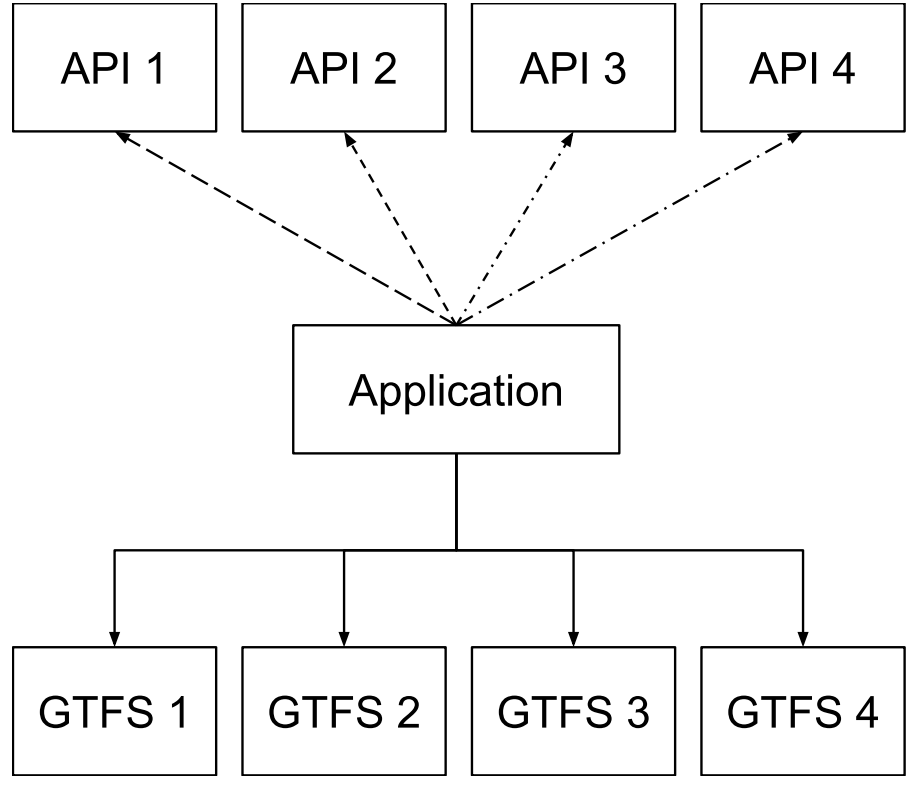
GEOSPATIAL PARTITIONING OF OPEN TRANSIT DATA

Harm Delva, Julián Andrés Rojas, Pieter-Jan Vandenberghe, Pieter Colpaert, and Ruben Verborgh

