A Seriation Based Framework to Visualize Multiple Aspects of Road Transport from GPS Trajectories

Alexandre Dubray Siegfried Nijssen Isabelle Thomas Pierre Schaus

International Conference on Intelligent Transportation, 2021

An introducing example

Imagine that we want to visualize the mean speed of the trucks, per municipality, for a given country.

An introducing example

Imagine that we want to visualize the mean speed of the trucks, per municipality, for a given country.

First you compute the mean speed of the trucks in each municipality.

An introducing example

Imagine that we want to visualize the mean speed of the trucks, per municipality, for a given country.

First you compute the mean speed of the trucks in each municipality.

Then a color is assigned to each municipality, depending on this mean speed (from light to dark)

An introducing example

Imagine that we want to visualize the mean speed of the trucks, per municipality, for a given country.

First you compute the mean speed of the trucks in each municipality.

Then a color is assigned to each municipality, depending on this mean speed (from light to dark)

Finally, the municipalities are shown on a map with their color.

An introducing example

Imagine that we want to visualize the mean speed of the trucks, per municipality, for a given country.

First you compute the mean speed of the trucks in each municipality.

Then a color is assigned to each municipality, depending on this mean speed (from light to dark)

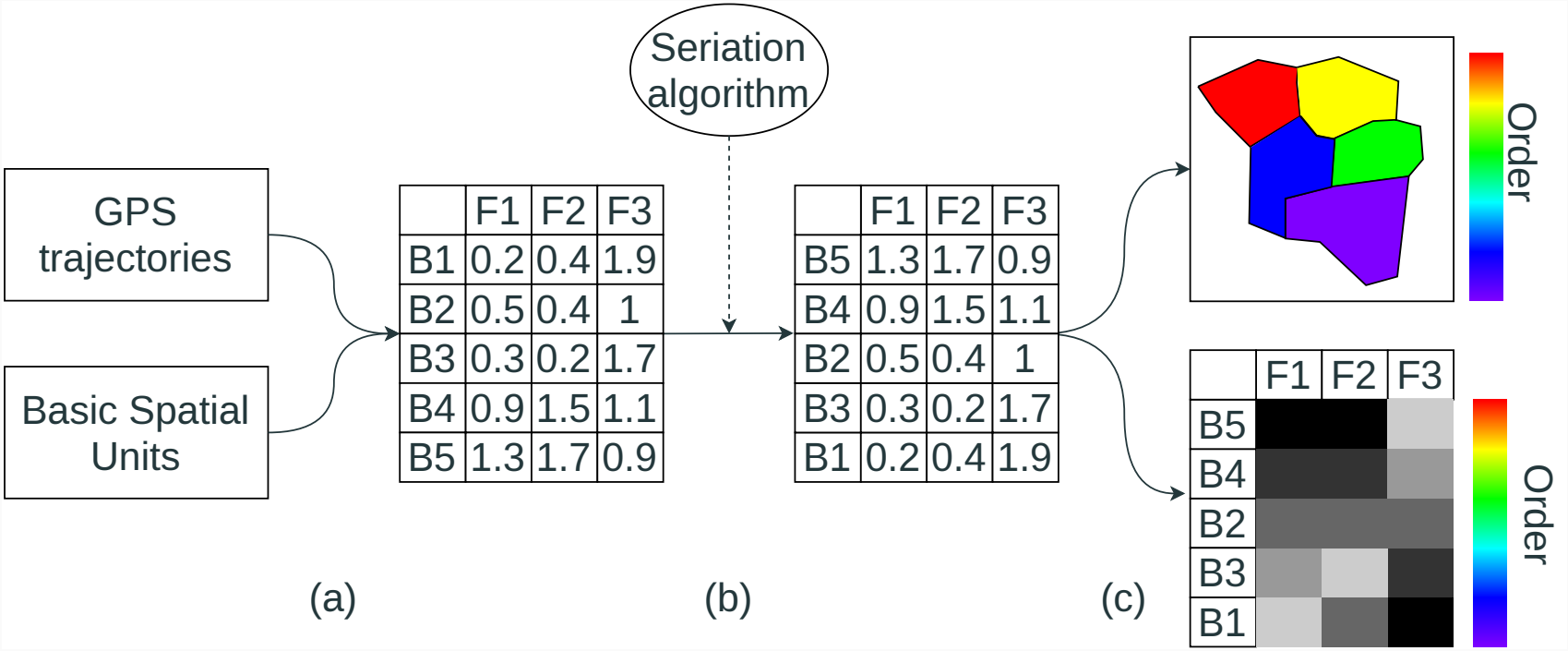
Finally, the municipalities are shown on a map with their color.

