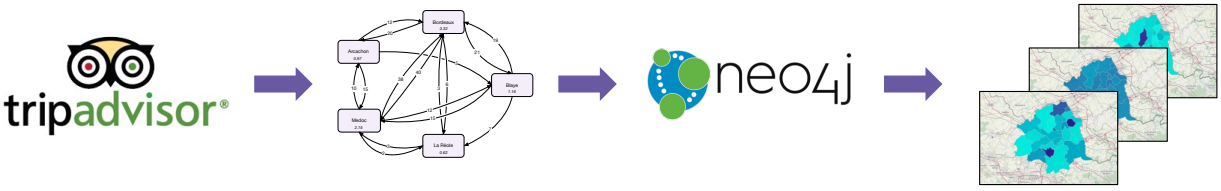
Neo4Tourism – a Framework for Mobility Analysis & Tourist Circulation on DB Oriented-Graph

Nicolas Travers
De Vinci Research Center (DVRC), Paris La Défense, France

Motivation

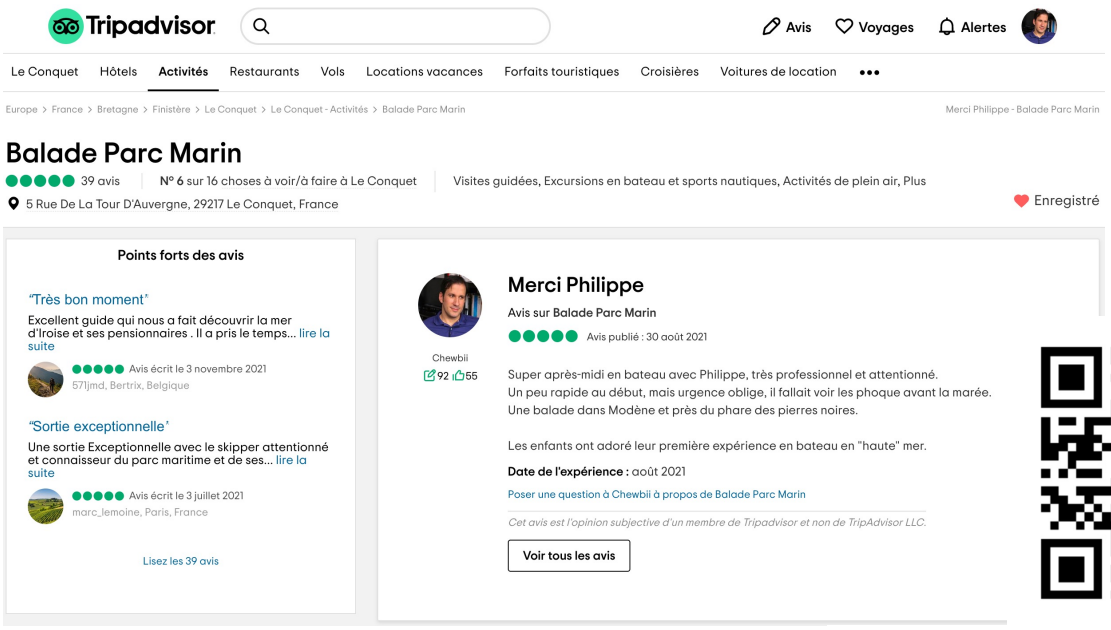
How to measure and compare tourists mobility and its evolution over time and space?



Context

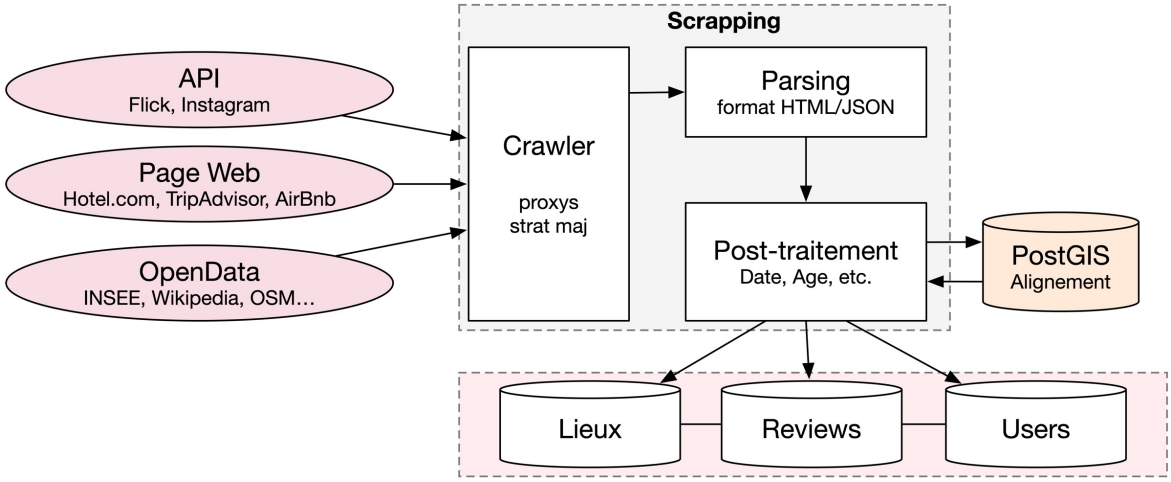
Research contracts with Cities: Tours (2017-2018), Bordeaux (2019-2020), Lille (2020-2021 & 2021-2024)
Partnership: EIREST Lab (*Equipe interdisciplinaire de recherches sur le tourisme*), Paris 1 Panthéon Sorbonne

Goal: Understand tourists behavior through footprints on e-tourism platforms (Tripadvisor, Booking, Airbnb, Flickr, Instagram, Twitter, etc.)