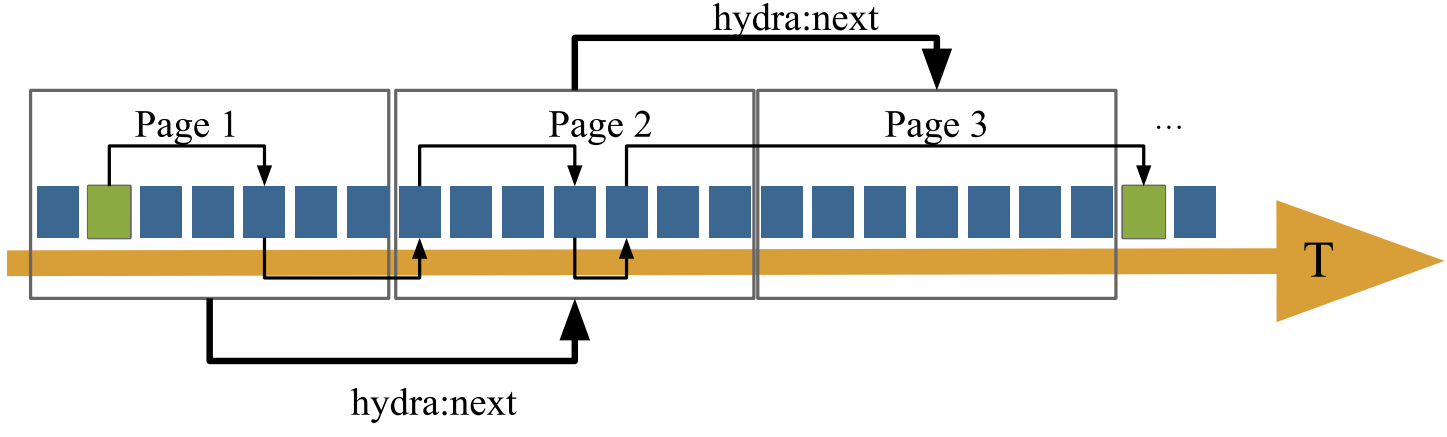
IDLab, Ghent University – imec

USING OPEN TRANSIT DATA

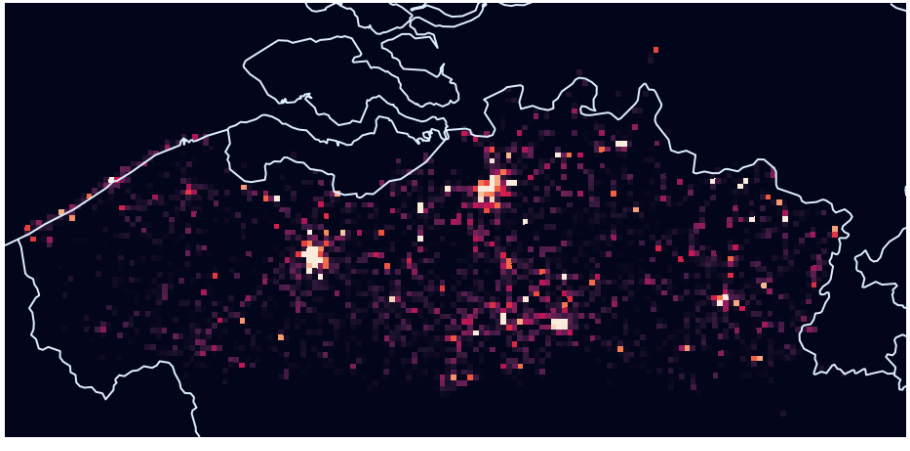
Would you rather interface with 4 APIs, or parse GBs of CSV files?