But how to extend that to handle multiple features at the same time?

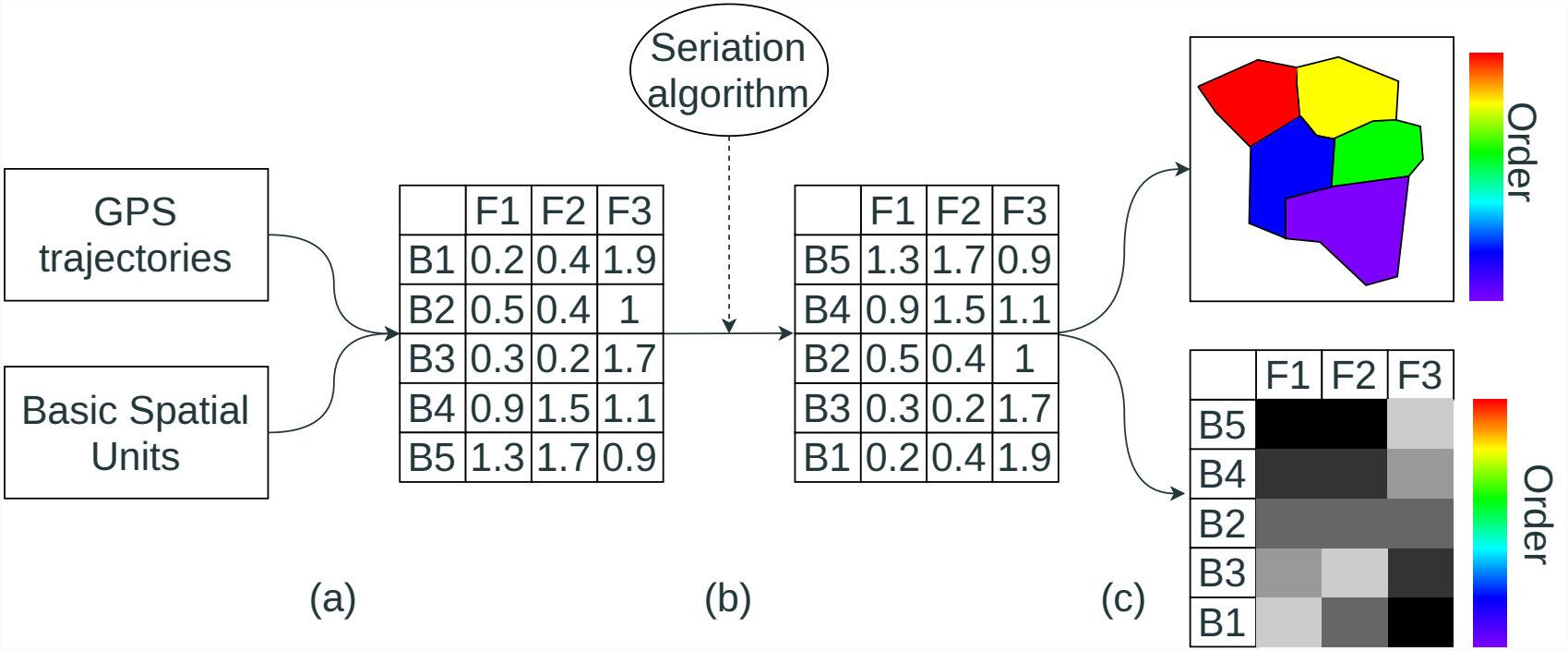
Existing solution

- One graph/map per computed feature
- Cluster in some way the studied areas before showing them on a map
  - Community detection, Self-Organizing Maps, ...
  - Hard to analyze the features
- In this work we propose a visualization that gives at one glance the geographical structure as well as feature interpretation