https://www.wooclap.com/NEO4TOURISM

Scraping Architecture



Scraping : Locations example

d	nom	url	rating	nbAvis	nbAvisRecup...	latitude	longitude	shape_gid	type	typeR	priceRange	roomNumber
245765	Place de la Bourse (Place Royale)	/Attraction_Review-g187079-d245765-Reviews-Place_de_la_Bourse_P...	4.5	5286	5363	44.83663	-0.58104	48328	Points of Interest & Landmarks	A		-1
8612874	Le Quatrieme Mur	/Restaurant_Review-g187079-d8612874-Reviews-Le_Quatrieme_Mur...	4	1707	3476	44.8...	-0.5...	48328		R	NULL	-1
10200972	La Cité du Vin	/Attraction_Review-g187079-d10200972-Reviews-La_Cite_du_Vin-Bo...	4	3116	3138	44.862453	-0.550281	48328	Architectural Buildings, Specialty...	A		-1
803318	L Entrecote	/Restaurant_Review-g187079-d803318-Reviews-L_Entrecote-Bordea...	4	2456	3118	44.84319	-0.574668	48328		R	NULL	-1
1441168	La Brasserie Bordelaise	/Restaurant_Review-g187079-d1441168-Reviews-La_Brasserie_Borde...	4	1704	2444	44.84108	-0.573023	48328		R	NULL	-1
219110	Rue Sainte-Catherine	/Attraction_Review-g187079-d219110-Reviews-Rue_Ste_Catherine-B...	4	2275	2317	44.83828	-0.58231	48328	Points of Interest & Landmarks	A		-1
8545109	Escape Hunt Bordeaux	/Attraction_Review-g187079-d8545109-Reviews-The_Escape_Hunt_E...	5	1936	2037	44.824783	-0.558088	48328	Room Escape Games	A		-1
9592207	Restaurant Les 4 Murs	/Restaurant_Review-g187079-d9592207-Reviews-Restaurant_Les_4_...	4	44	1007	44.8429	-0.5745	48328		R	NULL	-1
808476	Le Petit Commerce	/Restaurant_Review-g187079-d808476-Reviews-Le_Petit_Commerce...	3.5	1281	1299	44.84026	-0.57123	48328		R	NULL	-1
808972	Chez Peppone	/Restaurant_Review-g187079-d808972-Reviews-CHEZ_PEPPONE-Bor...	4	1102	1594	44.841858	-0.579675	48328		R	NULL	-1
196937	Ibis Bordeaux Centre Meriadeck	/Hotel_Review-g187079-d196937-Reviews-Ibis_Bordeaux_Centre_Me...	3.5	1492	1493	44.83606	-0.583402	48328		H	NULL	-1
219045	Cathédrale Saint-André	/Attraction_Review-g187079-d219045-Reviews-St_Andre_Cathedral_...	4.5	1436	1462	44.837093	-0.578348	48328	Sacred & Religious Sites	A		-1
952751	La Tupina	/Restaurant_Review-g187079-d952751-Reviews-La_Tupina-Bordeau...	3.5	1036	1388	44.83289	-0.562181	48328		R	NULL	-1
651293	InterContinental Bordeaux Le Gran...	/Hotel_Review-g187079-d651293-Reviews-Le_Grand_Hotel_de_Borde...	4.5	1526	1275	44.84258	-0.578464	48328	hotels, Hotel	H	241€ - 549€ (Selon les tarifs moy...	130
4311273	Le Pont Jacques Chaban Delmas	/Attraction_Review-g187079-d4311273-Reviews-Le_Pont_Jacques_Ch...	4.5	1179	1201	44.831997	-0.561162	48328	Points of Interest & Landmarks	A		-1
669310	Ibis Budget Bordeaux Centre Meria...	/Hotel_Review-g187079-d669310-Reviews-Ibis_Budget_Bordeaux_Ce...	3	1138	1140	44.83583	-0.584936	48328		H	NULL	-1
1325436	Restaurant Melodie	/Restaurant_Review-g187079-d1325436-Reviews-Restaurant_Melodi...	4.5	1099	1108	44.8406	-0.57047	48328	French, European	R	€€ - €€€	-1
809368	Chez Jean	/Restaurant_Review-g187079-d809368-Reviews-Chez_Jean-Bordeau...	4	845	1099	44.8407	-0.572009	48328		R	NULL	-1
2038188	Bordovino Wine Tasting Day Tours	/Attraction_Review-g187079-d2038188-Reviews-Bordovino_Wine_Ta...	5	1089	1094	44.84584	-0.57613	48328	Wine Tours & Tastings, Day Trips	A		-1
809368	Pont de Pierre	/Attraction_Review-g187079-d290953-Reviews-Pont_de_Pierre-Bord...	4.5	1066	1075	44.841377	-0.562415	48328	Points of Interest & Landmarks, B...	A		-1
219111	Grand Théâtre	/Attraction_Review-g187079-d219111-Reviews-Grand_Theatre_Oper...	4.5	1047	1053	44.832684	-0.56217	48328	Theaters, Architectural Buildings,...	A		-1
3489515	Peponne	/Restaurant_Review-g187079-d3489515-Reviews-Peponne-Bordeaux...	3.5	56	1031	44.841835	-0.579685	48328	italienne	R	NULL	-1
196932	Quality Hotel Bordeaux Centre	/Hotel_Review-g187079-d196932-Reviews-Quality_Hotel_St_Catherin...	4	1154	989	44.84053	-0.57407	48328	hotels, Hotel	H	116€ - 182€ (Selon les tarifs moy...	84
6728255	Rustic Vines	/Attraction_Review-g187079-d6728255-Reviews-Rustic_Vines-Borde...	5	968	973	44.844845	-0.579951	48328	Wine Tours & Tastings, Bike Tours	A		-1
1810068	Le Noailles	/Restaurant_Review-g187079-d1810068-Reviews-Le_Noailles-Bordea...	4	627	928	44.843376	-0.575564	48328		R	NULL	-1
2420614	Fufu	/Restaurant_Review-g187079-d2420614-Reviews-Fufu-Bordeaux_Gir...	4.5	799	889	44.841366	-0.572364	48328	Japanese, Asian, Soups	R	€	-1
5281513	Mama Shelter Bordeaux Restaurant	/Restaurant_Review-g187079-d5281513-Reviews-Mama_Shelter-Bor...	3.5	642	872	44.839886	-0.577318	48328		R	NULL	-1
4096472	Petit Mignon	/Restaurant_Review-g187079-d4096472-Reviews-Petit_Mignon-Bord...	4.5	725	869	44.84121	-0.572636	48328		R	NULL	-1
1109727	Adagio Bordeaux Gambetta	/Hotel_Review-g187079-d1109727-Reviews-Adagio_Bordeaux_Gamb...	4	987	856	44.839977	-0.582532	48328	hotels, Hotel	H	90€ - 132€ (Selon les tarifs moye...	111
781600	L Ombriere	/Restaurant_Review-g187079-d781600-Reviews-L_Ombriere-Bordea...	3	551	851	44.838665	-0.572935	48328		R	NULL	-1
196933	Hôtel Mercure Bordeaux	/Hotel_Review-g187079-d196933-Reviews-Mercure_Bordeaux_Centr...	3.5	1040	845	44.839134	-0.586448	48328	hotels, Hotel	H	90€ - 165€ (Selon les tarifs moye...	194
196935	Novotel Bordeaux Lac	/Hotel_Review-g187079-d196935-Reviews-Novotel_Bordeaux_Le_Lac...	4	1047	837	44.890648	-0.566261	48328	hotels, Hotel	H	84€ - 134€ (Selon les tarifs moye...	175
196936	Novotel Bordeaux Centre	/Hotel_Review-g187079-d196936-Reviews-Novotel_Bordeaux_Centre...	3.5	1035	827	44.836227	-0.583852	48328	hotels, Hotel	H	93€ - 152€ (Selon les tarifs moye...	137

Lieux

Alignement

Calculs

Post-traitement

Scraping : Review examples

id	idplace	titre	review	note	date_review	date_visit	langue
576459758	245765	Must See ?	One of the main attractions in Bordeaux and definitely worth seeing. Take t…	5	2018-04-29	2018-04-01	eng
576376524	245765	Magnificent place	Well this is the place to go in Bordeaux as it has this little something with th…	5	2018-04-29	2018-04-01	eng
576223693	245765	miroir d eau super ;;;A voir !!!	moment de détente on peu suivant le temps marcher nu pieds dans l eau G…	5	2018-04-28	2018-04-01	fra
576314935	245765	Красиво	ÐÑÐµÐ½Ñ Ð°ÑÐ°ÑÐ¸Ð²Ð¾ Ð¼ÐµÑÑÐ¾, Ð¾ÑÐ¾Ð±ÐµÐ½Ð½Ð¾ Ð²ÐµÑ… Ð²ÐµÑ…	5	2018-04-28	2018-04-01	rus
576248908	245765	Meravigliosa	Una piazza enorme e meravigliosa, che dà sul fiume. Palazzi lussuosi e sens…	5	2018-04-28	2018-04-01	ita
576058008	245765	Travail studieux	Réunion de travail dans endroit magnifique . J'en ai profité. Pour faire quelq…	4	2018-04-27	2018-04-01	fra
575974257	245765	Must see when in Bordeaux.	Really nice palace and river area, very tourist friendly with great vibes. Lovel…	4	2018-04-27	2018-04-01	eng
575960518	245765	Unas gran plazoleta	…rtar el tipo de clima en lugar es espectacular grande Magno interes…	4	2018-04-27	2018-04-01	spa
575867918	245765	ensemble architectural	une rèrèrence remplie d histoire et de pavés et face a la berge et le miroir d…	4	2018-04-26	2018-04-01	fra
575916206	245765	Buen lugar para tomar el pulso de la ciudad	Acabamos yendo los dos días que estuvimos a eso de las ocho que es cuan…	4	2018-04-26	2018-04-01	spa
575560663	245765	A voir de près et de loin, sur terre et depuis le Garonne	Bel ensemble architectural classique, le lieu est très vivant à proximité du m…	4	2018-04-25	2018-04-01	fra
575708497	245765	Ícono de Bordeaux	Ícono de Bordeaux que no puede pasarse por alto si se visita la ciudad. Está…	5	2018-04-25	2018-04-01	spa
575371903	245765	Bien joli espace historique	Sacré construction. C est vrai que l on se croirait a Saint-Petersbourg. Rien à…	4	2018-04-24	2018-04-01	fra
575347509	245765	Ein Muss in Bordeaux.	Zusammen mit dem Wasserspiegel definitiv ein Muss für alle, die in Bordeau…	5	2018-04-24	2017-10-01	deu
575250426	245765	Magnifique	Magnifique place avec ses immeubles tournés vers la Garonne et le miroir d…	5	2018-04-23	2018-04-01	fra
575065631	245765	Buraya yakın kalın	Bordo'da kalacaksanız buraya yakın kalmanızı öneririm. Hemen hemen her y…	5	2018-04-23	2018-03-01	eng
574969263	245765	Passage obligé sur les quais	Fait partie des monuments incontournables à visiter lors de votre passage à…	4	2018-04-22	2018-04-01	fra
574925904	245765	Lovely	Luckily the weather was fabulous for April. Picnic lunch and great photo opp…	4	2018-04-22	2018-04-01	eng
575015418	245765	Verrassende plek.	Leuk en verrassend object dat om een bepaalde tijd stoom dan weer water…	4	2018-04-22	2018-04-01	nld
574659412	245765	Mooi plein	Place de la Bourse is een flink uit de kluiten gewassen half rond plein, afgeb…	4	2018-04-21	2018-04-01	nld
574635467	245765	La grandeur et la splendeur bordelaise !	L unité architecturale des somptueux bâtiments qui compose la place de la…	5	2018-04-20	2018-04-01	fra
574620974	245765	toujours aussi magnifique	place toujours tres propre et bien entretenue c est un vrai plaisir de venir à…	5	2018-04-20	2017-08-01	fra
574586138	245765	Stunning	All ages enjoy this attraction. Simple idea and mesmerising. Great photos w…	5	2018-04-20	2018-04-01	eng
573974988	245765	Tout simplement magnifique	De passage à Bordeaux un arrêt s impose sur la place de la bourse de jour c…	5	2018-04-17	2018-04-01	fra

