# Publishing planned, live and historical public transport data on the Web with the Linked Connections Framework

**Julián Rojas**, Harm Delva, Pieter Colpaert, Ruben Verborgh

Ghent University – imec

**3nd International Workshop on Semantics and the Web for Transport**

September 2021

GHENT UNIVERSITY

imec

# How is public transport data published today?

**Data dump**
**(GTFS, NeTEx, GTFS-RT, SIRI, etc)**

**(Route planning) API on agency's server**

A data reuser can:
(i) download a data dump or;
(ii) ask a question to the agency's server (via an API)

# How is public transport data published today?

low cost publishing
high flexibility for reusers

Outdated from creation
high integration costs

**Data dump**
**(GTFS, NeTEx, GTFS-RT, SIRI, etc)**

low cost access
up to date data

high scalability costs
limited flexibility for reusers

**(Route planning) API on agency's server**

A data reuser can:
(i) download a data dump or;
(ii) ask a question to the agency's server (via an API)

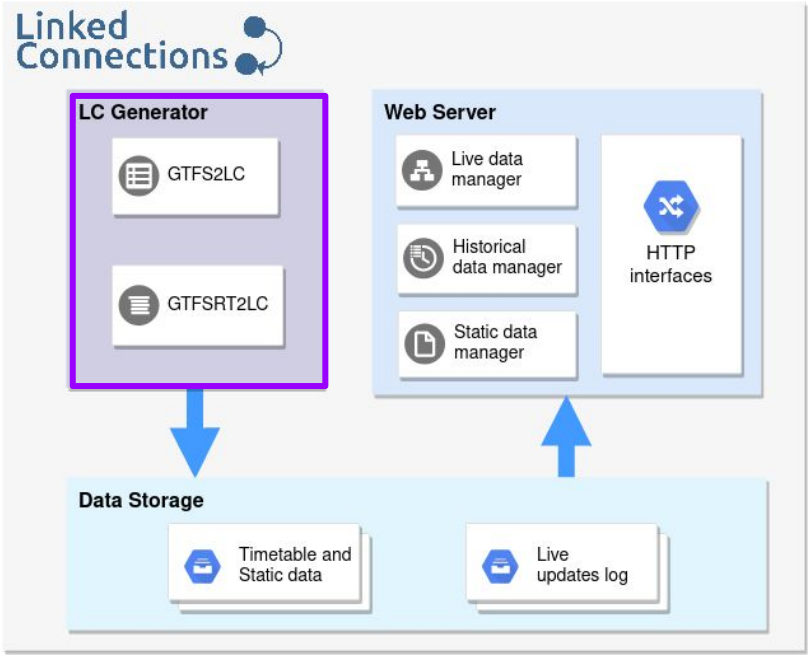
# Linked Connections as an *In-between* alternative that takes the best of both worlds

low cost publishing
high flexibility for reusers

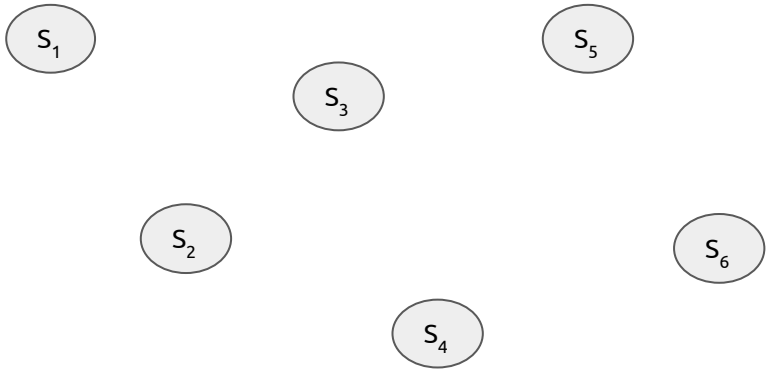
low cost access
up to date data


Serve data fragments via low-cost interfaces that achieve a **reasonable trade-off** in terms of **query processing** and **data integration** efforts for publishers and reusers

