DIGITAL TOURISM FROM TRIPADVISOR CIRCULATION GRAPHS TO BEHAVIOR EXTRACTION AND CHARACTERIZATION



Nicolas Travers

ADBIS'26 Tutorial

DE VINCI
RESEARCH
CENTER

NICOLAS TRAVERS

Full Professor of Computer Science at ESILV

Co-director of the DVRC laboratory

Habilitated (HDR) in Computer Science from Sorbonne University in 2018. PhD thesis from UVSQ-Paris Saclay in 2006.

Main topics: database management and optimization. Specialization:

- Distributed databases,
- Graph databases,
- Information Retrieval
- Recommendation systems
- Ad hoc databases design in a transdisciplinary approach: Digital Tourism & Geography, Knowledge Management & History, IR & Musicology or Multimodal libraries.



DATA : E-TOURISM

- Tripadvisor
- Booking
- Airbnb
- Flickr
- Hotels.com

~1M reviews

650 k reviews

28 k reviews

304 k reviews

Digital Service Acts – Art 40 – En cours
=> Google Maps

HOW TO EXTRACT TOURISTS' BEHAVIOR FROM TOURISM DIGITAL FOOTPRINT?

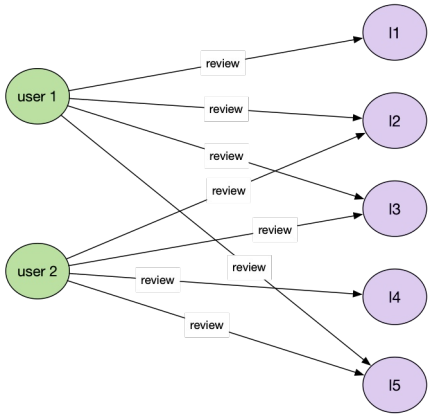
Or:

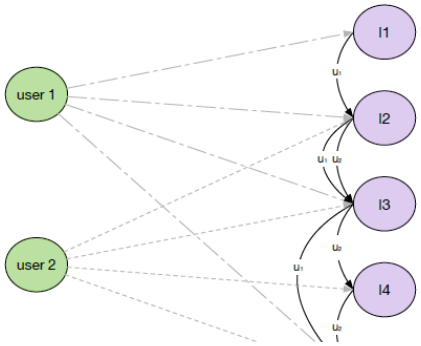
How to extract knowledge from graphs?

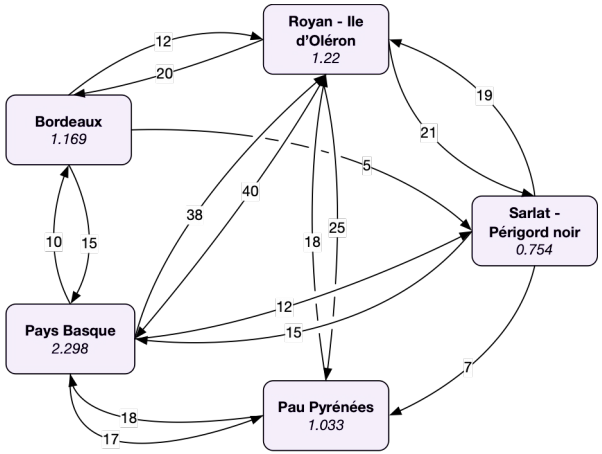


LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR

GRAPH DATA MODELING
ANALYSIS OF TOURIST TRAFFIC







- Comparison / trends
- Population comparison

Aggregation of graphs with *Cypher*

```
MATCH (t1:Town) -[t:trip]-> (t2:Town)
MERGE (c1:City{name:t1.city})
MERGE (c2:City{name:t2.city})
MERGE (c1)-[ct:trip{year:t.year, nat:t.nat}]->(c2)
ON CREATE SET ct.NB=t.NB
ON MATCH SET ct.NB=ct.NB+t.NB
```

The **E-khonsou** platform
(ReactJS&nodeJS)
https://mel.dvrc.fr

GRAPH MINING & IA
INTRAMUROS STUDY - MEL

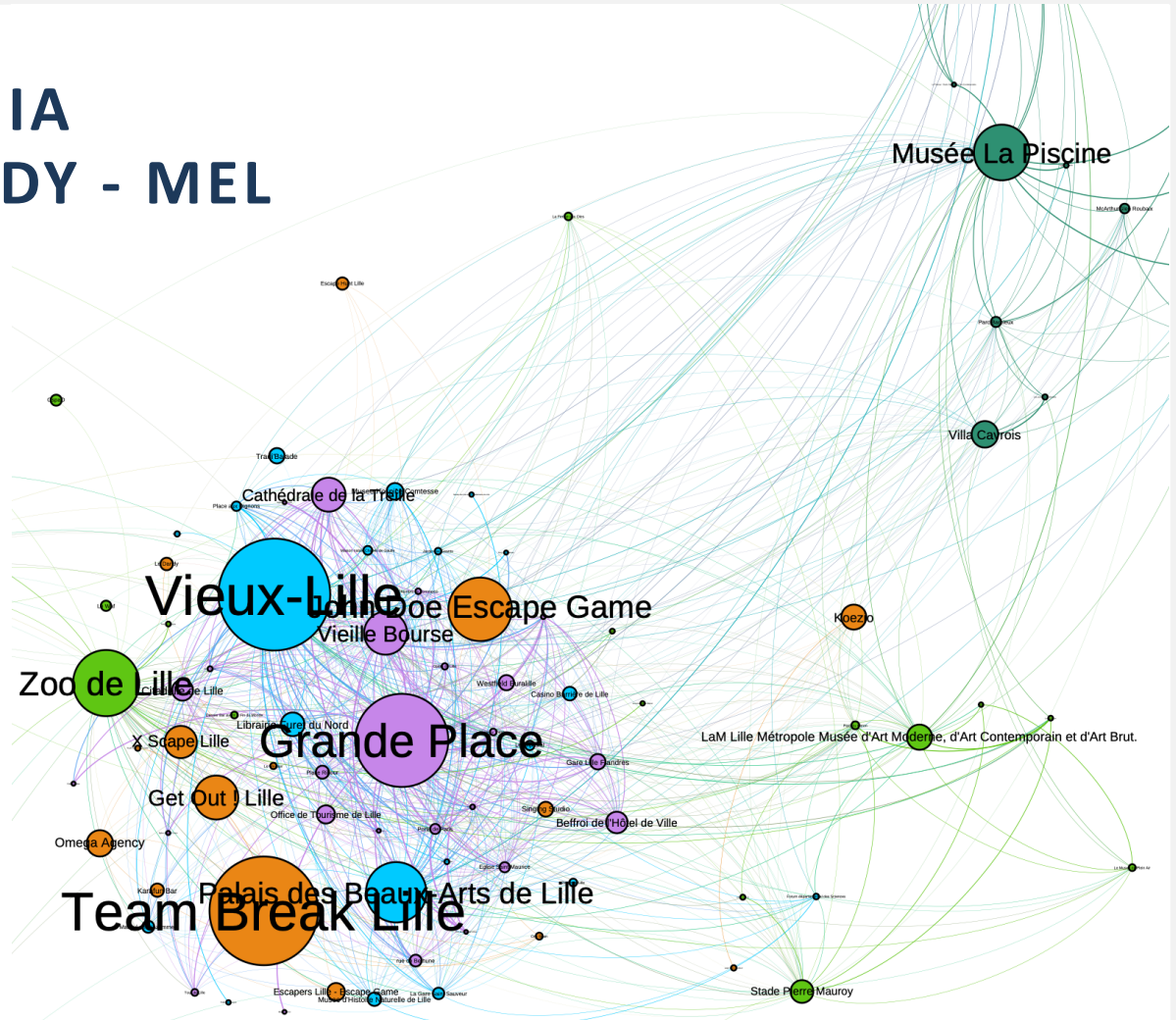
High connectivity between locations

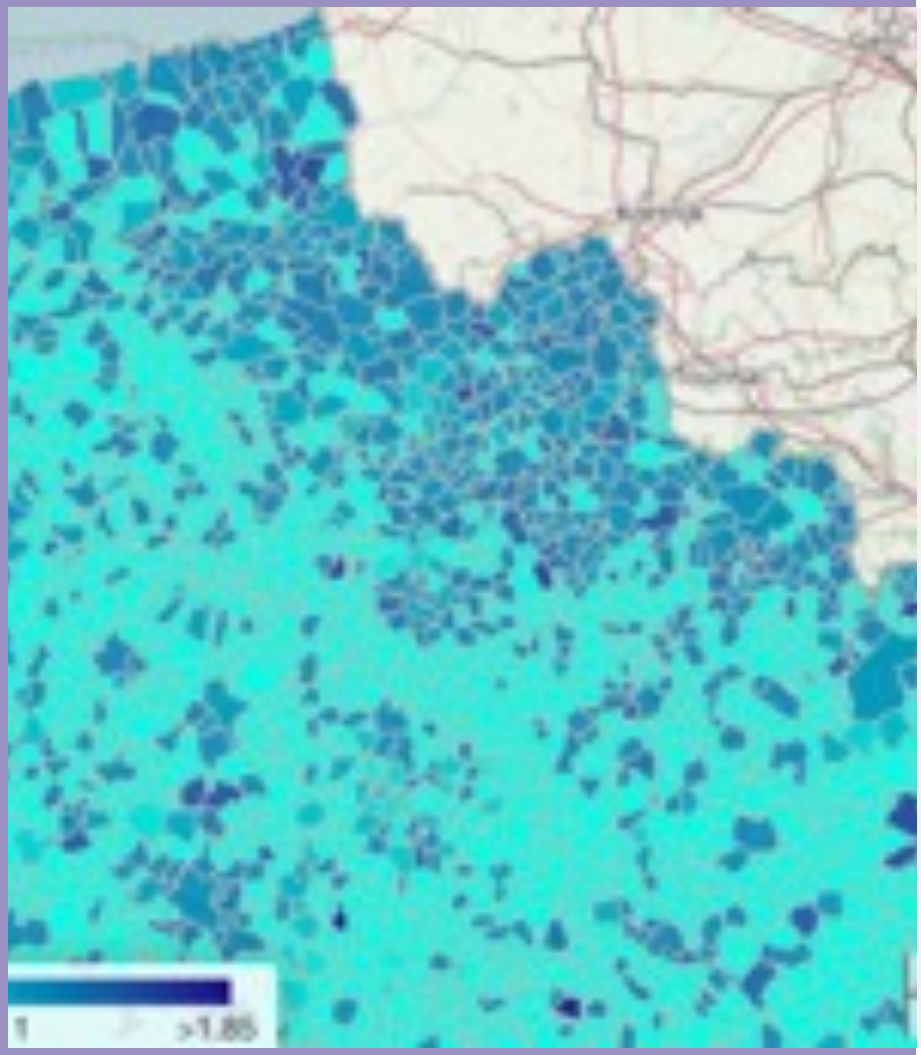
Gare Lille Flandres - Vieille Bourse - Grand'Place - Office de Tourisme/Palais Rihour - Citadelle de Lille - Cathédrale Notre Dame de la Treille – Euralille

