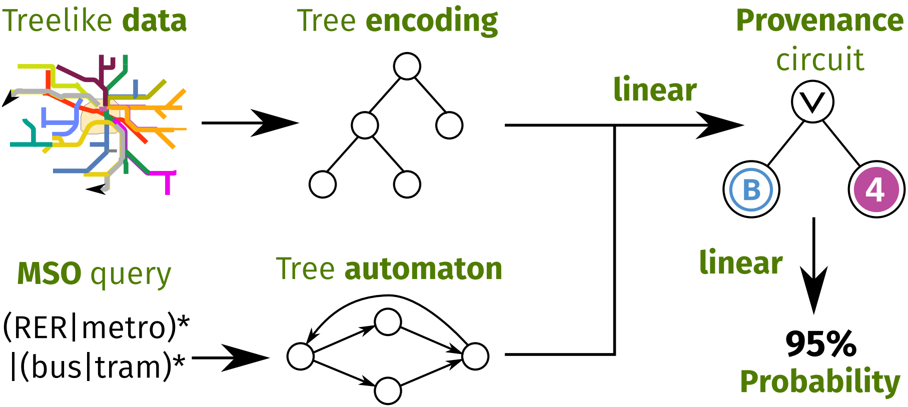
Tree **automaton**



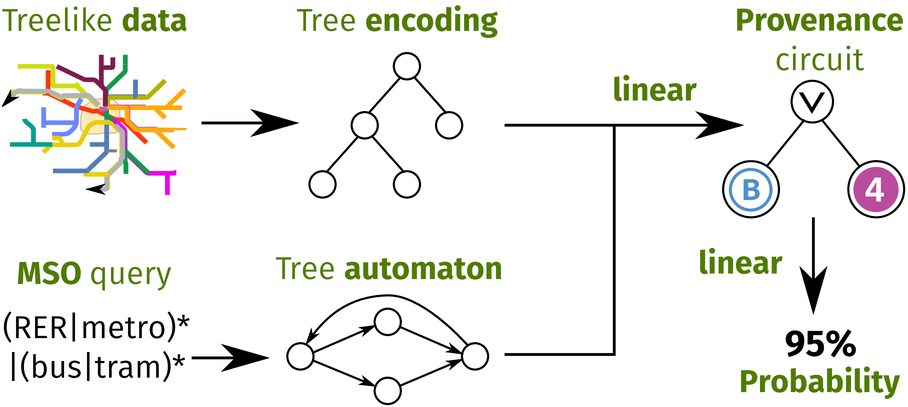
Tractability on treelike instances



Tractability on treelike instances



Tractability on treelike instances



Theorem

For any fixed Boolean MSO query q and $k \in \mathbb{N}$,
given a database D of treewidth $\leq k$ with independent probabilities,
we can compute in linear time the probability that D satisfies q

Lower bound

What can we do for unbounded-treewidth instances?

Lower bound

What can we do for unbounded-treewidth instances? … not much.

Lower bound

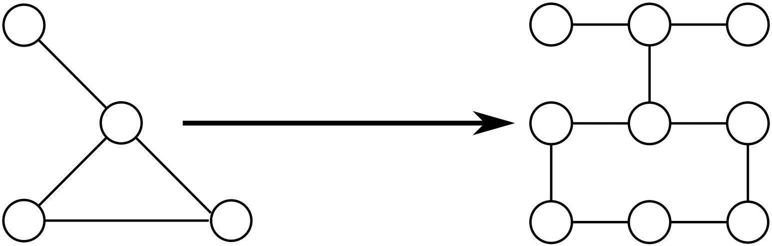
Theorem

For any graph signature σ, there is a first-order query $\boldsymbol{q}$ such that for any constructible unbounded-treewidth class $\mathcal{I}$, probability evaluation of $\boldsymbol{q}$ on $\mathcal{I}$ is #P-hard under RP reductions

Lower bound

Theorem

For any graph signature σ, there is a first-order query $\mathbf{q}$ such that for any constructible unbounded-treewidth class $\mathcal{I}$, probability evaluation of $\mathbf{q}$ on $\mathcal{I}$ is #P-hard under RP reductions