# How does Linked Connections work?

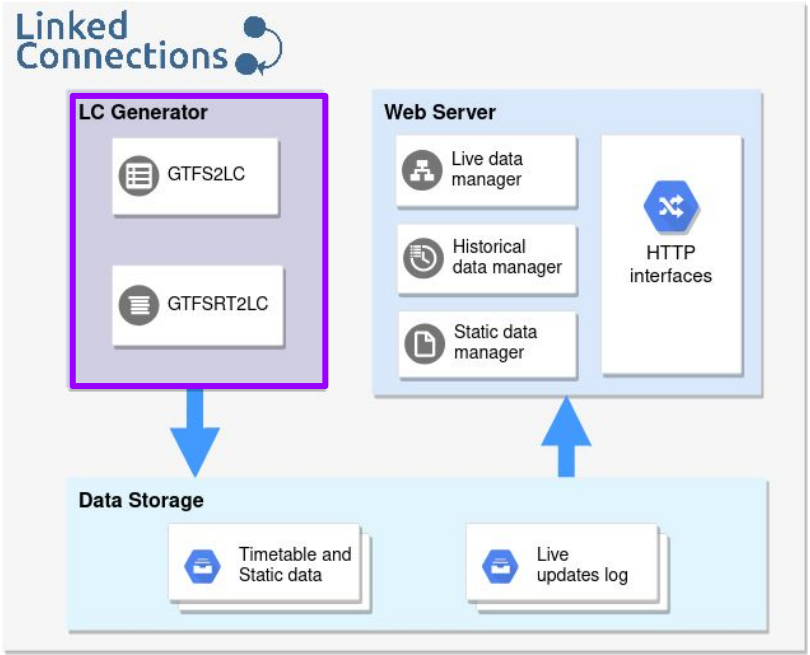



Set of **`gtfs:Stops`***: $S_1$… $S_6$



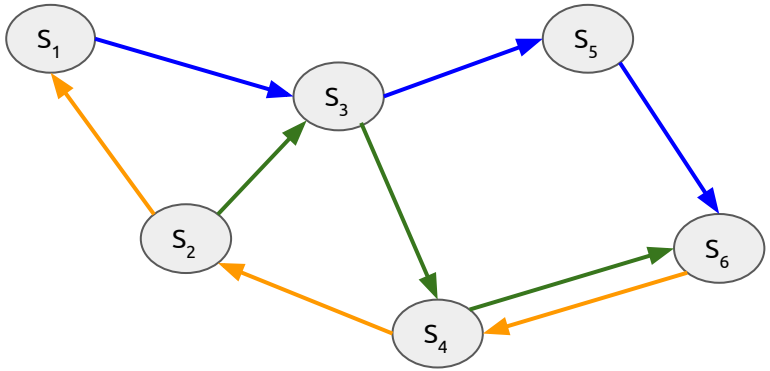

* @prefix gtfs: <http://vocab.gtfs.org/terms#>

# How does Linked Connections work?



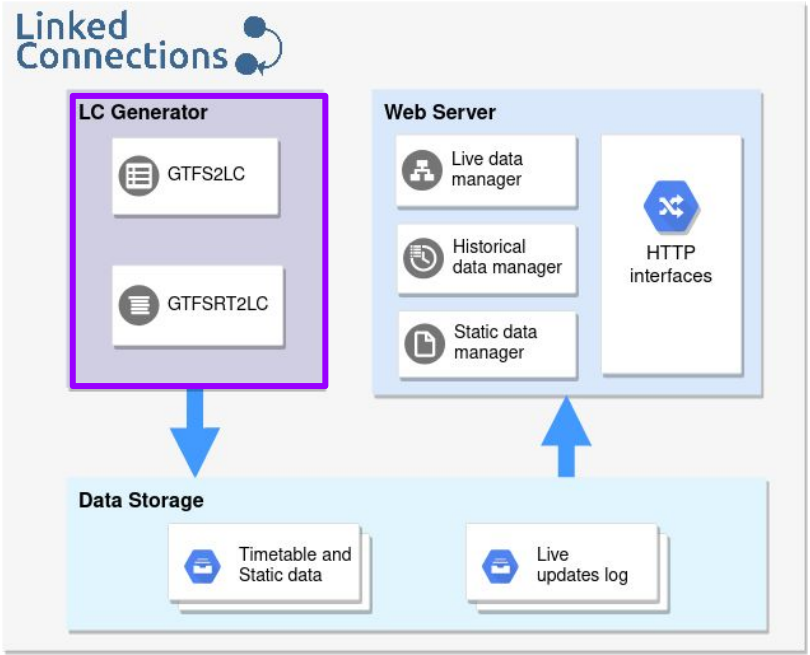

Set of **gtfs:Stops***: $S_1$... $S_6$

Set of **gtfs:Routes***: $R_1$, $R_2$, $R_3$




* @prefix gtfs: <http://vocab.gtfs.org/terms#>

# How does Linked Connections work?

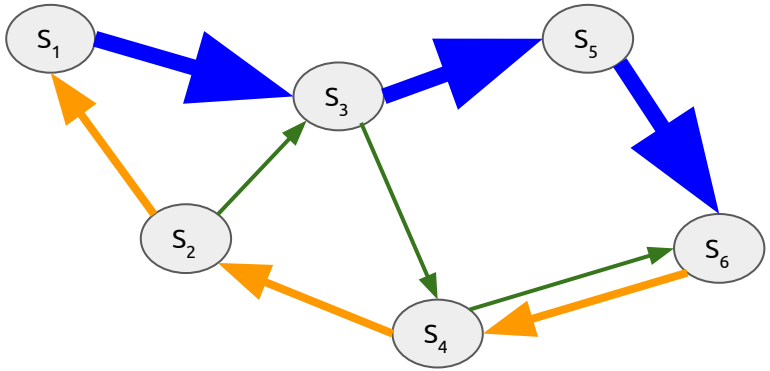



Set of **gtfs:Stops***: $S_1$... $S_6$

Set of **gtfs:Routes***: $R_1$, $R_2$, $R_3$

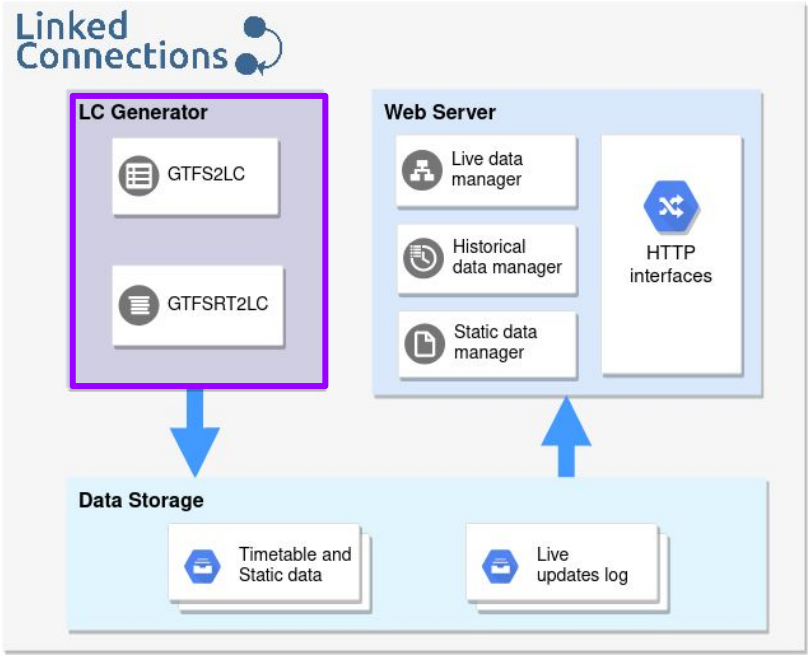
Set of **gtfs:Trips***:

- $T_1$, $T_2$, $T_3$, $T_4$, $T_5$
- $T_6$, $T_7$
- $T_8$, $T_9$, $T_{10}$




* @prefix gtfs: <http://vocab.gtfs.org/terms#>

# How does Linked Connections work?

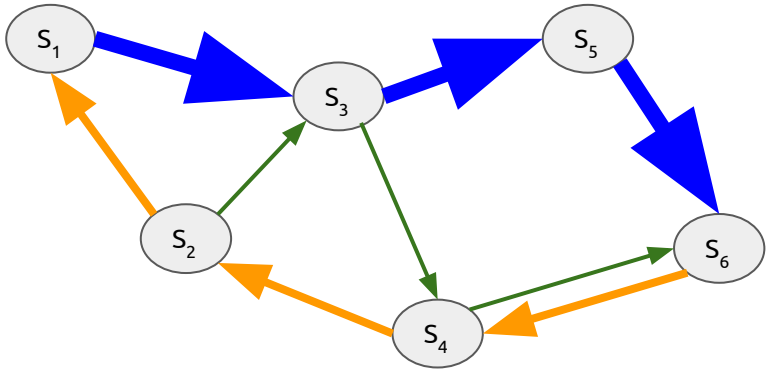



Set of **gtfs:Stops***: $S_1$... $S_6$

Set of **gtfs:Routes***: $R_1$, $R_2$, $R_3$

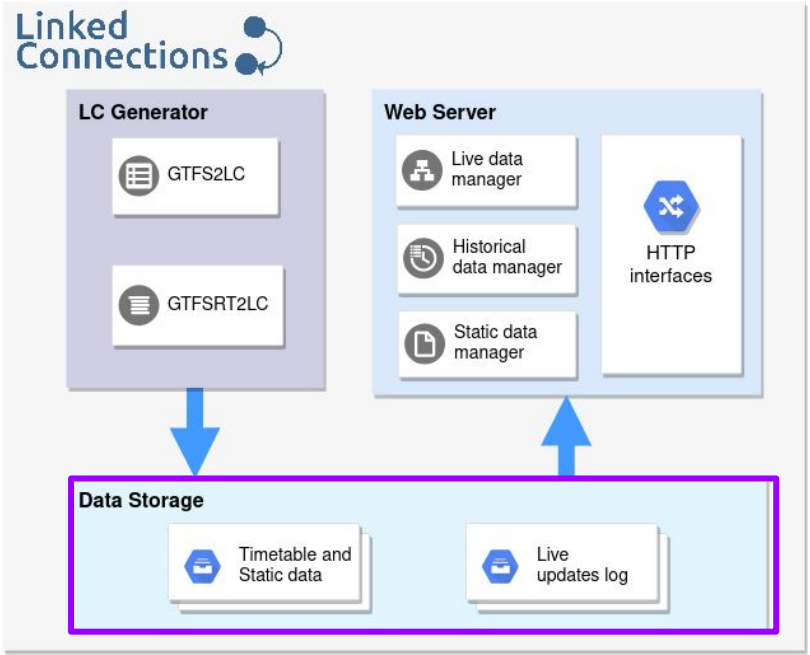
Set of **gtfs:Trips***:
- $T_1$, $T_2$, $T_3$, $T_4$, $T_5$
- $T_6$, $T_7$
- $T_8$, $T_9$, $T_{10}$