Vieux Lille – Musée Hospice Comtesse – Maison natale du Général de Gaulle – Librairie du Furet du Nord – Palais des Beaux Arts de Lille – Musée d'Histoire Naturelle de Lille - Tradi Balade

La Piscine de Roubaix – McArthurGlen – Parc Barbieux – La Manufacture – l'Usine – Villa Cavrois – Les Jardins Marc Stevens

➢ Dedicated to locations, no global view



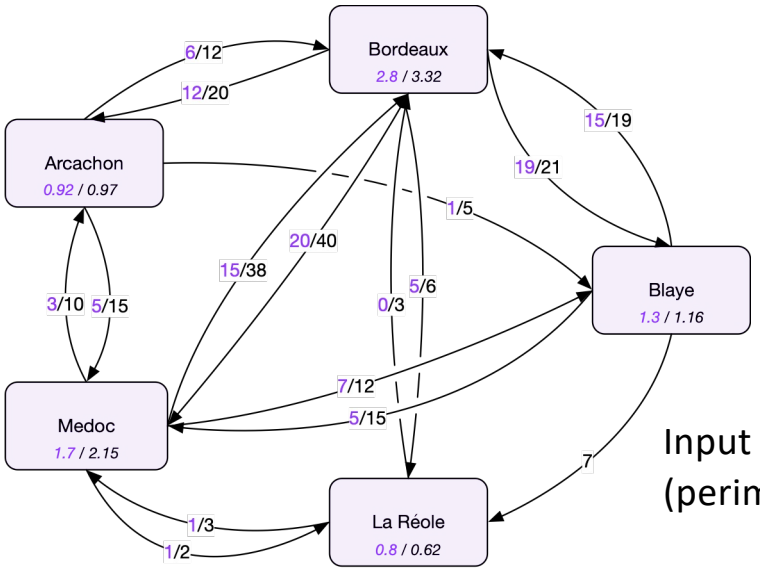


HOT POINTS IN THE CIRCULATION GRAPH

Indicators for Measuring Tourist Mobility [WISE'20]

CIRCULATION GRAPH MANIPULATION

Circulation graph aggregation AC(V',E')



Compute PageRanks with *gds* (with filters)

```
CALL gds.graph.create.cypher("CypherProjection",
   "MATCH (c:City{department:'Gironde'}) RETURN id(c) as id",
   "MATCH (c1:City)-[t:trip{year:2018,nationality:'USA'}]->(c2:City)
    RETURN id(c1) as from, id(c2) as to, sum(toFloat(t.NB)) as weight")
CALL gds.pageRank.stream("CypherProjection",
   {dampingFactor:0.85,iterations:50,weightProperty:true}) YIELD node, score
   RETURN node.city, sum(score) as score;
```

Filters on edges

$$ac^{USA,2018}_{cities,year\&nat} = AC(cities, year\&nat, \sigma_{nat="USA" \wedge year=2018})$$

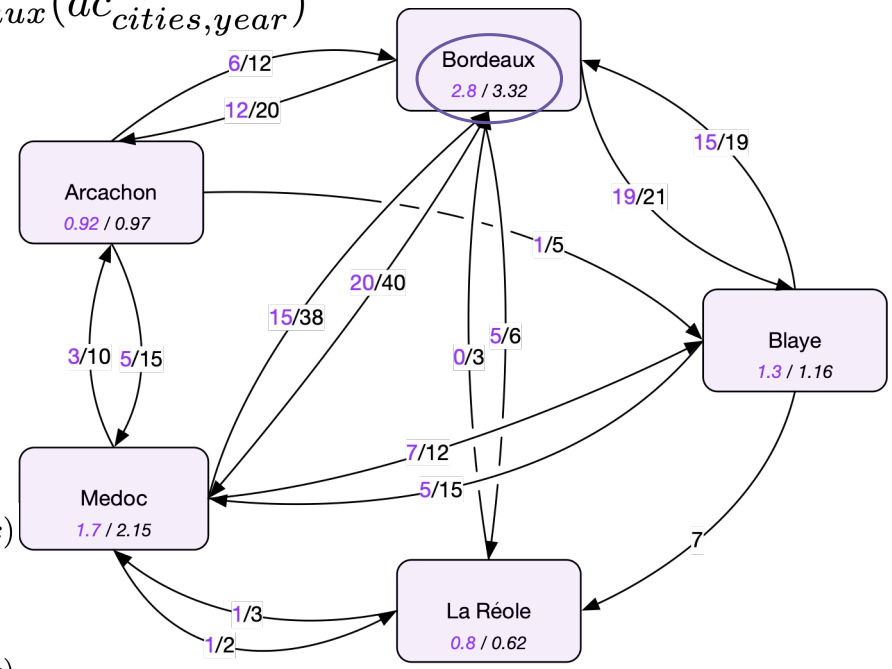
Input graph (perimeter)

Level of aggregation

TCF – EXAMPLE

$$TCF^{USA,2018}_{cities,year}(AC, Bordeaux) = \frac{PR_{Bordeaux}(ac^{USA,2018}_{cities,year})}{PR_{Bordeaux}(ac^{-,2018}_{cities,year})}$$

Ex: Impact factor of the **US** population in 2018 at Bordeaux (at city scale)



Main Goal

- TCF > 1

  Population *p* has a greater impact on *v than others*

- Compare TCF evolution:

$$TCF^{USA,2017}_{cities,year}(AC, Bordeaux) < TCF^{USA,2018}_{cities,year}(AC, Bordeaux)$$

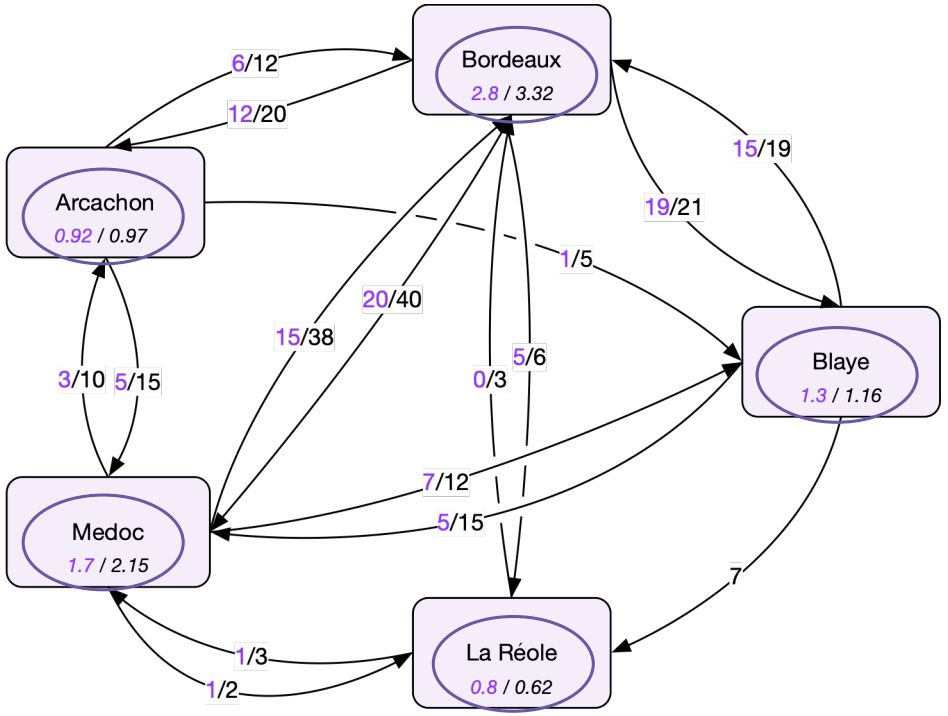
- Compare populations:

$$TCF^{French,2018}_{cities,year}(AC, Bordeaux) < TCF^{USA,2018}_{cities,year}(AC, Bordeaux)$$

LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR

GCF – THE GLOBAL CIRCULATION FACTOR

- Measures the impact of a population on the whole graph
- Mean value of TCFs for AC of a given population *p* with context *f*



Main Goal

- GCF > 1
  - Population *p* circulates more than others on the whole graph
- Comparison over years and populations

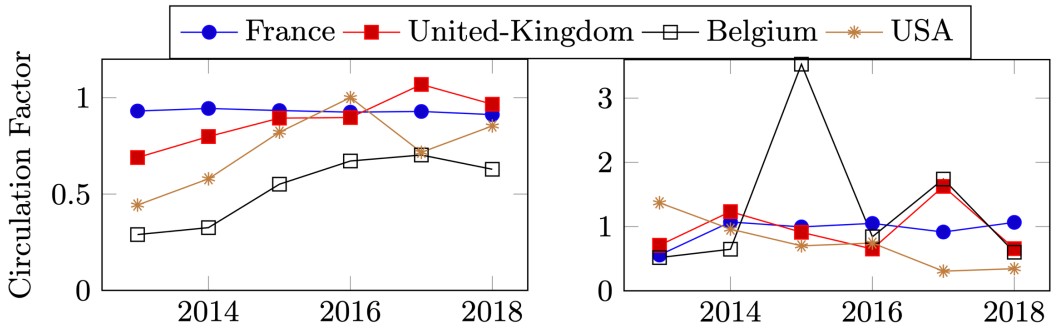
LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR

EXPERIMENTS – TCF EVOLU

Area: Department (Gironde)

