INSTITUT VEDECOM
DU VÉHICULE DÉCARBONÉ ET COMMUNICANT ET DE SA MOBILITÉ

UNIVERSITÉ DE VERSAILLES ST-QUENTIN-EN-YVELINES

université PARIS-SACLAY

david
données et algorithmes
pour une ville intelligente et durable

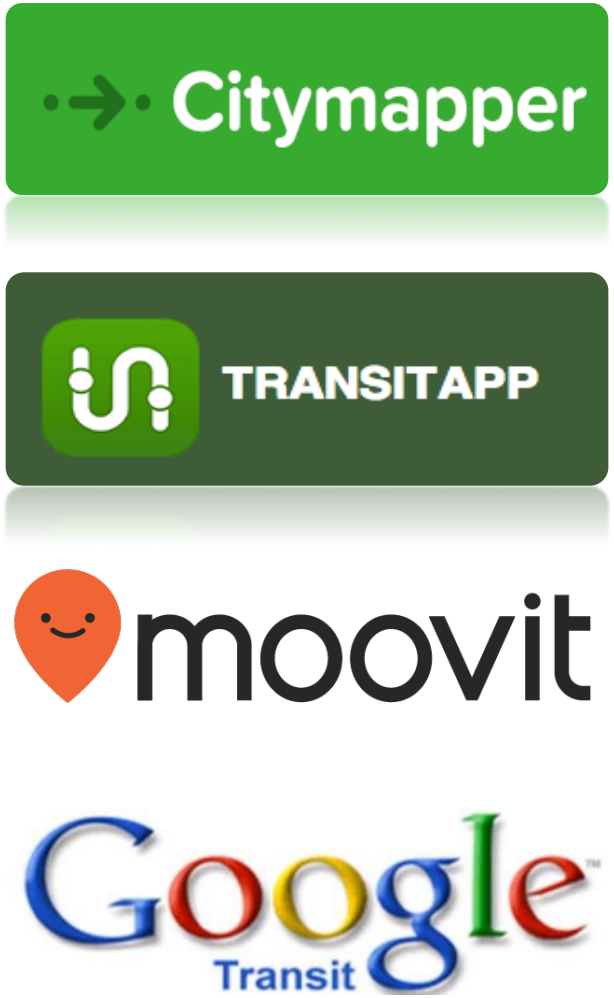
Link++: Adaptive Linking of Multiple Transportation Networks

Authors:

Ali MASRI, Karine ZEITOUNI & Zoubida KEDAD

Introduction

- Transportation data integration → improves transportation services e.g. trip planning
- Mainly focused on public transportation services
- Individual and collective transport services e.g. dynamic car pooling are not taken into account or not fully integrated
- Integrating such services will boost the performance of existing solutions



The Integration Problem

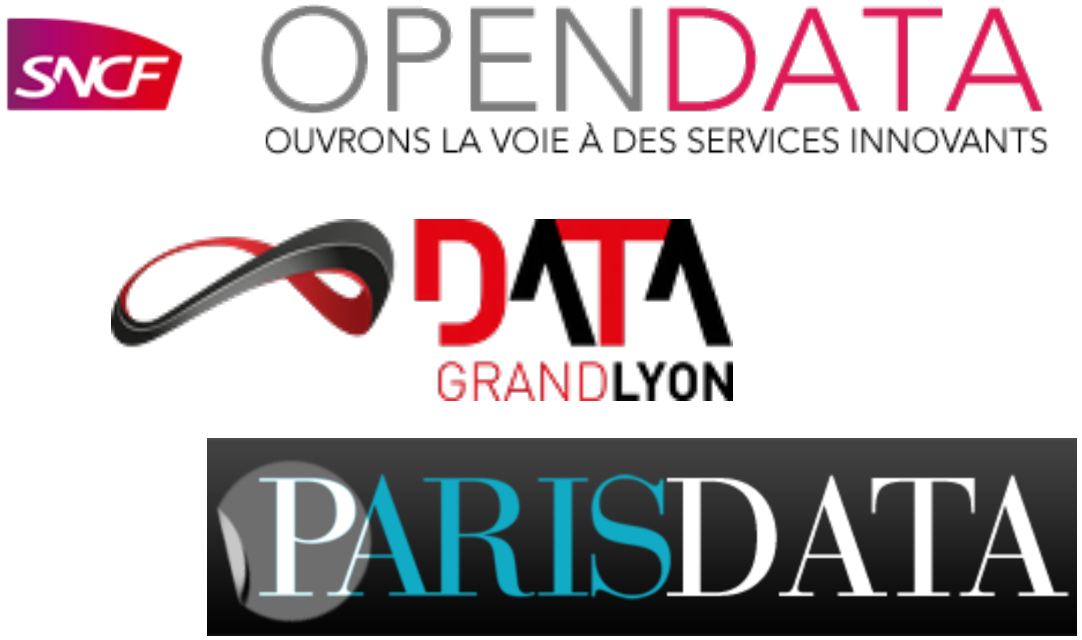
- Heterogeneous data representations
  - Many formats and representation models
- Spatiotemporal feature
- Dynamicity of the network
  - e.g. Stops may appear/disappear in real-time
- Dynamicity on the network
  - unexpected events such as alerts

Problem

How to achieve the integration of the transportation data sources?

Open Transportation Data

- Companies are moving towards publishing their data for public
- This allows
  - Better market visibility
  - Support for better transportation applications
- Multiple formats
  - GTFS
  - CSV
  - Shape
- Connecting these data together allows easier access and richer information **– Data Interlinking**



66 jeux de données

Trier par Dernière modification

Filtres

Trouver un jeu de données...