Set of **lc:Connections****:
- $T_1$: $C_{S1-S3}$, $C_{S3-S5}$, $C_{S5-S6}$ (departing at 07:00)
- $T_6$: $C_{S2-S3}$, $C_{S3-S4}$, $C_{S4-S6}$ (departing at 07:10)
- $T_8$: $C_{S6-S4}$, $C_{S4-S2}$, $C_{S2-S1}$ (departing at 07:30)



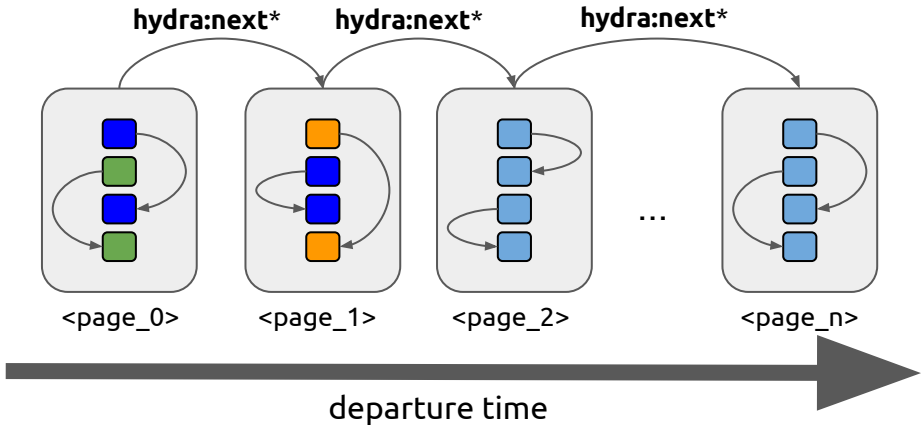

* @prefix gtfs: <http://vocab.gtfs.org/terms#>
** @prefix lc: <http://semweb.mmlab.be/ns/linkedconnections#>

# How does Linked Connections work?



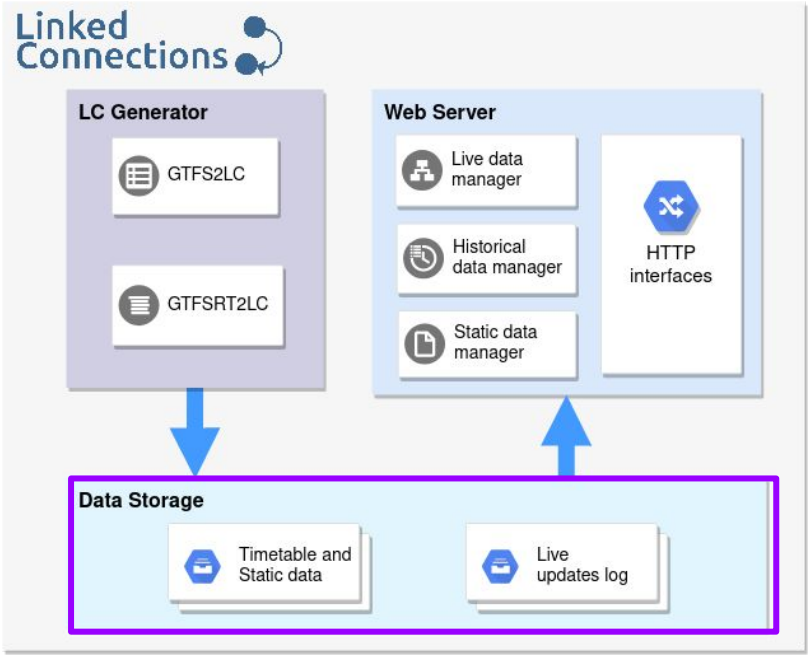

Set of **`lc:Connections`****:

- $T_1$: $C_{S1\text{-}S3}$, $C_{S3\text{-}S5}$, $C_{S5\text{-}S6}$ (departing at 07:00)
- $T_6$: $C_{S2\text{-}S3}$, $C_{S3\text{-}S4}$, $C_{S4\text{-}S6}$ (departing at 07:10)
- $T_8$: $C_{S6\text{-}S4}$, $C_{S4\text{-}S2}$, $C_{S2\text{-}S1}$ (departing at 07:30)



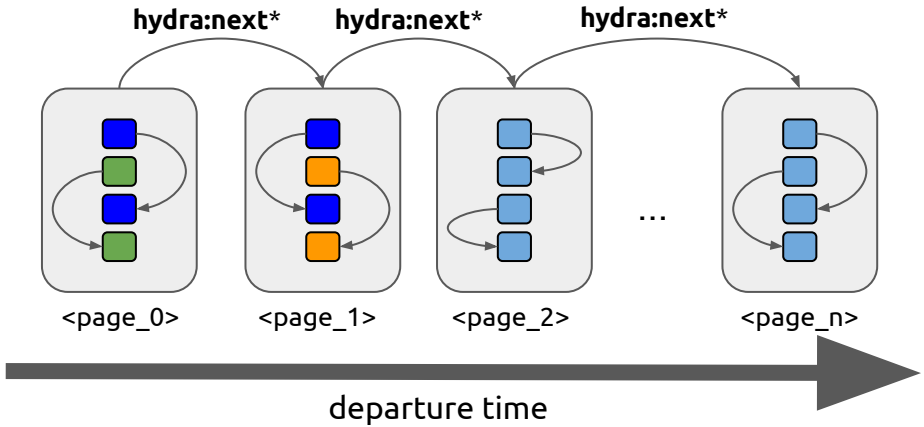

* @prefix lc: <http://semweb.mmlab.be/ns/linkedconnections#>
** @prefix hydra: <http://www.w3.org/ns/hydra/core#>