Overview of the framework



Overview of the framework

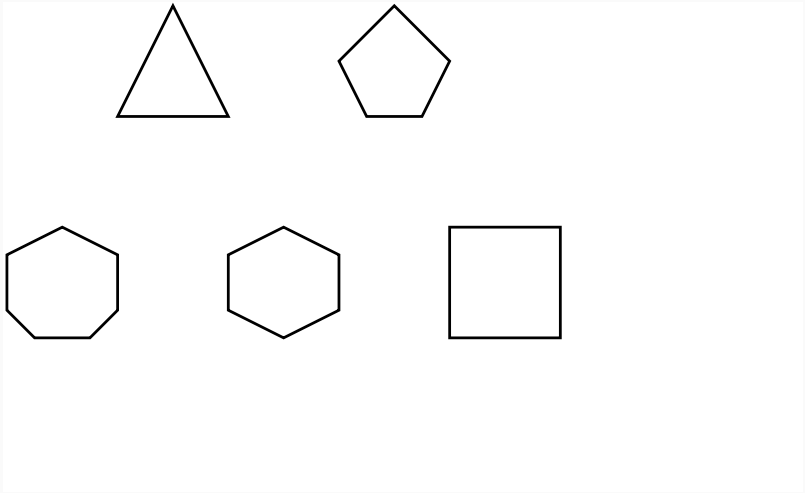


Two main questions

1. How to order the spatial units?
2. How to choose their color?

Ordering the units: Seriation

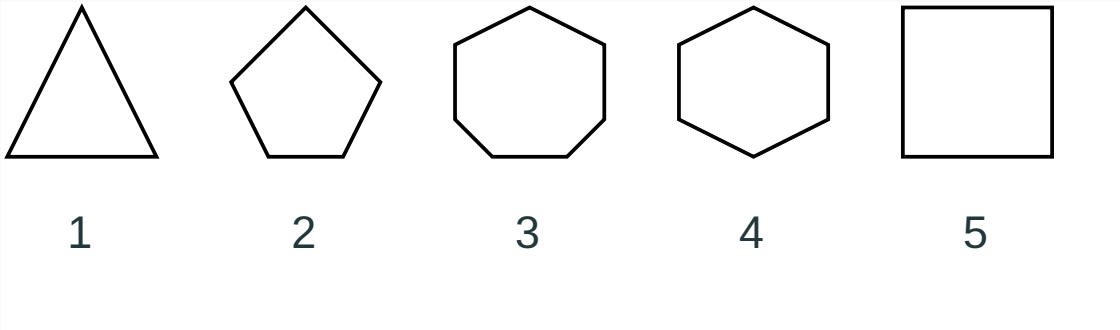
The goal of a seriation algorithm is to find a linear order of some objects so that similar objects are near each other. As an example, let us assume that we want to order these polygons.



The dissimilarity between two polygons is the difference between the number of edges they have.

Ordering the units: Seriation

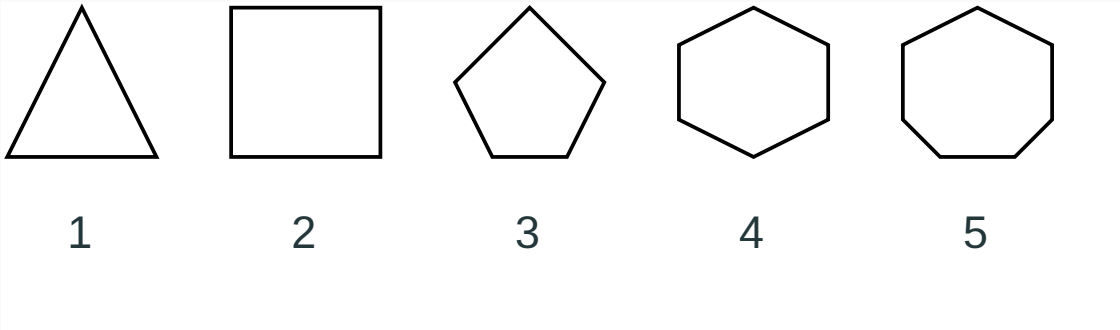
A possible ordering is the following



The score of the ordering is computed as the sum of the successive dissimilarity. This ordering has a score of $2 + 2 + 1 + 2 = 7$

Ordering the units: Seriation

The goal of a seriation algorithm is to find an ordering that minimize this score. In our example, the optimal ordering is the following



with a score of 4.