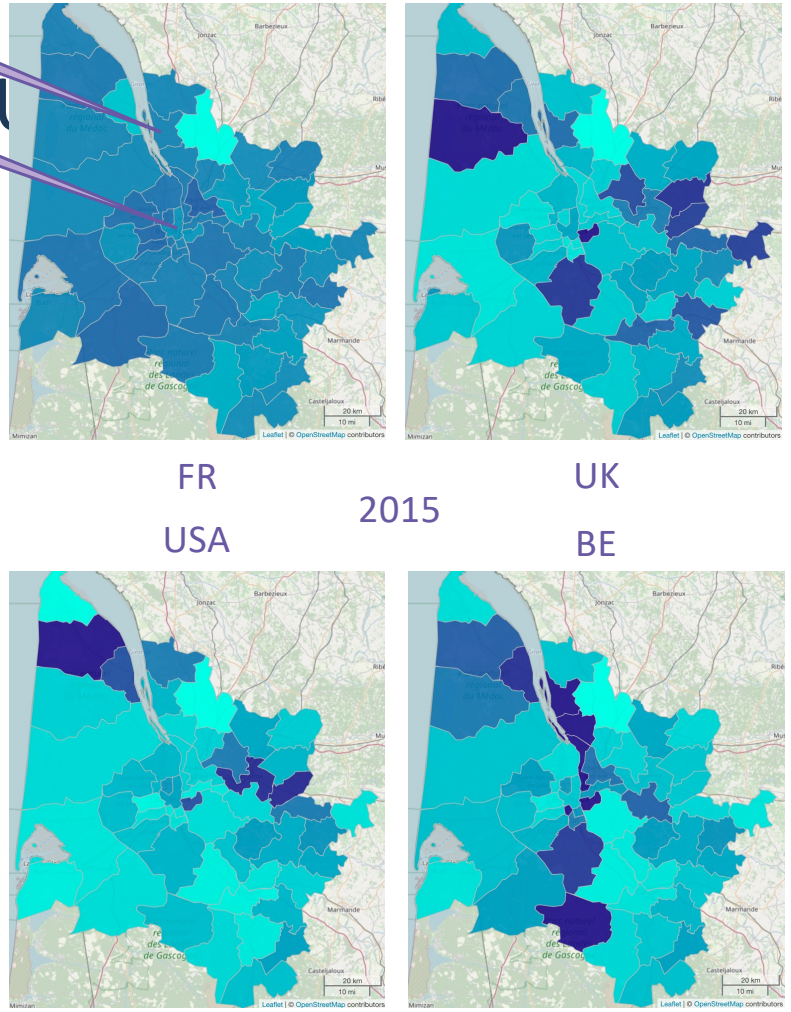
Aggregation level: Cities

⇒ *TCF at small scales - events detection*



Bordeaux

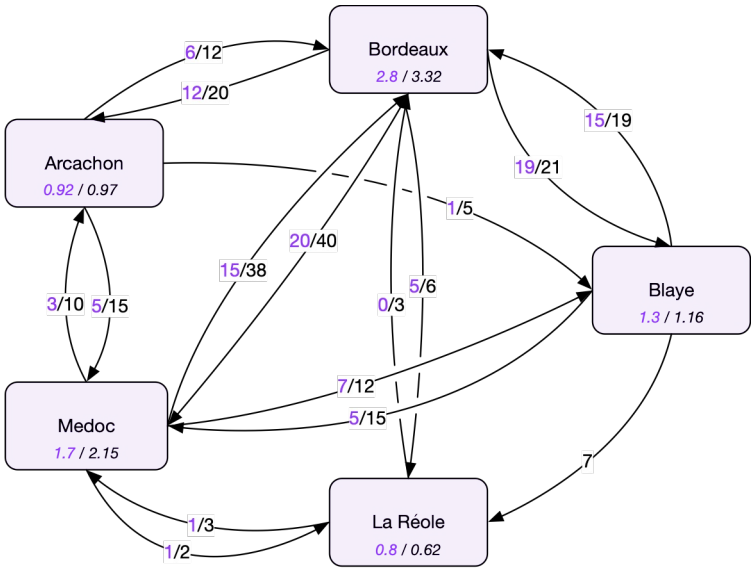
Blaye

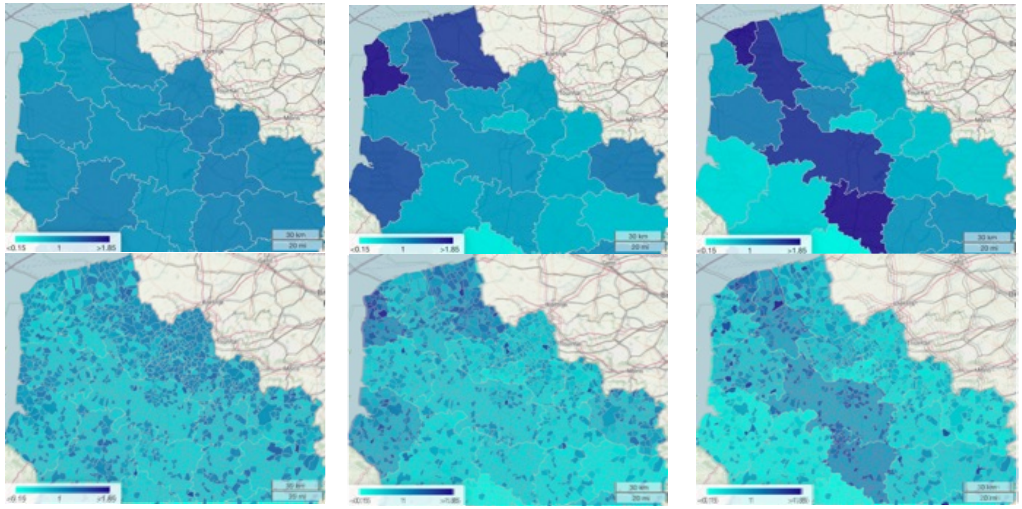


LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR

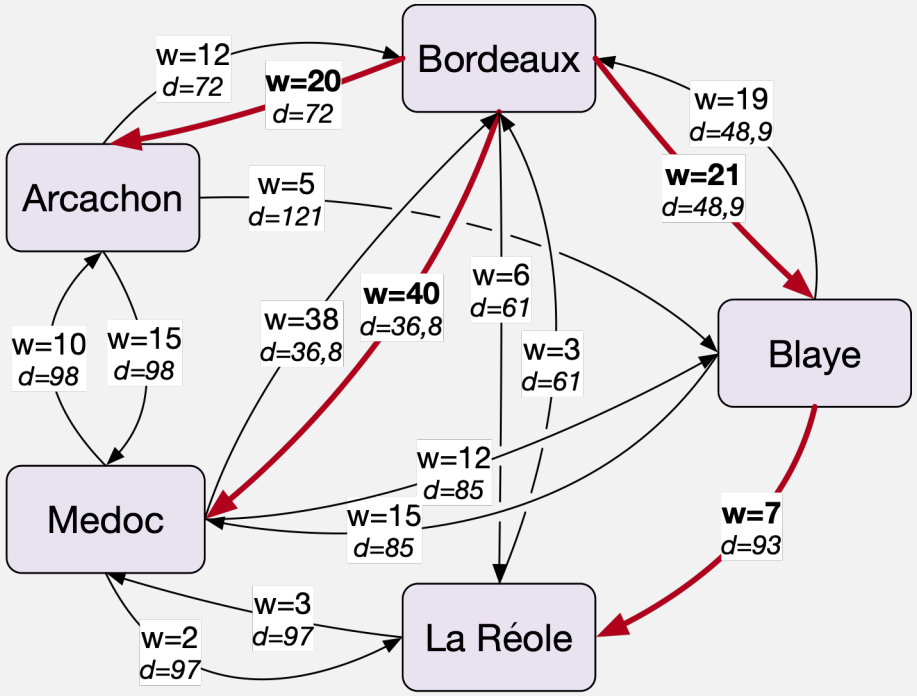
GRAPH MINING & IA
ANALYSIS OF TOURIST TRAFFIC

- Comparison of developments / trends
- Population comparison





French ***Belgium*** ***English***



TOURIST PROPAGATION IN MULTI-WEIGHTED CIRCULATION GRAPHS

Propagation Measure on Circulation Graphs for Tourism Behavior Analysis [SAC'22]

REMOTENESS INFLUENCE FACTOR

$$RIF(AC, s) = \sum_{n \in \mathcal{V}} \left(\frac{\log_{dist_{max}} \left(\sum_{e \in path(s,n)} dist(e) \right)}{|\mathcal{V}| - 1} \times \frac{\frac{1}{1+WBC(n)}}{|\mathcal{V}|} \right)$$

i.e., Max WBC(AC)

Normalized log(distance) from source **s**
Remoteness

Impact of nodes in AC
Influence

LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR

RIF COMPUTATION ON A GRAPH

Algorithm 1 Computation of the Remoteness Influence Factor

Require: $\mathtt{AC}(\mathcal{V}, \mathcal{E}(w, d))$ a graph, $s \in \mathcal{V}$ is the source node of the graph

1: **function** REMOTENESS(AC, s)
2: wBetweennessCentrality = `WeightedBetweennessCentrality(`$AC(w)$`)`
3: distancePairs = `Dijkstra(`$AC(d)$, s`)`
4: max_dist = max(distancePairs)
5: **for** $n \in \mathcal{V} - s$ **do**
6: $rif = rif + \log_{max_dist}(distancePairs[s][n]) \times \frac{1}{1+wBetweennessCentrality[n]}$
7: **end for**
8: **return** $\frac{rif}{(|\mathcal{V}|-1)\times|\mathcal{V}|}$
9: **end function**

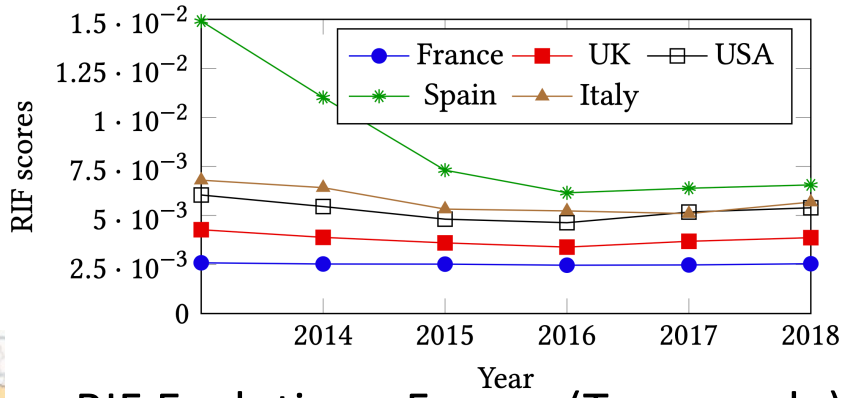
Complexity: $O|\mathcal{V}|^3 + O(|\mathcal{V}|^2 \log|\mathcal{V}|) + O(V)$

$\Rightarrow$ Maximum Spanning Tree (MST)
Kruskal algorithm [AMS'56]