lieu

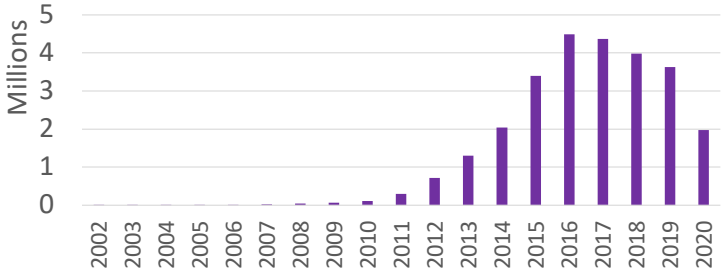
Post-traitement

nat

The Tripadvisor Database (since 2002)

	Nb locations	Nb Reviews	Nb Users
France	284 988	26 992 448	7 648 165
Nouvelle-Aquitaine	31 895	2 787 809	1 120 863
Gironde	8 475	806 224	375 118
Hauts-de-France	13 650	1 153 852	499 258
Nord	5 641	439 477	202 516

Distribution of reviews in France per year



Plan

- Context & Tripdavisor dataset
- Related Work
  - Graphs & Circulation
- Circulation Graph
  - Graph Data Model
  - Graph Data Manipulations
  - Integration with Neo4j
- The Circulation Factor
  - TCF & GCF
  - Experiments
  - PageRank vs Betweenness Centrality
- Tourists Propagation
  - Maximum Spanning Trees
  - Experiments
- Conclusion & Perspectives

Related Work: Graphs and circulation?

Graphs community extraction

- HCS [Hartuv00], Louvain [Blondel08], Label Propagation [Raghavan07], Chameleon [Karypis99]
- *Groups of nodes, no circulation*

Trees

- Spanning Trees [Graham85], Maximum Frequent Sub-Graph [Hua04]
- *Main path, main trend*

Graphs centrality

- Closeness [Das18], Betweenness [Das18], Degree [Das18], Eigen [Das18], PageRank [LangVille07]
- *Hardly comparable*

Tourists flow extraction

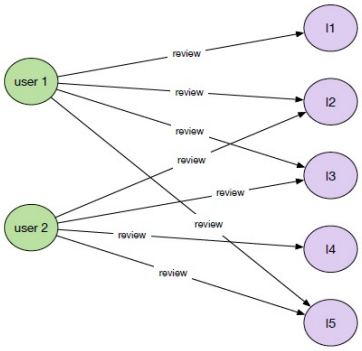
- Flow visualization [Chua16], pattern mining [Vu15], POI extraction [Spyrou17], Kernel density [Sun13]
- *Static and hardly flexible*

Plan

- Context & Tripdavisor dataset
- Related Work
  - Graphs & Circulation
- Circulation Graph
  - Graph Data Model
  - Graph Data Manipulations
  - Integration with Neo4j
- The Circulation Factor
  - TCF & GCF
  - Experiments
  - PageRank vs Betweenness Centrality
- Tourists Propagation
  - Maximum Spanning Trees
  - Experiments
- Conclusion & Perspectives

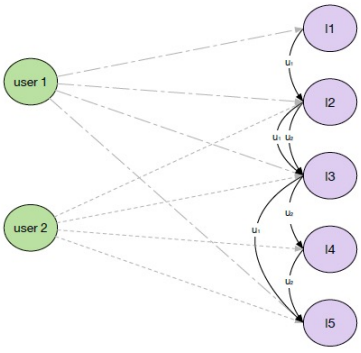
Circulation Graph Model

Bi-partite graph C(V, E)

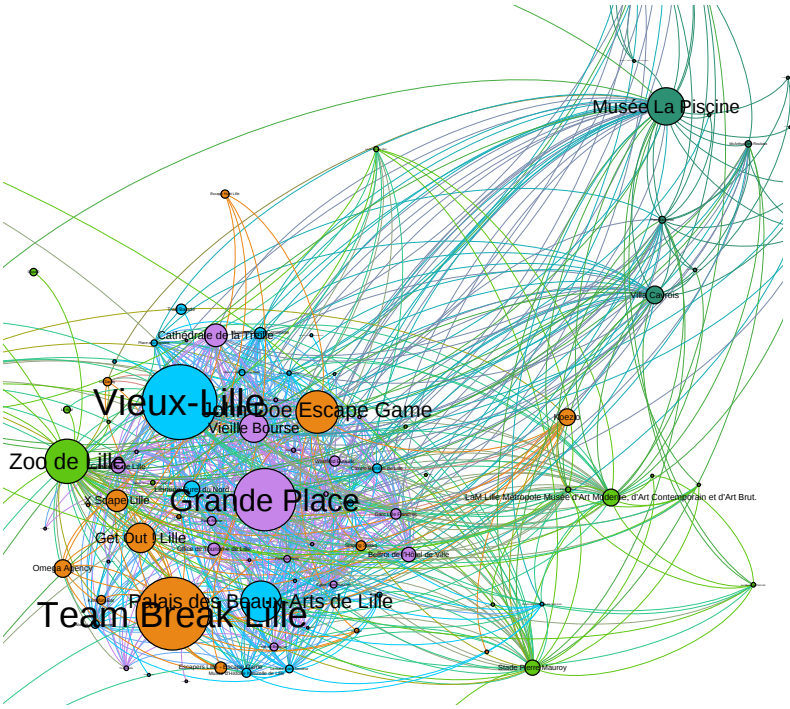


Geolocalized locations
Linked to administrative areas (***GADM***)

Circulation graph C(V', E')



(t2 − t1 < **7 days**)
[Gössling18]



High connectivity between <u>local</u> locations

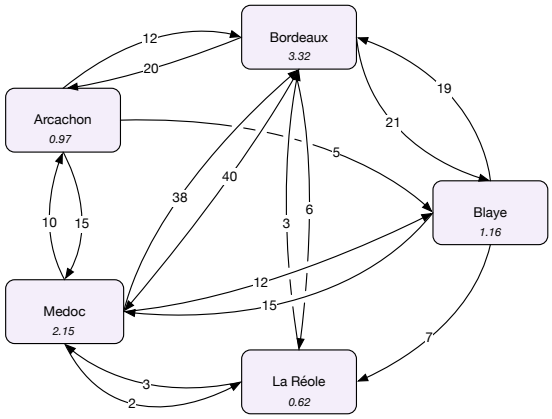
Gare Lille Flandres - Vieille Bourse - Grand'Place - Office de Tourisme/Palais Rihour - Citadelle de Lille - Cathédrale Notre Dame de la Treille – Euralille

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

Circulation Graph Manipulation

Circulation graph aggregation AC(V',E') => Study global circulation with zones



- Aggregate nodes on areas (GADM)
  - Countries, regions, districts, cities…
- Aggregate edges on time & user properties
  - Year/month, nationality, age…

Circulation Graph Manipulation

Applying graph-database "geodesic" operations

- Filter nodes on geographic locality: region, department, etc.
- Filter edges on user properties: nationality, time, etc.
- Aggregate edges