Ordering the units: Seriation

In general

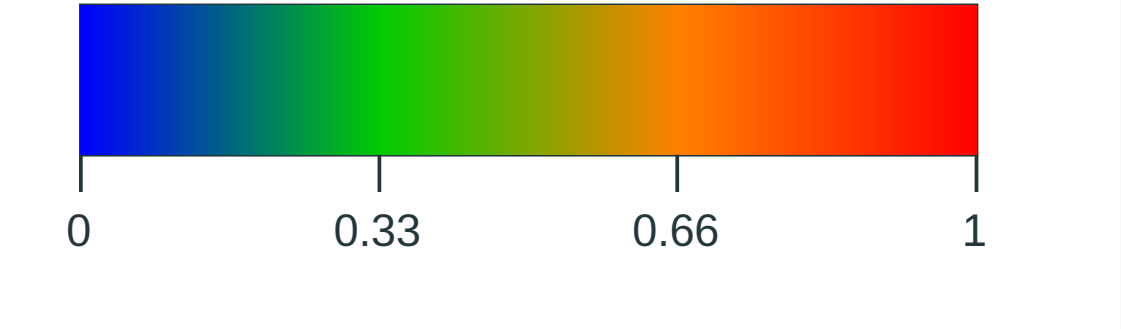
- The BSUs have more than one feature.
- The dissimilarity function depends on the type of the features.
- Multiple methods exist to solve the seriation problem. In this work we consider two of the most popular
  1. Optimal-Leaf-Ordering[1]
  2. Modelization as a Travelling Salesman Problem[2]

[1] Ziv Bar-Joseph, David K Gifford, and Tommi S Jaakkola. "Fast optimal leaf ordering for hierarchical clustering". In: *Bioinformatics* (2001).

[2] Gilbert Laporte. "The seriation problem and the travelling salesman problem". In: *Journal of Computational and Applied Mathematics* (1978).