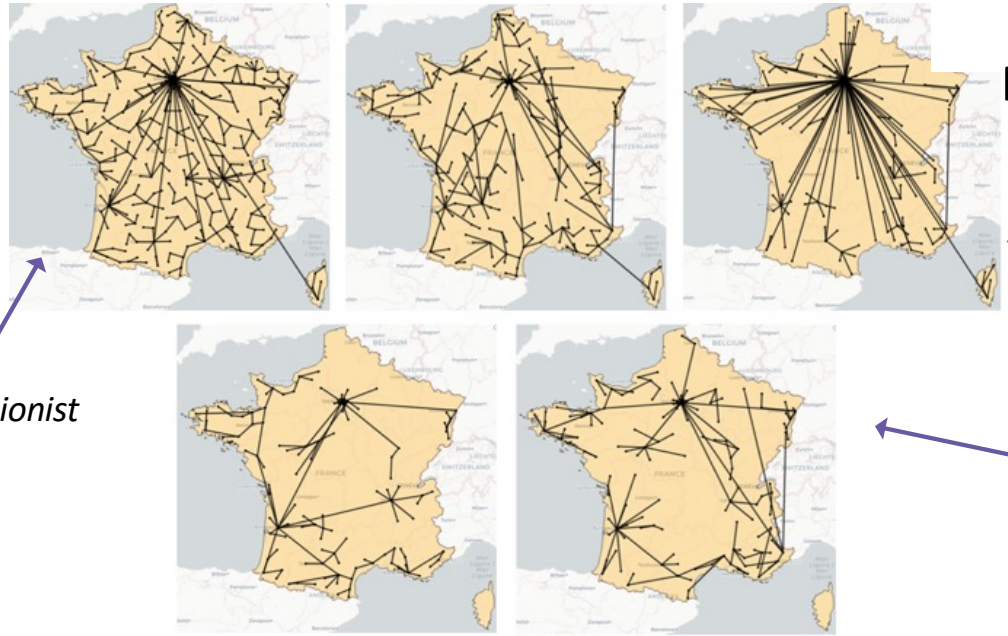
$O|\mathcal{V}|^2$

LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR

GRAPH MINING & IA
MESURE DE PROPAGATION



RIF Evolution - France (Town-scale)



excursionist

Day trippers

Local trippers

LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR

AUTHENTICITY PERCEPTION IN GRAPHS



Authenticity perceptions - step 0 (before visiting Italy)



Authenticity perceptions - step 1 (during the visit to Italy)



Authenticity perceptions - step 2 (after visiting Italy)

Authenticity goes digital: a big data analysis of the influence of the country of origin and authenticity perceptions on TripAdvisor ethnic restaurant reviews [Academy of Marketing Science'20]

AUTHENTICITY

Authenticity = Perception of a representative experience of the destination

Study on the evolution of this perception of Italian food

Study on the experience before/during/after the trip to Italy

Time series extraction / Pattern Mining from the graph

Cypher query:

```
MATCH p1 = (b1) -[*]-> (bn) -[u]-> (it1),
           p2 = (it1) -[*]-> (itn) --> (a1),
           p3 = (a1) -[*]-> (an)

WHERE ALL(r in nodes(p1) where r.country<>"Italy") AND
      ALL(r in nodes(p2) where r.country="Italy") AND
      ALL(r in nodes(p3) where r.country<>"Italy") AND
      ALL(rev in relationships(p1,p2,p3)
          WHERE rev.user = u.user)

RETURN u as user, AVG(p1.note) AS before,
AVG(p2.note) AS during, AVG(p3.note) as after
```

Authenticity goes digital: a big data analysis of the influence of the country of origin and authenticity perceptions on TripAdvisor ethnic restaurant reviews [Academy of Marketing Science'20]

AUTHENTICITY

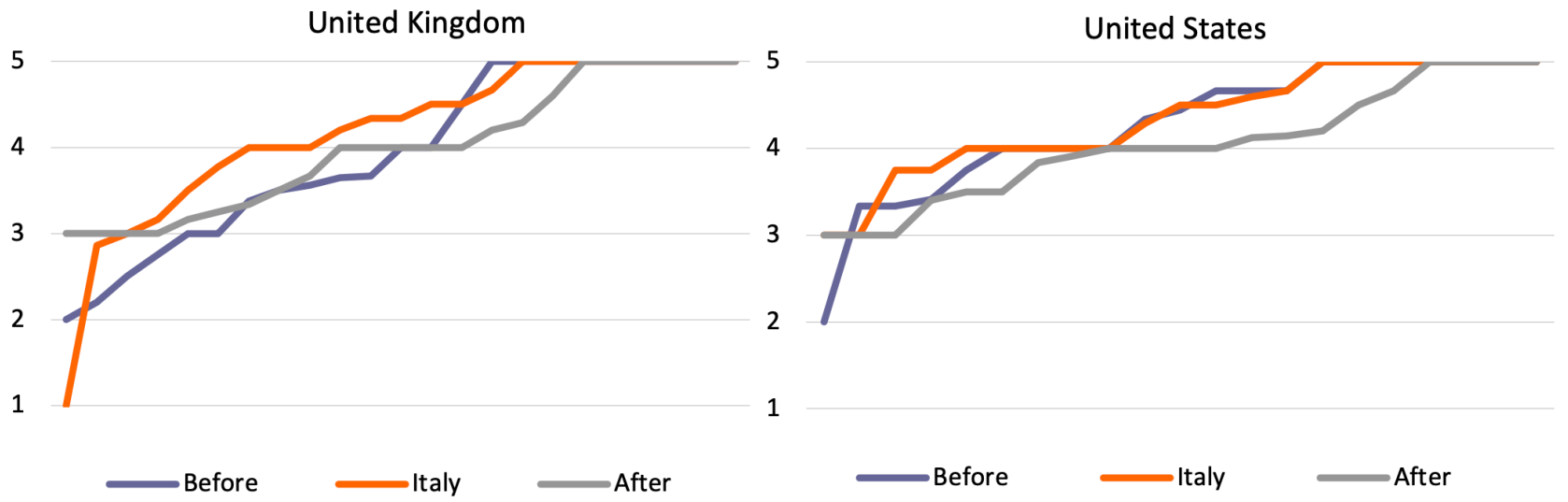
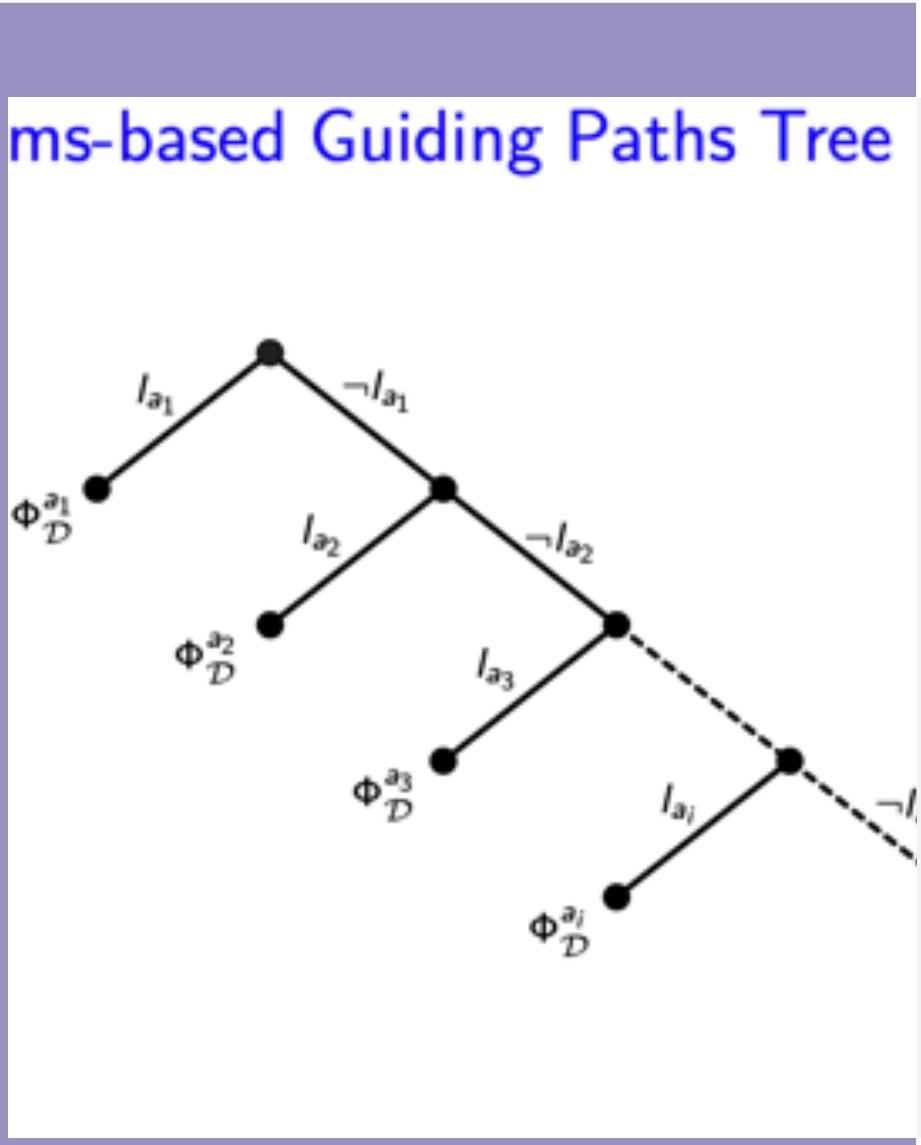


Figure 1 – Distribution of authenticity vectors from UK and USA citizens on Italian restaurants



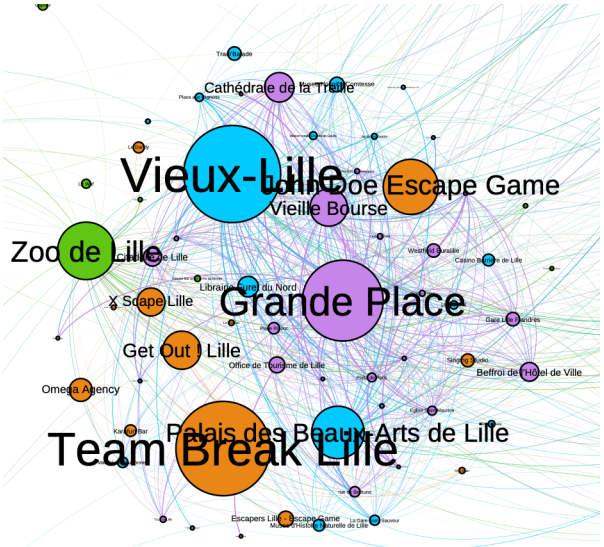
MINING EFFICIENTLY ITEMSETS FROM THE GRAPH

A Distributed SAT-based Framework for Closed Frequent Itemset Mining [ADMA'22]