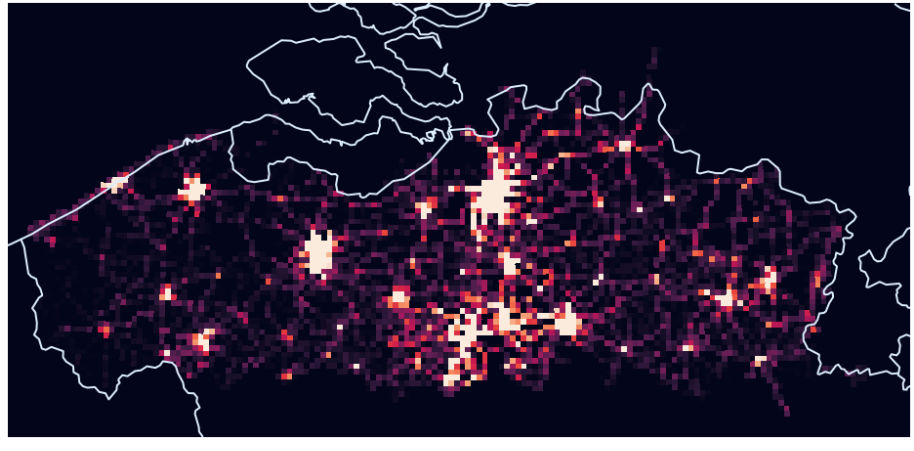
LINKED CONNECTIONS



DEMAND VARIES BY REGION



Where people want to go



Where the buses are

Context

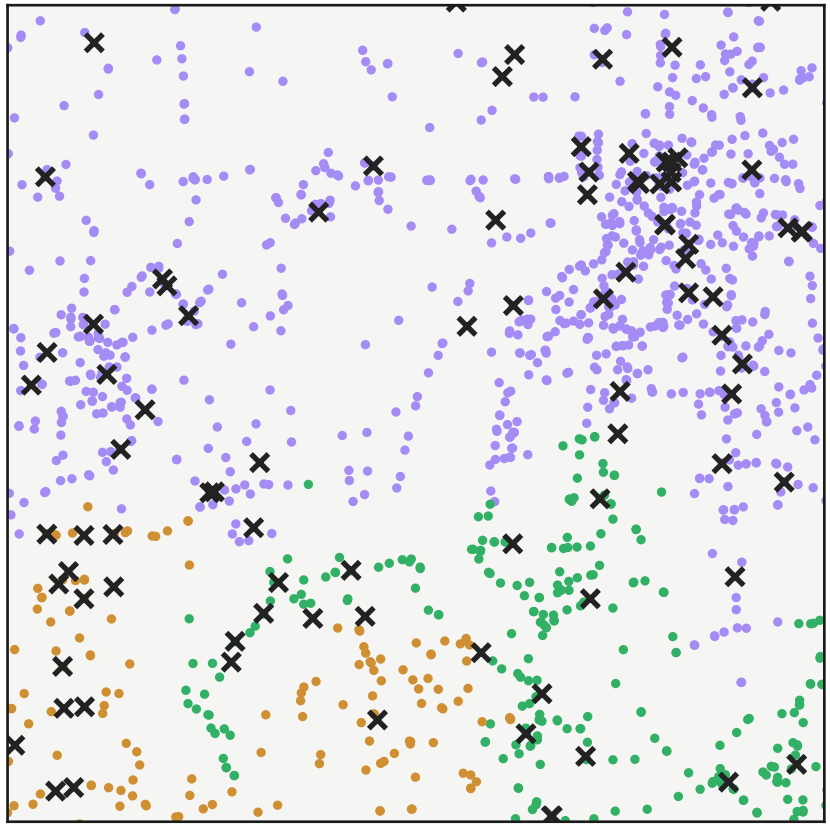
- **Publishing the data**

Fragmenting the data