The colors

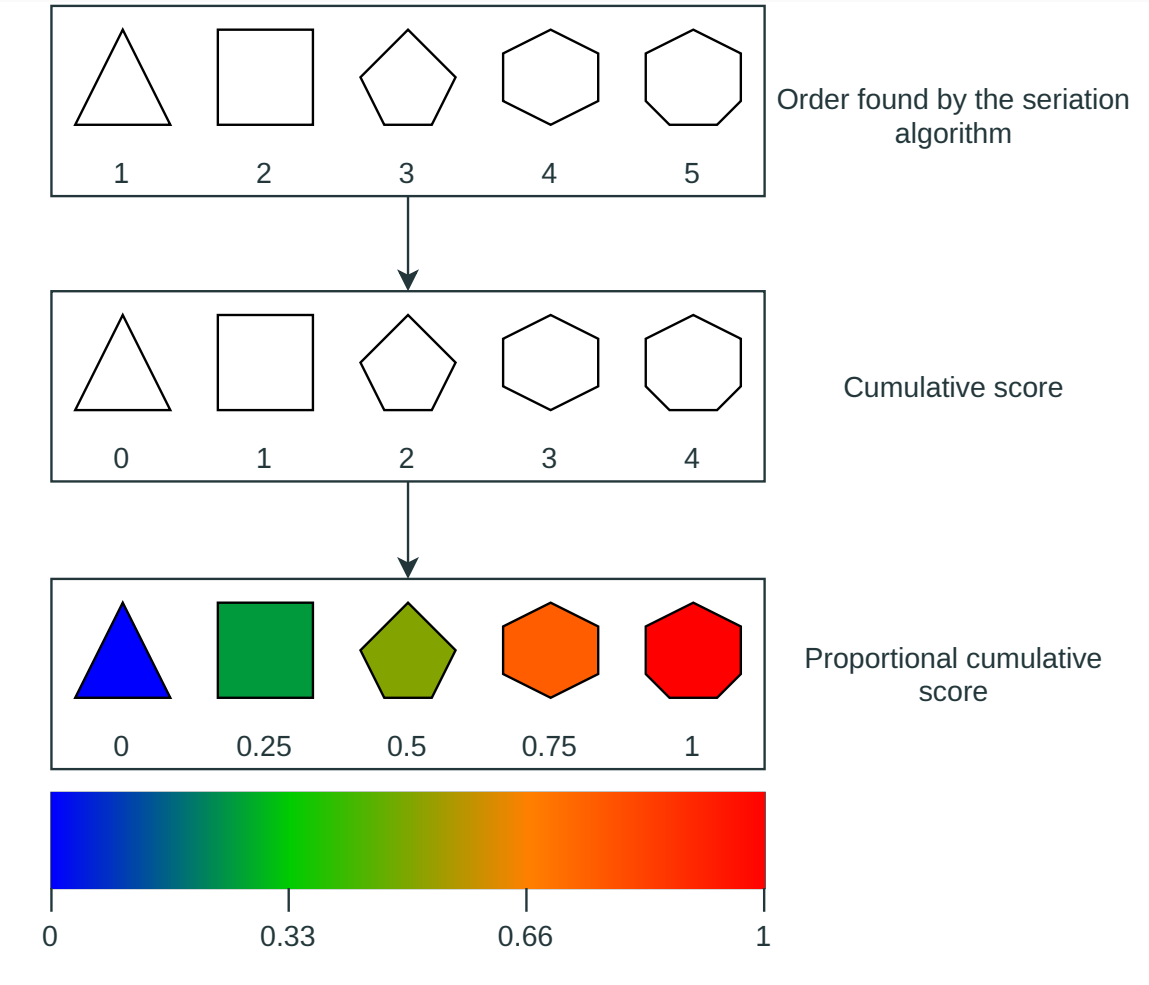
A linear color scale is created using n different colors evenly spaced. For instance, with four colors: blue; green; orange and red, the following color scale is obtained.



- Each of the n colors has a value between 0 (the first) and 1 (the last) at regular intervals. In the above color scale, the green is at 0.33 and orange at 0.66.
- The intermediate colors are linearly interpolated

The colors

To assign a color to the BSUs, first their cumulative score (w.r.t the ordering) is computed why is then divided by the total score of the ordering.



The colors

- If two BSUs have similar color, they have close features.

The colors

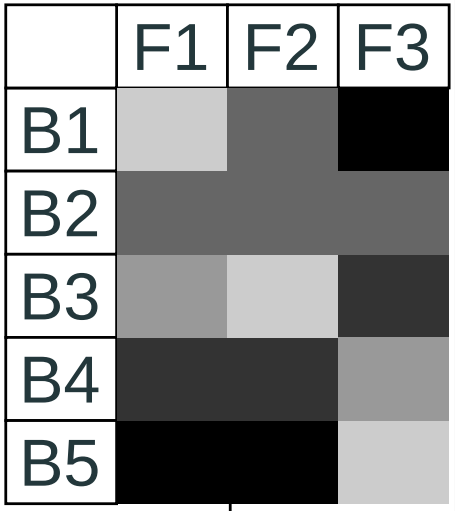
- If two BSUs have similar color, they have close features.
- Seriation only provides "local" guarantees, thus a multicolor scale is needed

The colors

- If two BSUs have similar color, they have close features.
- Seriation only provides "local" guarantees, thus a multicolor scale is needed
- Some colors of the scale might be unassigned in case of big gaps between the BSUs.