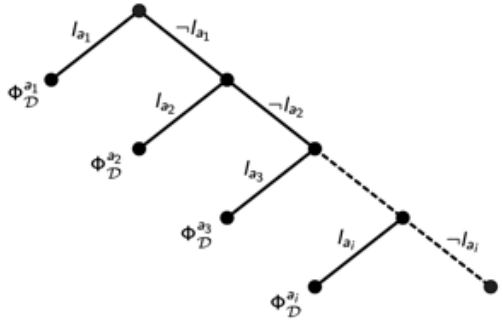
FREQUENT ITEMSETS MINING

Cover({Vieux-Lille, Grand Place}, D)= {T_1, T_2, T_5}

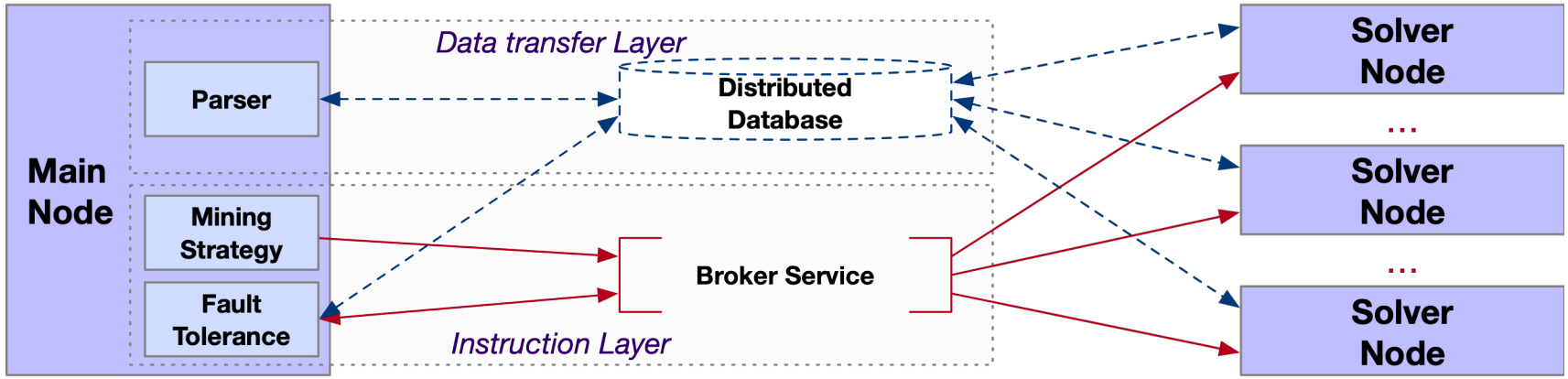
Decomposition-based **SAT** Approach



Items-based Guiding Paths Tree



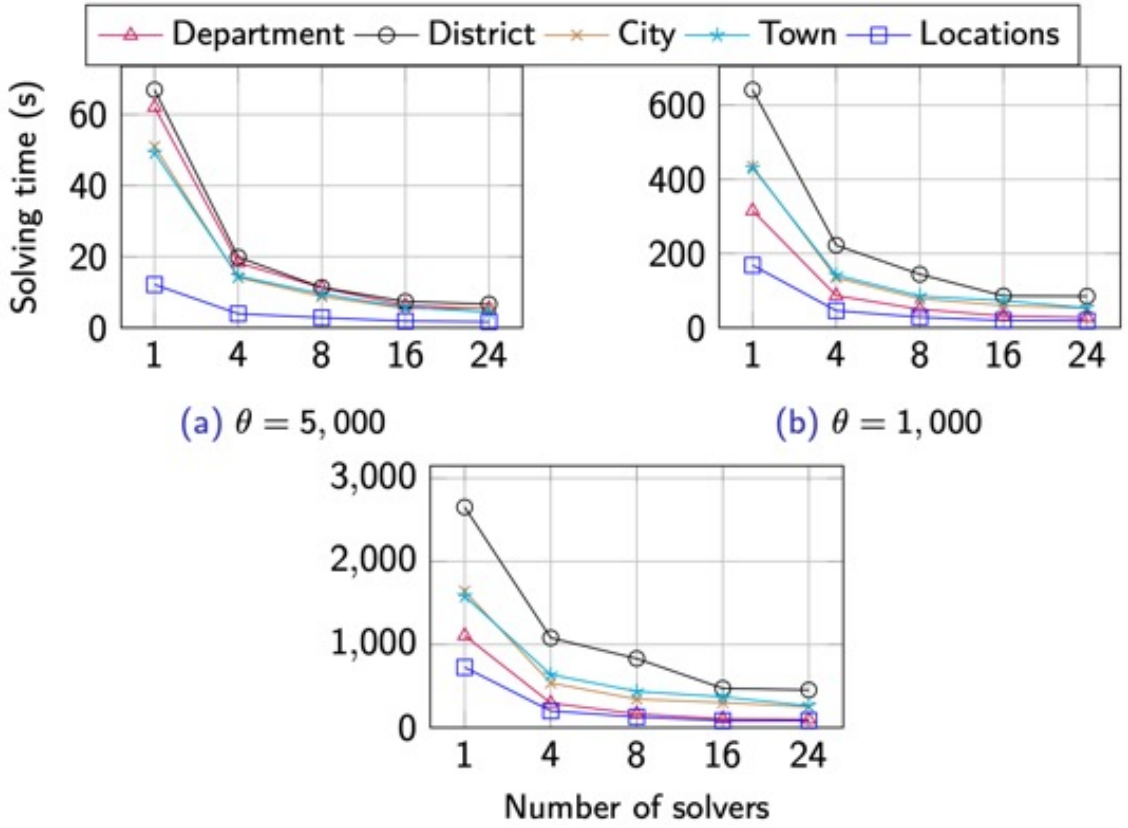
DISTRIBUTED ARCHITECTURE FOR DECOMPOSED ITEMSETS



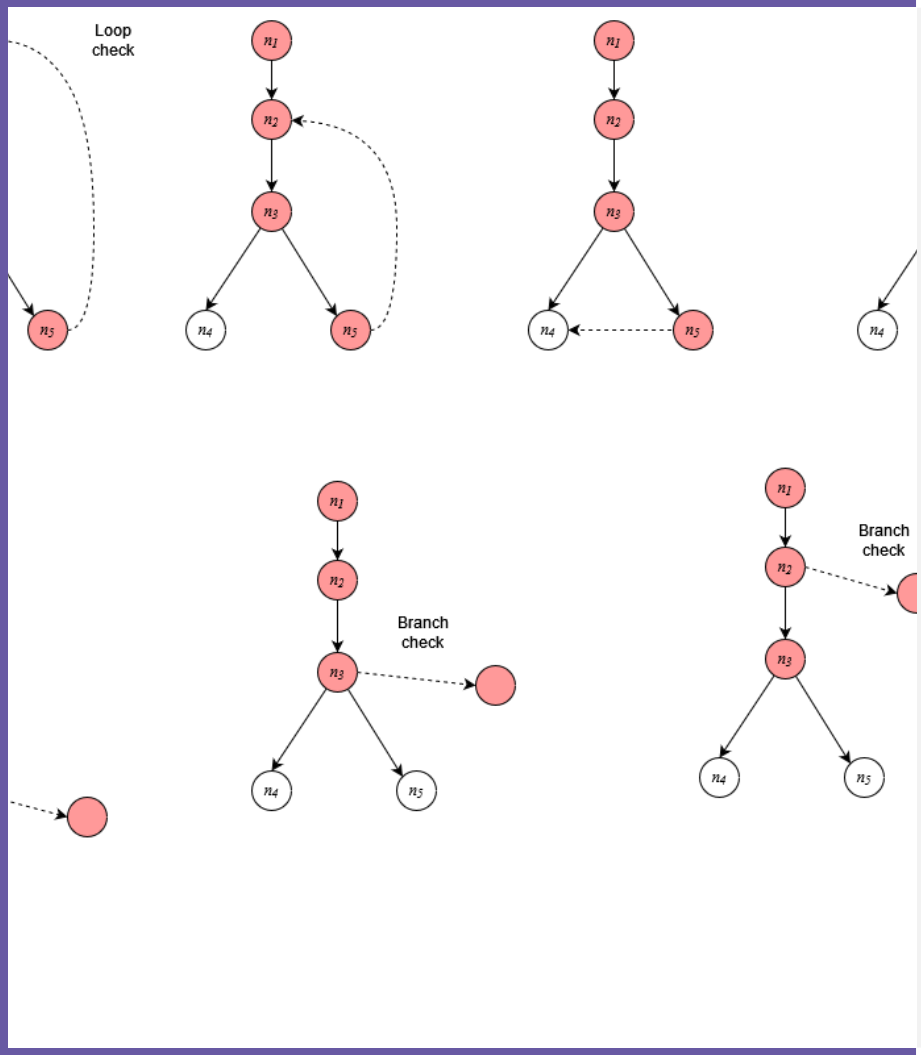
Steps :

1. **Encoding** transactions DB -> build the *queue*
2. **Distribute encoded items** to mine to the broker
3. **Distributed solving**
4. Store mined solutions (models) in **MongoDB**
5. Check completion / **fault tolerance**
6. Provide **mined models** from the db (mongodb)

FREQUENT ITEMSETS MINING



LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR



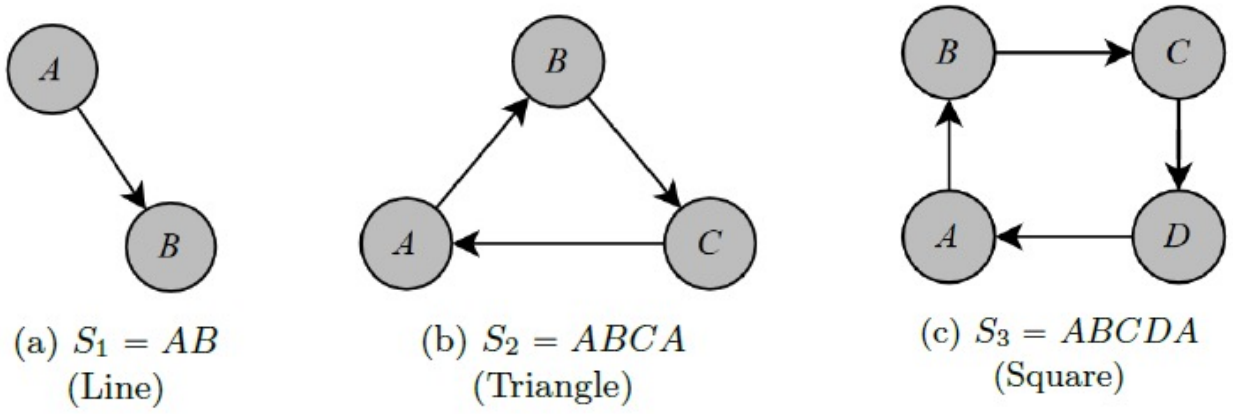
EFFICIENT PATTERN EXTRACTION

(Master thesis Muhamad & Koumudi)

(Master thesis Muhamad & Koumudi)