# How does Linked Connections work?



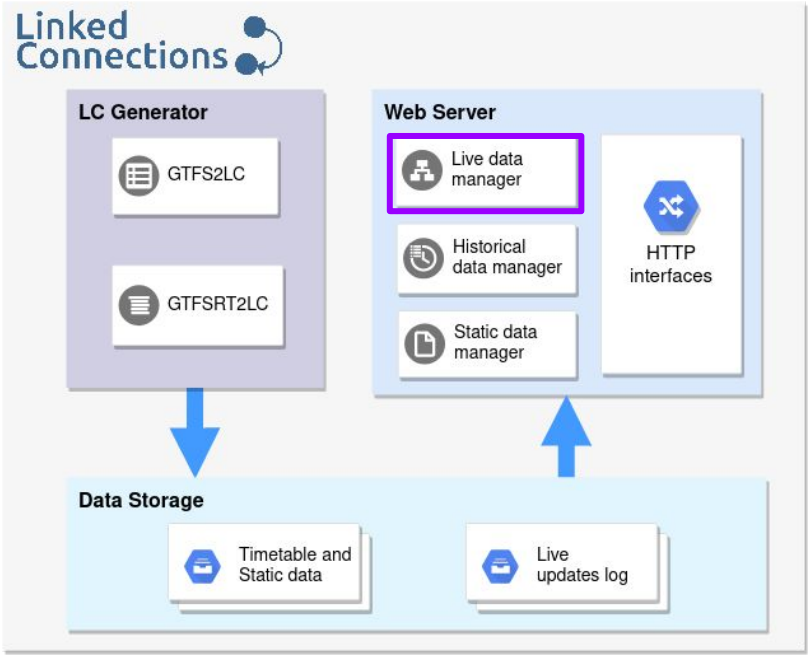

Set of **`lc:Connections`****:

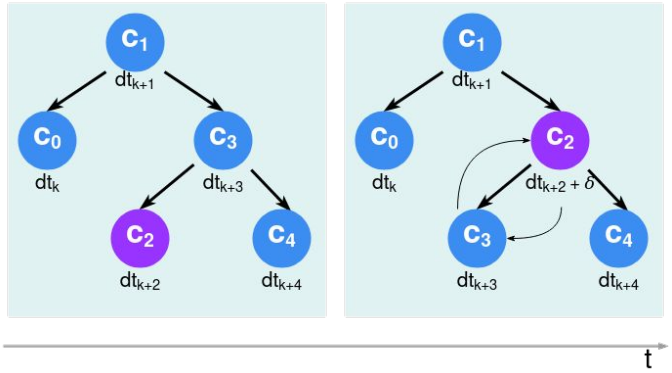
- $T_1$: $C_{S1\text{-}S3}$, $C_{S3\text{-}S5}$, $C_{S5\text{-}S6}$ (departing at 07:00)
- $T_6$: $C_{S2\text{-}S3}$, $C_{S3\text{-}S4}$, $C_{S4\text{-}S6}$ (departing at 07:10)
- $T_8$: $C_{S6\text{-}S4}$, $C_{S4\text{-}S2}$, $C_{S2\text{-}S1}$ (departing at 07:30)

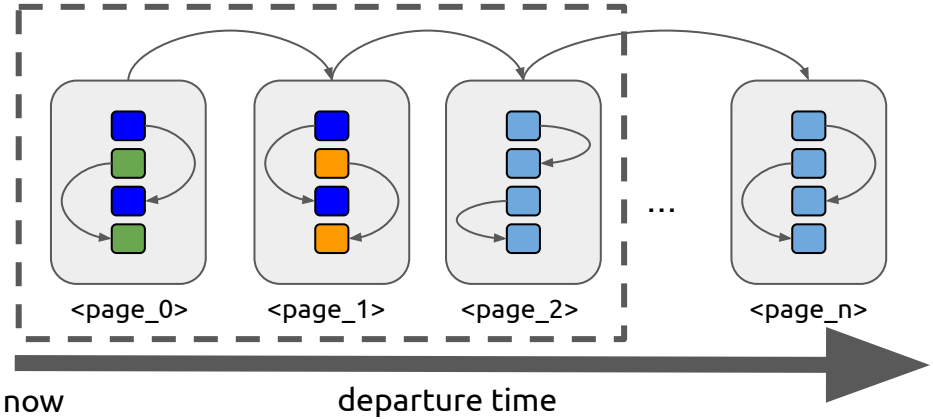



* @prefix lc: <http://semweb.mmlab.be/ns/linkedconnections#>
** @prefix hydra: <http://www.w3.org/ns/hydra/core#>

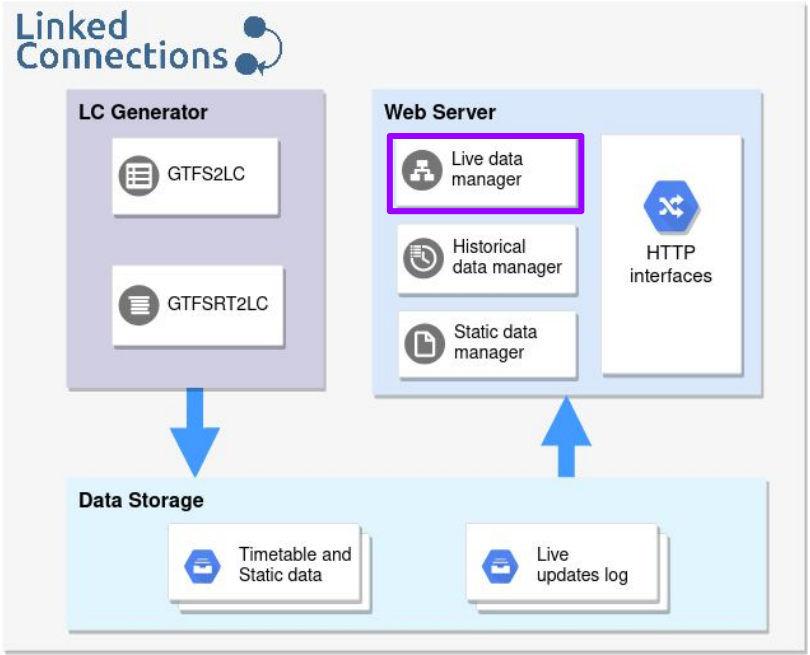
# Serving **live data** efficiently with an AVL tree