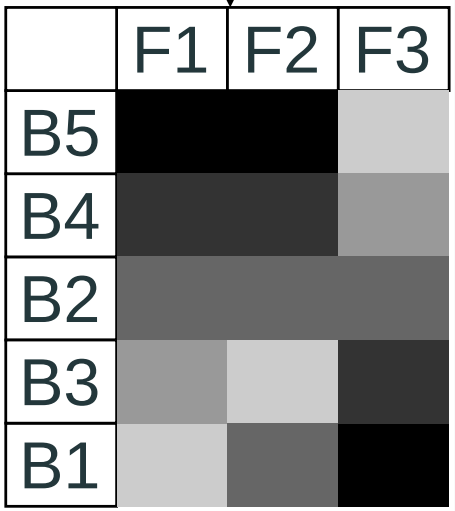
The heatmap

	F1	F2	F3
B1	0.2	0.4	1.9
B2	0.5	0.4	1
B3	0.3	0.2	1.7
B4	0.9	1.5	1.1
B5	1.3	1.7	0.9



Seriation

	F1	F2	F3
B5	1.3	1.7	0.9
B4	0.9	1.5	1.1
B2	0.5	0.4	1
B3	0.3	0.2	1.7
B1	0.2	0.4	1.9

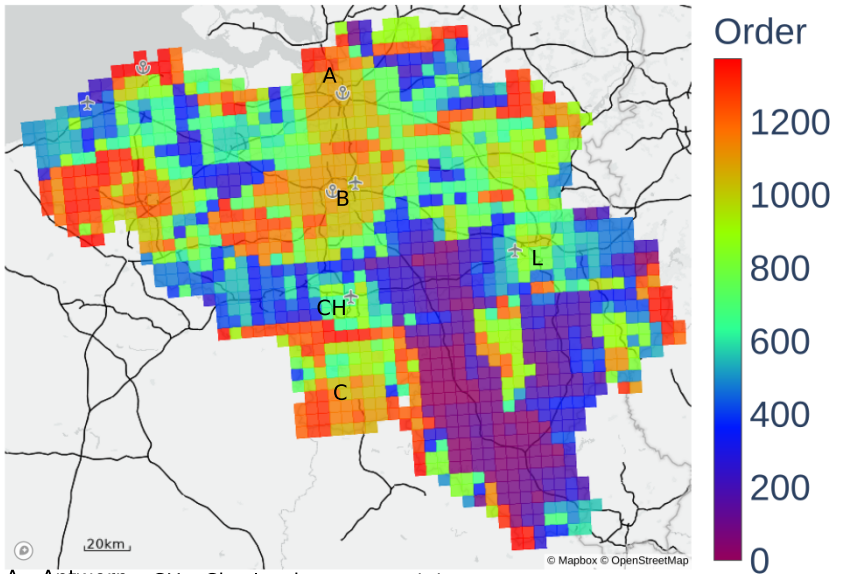


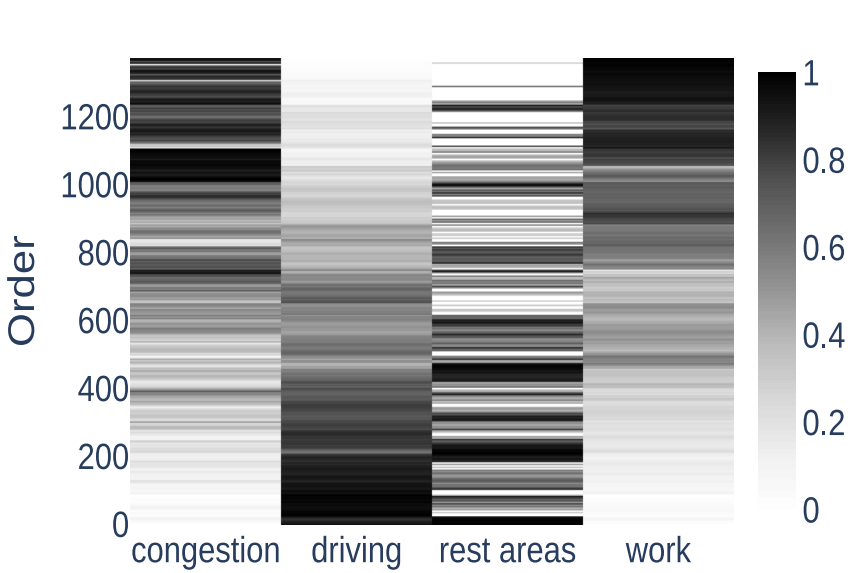
Results

The data set

- Trajectories of heavy good vehicles
- 1 day of data of all trucks that were on the Belgian territory
- Roughly $90,000$ vehicles and 30 millions GPS ping
- BSU are 5 km by 5 km grid cells provided by Eurostat
- Features computed for each cell is the proportions of time passed in the following states: Driving; Resting; Work-related actions; In congestion.