Vue

Analyse 50

Carte 9

Modifié

2014 25

2015 33

2016 8

Producteur

Transilien 14

ésion & ressources 11

Régularité mensuelle TER

Régularité mensuelle TER depuis janvier 2013.

Producteur TER
Licence SNCF Open Data
Données 720

Régularité TER Train

Tableau
Analyse
Export

Lettres de suite des audits de sécurité des établissements ferroviaires

Lettres de suite des audits d'établissements ferroviaires impliquant la Sécurité de l'Exploitation Ferroviaire.

Producteur Direction Sécurité & Qualité de Service Ferroviaire
Licence SNCF Open Data
Données 121

Sécurité Audits Etablissements ferroviaires

Tableau
Analyse
Export

Horaires des lignes TER

Horaires des lignes TER au format GTFS.

Producteur TER
Licence SNCF Open Data
Données 1

Horaire Train Gare de voyageurs GTFS France

Tableau
Export

Horaires des lignes Intercités

Horaires des lignes Intercités au format GTFS.

Producteur Intercités
Licence SNCF Open Data
Données 1

Horaire Train Gare de voyageurs GTFS France

Tableau
Export

Horaires des Tram-Train TER Pays

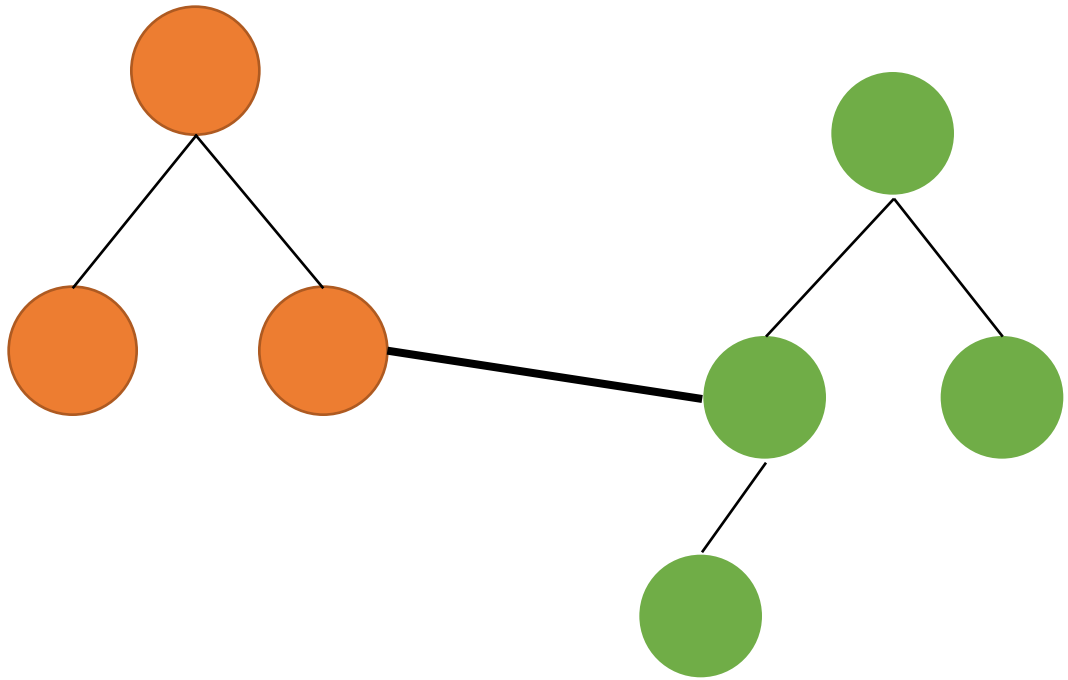
Tableau

Incidents de sécurité

Tableau

What is data interlinking?

- Finding relations between entities in different datasets
- Discovery based on a linkage rule defined by
  - Preprocessing functions
  - Distance functions
  - Threshold
- If a rule is valid then a link is created



Related Work

- MELINDA [Scharffe et al. 2011]
  - A framework for integrating open data
- DataLift [Scharffe et al. 2012]
  - A platform for transforming, interlinking and publishing open data
- GeoKnow [Athanasiou et al.]
  - Same objective as DataLift
  - More specific to geospatial data
  - Column indexing, faster interlinking and querying
- LinkedGeoData [Auer et al.]
  - An approach to integrate and publish OSM data as open data
- [Silk] and [LIMES]
  - Interlinking tools

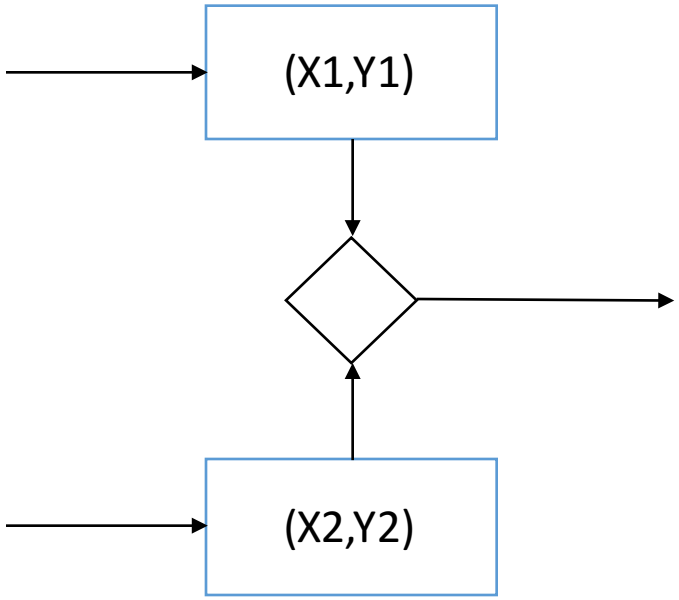
Example – Transportation Data

Linking Rule

- Get the coordinate of both locations
- Calculate the geometric distance
- Link if distance < threshold