Filters on edges

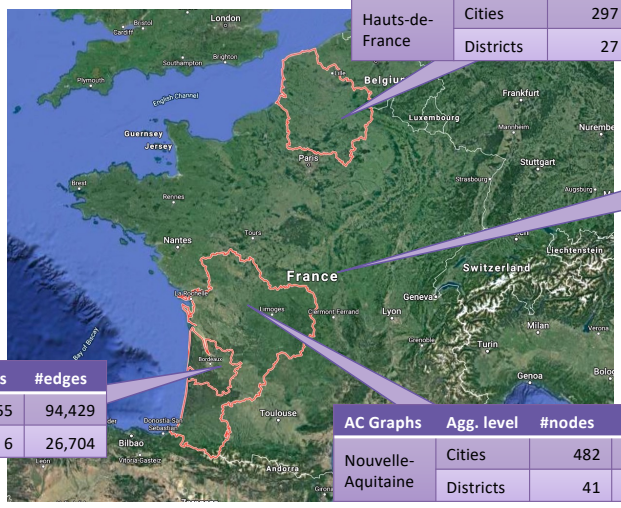
$$ac^{USA,2018}_{cities,year\&nat} = AC(cities, year\&nat, \sigma_{nat="USA" \wedge year=2018})$$

Input graph (perimeter)

Level of aggregation

The Tripadvisor Dataset to the Circulation Graph

tripadvisor®



AC Graphs	Agg. level	#nodes	#edges
Hauts-de-France	Cities	297	153,345
	Districts	27	75,429

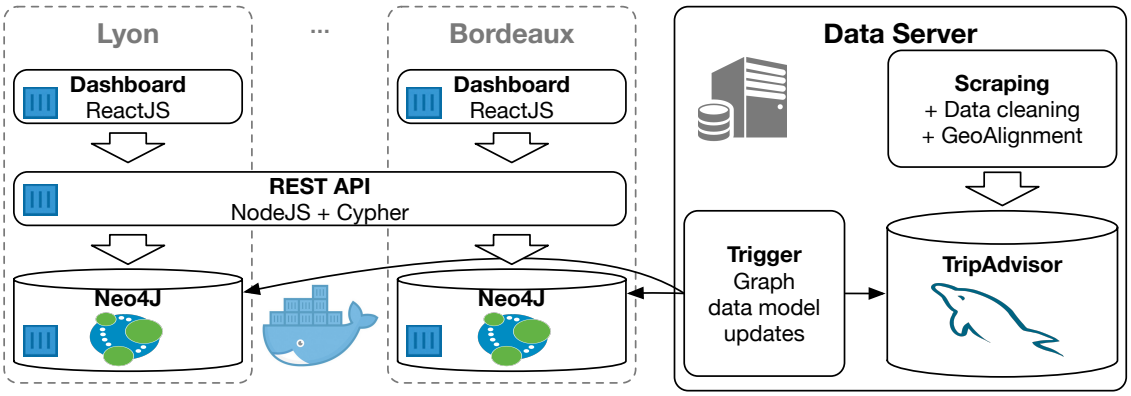
AC Graphs	Agg. level	#nodes	#edges
France	Towns	16 824	5 596 001
	Cities	3 678	4 725 402
	Districts	350	1 858 446

AC Graphs	Agg. level	#nodes	#edges
Gironde	Cities	55	94,429
	Districts	6	26,704

AC Graphs	Agg. level	#nodes	#edges
Nouvelle-Aquitaine	Cities	482	382,266
	Districts	41	170,403

The Neo4Tourism Framework
With Gaël Chareyron & Ugo Quelhas

Construction of the circulation graph with Neo4j [WISE19, BDA19]



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

The Neo4Tourism Framework

Automatic geodesic aggregation of graphs (AC on cities) 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
```

$ac_{cities,year\&nat}$

Automatic subgraphs extraction and computation with *Cypher projections* using *gds*