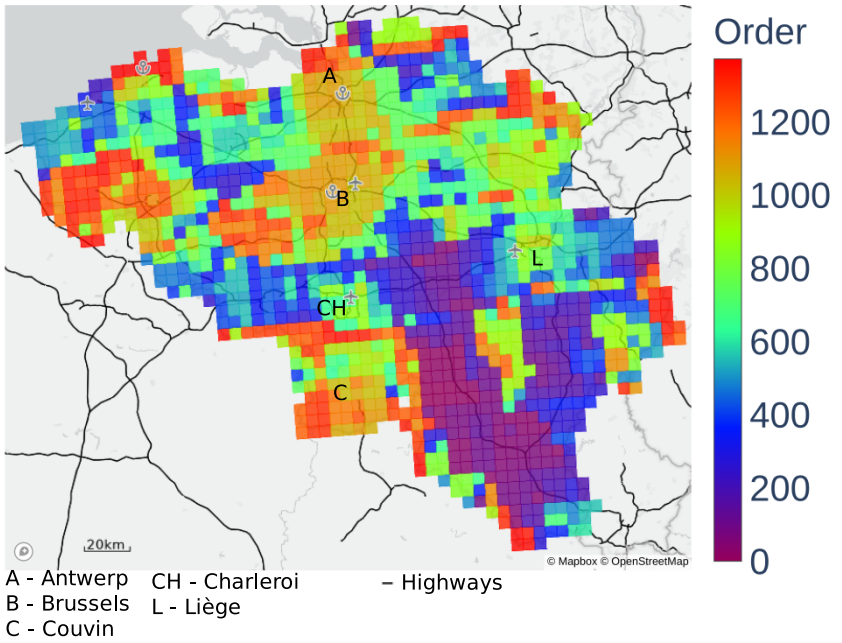
Final visualization

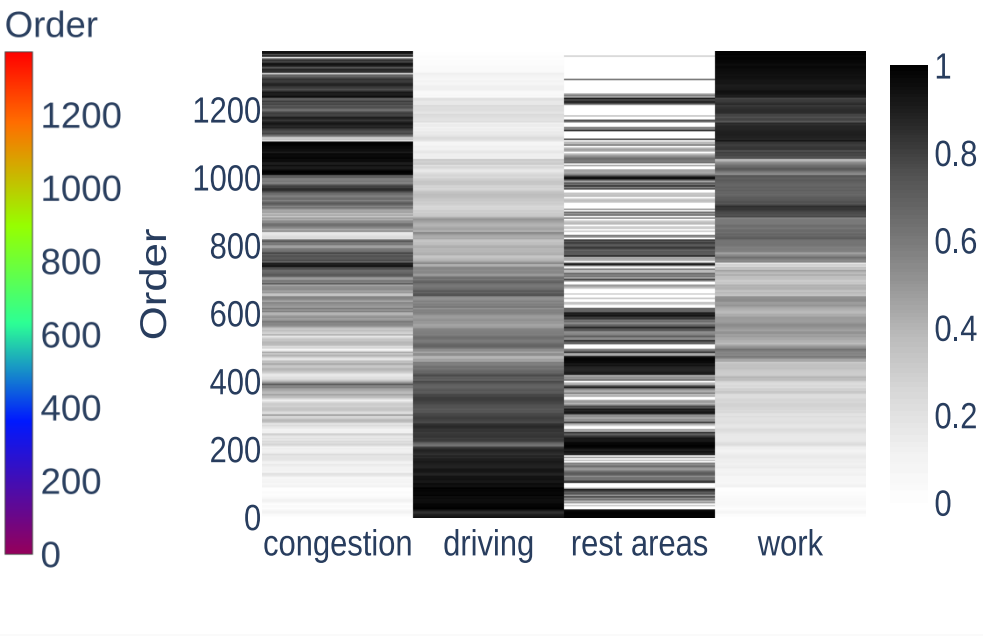




What can we infer on the geography of Belgium from this visualization ?