Bike Sharing Station

Bus Stop



<BikeStation1><owl:sameAs><BusStop1>

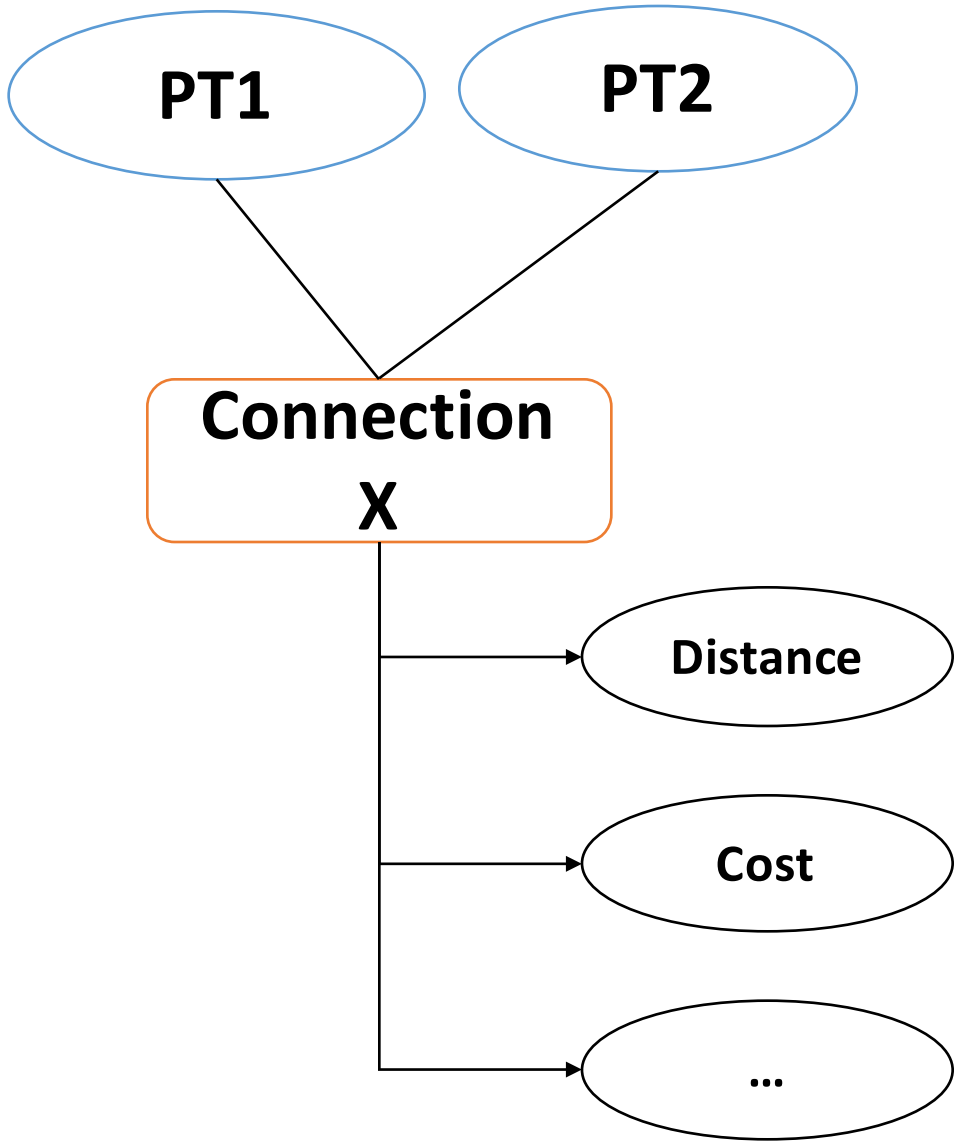
Is this enough?

Problems

- Limited distance functions
- Insufficient information level (output)
- Static links