Findings

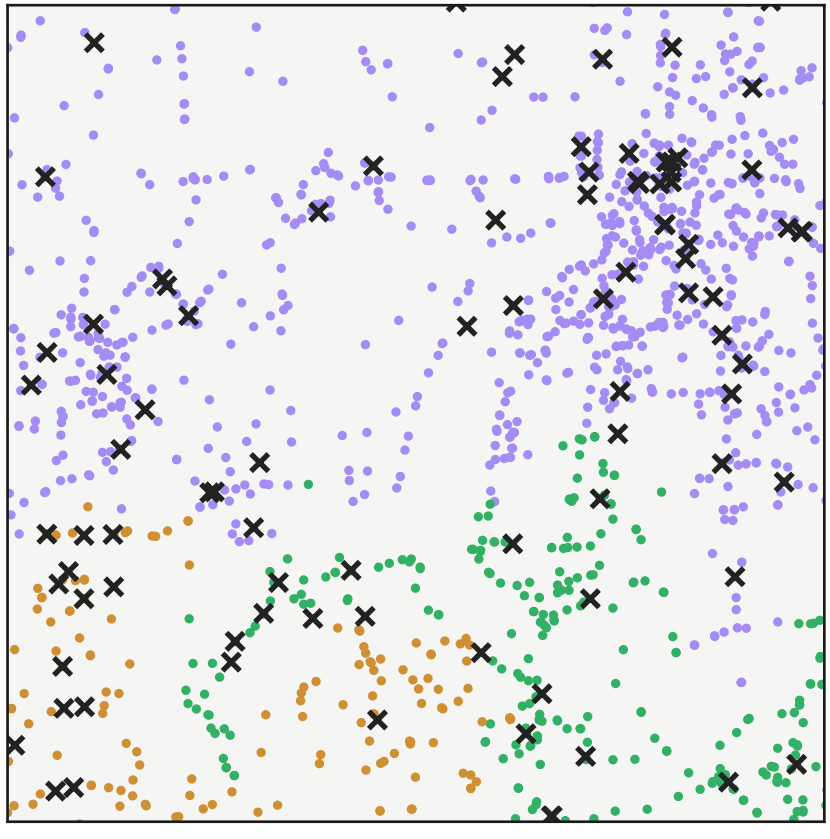
Takeaways

SETS OF STOPS ARE RIGID



What if a new stop is added?

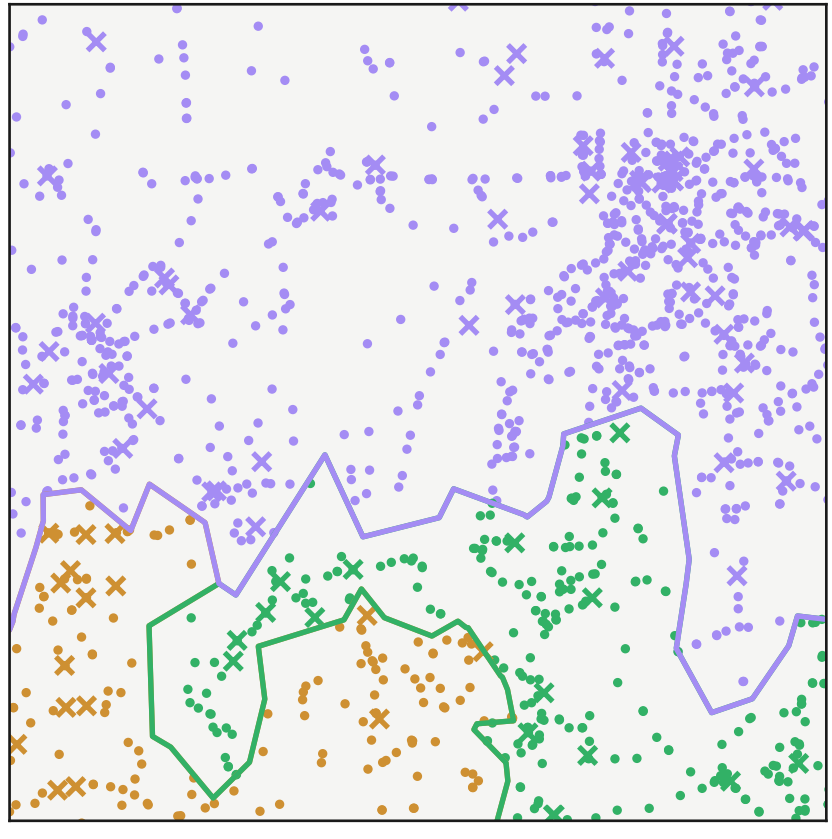
SETS OF STOPS ARE RIGID



What if a new stop is added?