```
CALL gds.graph.create.cypher("Gironde_USA_2018",
  "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("Gironde_USA_2018",
  {dampingFactor:0.85,iterations:50,weightProperty:true}) YIELD node, score
  RETURN node.city, sum(score) as score;
```

$ac^{USA,2018}_{cities \in Gironde,year\&nat}$

Neo4Tourism (WISE20)

Neo4j - GDS

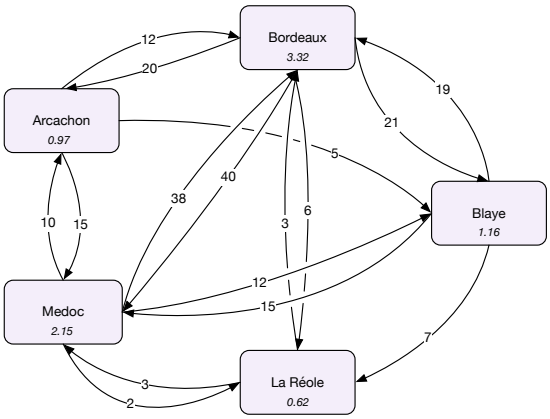
Plan

- Tripdavisor dataset
- Related Work
  - Graphs & Circulation
- Circulation Graph
  - Graph Data Model
  - Graph Data Manipulations
  - Integration with Neo4j
- The Circulation Factor
  - TCF & GCF
  - Experiments
  - PageRank vs Betweenness Centrality
- Tourists Propagation
  - Maximum Spanning Trees
  - Experiments
- Conclusion & Perspectives

Extracting centrality of circulation

Centrality of nodes (areas)

- Witness of mutual activity in the area and interconnections
- Highlight central areas in tourist trips
- PageRank centrality [LangVille07]
  - Markov Chain process
  - Simulate random walks
  - Graph topology & weights dependent
    - Two AC graphs cannot be compared…



TCF – The Transient Circulation Factor
With Sonia Djebali & Nicolas Loas

[WISE20]

TCF:
- Measures the impact of a population *p* on a node *v* of AC
- Is the ratio of centralities of *v* (*PageRank* here)
  - For a context *f* (i.e., year)
  - Between *p* and the whole graph filtered by *f*
  - ➢ *PageRank normalization*

$$TCF^{p,f}_{n,e}(AC,\nu) = \frac{PR_\nu(ac^{p,f}_{n,e})}{PR_\nu(ac^{-,f}_{n,e})}$$

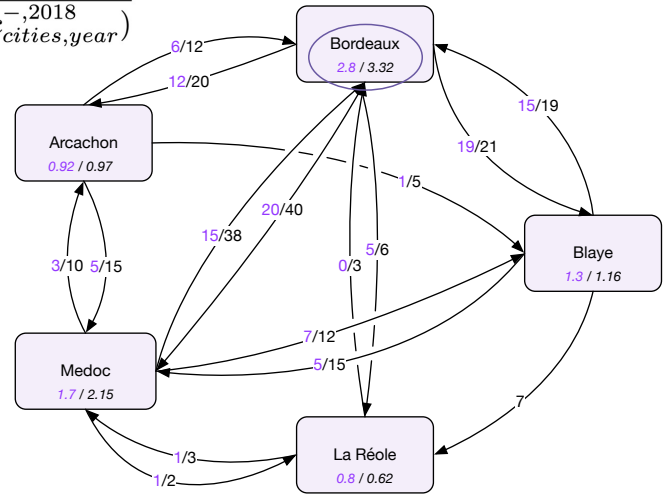
$$TCF^{USA,2018}_{cities,year}(AC, Bordeaux) = \frac{PR_{Bordeaux}(ac^{USA,2018}_{cities,year})}{PR_{Bordeaux}(ac^{-,2018}_{cities,year})}$$

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

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:

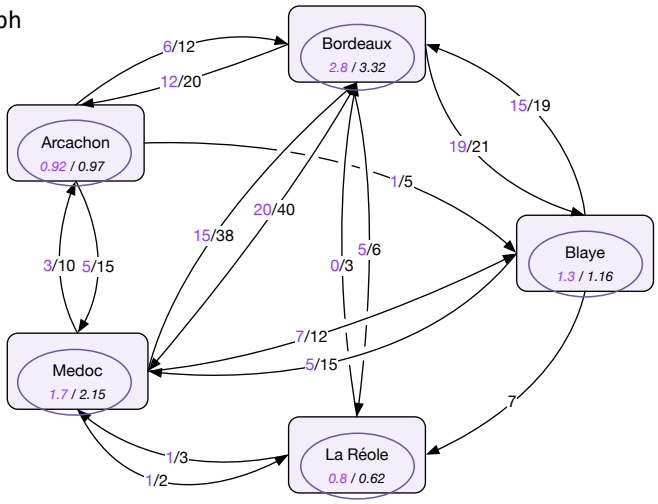
$$TCF^{USA,2017}_{cities,year}(AC, Bordeaux) < TCF^{USA,2018}_{cities,year}(AC, Bordeaux)$$

- Compare populations:

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

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*

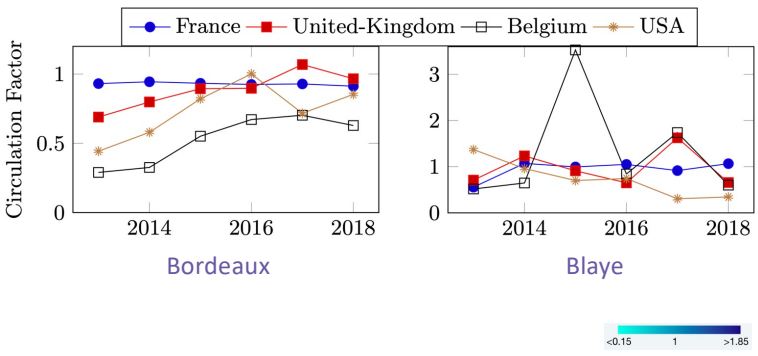


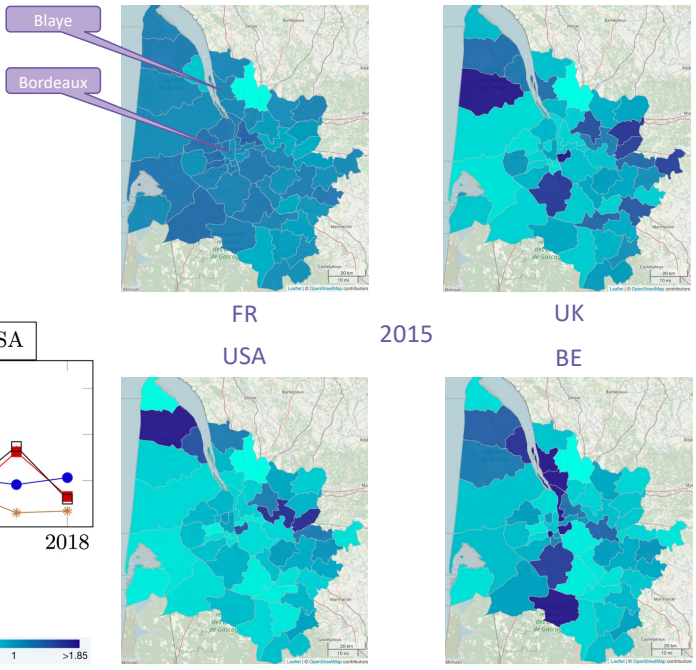
Main Goal

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

Experiments – TCF Evolution

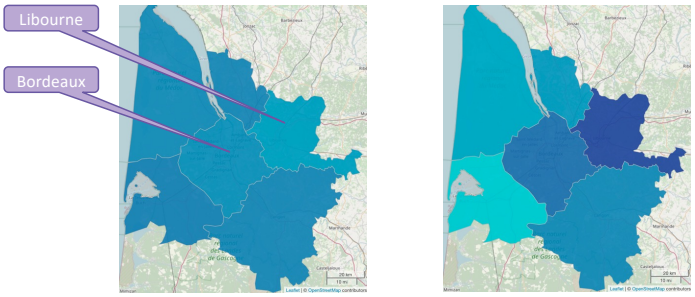
Area: Department (Gironde)
Aggregation level: Cities
⇒ *TCF at small scales - events detection*

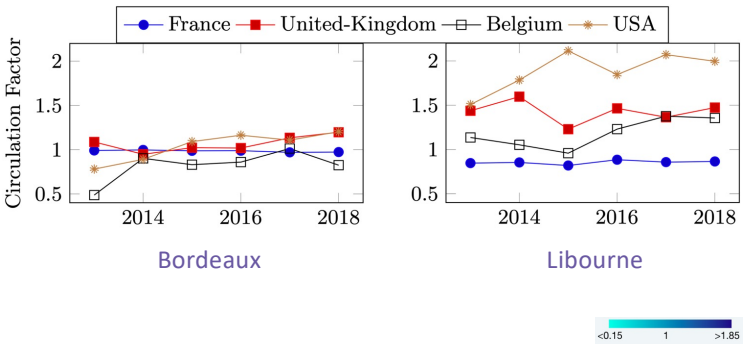