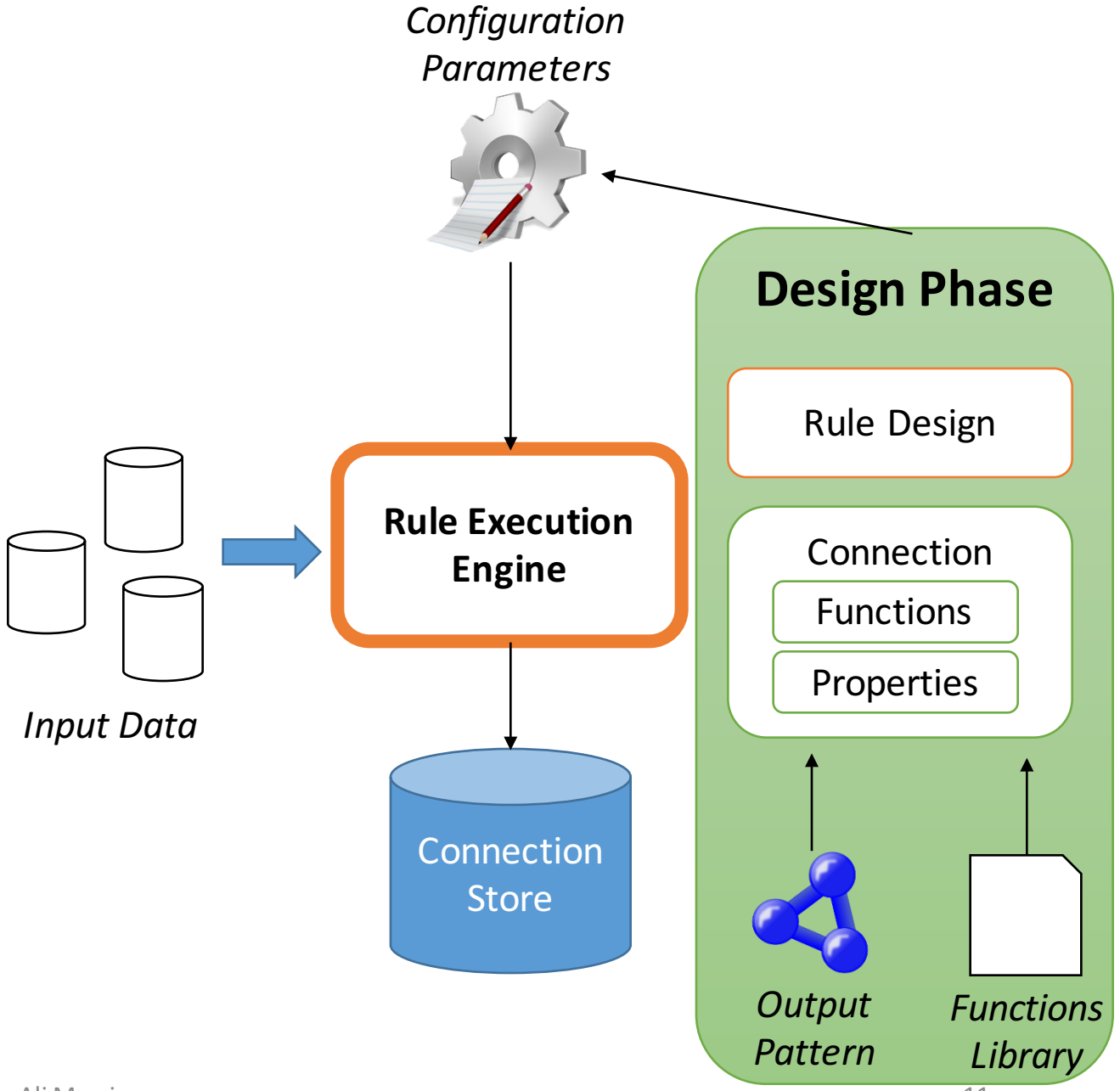
Introducing New Transportation Connections

- Existing approaches discover sameAs links
- Dedicated to equivalence matching
- We want a link that is able to
  - Define a connection between two transportation points of transfer
  - Affected by the status of transportation system
    - Time schedule
    - Operating days
    - Real-time status
  - Capable of expressing information



Link++

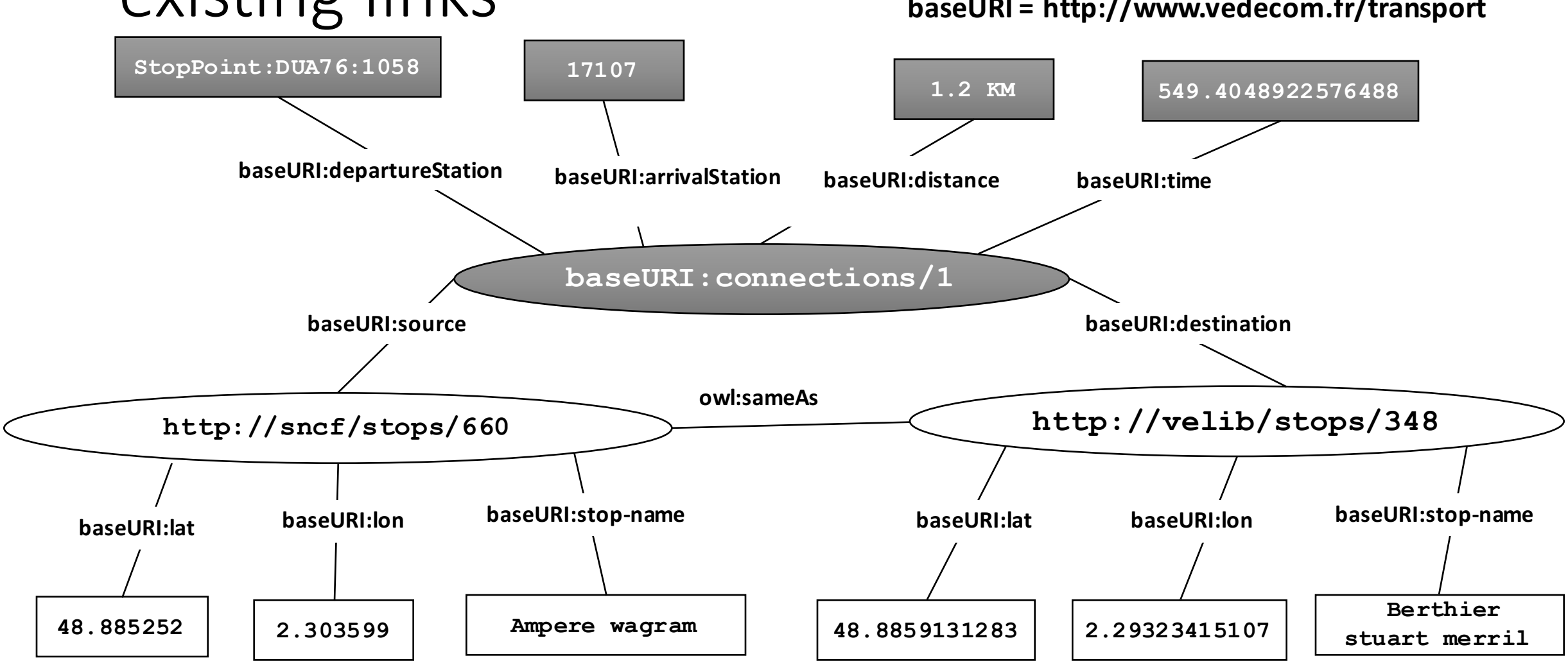
- Consists of two phases
  - Design phase
  - Execution phase
- Enables users to define
  - Custom functions
  - External libraries
  - Custom linking rules
  - Custom output
- Can be customized to fulfill any interlinking problem