Changes are needed

LET USERS CLASSIFY THE STOPS



Context

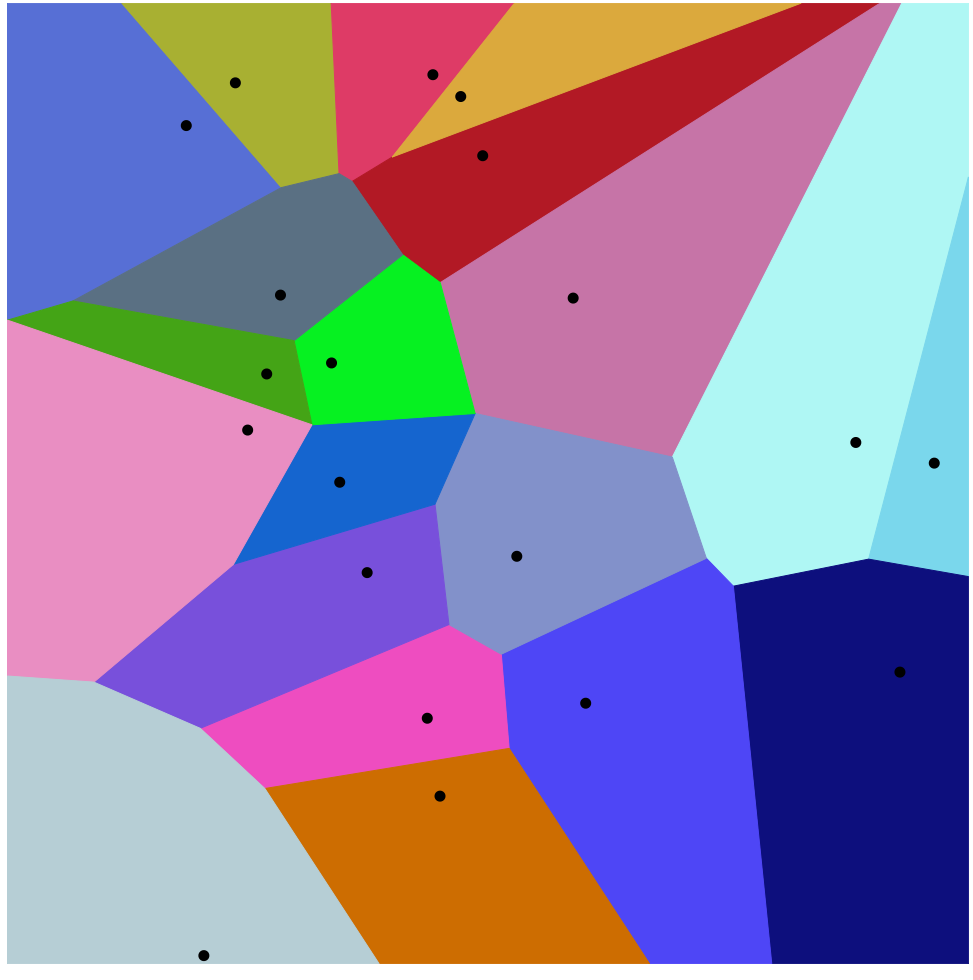
Publishing the data

- **Fragmenting the data**

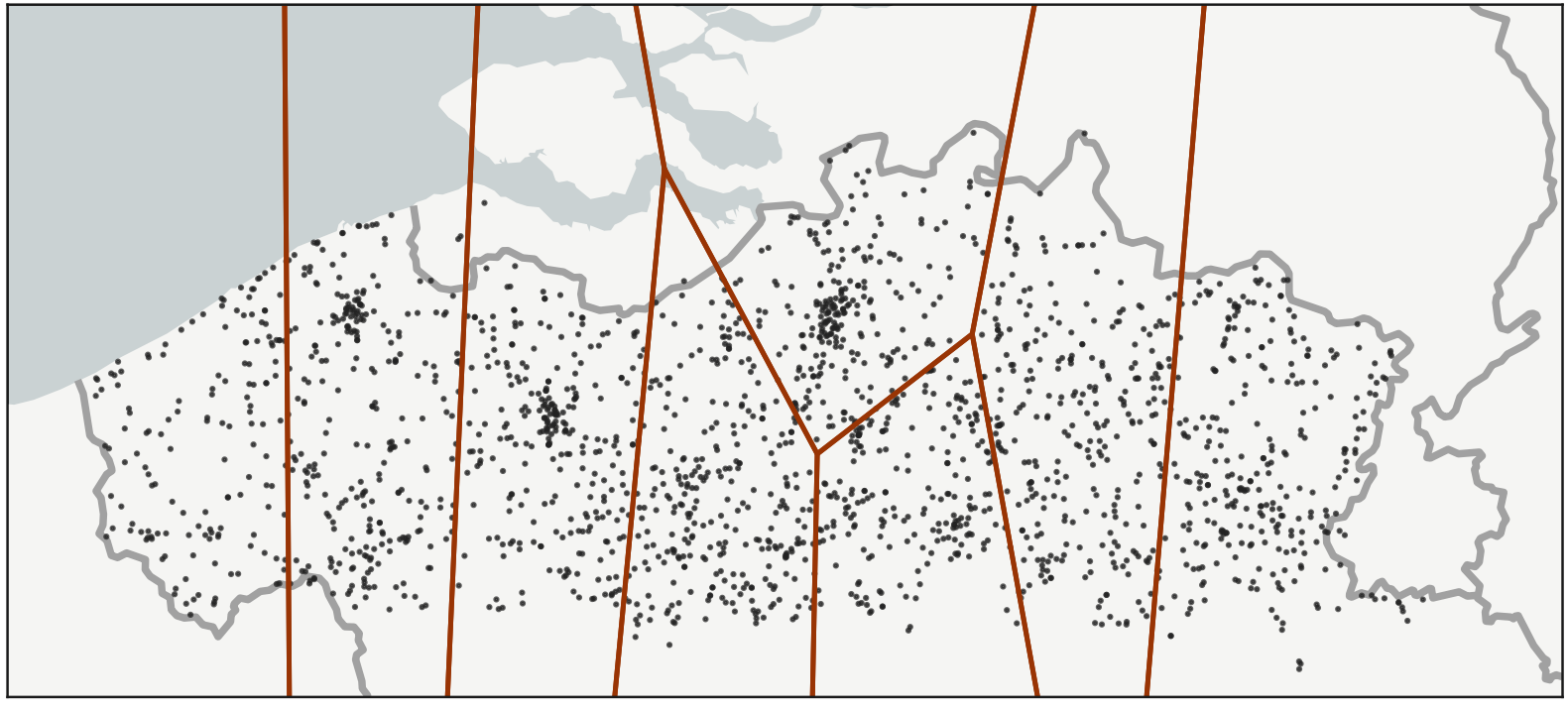
Findings