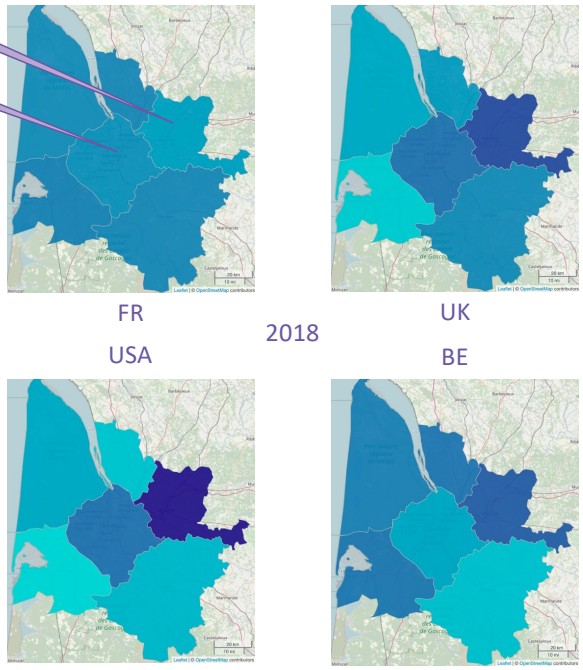


Experiments – TCF Evolution

Area: Department (Gironde)
Aggregation level: Districts
⇒ *TCF at large scales - tendencies extraction*

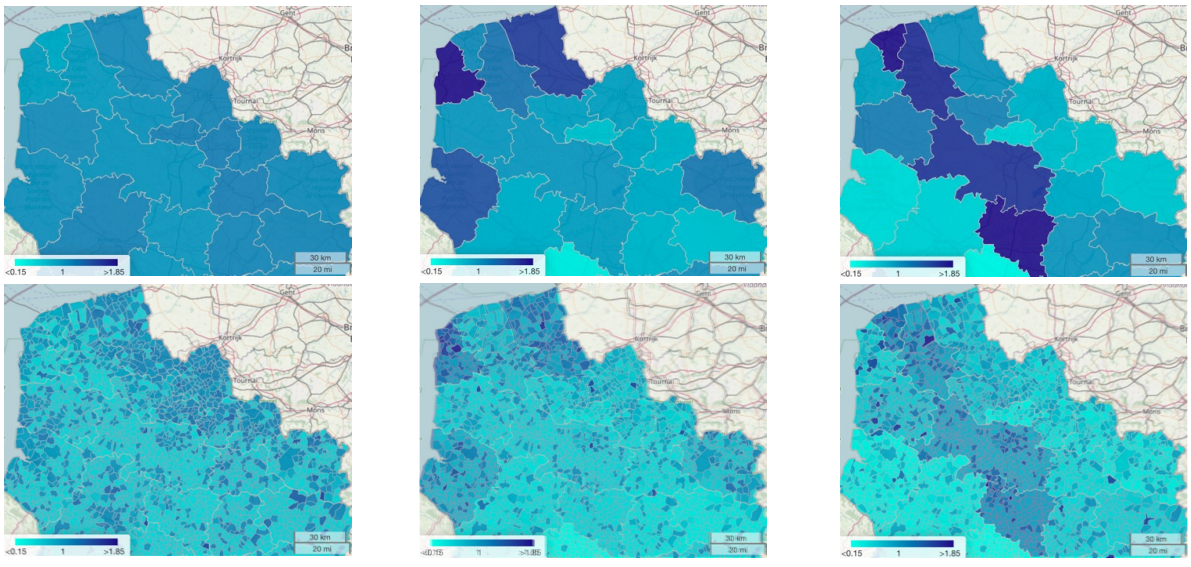






Experiments – TCF comparison

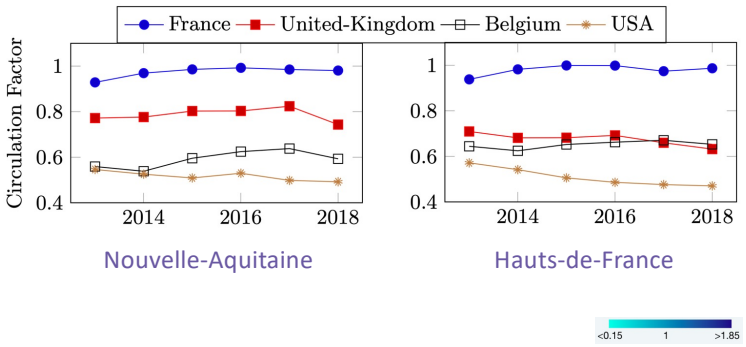
Hauts-de-France - 2019

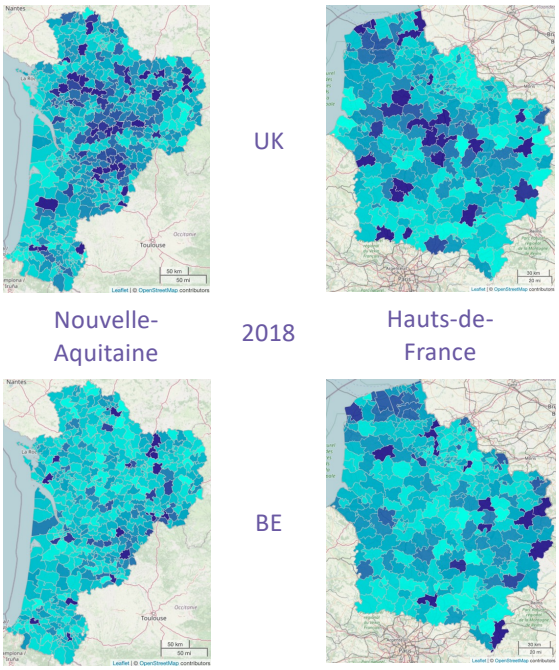


French ***Belgium*** ***English***

Experiments – GCF Evolution

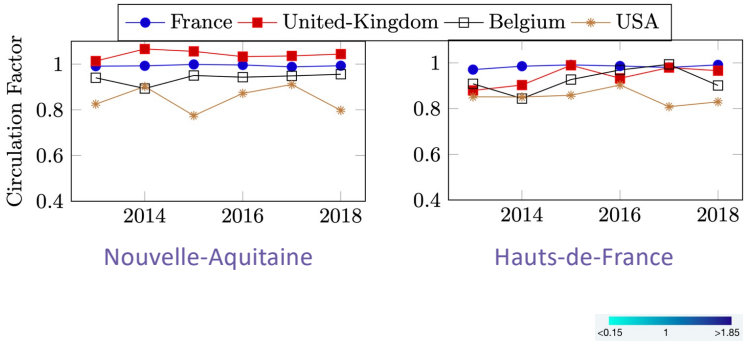
Area: Region
Aggregation level: Cities
⇒ *GCF at small scales - global interest of an area*

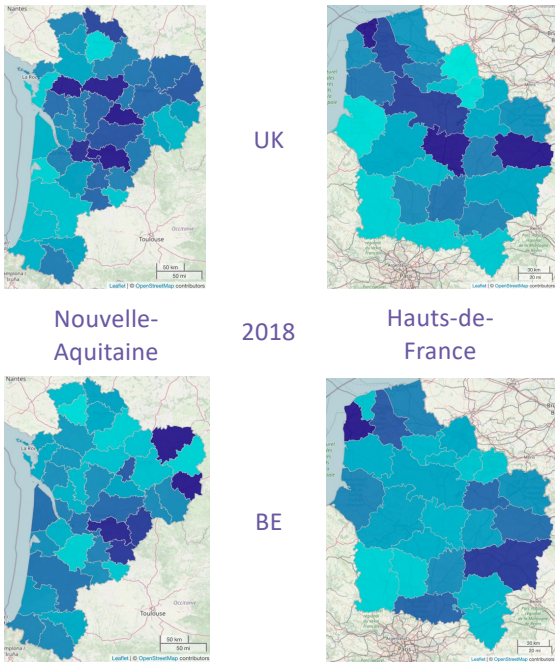




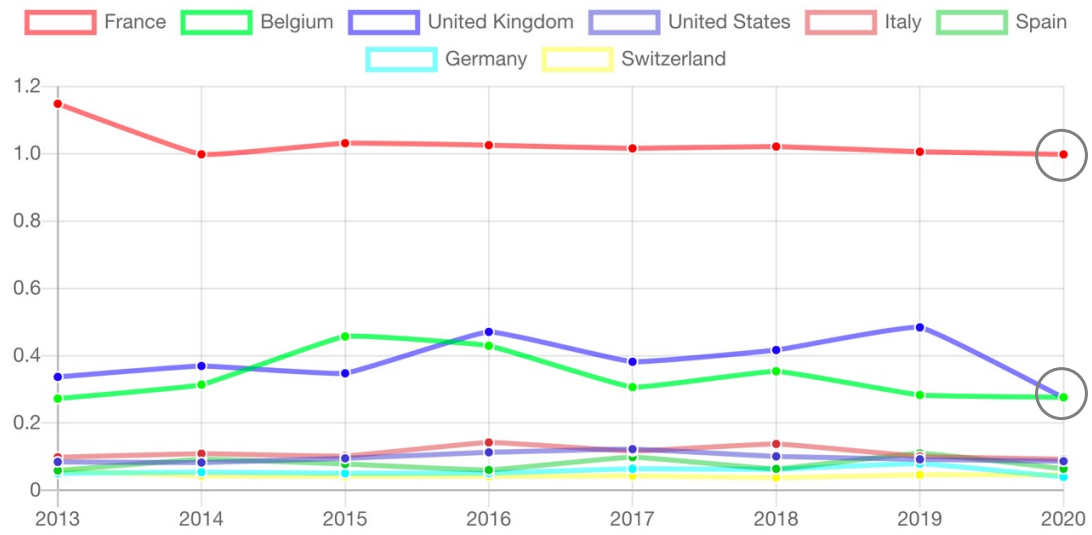
Experiments – GCF Evolution

Area: Region
Aggregation level: District
⇒ *GCF at large scales - global interest of an area*

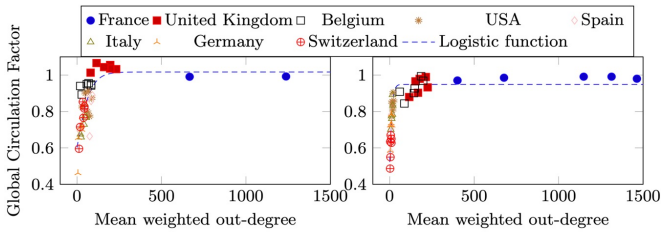




Pandemy Impact

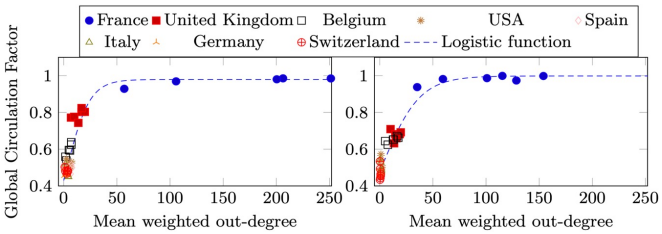


Experiments – GCF *vs* Degree Centrality



Nouvelle-Aquitaine Hauts-de-France

Logistic regression *f* of regions' GCF at *district* scale



Nouvelle-Aquitaine Hauts-de-France

Logistic regression *f* of regions' GCF at *city* scale

$$f(x) = \frac{L}{1+e^{-k(x-x_0)}}$$

Parameters	NA cities	HF cities	NA districts	HF district
L	9.791×10^{-1}	9.984×10^{-1}	1.017×10^{0}	9.488×10^{-1}
k	9.215×10^{-2}	5.611×10^{-1}	2.250×10^{-2}	1.062×10^{-1}
x_0	2.667×10^{0}	1.598×10^{0}	-1.653×10^{1}	-1.937×10^{0}
MSE	1.620×10^{-1}	7.229×10^{-2}	2.871×10^{-1}	1.231×10^{-1}
MAE	4.540×10^{-2}	3.165×10^{-2}	5.660×10^{-2}	4.366×10^{-2}
MAPE	8.204%	5.688%	7.163%	5.753%

Circulation Factor – PageRank vs Betweenness Centrality
With Flavien Galbez

- PageRank:
  - Mutual influence convergence between nodes
- Betweenness Centrality:
  - Nodes used by shortest paths have more weight
  - More flow oriented

Related Work

LEE S. et al. 2013, **Evaluating spatial centrality for integrated tourism management in rural areas using GIS and network analysis**, Tourism Management 34, 14-24, Seoul National University, Seoul, Republic of Korea

BHOGARAM P. et al. 2020, **Optimal and Critical Path Analysis of State Transportation Network Using Neo4J**, International Journal of Urban and Civil Engineering, Vol:14, No:10, 2020, World Academy of Science, Engineering and Technology.

SHIH H., 2005. **Network characteristics of drive tourism destinations: An application of network analysis in tourism**, Tourism Management 27, 1029–1039, Taiwan