Interlinking Tools - Outputs

	Techniques	Output	Domain
RKB-CRS	String	owl:sameAs	Publications
GNAT	String, Similarity-propagation	owl:sameAs	Music
ODD-Linker	String	linkset	Independent
RDF-AI	String, WordNet	alignment format	Independent
Silk	String, Numerical, Date	owl:sameAs, user-specified	Independent
LIMES	String, Geographical, Numerical, Date	owl:sameAs, user-specified	Independent
Link++	User defined	User defined	Independent

Comparison between our connections and existing links



Evaluation

Railway services
http://www.sncf.com/
1067 train stations

Find connections between two different modes of transportation to enable Multimodality

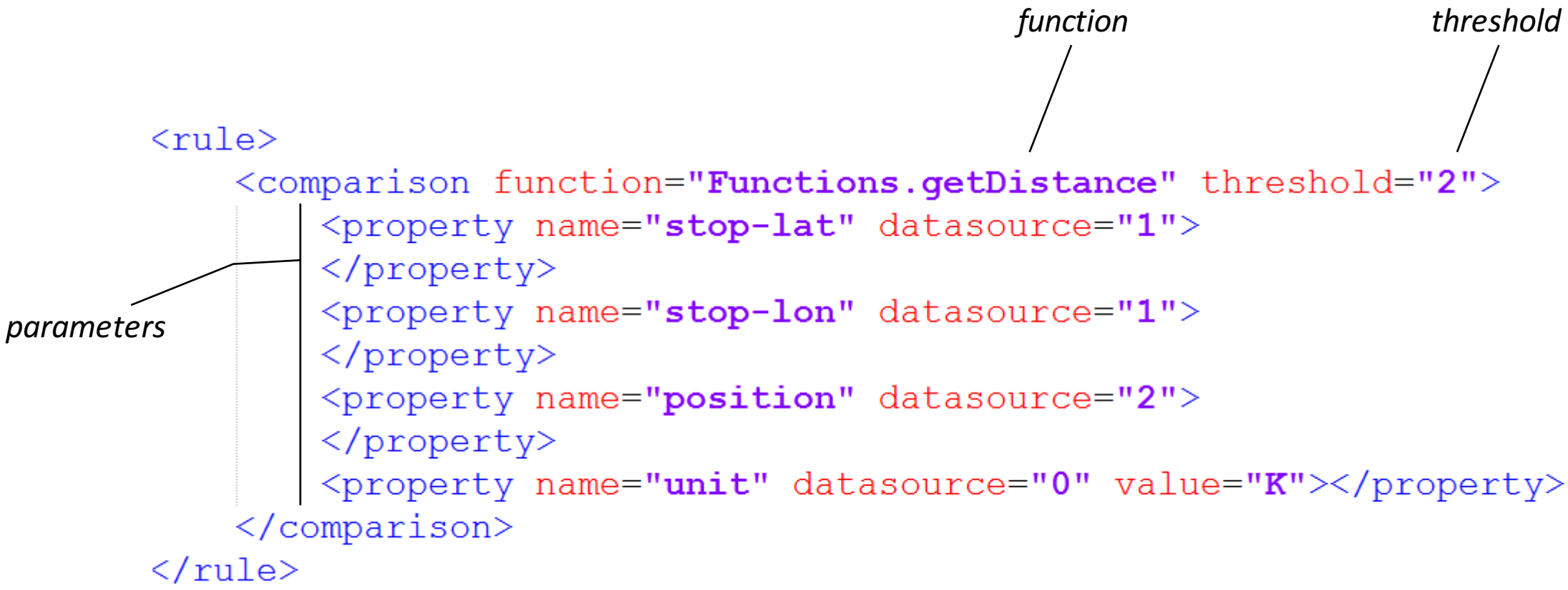
Bike sharing services
http://www.velib.fr
1242 bike stations

VELIB dataset http://opendata.paris.fr/explore/dataset/stations-velib-disponibilites-en-temps-reel
SNCF dataset http://gtfs.s3.amazonaws.com/transilien-archiver_20160202_0115.zip