FROM PATTERNS (SUBGRAPHS) TO SIGNATURES

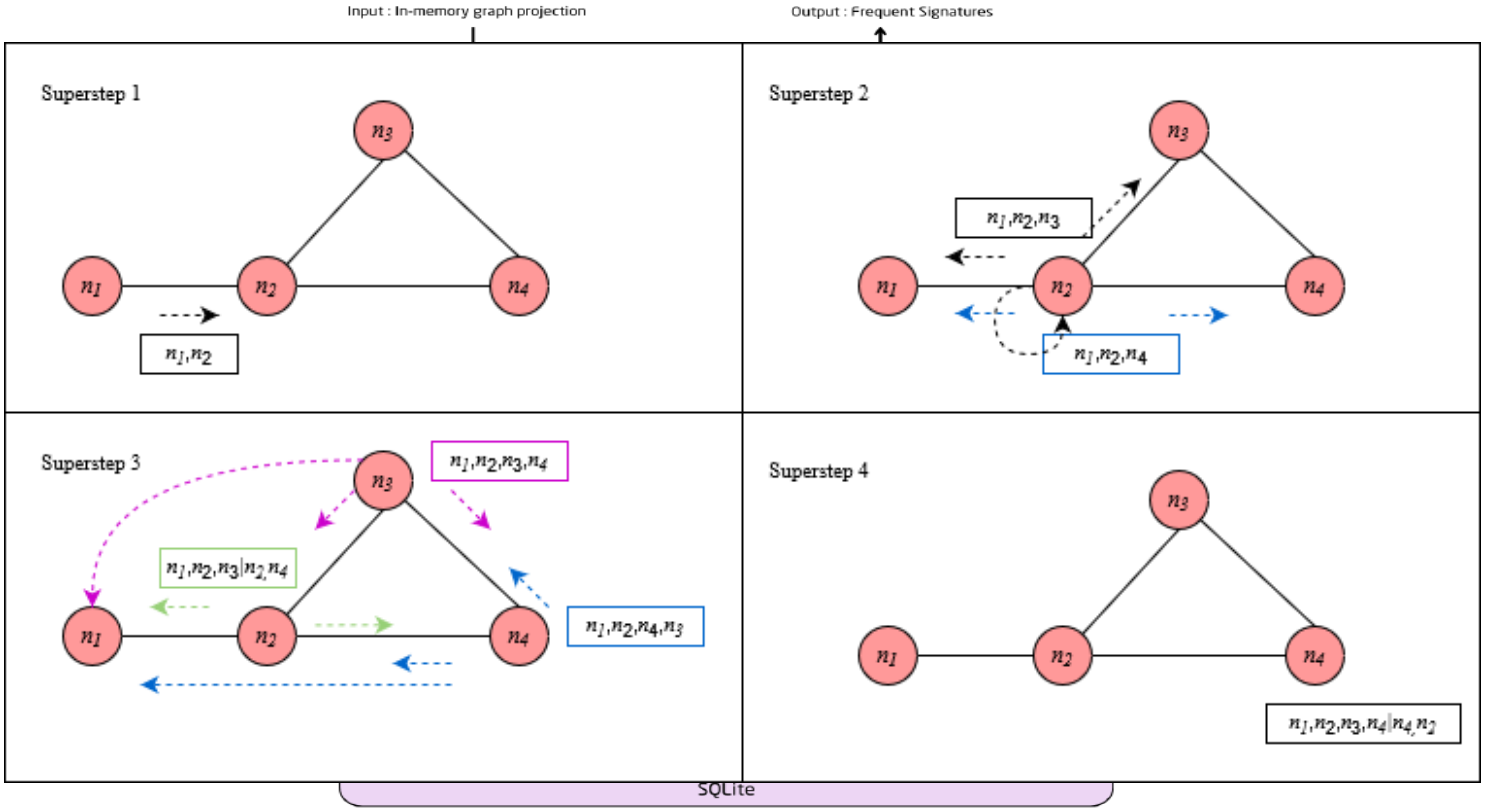
Proposed approach to extract efficiently frequent signatures in circulation graphs



Different Signature Types

(Master thesis Muhamad & Koumudi)

PREGEL SIGNATURES



(Master thesis Muhamad & Koumudi)

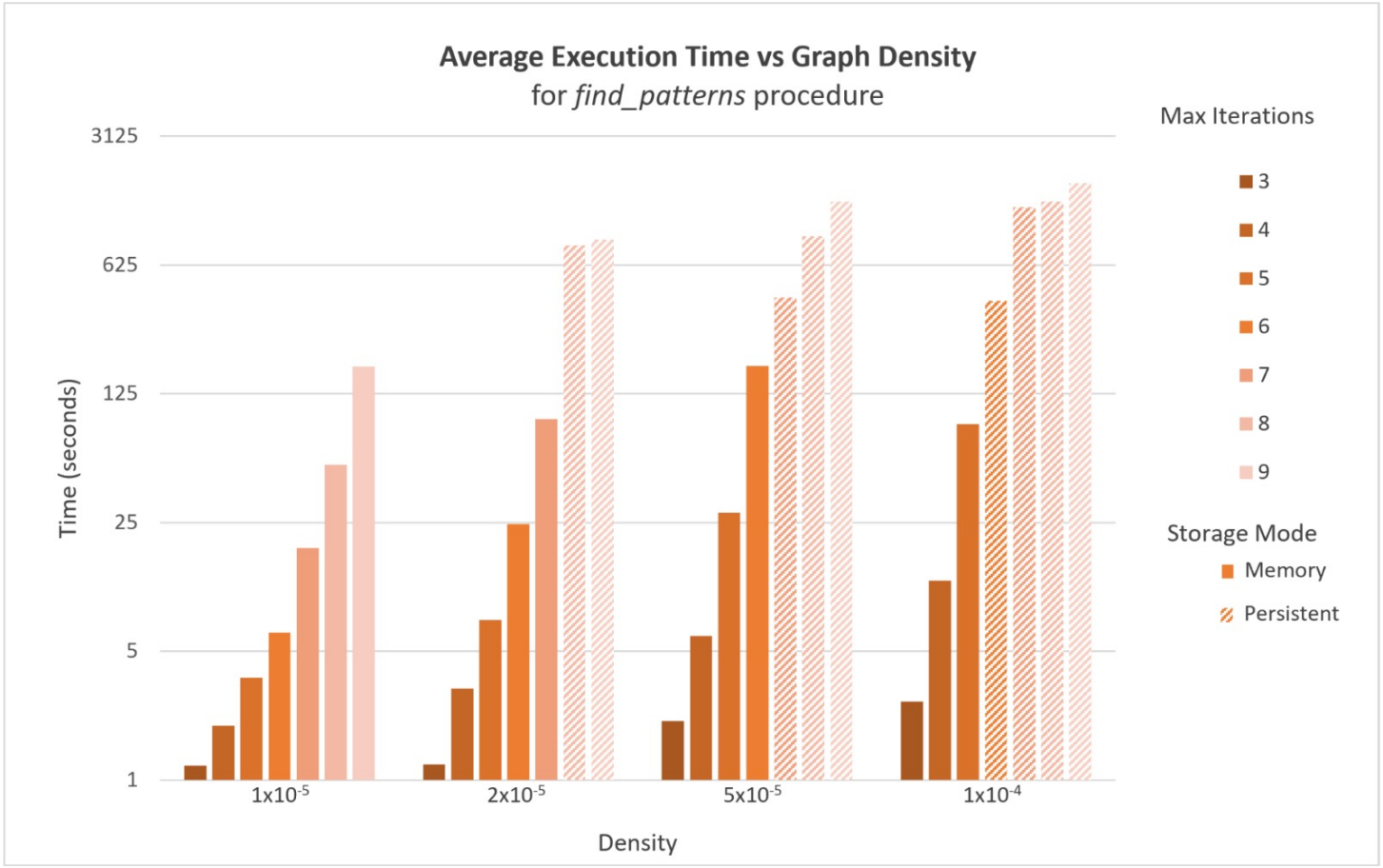
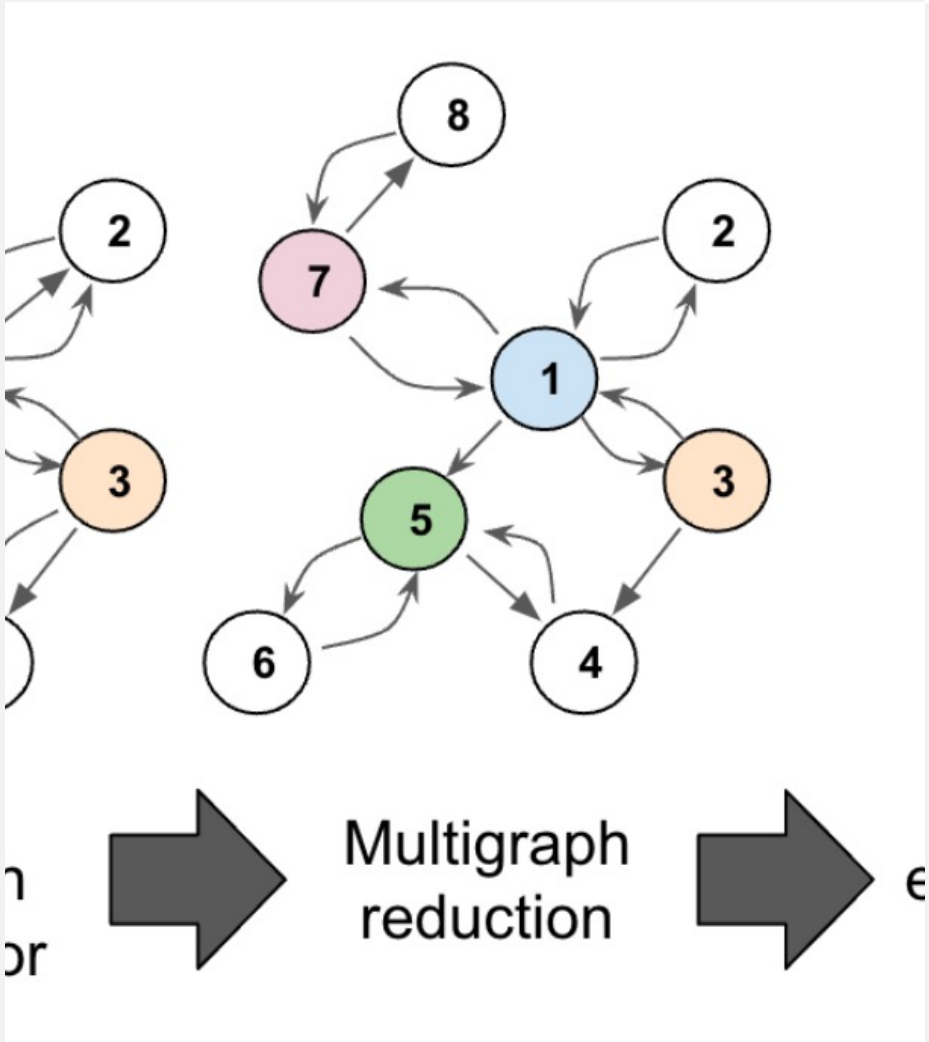