Final visualization





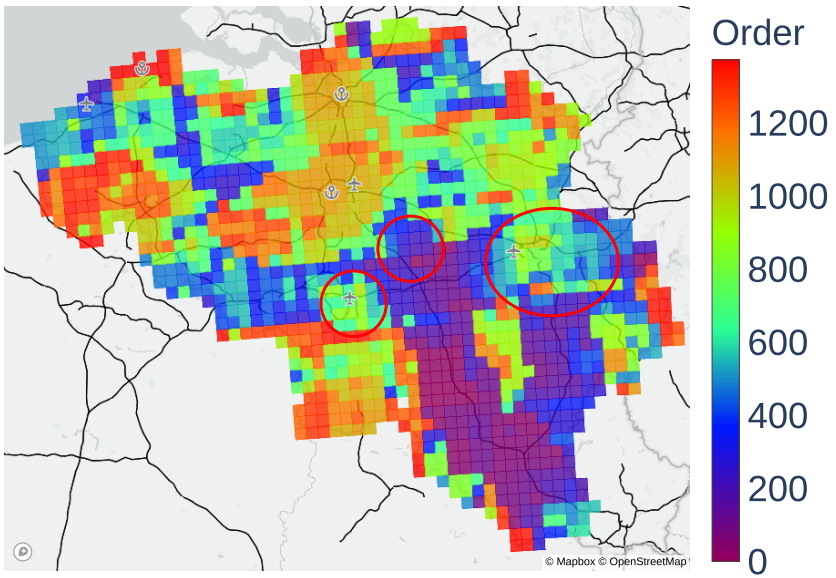
Three zones that match the structure of the country

1. Purple and blue : mainly driving
2. Orange-red: Proportion of work and congestion much higher
3. Green-yellow: in between

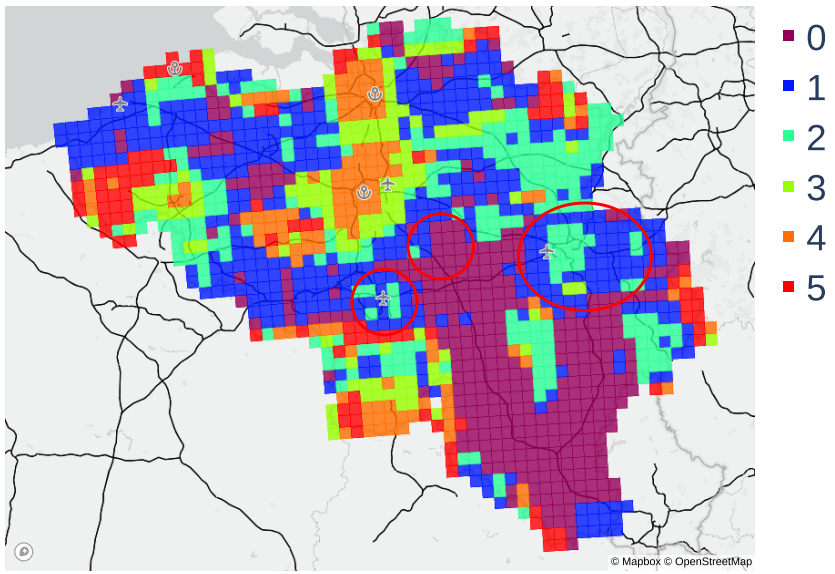
Comparison

Comparison of our method and Self-Organizing Maps

- Overall the same structure
- Some areas are, however, less distinguishable
- No interpretation of the clusters in terms of features



Our method



Self-Organizing Maps

Conclusion & Future work

The propose framework allows to quickly see

- The geographical structure given the features
- How this structure translates in term of features

Conclusion & Future work

The propose framework allows to quickly see

- The geographical structure given the features
- How this structure translates in term of features

In the future, we could

- Integrate geographical informations in the seriation process
- Add time elements in the visualization