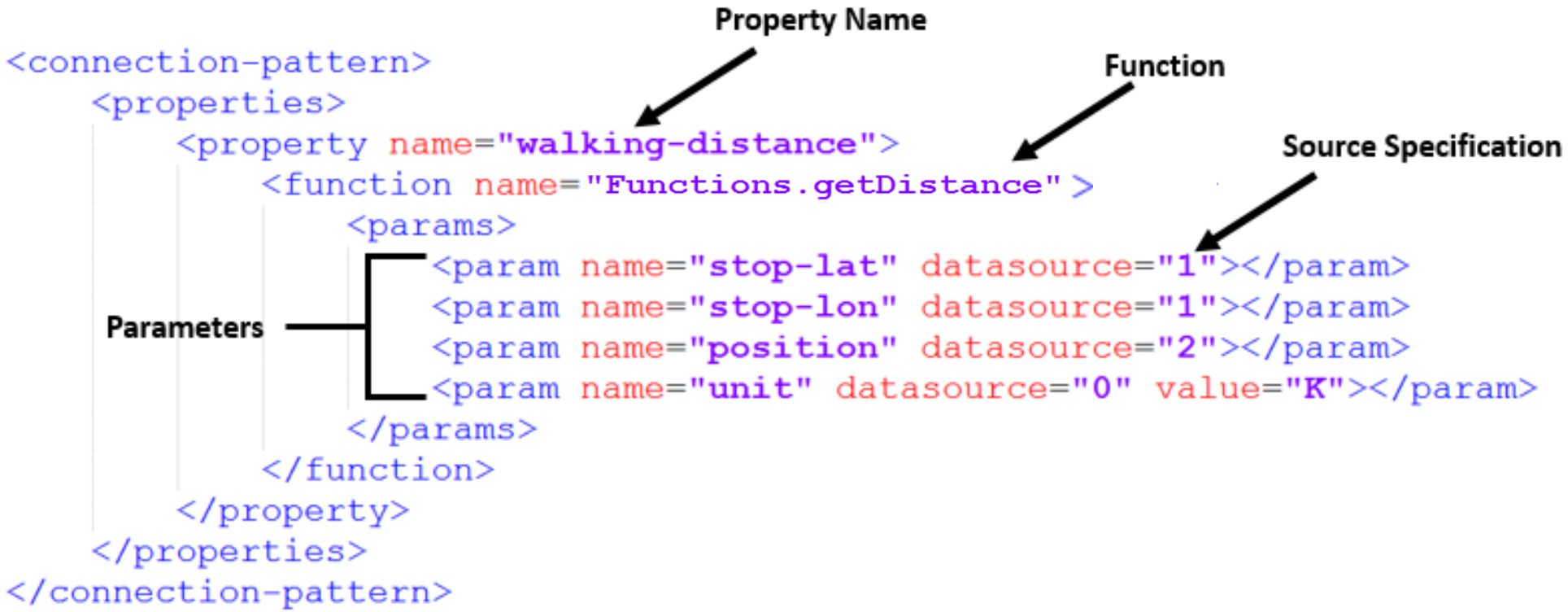
Process

1. Transform data into RDF
   - DataLift framework [Scharffe et al. 2012]
2. Define custom functions
   - Google distance API
   - More precise results
3. Define linking rule
   - Bike to bike station: If a cycling path exists within three kilometers, create a connection
   - Bike to train station: If a walking path exists from one stop to another within one kilometer, create a connection
4. Define output pattern

Linking Rule



Output Pattern



Connection Generation

3643 new connections between SNCF and VELIB

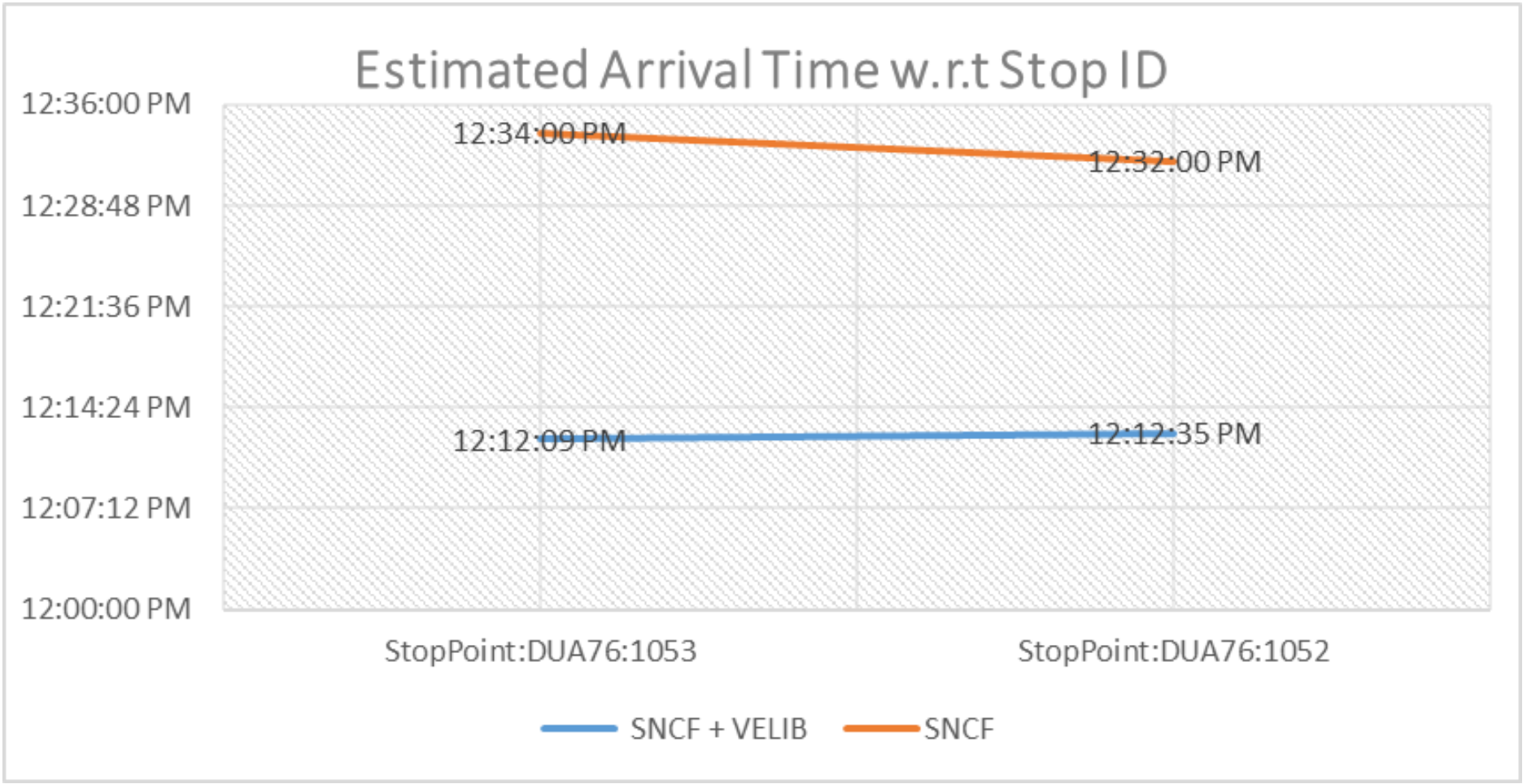
319594 internal connections between VELIB bike stations