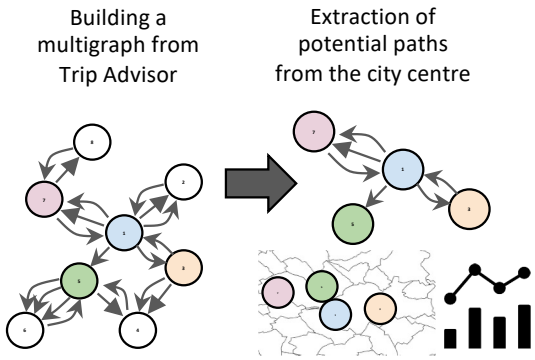
Figure 5.3: Average Execution Time vs Graph Density for Synthetic Graph Data



TOURISTIC SEQUENTIAL PATTERNS

SeqPatTour: sequential patterns-based metrics for better understanding the tourism behaviour [Quality & Quantity'25]

TOURISTIC SEQUENTIAL PATTERNS

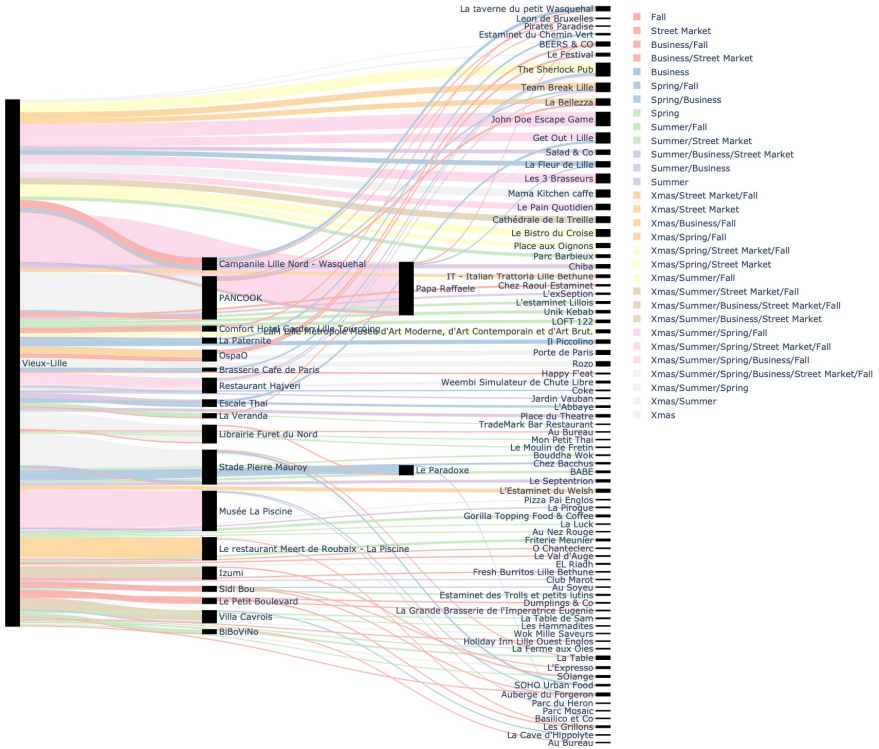


Sequential patterns analysis – graph metrics:

- Path Preference Index $$PPI(s) = \prod_{i=0}^{|s|} \frac{\omega(a_i, a_{i+1})}{\omega(a_i)}$$
- Path Satisfaction Index $$PSI(s) = \frac{\sum_{i=0}^{|s|} rating(a_i) \times nbPosts(a_i)}{\sum_{i=0}^{|s|} nbPosts(a_i)}$$
- Location-Type Diversity Index $$H(cat(s)) = -\sum^{|cat(s)|} P_i * \log_2(P_i)$$
- Covered Distance by Path $$CDP(s) = \sum_{i=0}^{|s|-1} dist(a_i, a_{i+1})$$
- Diversity by Distance Covered Index $$DDI(s) = \frac{e^{\delta \times LDI(s)}}{CDP(s)}$$
- Nationality Diversity Index $$NDI(s) = \frac{\sum_{i=0}^{|s|} gini(i)}{|Countries|}$$

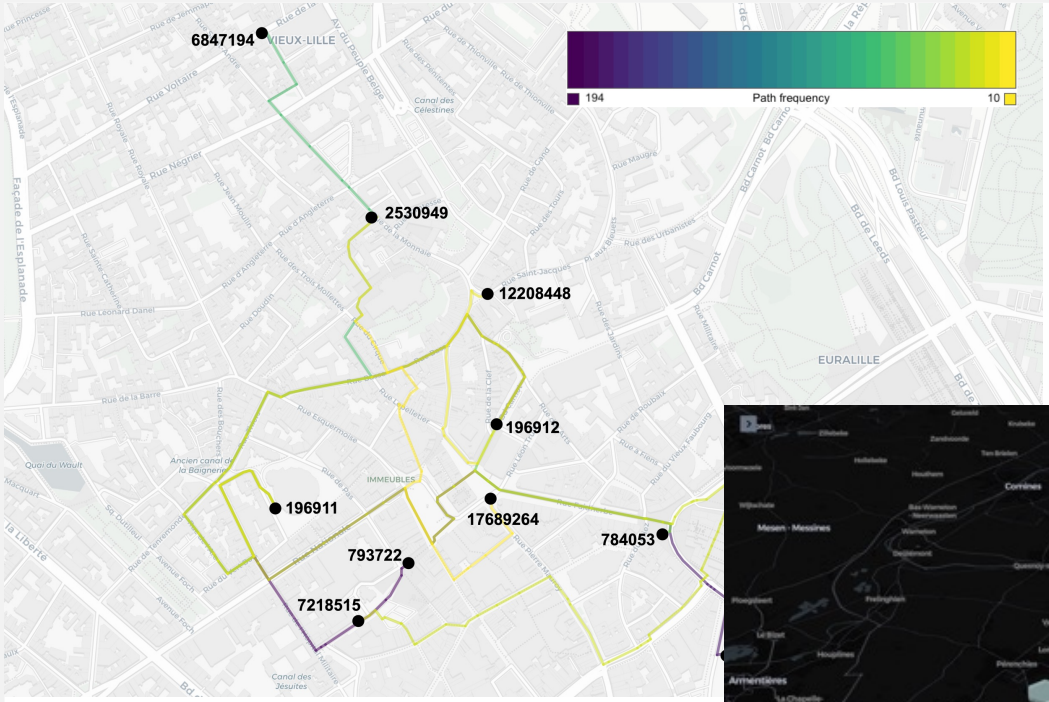
TOURISTIC SEQUENTIAL PATTERNS





DE VINCI HIGHER EDUCATION

SeqPatTour: sequential patterns-based metrics for better understanding the tourism behaviour [Quality & Quantity'25]





OVERTOURISM

From Flows to Graphs: Data-Driven Insights on Latent Overtourism with Frequent Pattern Mining [DOING@ADBIS'25]

Overtourism describes destinations where hosts or guests, locals or visitors feel that there are too many visitors and that the quality of life in the area, or the *quality of the experience* has deteriorated unacceptably.

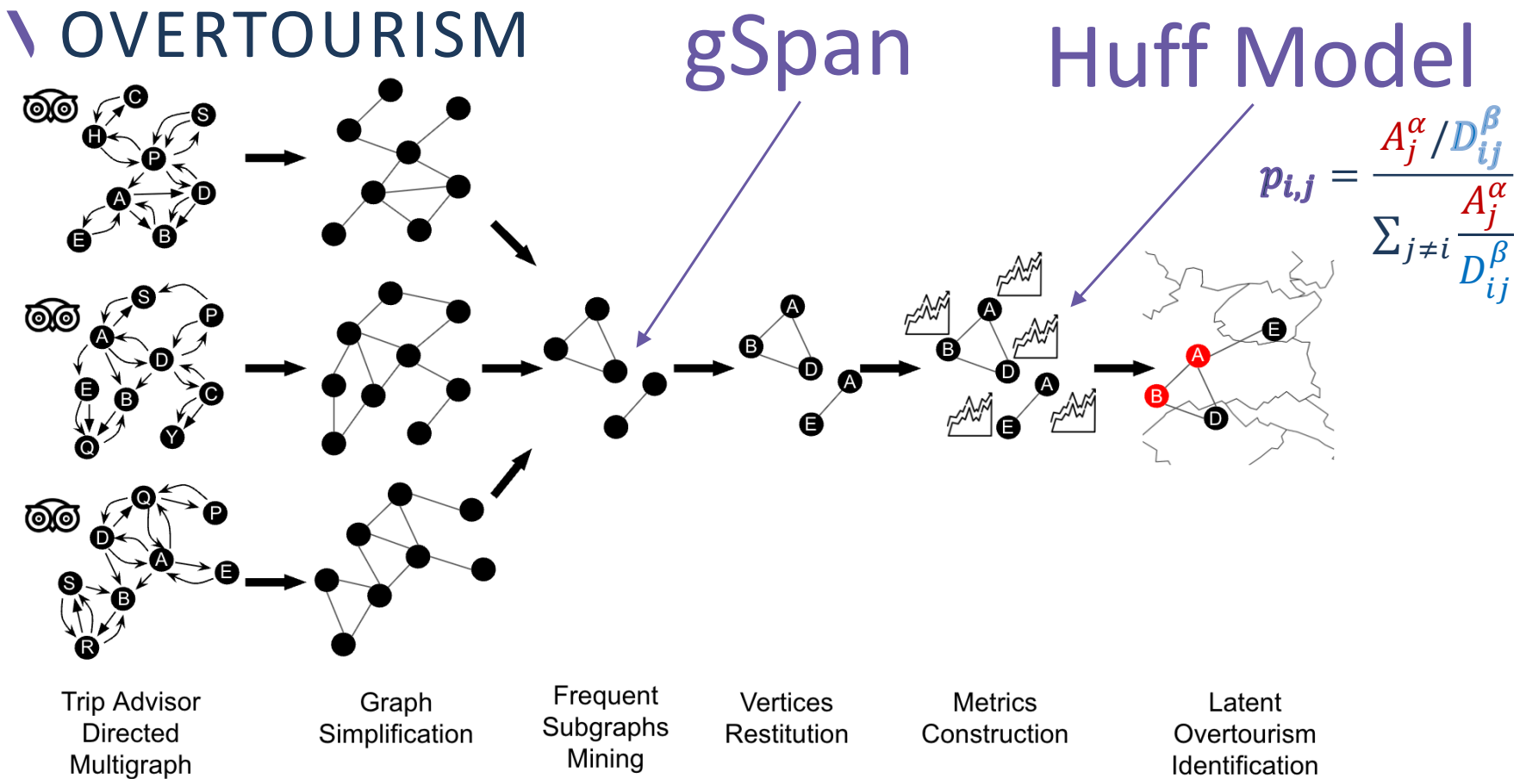
[Butler R., Overtourism, 2019]

Latent overtourism refers to destinations showing early warning signs before tipping points