Takeaways

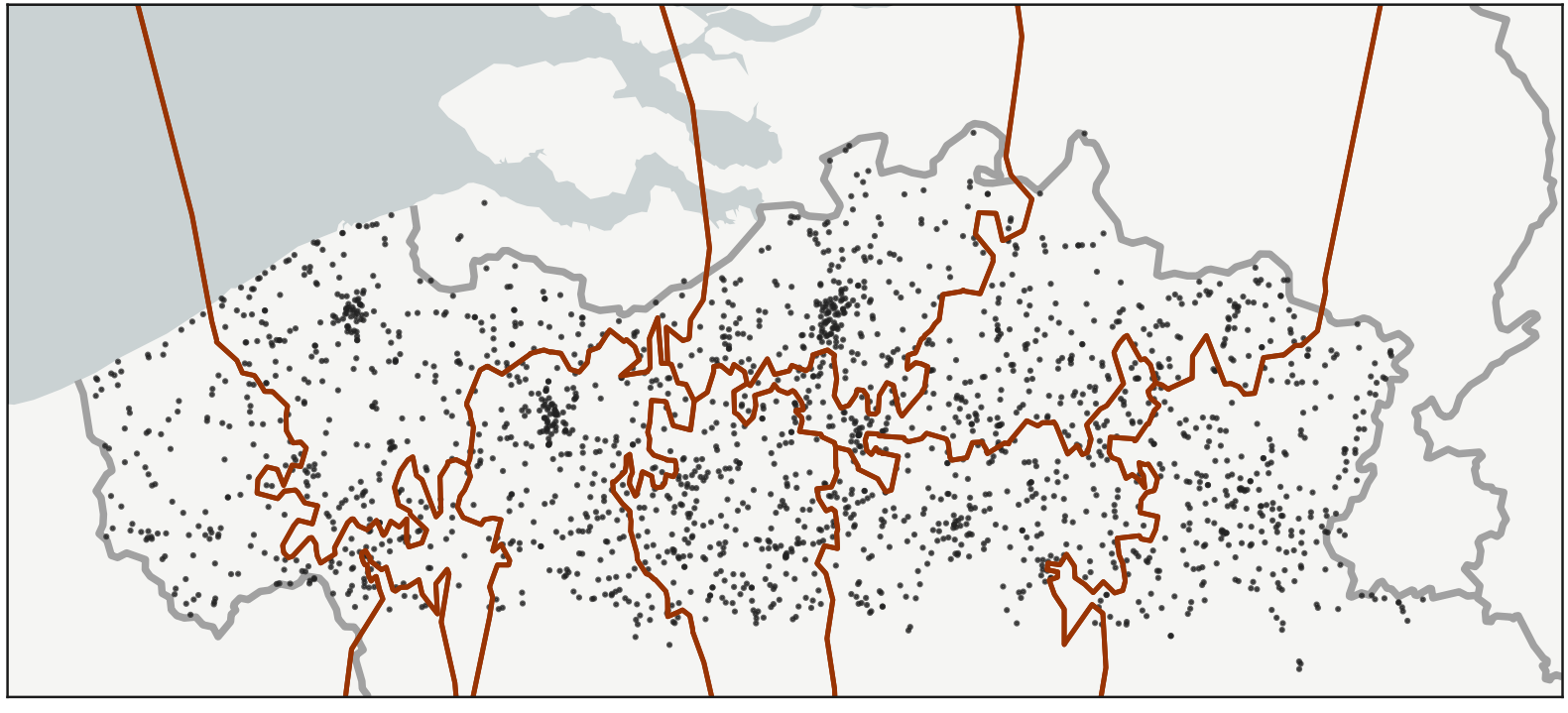
VORONOI DIAGRAMS



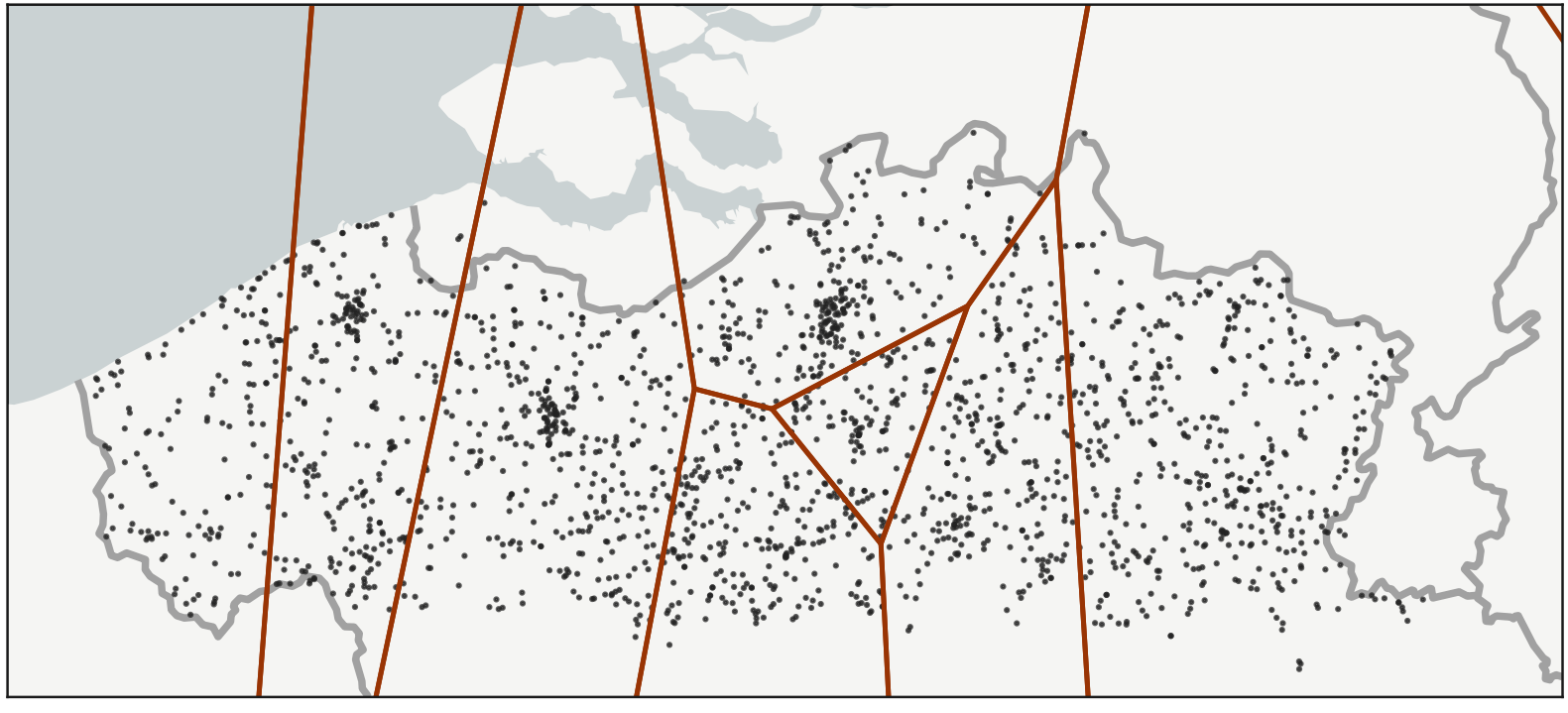
K-MEANS



METIS



VORONOI CELLS AROUND HUBS



Context

Publishing the data