SNCF vs SNCF+VELIB

Minimum Estimated Arrival Time Problem

Departure Station “StopPoint-DUA76-356”

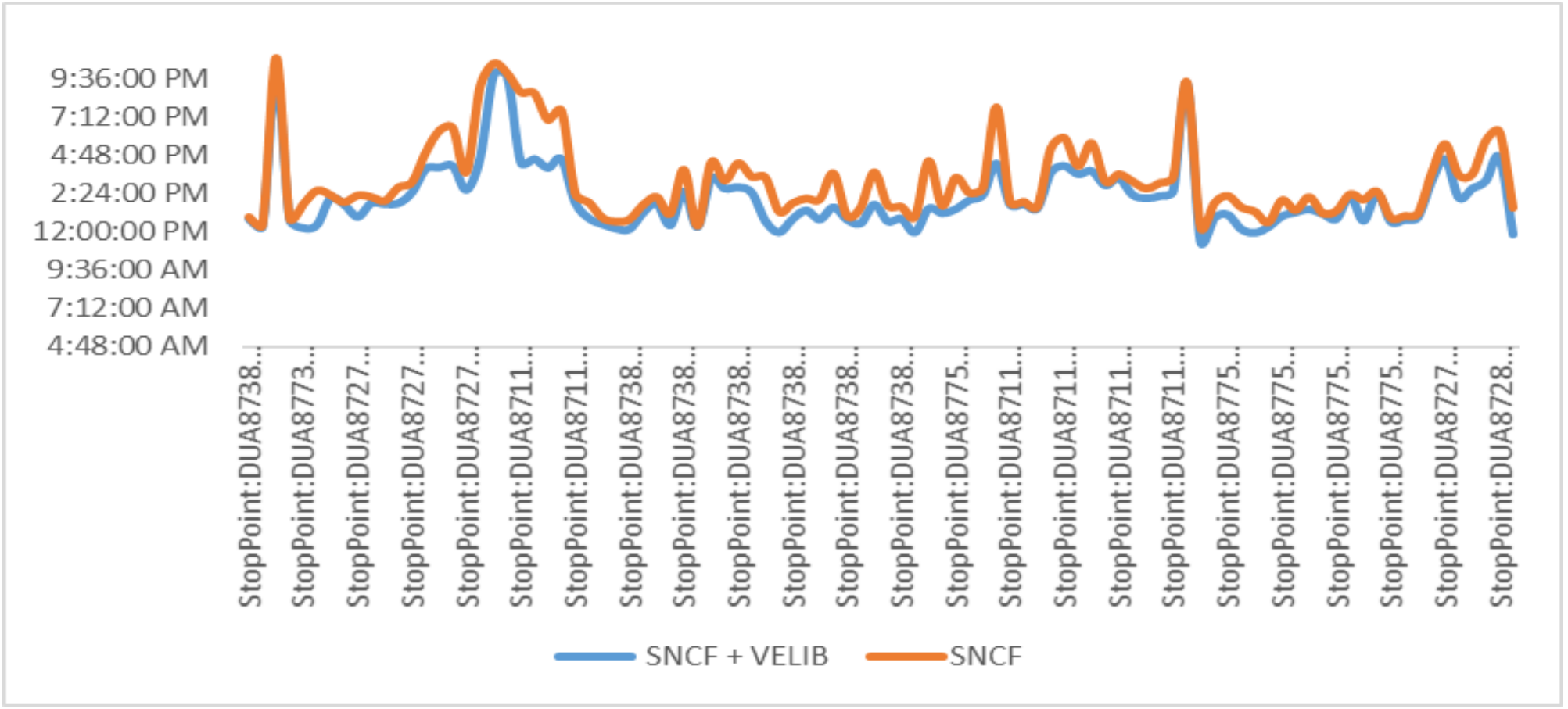
Departure Time 14:00



SNCF vs SNCF+VELIB

Departure Station "StopPoint-DUA8739300"

Departure Time 14:00



Link++ [SNCF-VELIB-GEOMETRIC]

File About

Toolbar

Load data sources | Add functions file | Add library | Compile functions | Load linking rule | Load output pattern | Generate connections | Stop | Save | Save All