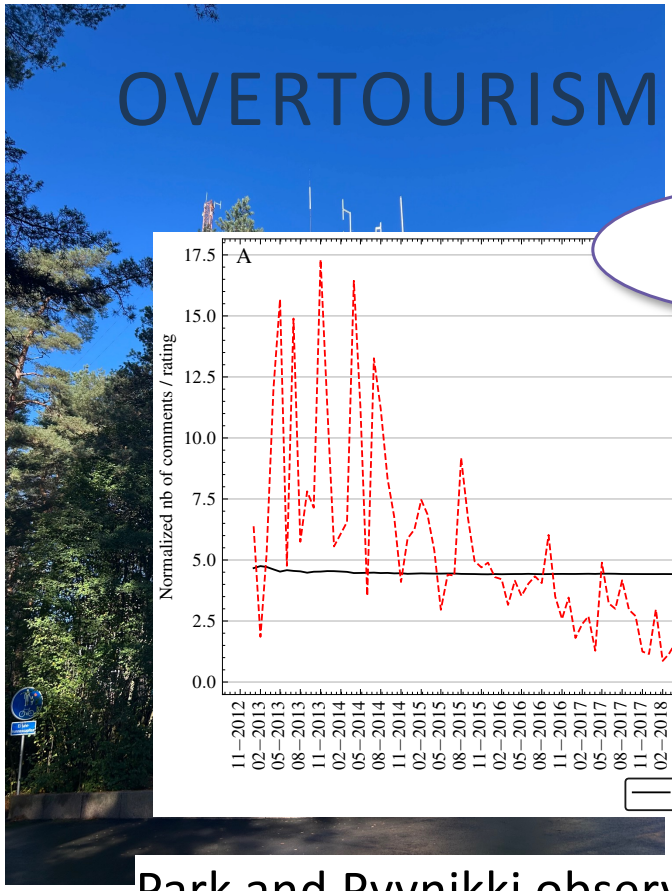
[Adie, Falk & Savioli, Current issues in Tourism, 2020]



https://www.solimarinternational.com/what-is-overtourism-and-how-can-we-prevent-it/



LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR



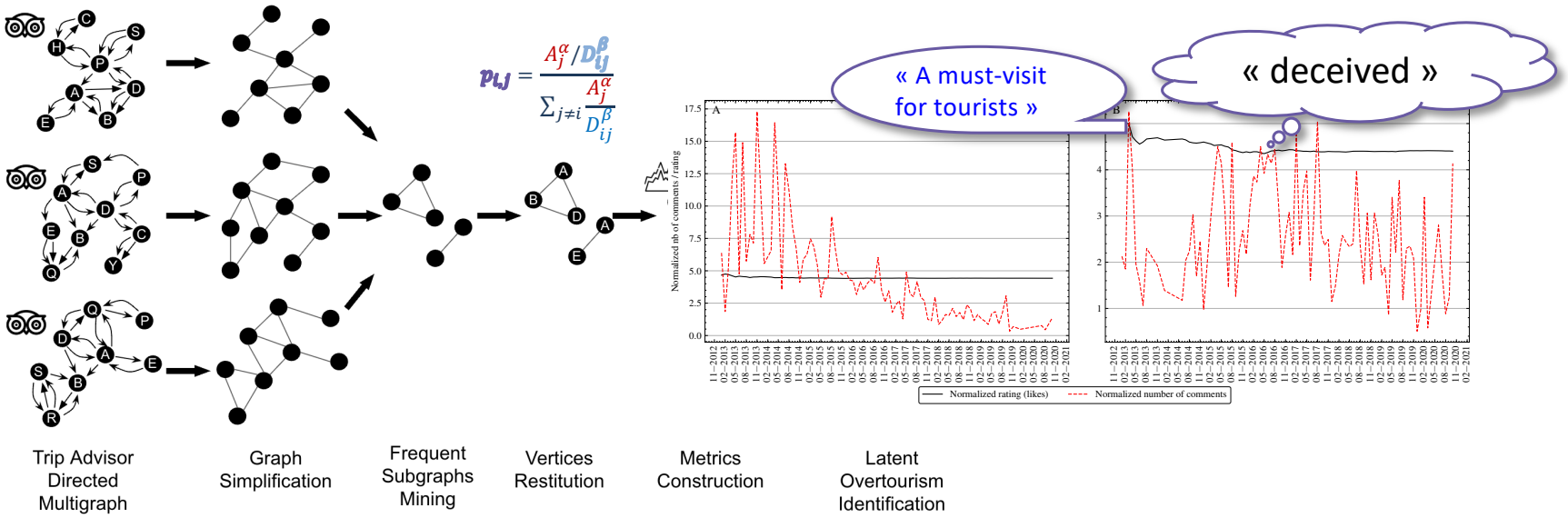
« A must-visit for tourists »

« deceived »

Park and Pyynikki observation tower

Pyynikin Nakotorninkahvila

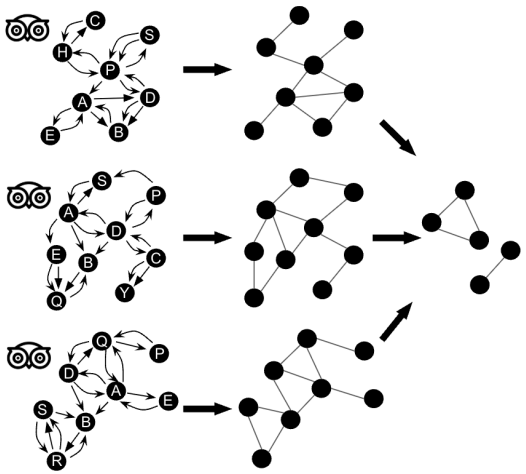
OVERTOURISM

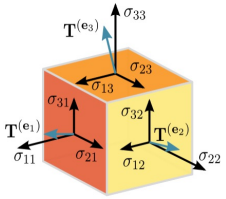


NEXT POI PREDICTION

(MASTER THESIS LAURA & RUBESH & AURELIEN)

- Prediction of the next POI
- Based on context-aware frequent sequential patterns





Tensor Transformation:
Sequential patterns
+ context

LSTM
Predictions:
Next POI

NUMEROUS COLLABORATIONS IN DIGITAL TOURISM



Hugo Alatrista-Salas
(DVRC, DVHE)



Gaël Chareyron
(DVRC, DVHE)



Sonia Djebali
(DVRC, DVHE)



Saïd Jabbour
(Cril, Univ. Artois)



Sébastien Jacquot
(EIREST, Univ. Panthéon Sorbonne)



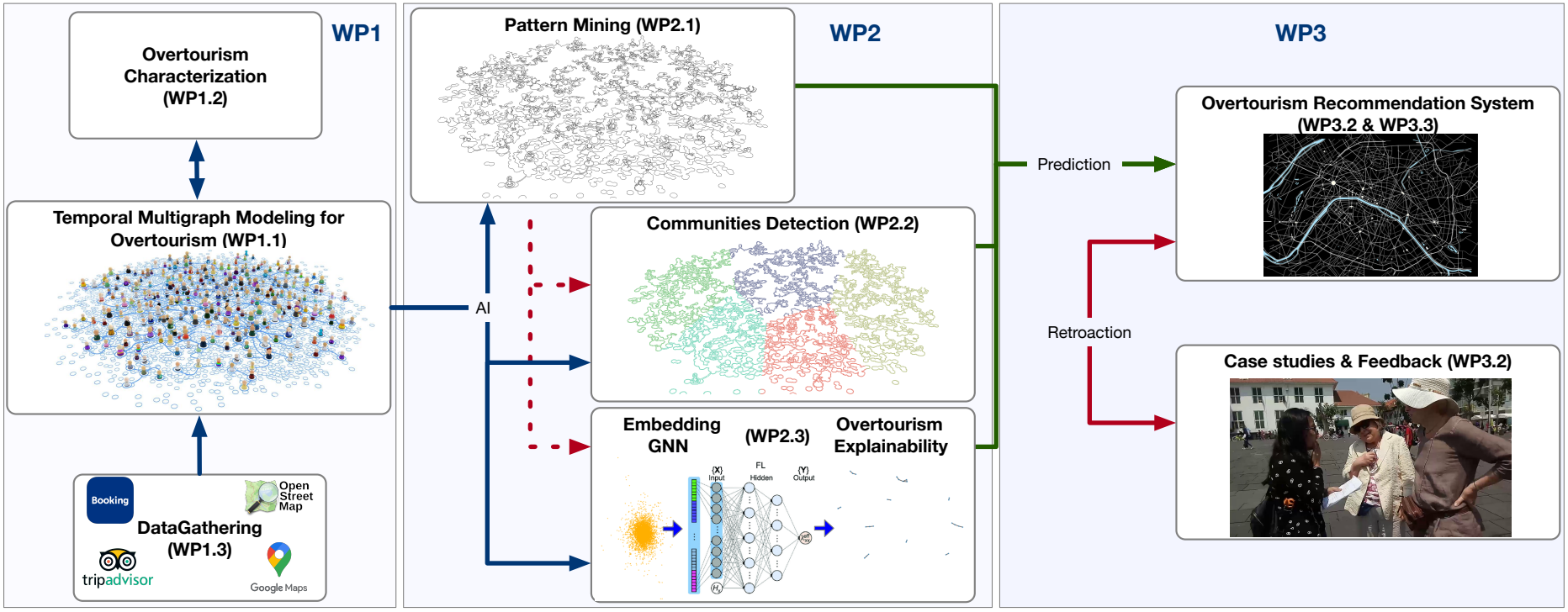
Imen Ouled-Dlala
(DVRC, DVHE)



Nicolas Travers
(DVRC, DVHE)

And also, thanks to:
Laura Aboukrat, ESILV
Elena Chatzopoulou, DVRC, EMLV
Koumudi Ganepola, ESILV
Nicolas Loas, ESILV
Julien Martin-Prin, ESILV
Mélanie Mondo, EIREST, Panthéon Sorbonne
Hugo Prevoteau, ESILV
Hugo Quehlas, ESILV
Rubesh Prabakran, ESILV
And more to come!

PERSPECTIVES
STUDY ON OVERTOURISM



LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR

PERSPECTIVES

Problematic:

How can we design an interactive tourist recommendation system that adapts the routes to the user's profile (preferences, time, budget, etc.) while providing clear explanations to help them make their choice?

- **Interactive personalization**: adapt the routes according to the profile, preferences and constraints (time available, distance, schedules, budget).
- **Geo-spatial encoding and LLM**: representating spatiality & temporality in LLMs
- **Explainable recommendations:** propose courses with simple and understandable justifications (combine LLM + structured model (hybrid))

Wang, Y., Sun, H., & Zhang, S. (2025). LLM as Explainable Re-Ranker for Recommendation Systems.

Preprint, cs.IR, Dec 2025.

LIKE,SHARE, FOLLOW ! @poledevinci | DEVINCI.FR