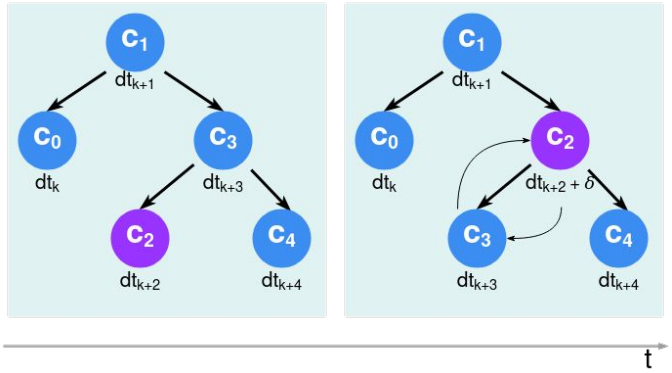


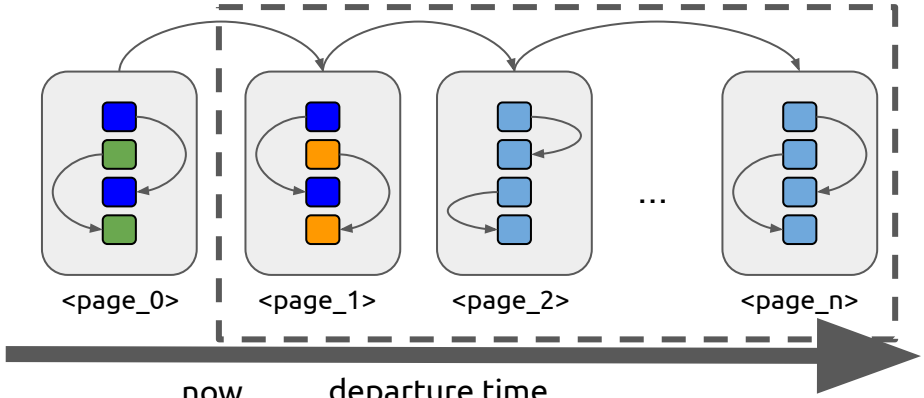



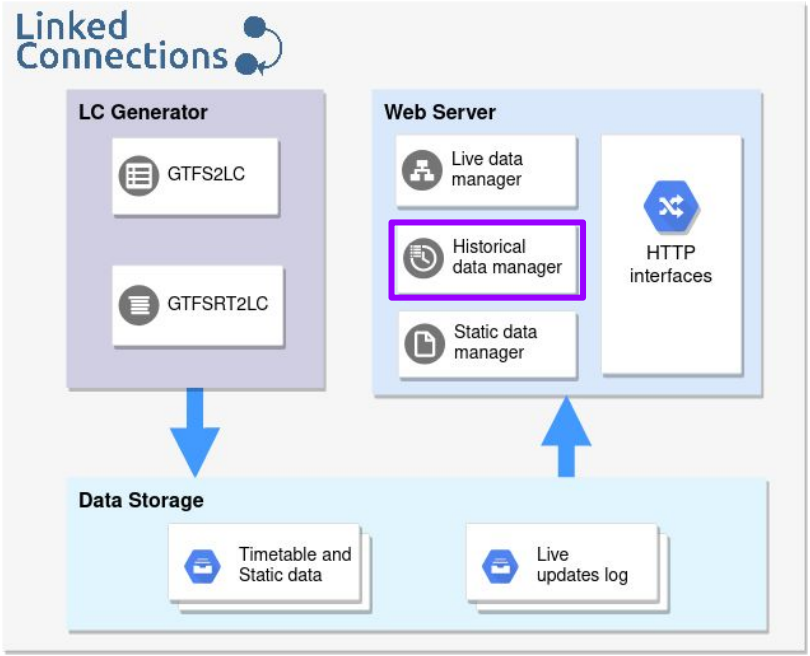
# Serving **live data** efficiently with an AVL tree



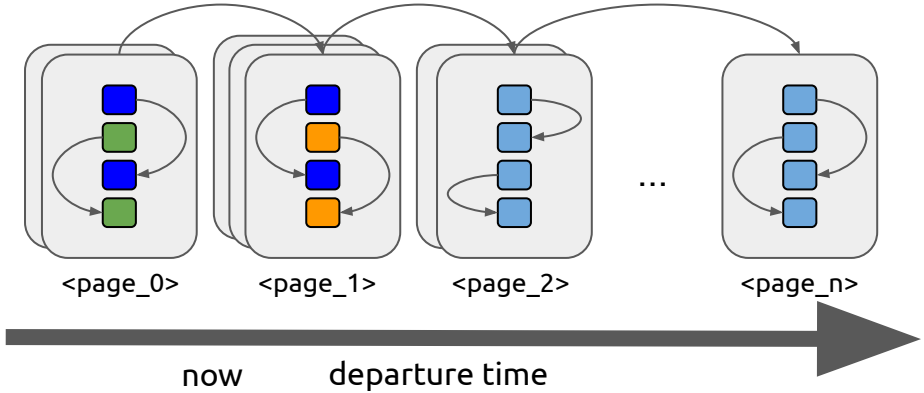



# Serving **historical data** with Memento




```
GET /connections?departureTime=2021-09-06T08:00:00.000Z
HTTP/1.1

Host: agency.org
Accept-Datetime: Mon, 06 Sep 2021 07:35:00 GMT
Connection: close
```



# Three main aspects were investigated*:

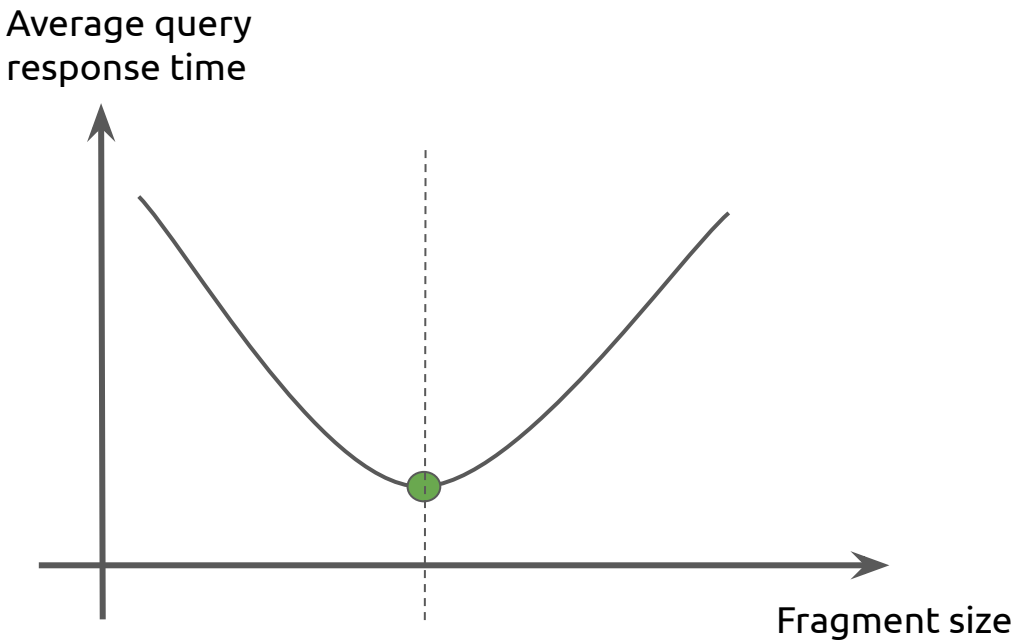
1. Optimal *fragment size* for route planning querying

2. Correlation of *query performance* with graph network properties

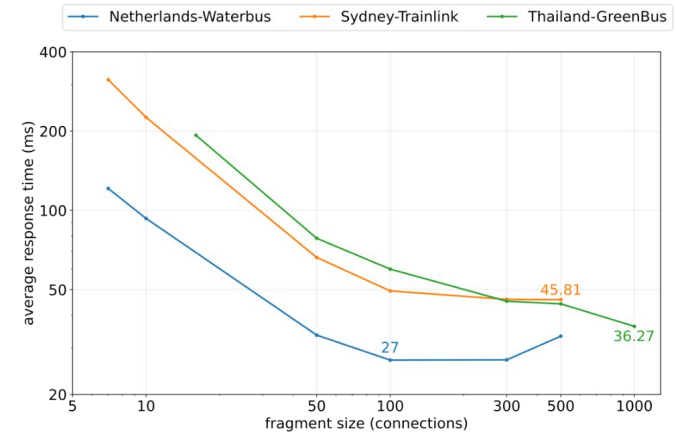
3. *Cost-efficiency* in relation to traditional non-semantic solution