KISS C., BICHLER M., 2008. **Identification of influencers — Measuring influence in customer networks,** Decision Support Systems 46, 233-253, Internet-based Information Systems, Department of Informatics, TU München, Germany

Betweenness Centrality in Neo4j – GDS

input: directed graph $G = (V, E)$
data: queue Q, stack S (both initially empty)
and for all $v \in V$:
$dist[v]$: distance from source
$Pred[v]$: list of predecessors on shortest paths from source
$\sigma[v]$: number of shortest paths from source to $v \in V$
$\delta[v]$: dependency of source on $v \in V$
output: betweenness $c_B[v]$ for all $v \in V$ (initialized to 0)

for $s \in V$ **do**
▼ **single-source shortest-paths problem**
▼ **initialization**
for $w \in V$ **do** $Pred[w] \leftarrow$ empty list
for $t \in V$ **do** $dist[t] \leftarrow \infty$; $\sigma[t] \leftarrow 0$
$dist[s] \leftarrow 0$; $\sigma[s] \leftarrow 1$
enqueue $s \rightarrow Q$
while Q *not empty* **do**
dequeue $v \leftarrow Q$; push $v \rightarrow S$
foreach *vertex* w *such that* $(v, w) \in E$ **do**
▼ **path discovery** // — w found for the first time?
if $dist[w] = \infty$ **then**
$dist[w] \leftarrow dist[v] + 1$
enqueue $w \rightarrow Q$
▼ **path counting** // — edge (v, w) on a shortest path?
if $dist[w] = dist[v] + 1$ **then**
$\sigma[w] \leftarrow \sigma[w] + \sigma[v]$
append $v \rightarrow Pred[w]$
▼ **accumulation** // — back-propagation of dependencies
for $v \in V$ **do** $\delta[v] \leftarrow 0$
while S *not empty* **do**
pop $w \leftarrow S$
for $v \in Pred[w]$ **do** $\delta[v] \leftarrow \delta[v] + \frac{\sigma[v]}{\sigma[w]} \cdot (1 + \delta[w])$
if $w \neq s$ **then** $c_B[w] \leftarrow c_B[w] + \delta[w]$

Unweighted

input: directed graph $G = (V, E)$ with edge lengths $\lambda : E \rightarrow \mathbb{R}_{>0}$
data: priority queue Q with keys $dist[\cdot]$
▼ **single-source shortest-paths problem**
▶ **initialization**
while Q *not empty* **do**
extract $v \leftarrow Q$ with minimum $dist[v]$; push $v \rightarrow S$
foreach *vertex* w *such that* $(v, w) \in E$ **do**
▼ **path discovery** // — shorter path to w?
if $dist[w] > dist[v] + \lambda(v, w)$ **then**
$dist[w] \leftarrow dist[v] + \lambda(v, w)$
insert/update $w \rightarrow Q$ with new key; $\sigma[w] \leftarrow 0$;
$Pred[w] \leftarrow$ empty list
▼ **path counting**
if $dist[w] = dist[v] + \lambda(v, w)$ **then**
$\sigma[w] \leftarrow \sigma[w] + \sigma[v]$
append $v \rightarrow Pred[w]$

Weighted