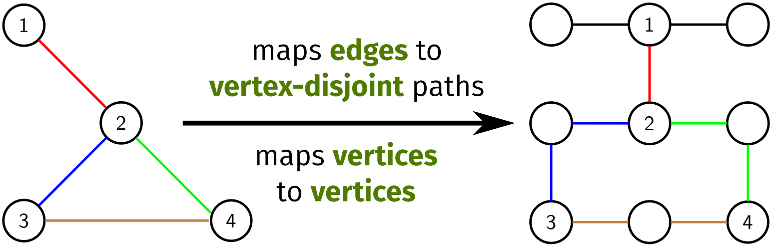


Proof idea: extract instances of a hard problem as topological minors using recent polynomial bounds [Chekuri and Chuzhoy, 2014]

Lower bound

Theorem

For any graph signature σ, there is a first-order query $\boldsymbol{q}$ such that for any constructible unbounded-treewidth class $\mathcal{I}$, probability evaluation of $\boldsymbol{q}$ on $\mathcal{I}$ is #P-hard under RP reductions



Proof idea: extract instances of a hard problem as topological minors using recent polynomial bounds [Chekuri and Chuzhoy, 2014]

Future and ongoing work

- Improving the lower bound:
  - From graphs to arbitrary arity databases
  - From FO down to unions of conjunctive queries with $\neq$

Future and ongoing work

- Improving the lower bound:
  - From graphs to arbitrary arity databases
  - From FO down to unions of conjunctive queries with $\neq$

- Complexity in query and database — currently $\Omega\left(2^{2^{\cdot^{\cdot^{\cdot^{2^{|Q|}}}}}} \times |D|\right)$

  $\rightarrow$ Which queries can efficiently be compiled to automata?

Future and ongoing work

- Improving the lower bound:
  - From graphs to arbitrary arity databases
  - From FO down to unions of conjunctive queries with $\neq$

- Complexity in query and database — currently $\Omega\left(2^{2^{\cdot^{\cdot^{\cdot^{2^{|Q|}}}}}} \times |D|\right)$
  $\rightarrow$ Which queries can efficiently be compiled to automata?

- Non-Boolean queries: efficient enumeration of query results?

Future and ongoing work

- Improving the lower bound:
    - From graphs to arbitrary arity databases
    - From FO down to unions of conjunctive queries with $\neq$
- Complexity in query and database — currently $\Omega\left(2^{2^{\cdot^{\cdot^{\cdot^{2^{|Q|}}}}}} \times |D|\right)$
  $\rightarrow$ Which queries can efficiently be compiled to automata?
- Non-Boolean queries: efficient enumeration of query results?
- Other tasks: probabilistic conditioning
  “Knowing that I’m here, what’s the probability that RER B is up?”

Future and ongoing work

- Improving the lower bound:
  - From graphs to arbitrary arity databases
  - From FO down to unions of conjunctive queries with $\neq$

- Complexity in query and database — currently $\Omega\left(2^{2^{\cdot^{\cdot^{\cdot^{2^{|Q|}}}}}} \times |D|\right)$
  $\rightarrow$ Which queries can efficiently be compiled to automata?

- Non-Boolean queries: efficient enumeration of query results?

- Other tasks: probabilistic conditioning
  "Knowing that I'm here, what's the probability that RER B is up?"

Thanks for your attention!

References I

Chekuri, C. and Chuzhoy, J. (2014).
Polynomial bounds for the grid-minor theorem.
In *STOC*.

Courcelle, B. (1990).
The monadic second-order logic of graphs. I. Recognizable sets of finite graphs.
Inf. Comput., 85(1).

Image credits

- Slide 2:
  - `https://commons.wikimedia.org/wiki/File:Paris_Metro_map.svg` (cropped), user Umx on Wikimedia Commons, public domain
  - `http://www.parisvoyage.com/images/cartoon18.jpg`, ParisVoyage, fair use
  - `http://www.vianavigo.com/fileadmin/galerie/pdf/CGU_t_.pdf` (cropped), RATP, fair use
- Slides 4 and 5: `https://commons.wikimedia.org/wiki/File:Carte_Transilien_RER_sch%C3%A9matique.svg` (modified), user Benjamin Smith on Wikimedia Commons, license CC BY-SA 4.0 international.