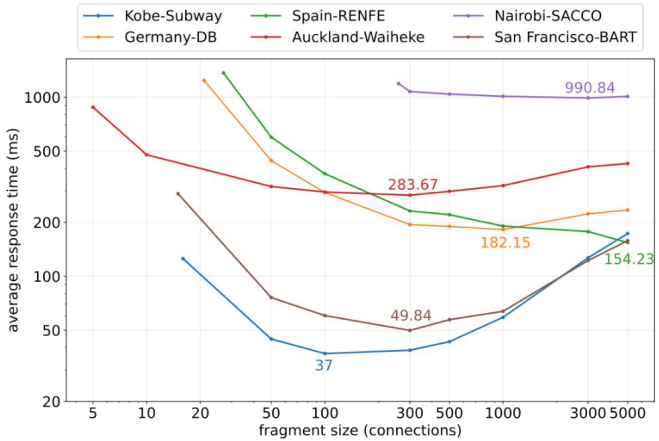
* All experiments and data are available at https://github.com/julianrojas87/lc-evaluation-swj

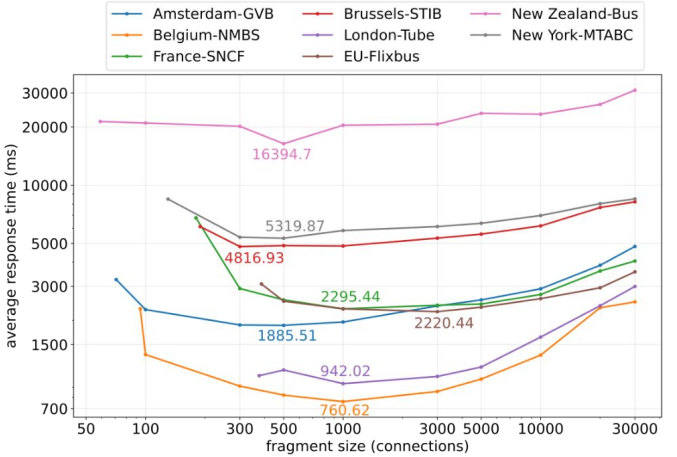
# 1. Optimal fragment size for route planning querying

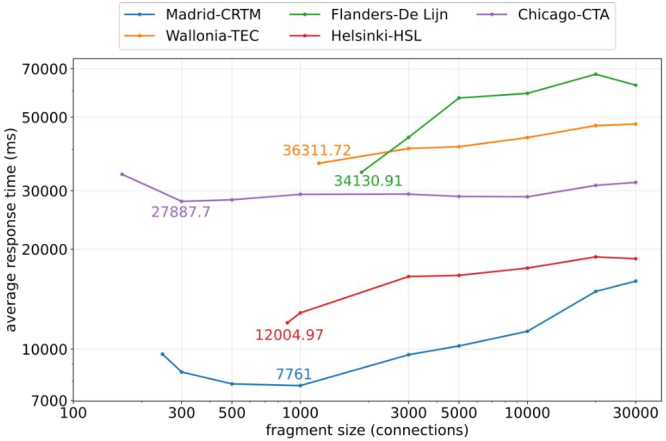


# 1. Optimal fragment size for route planning querying

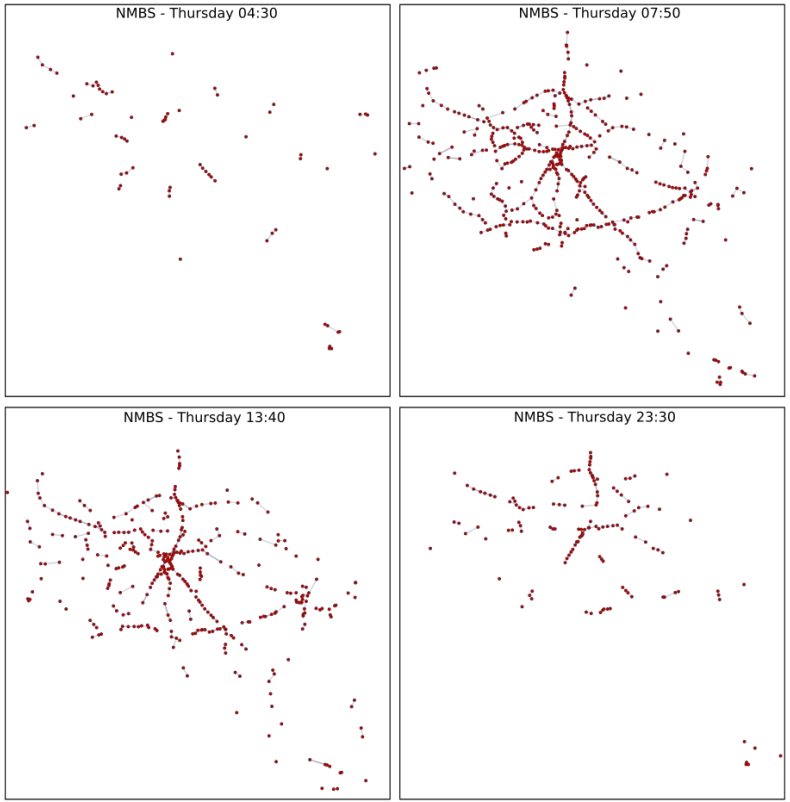


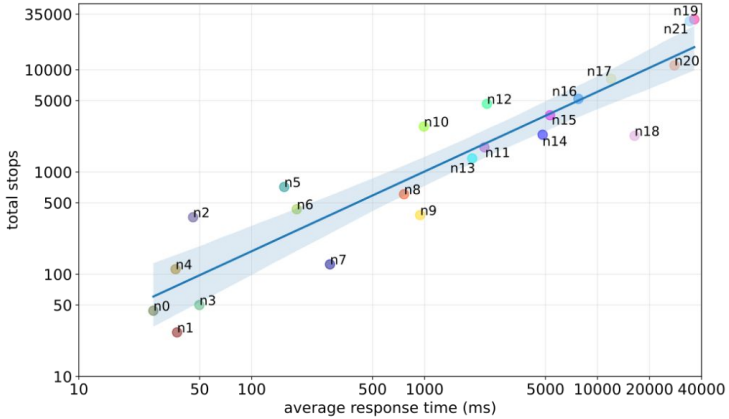
# 2. Correlation of query performance with graph network properties