Fragmenting the data

- **Findings**

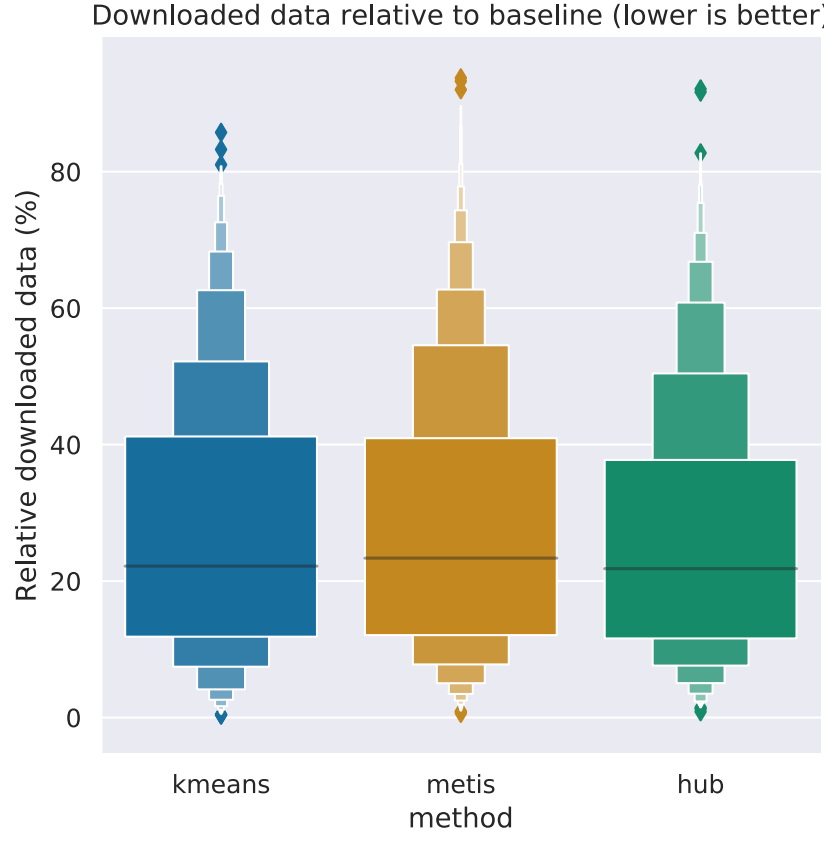
  Takeaways

SETUP

- Client-side route planning
- 3 methods + baseline
- 4, 8, 16, 32 partitions
- 5000 actual queries
- 3 metrics