Project Structure

- sncf-velib-geometric
  - output
    - output 2016-09-30 11-17-57.ttl
  - functions
    - Functions.java
  - datasources
    - ds1-sncf-stops.ttl
    - ds2-velib.ttl
  - classes
    - Functions.class
  - linking-rule.xml
  - output-pattern.xml
  - project.lpp

output-pattern.xml | output 2016-09-30 11-17-57.ttl | Functions.java | ds1-sncf-stops.ttl | ds2-velib.ttl | linking-rule.xml | project.lpp

```
1  <http://www.vedecom.fr/connection1> a <http://www.vedecom.fr/connection>;
2       <http://www.vedecom.fr/source> <http://localhost:9091/linking-cities/sncf-a-csv/15>;
3       <http://www.vedecom.fr/target> <http://velib/stops/433>;
4       <source-id> "StopArea:DUA8775859";
5       <target-id> "07016 - TOUR MAUBOURG UNIVERSITE";
6       <distance> "1.9405484411018414";
7       <time> "1400.2889560706597".
8  <http://www.vedecom.fr/connection2> a <http://www.vedecom.fr/connection>;
9       <http://www.vedecom.fr/source> <http://localhost:9091/linking-cities/sncf-a-csv/15>;
10      <http://www.vedecom.fr/target> <http://velib/stops/1221>;
11      <source-id> "StopArea:DUA8775859";
12      <target-id> "08032 - MATIGNON";
13      <distance> "1.192232231021673";
14      <time> "860.3081432088387".
15 <http://www.vedecom.fr/connection3> a <http://www.vedecom.fr/connection>;
16      <http://www.vedecom.fr/source> <http://localhost:9091/linking-cities/sncf-a-csv/15>;
17      <http://www.vedecom.fr/target> <http://velib/stops/1027>;
18      <source-id> "StopArea:DUA8775859";
19      <target-id> "17013 - LEGENDRE";
20      <distance> "1.7802715899663113";
```

Log

```
[9/30/16 11:17 AM]: Functions compiled successfully - 0:00:00.650
[9/30/16 11:17 AM]: Parsing linking rule...
[9/30/16 11:17 AM]: Linking rule parsed successfully
[9/30/16 11:17 AM]: Parsing output pattern
[9/30/16 11:17 AM]: Output pattern parsed successfully
[9/30/16 11:17 AM]: Generating connections...
[9/30/16 11:18 AM]: Process terminated - 0:00:17.578
```

Project Information

Source datasource: ds1-sncf-stops.ttl

Target datasource: ds2-velib.ttl

More information

Discussion

- Transportation data integration is important to enable Multimodality
- Existing approaches focus on interlinking datasets based on equivalence matching
- Generated links lack the properties relevant to a transportation link
- Our approach enables interlinking beyond equivalence matching
  - Custom functions, rules and output patterns
  - Provide a rich semantic connection between entities

References

[Scharffe et al. 2011] Scharffe, François, and Jérôme Euzenat. **"MeLinDa: an interlinking framework for the web of data."** arXiv preprint arXiv:1107.4502 (2011).

[Scharffe et al. 2012] Scharffe, François, et al. **"Enabling linked data publication with the Datalift platform."** Proc. AAAI workshop on semantic cities. 2012.

[Athanasiou et al.] Athanasiou, Spiros, et al. **"GeoKnow: Making the web an exploratory place for geospatial knowledge."** ERCIM News 96 (2014): 12-13.

[Auer et al.] Auer, Sören, Jens Lehmann, and Sebastian Hellmann. **Linkedgeodata: Adding a spatial dimension to the web of data**. Springer Berlin Heidelberg, 2009.

[Hong-Hai et al.] Do, Hong-Hai, and Erhard Rahm. **"COMA: a system for flexible combination of schema matching approaches."** Proceedings of the 28th international conference on Very Large Data Bases. VLDB Endowment, 2002.

[Brauner et al.] Brauner, Intrator, Freitas - 2007 - **An instance-based approach for matching export schemas of geographical database Web services**.

[LIMES] Ngomo, Axel-Cyrille Ngonga, and Sören Auer. **"Limes-a time-efficient approach for large-scale link discovery on the web of data."** integration 15 (2011): 3.

References

[SILK] Volz, Julius, et al**. "Silk-A Link Discovery Framework for the Web of Data."** LDOW 538 (2009).

[Rahm et al.] Rahm, Erhard, and Philip A. Bernstein**. "A survey of approaches to automatic schema matching."** the VLDB Journal 10.4 (2001): 334-350.

[Volz2004] Volz, Steffen, and Volker Walter. **"Linking different geospatial databases by explicit relations."** Proceedings of the XXth ISPRS Congress, Comm. IV, Istanbul, Turkey. 2004.

[Vaccari2009] Vaccari, Lorenzino, Pavel Shvaiko, and Maurizio Marchese. **"A geo-service semantic integration in spatial data infrastructures."** International Journal of Spatial Data Infrastructures Research 4 (2009): 24-51.

[Feliachia] Feliachia, Abdelfettah, Nathalie Abadieb, and Fayçal Hamdic. **"Matching and Visualizing Thematic Linked Data: An Approach Based on Geographic Reference Data.“** semantic-web-journal.net

[Jain2010] Jain, Prateek, et al. **"Ontology alignment for linked open data."** International Semantic Web Conference. Springer Berlin Heidelberg, 2010.


תודה
Dankie
Gracias
Спасибо
شكراً
Merci
Takk
Köszönjük
Terima kasih
Grazie
Dziękujemy
Dėkojame
Ďakujeme
Vielen Dank
Paldies
Täname teid
Kiitos
谢谢
Thank You
Tak
感謝您
Obrigado
Teşekkür Ederiz
감사합니다
Σας ευχαριστούμε
ขอบคุณ
Bedankt
Děkujeme vám
ありがとうございます
Tack