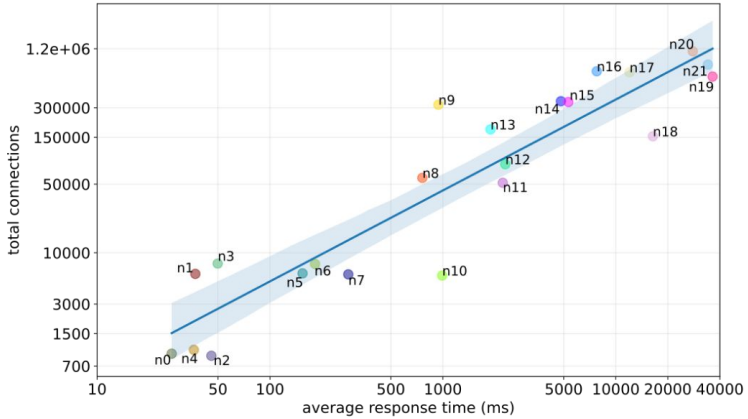


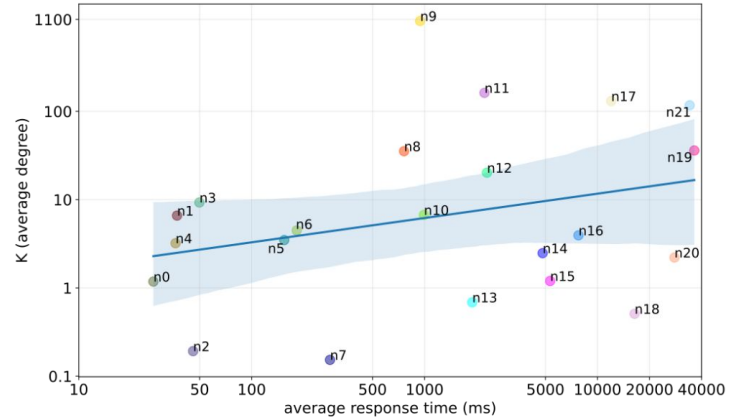

Transport network as a **Time-Varying Graph**.
We measured:

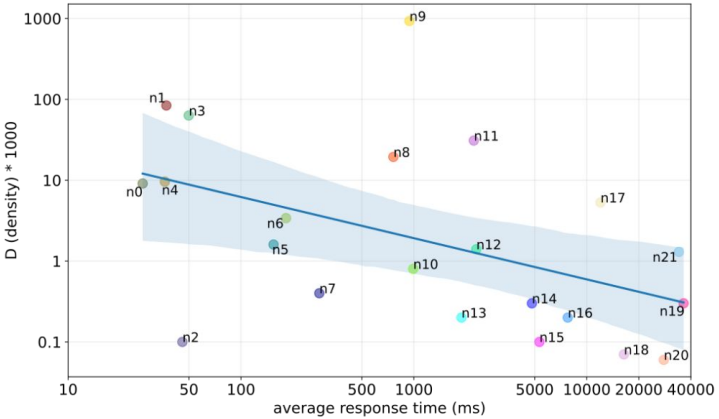
- Size (in terms of stops and connections)
- Average Degree
- Density
- Clustering Coefficient
- Average Connection Duration

# 2. Correlation of query performance with graph network properties

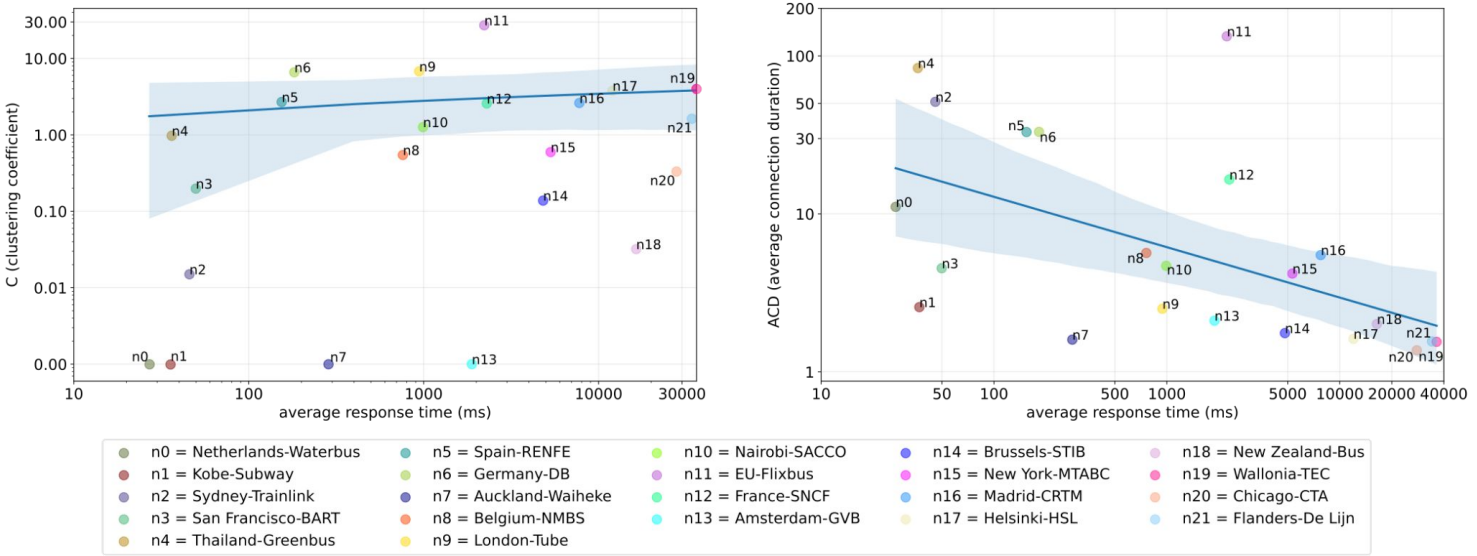


# 2. Correlation of query performance with graph network properties



# 3. Cost-efficiency in relation to traditional non-semantic solution

We measured:

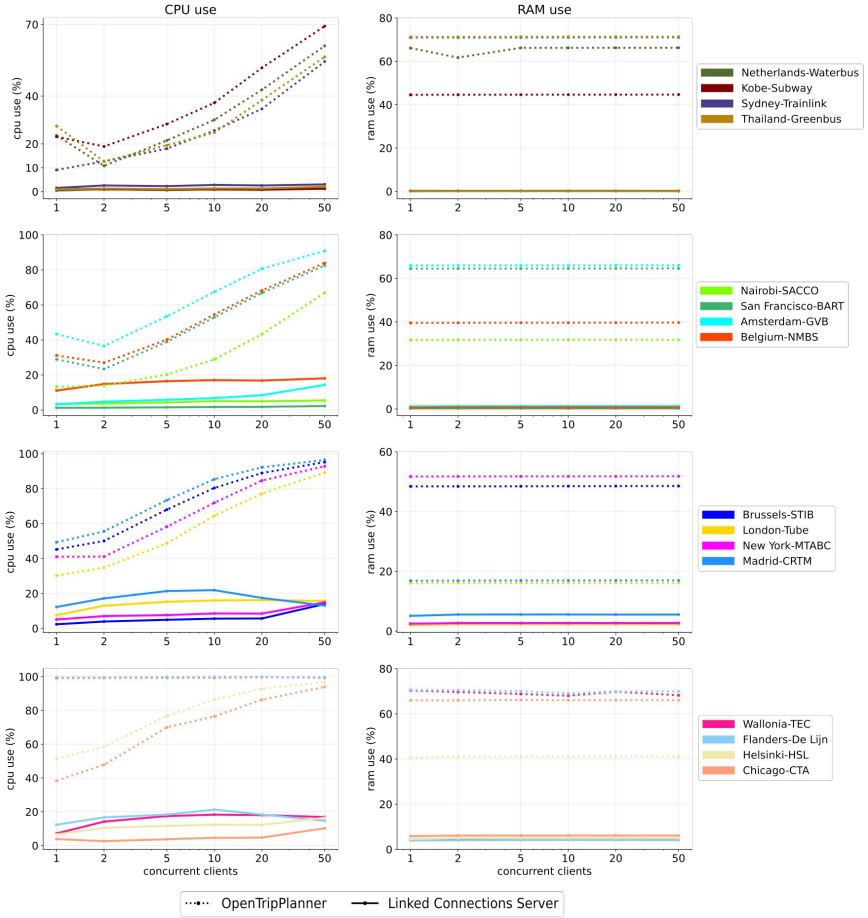
- query response time and;
- server-side CPU and RAM use