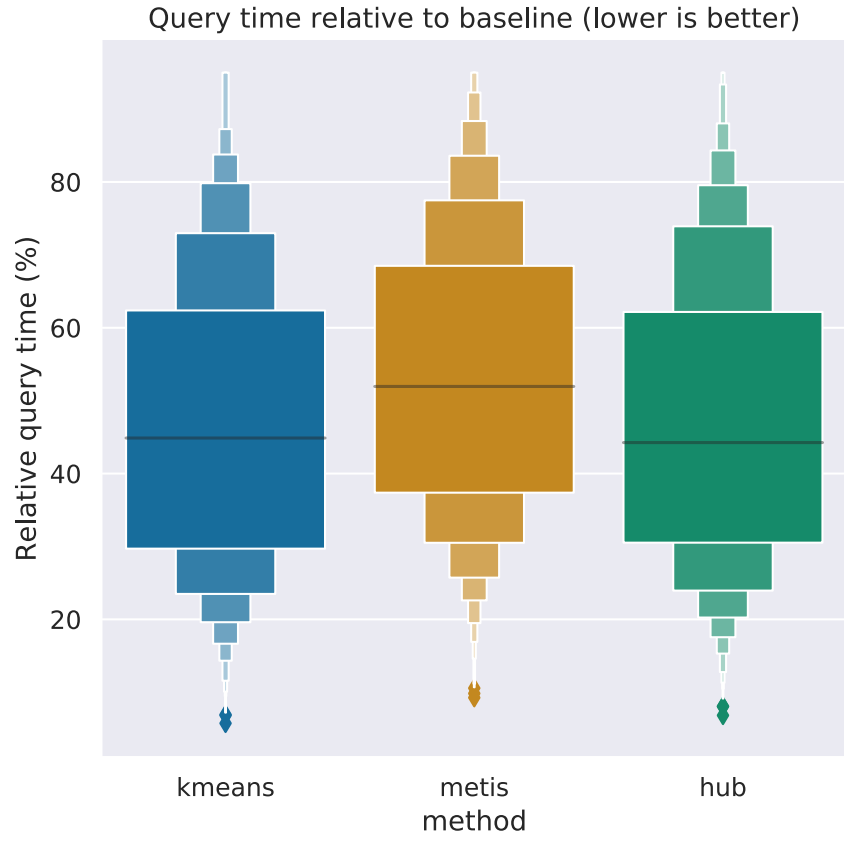
DOWNLOADED DATA

Just a few % difference



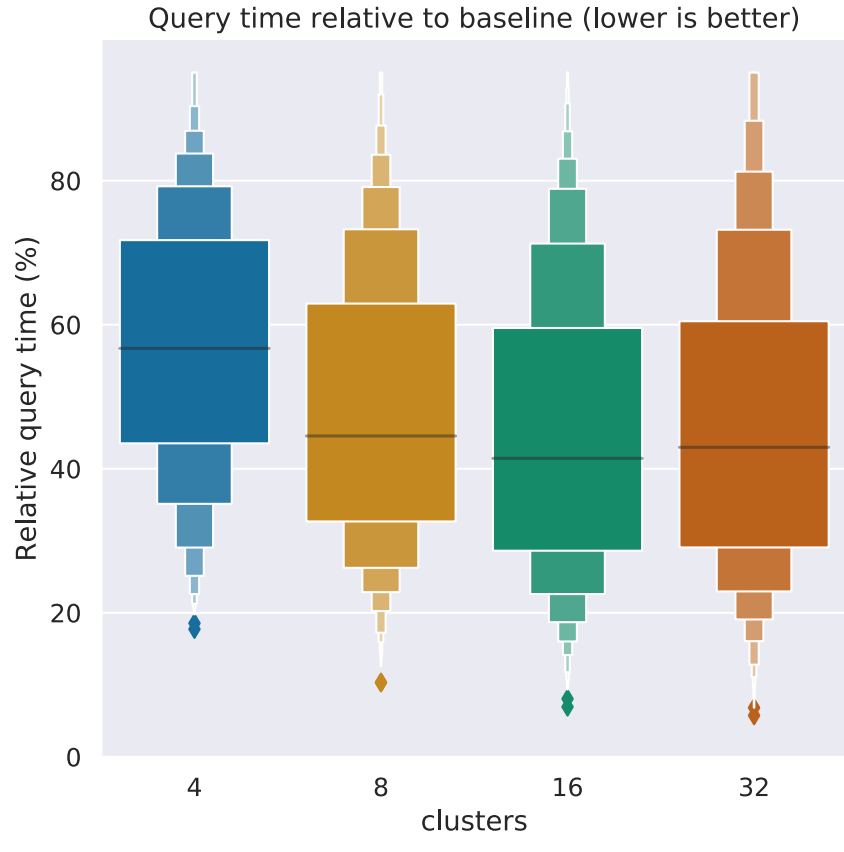
EXECUTION TIMES

Complex shapes are hard to use

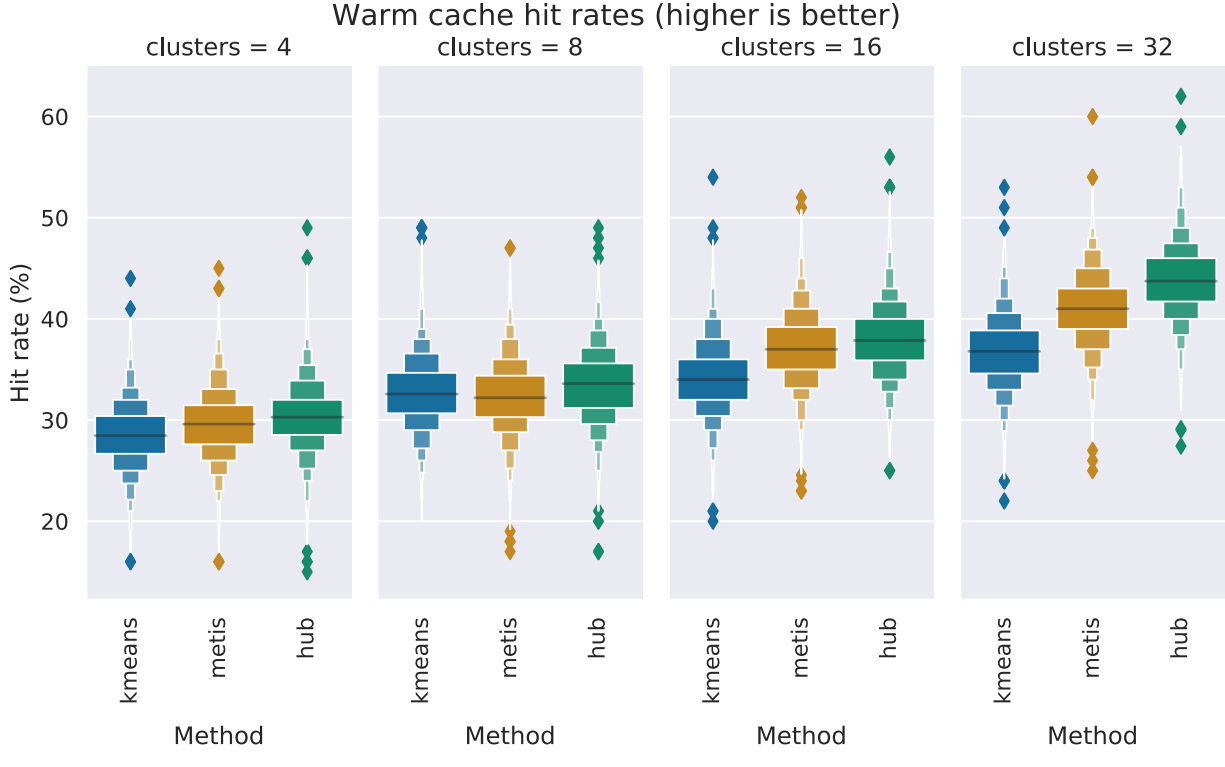


EXECUTION TIMES

More clusters is not better



CACHE HIT RATES



What's happening to k-means?

It doesn't matter how you partition your data, just do it.

hdelva.be/slides/icwe2020
hdelva.be/articles/geospatial-linked-connections