```
// BC forward traversal
while (!forwardNodes.isEmpty()) {
    long node = forwardNodes.remove();
    backwardNodes.push(node);
    int distanceNode = distance.get(node);

    localRelationshipIterator.forEachRelationship(node, (source, target) -> {
        if (distance.get(target) < 0) {
            forwardNodes.add(target);
            distance.set(target, distanceNode + 1);
        }

        if (distance.get(target) == distanceNode + 1) {
            sigma.addTo(target, sigma.get(source));
            append(target, source);
        }
        return true;
    });
}
```

Requires a *Pregel* implementation

Plan

- Tripdavisor dataset
- Related Work
  - Graphs & Circulation
- Circulation Graph
  - Graph Data Model
  - Graph Data Manipulations
  - Integration with Neo4j
- The Circulation Factor
  - TCF & GCF
  - Experiments
  - PageRank vs Betweenness Centrality
- Tourists Propagation
  - Maximum Spanning Trees
  - Experiments
- Conclusion & Perspectives

How to Analyze Tourists' Propagation?

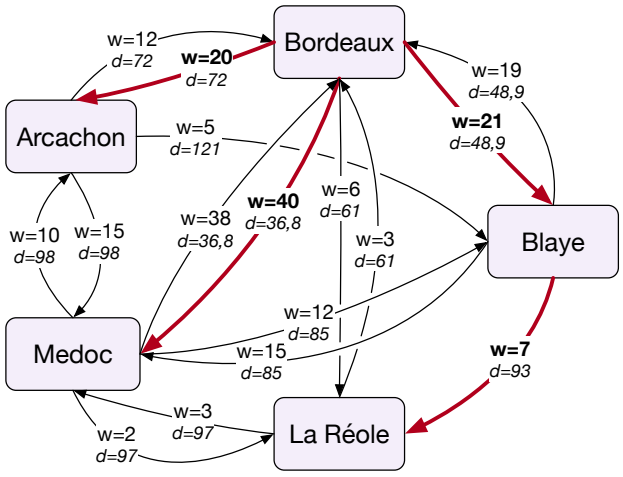
Is it possible to use the Circulation Graph to understand propagation?

- Is there any logic?

Must take into account both topology & weights

- But also, distance!

➢ Multi-weighted aggregated graphs



MST – How to compare both topologies and Remoteness?

- **Tree Edit Distance** - # of nodes' interchange
  - No viable comparison
- **Tree Hierarchy** - # of leafs vs links & Betweenness Centrality

$$T_h = \frac{L}{2 \times m \times BC_{max}}$$

=> Only dedicated to star vs lines topologies & no notions of distances.

Propagation Measure on Circulation Graphs for Tourism Behavior Analysis

With Sonia Djebali & Hugo Prevoteau (submitted at SAC'22 / SIGPLAN)

Definition : Remoteness Influence Factor (RIF)

*Consider a Multi-weighted graph AC(V, E(w, d)) the RIF measures the **remoteness** of vertices combined with their **influences** in AC. For each node n ∈ V, it computes its normalized distance from a source* **s**, *combined with the inverse of its centrality BC(n). It is defined as:*

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

i.e., Max WBC(AC)

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

Impact of nodes in AC

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**

Issue: complexity – $O(|V|^3)$

Propagation Measure on Circulation Graphs for Tourism Behavior Analysis

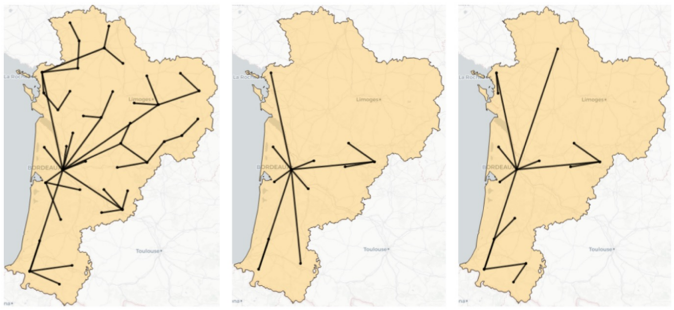
With Sonia Djebali & Hugo Prevoteau (submitted at SAC'22 / SIGPLAN)

Minimum/Maximum Spanning Tree (MST)

➡️ Reflects the traffic flow and hierarchy in the underlying system [Stam et al. 2014]

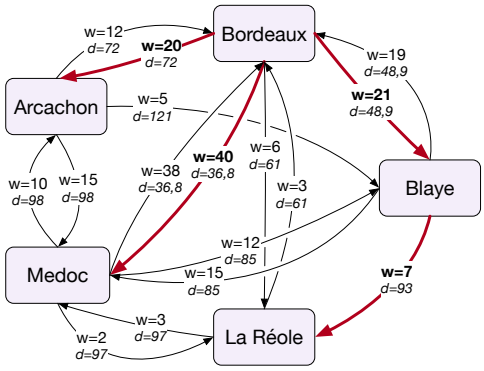
Apply RIF on the MST to reduce complexity

Complexity: $O(|V|^2)$



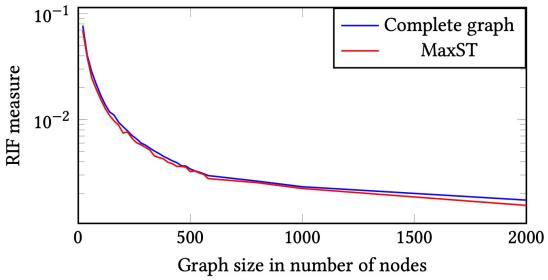
French, American and Spanish Maximum Spanning Trees in 2018
Nouvelle-Aquitaine (district scale)

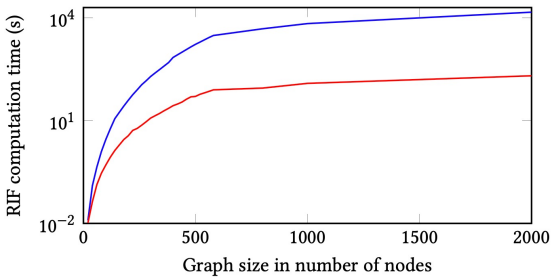
Remoteness Influence Factor - Example



$$RIF_{Fig1b} = \underbrace{\frac{log_{141.9}72}{4} \times \frac{1}{5}}_{Arcachon} + \underbrace{\frac{log_{141.9}36.8}{4} \times \frac{1}{5}}_{Médoc} + \underbrace{\frac{log_{141.9}48.4}{4} \times \frac{1}{5}}_{Blaye} + \underbrace{\frac{log_{141.9}141.9}{4} \times \frac{1}{5}}_{La\ Réole}$$

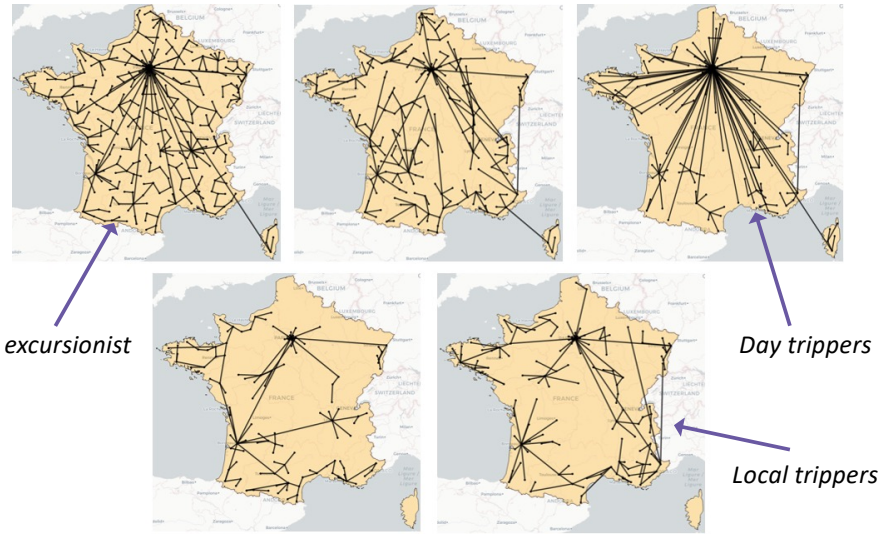
RIF approximation (Graph vs MST)

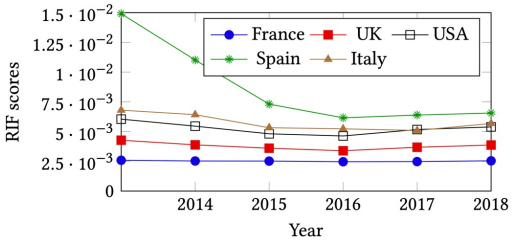




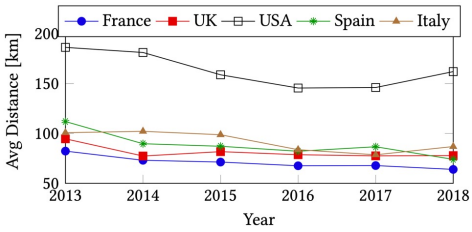
Graphs size	MSE	MAE	MAPE
200 Nodes	6.96×10^{-8}	2.39×10^{-4}	5.53%
500 Nodes	7.20×10^{-9}	7.92×10^{-5}	4.50%
1000 Nodes	5.89×10^{-10}	8.20×10^{-6}	3.92%
2000 Nodes	2.57×10^{-11}	3.32×10^{-6}	3.17%

French, English, American, Spanish and Italian MST in 2018 over France



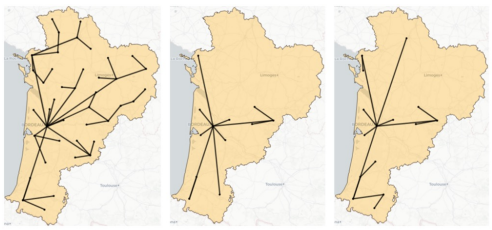


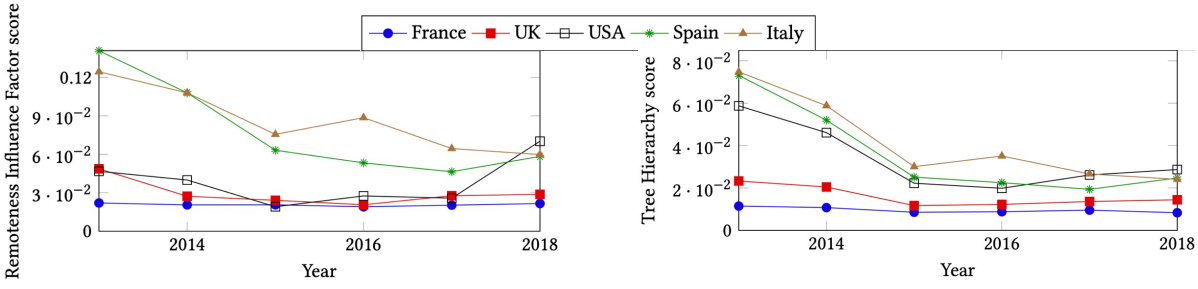
RIF Evolution - France (Town-scale)



Average geodesic distance (Town-scale)

Experiments - RIF vs Tree Hierarchy





RIF Evolution

Tree Hierarchy evolution

Nouvelle-Aquitaine (District-scale)

Conclusions & perspectives

Neo4Tourism: a methodology to produce and manipulate circulation graphs

- Digital traces from Social Networks
- Aggregation and filters on **circulation graphs** with Neo4j
- A **Circulation Factor** to make centralities comparable on space and time
- An automatic **Maximum Spanning Tree** extraction methodology dedicated to spatiotemporal graphs,
- The **Remoteness Influence Factor** (RIF), a new propagation measure

Perspectives

- Thorough comparison of TCFs/GCFs between PageRank *vs* Betweenness centralities (Pregel Implementation)
- Prediction model for each population (takes local & global tendencies = multiplex graphs)
- Pattern Mining of MSTs & scalability
- Clustering of nodes to analyze “touristic zones”

RGPD & Scrapping

- Tricky
  - DCP (*Données à caractères Personnel*) vs Research
- Points to keep in mind
  1. *Legal base*: research analysis context (research contract / public interest)
  2. *Clear finality*: data analysis, enhance knowledge (no decision making)
  3. *Anonymization*: at least pseudo-anonymization, but correlation
     1. When aggregate in the circulation graph = no DCP = no RGPD issue
  4. *Data conservation: "unlimited" for research purpose (patrimoine)*
  5. *Criticality/Impact*: if data is stolen = no more information than available online ➡️ low criticality