of

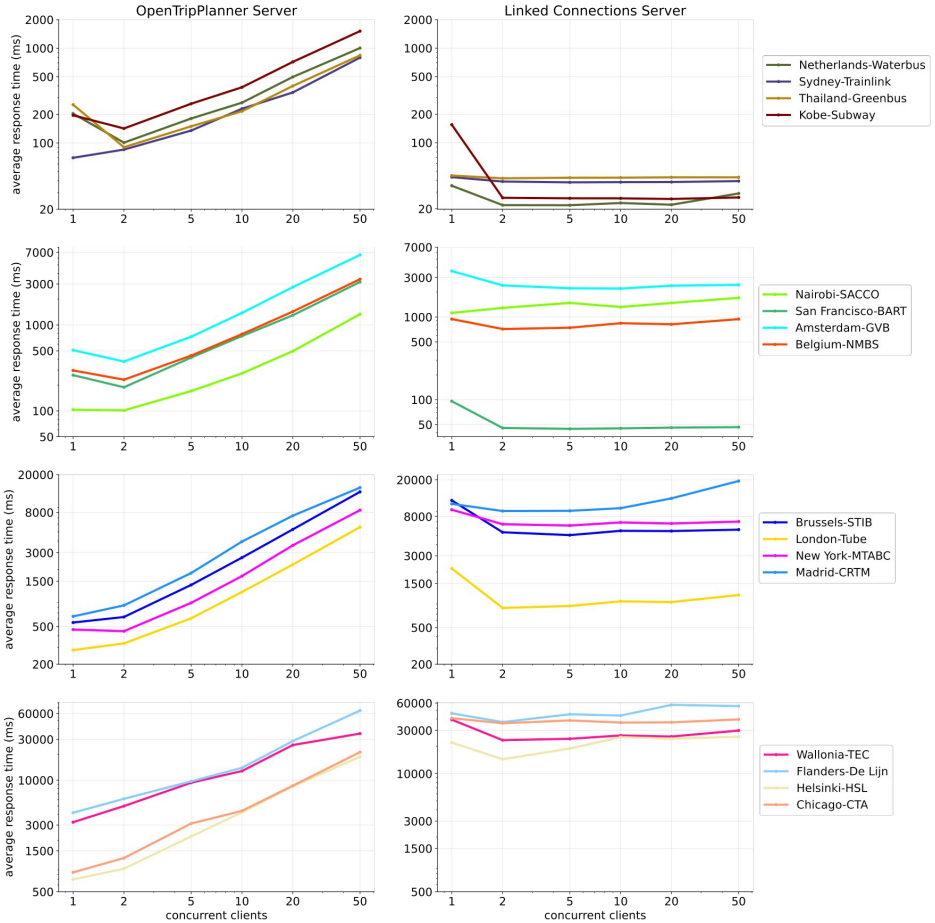
(i) **Linked Connections Server** vs **OpenTripPlanner***
(ii) **live** and **historical** queries in Linked Connections

* OpenTripPlanner: https://www.opentripplanner.org/

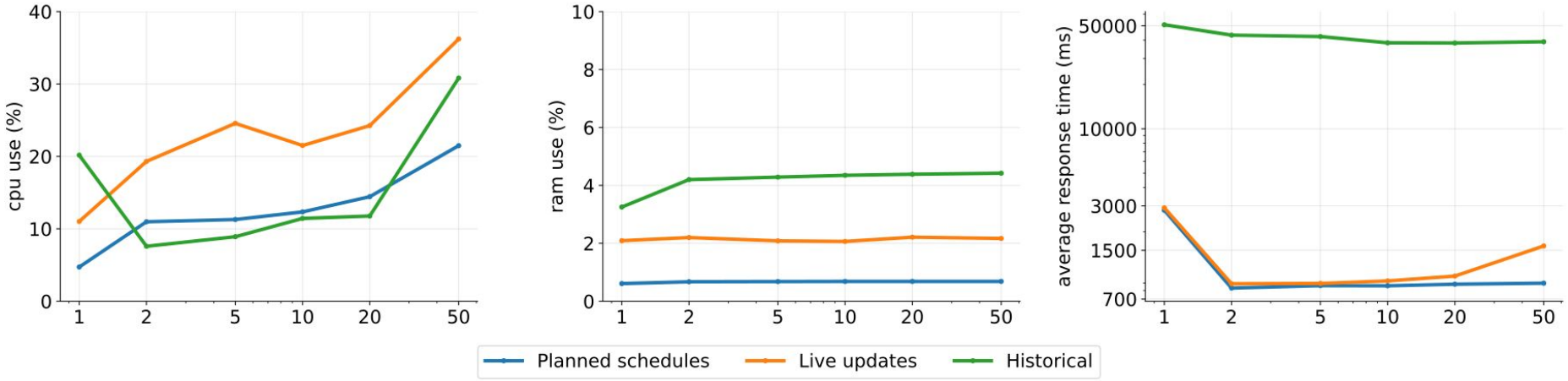
# 3. Cost-efficiency in relation to traditional non-semantic solution



# 3. Cost-efficiency in relation to traditional non-semantic solution



# 3. Cost-efficiency of live and historical queries



# Conclusions

Semantic technologies can be used efficiently to describe not only **domain specific** data but also the **(Web) interfaces** that give access to it.

The optimal fragment size of a Linked Connections dataset is highly **correlated** with the **network size, density** and **average connection duration.**

Linked Connections is more **cost-efficient** than traditional solutions and for **smaller networks** (< ~1000 stops) can **outperform** traditional solutions.

# Future work

**Optimization** of query performance on **larger networks** by relying on **geospatial fragmentations** created based on network properties.

**SPARQL** query execution clients over Linked Connections interfaces for supporting other use cases -> **Comunica**

**Stream-based** data architectures for efficient historical querying and archiving -> **LDES**, **TREE**

# Publishing planned, live and historical public transport data on the Web with the Linked Connections Framework

**Julián Rojas**, Harm Delva, Pieter Colpaert, Ruben Verborgh

Ghent University – imec

**@julianr1987**
julianandres.rojasmelendez@ugent.be

**3nd International Workshop on Semantics and the Web for Transport**

September 2021

GHENT UNIVERSITY

imec