Analytical methodology for assessing similarities/differences between Ports

Renaud Di Francesco, Françoise Bahoken

Analytical methodology for assessing similarities/differences between Ports

Renaud Di Francesco[1], PhD; Prof. Françoise Bahoken[2], PhD.
1. Fellow of the IET, IAME Member;
2. Université Gustave Eiffel/AME-Splott
rdifrancesco@ymail.com

Abstract: The analysis of similarities and differences between ports at the global level can help the Maritime ecosystem for different purposes, including business planning, operation, and regulation. Our methodology is demonstrated in a particular case of interest as the "short sea shipping" traffic through a harbour. However, no structural limit to the scope of adaptation is introduced. The new methodology proposed for Ports builds on a recent framework of Matrices of Sets, where the computational efficiency of matrices is combined with the versatility and flexibility of sets to capture quantitative and qualitative information. Origin Destination (OD) Matrices of Sets (odMS) are cross tables where the value M_{ij} placed at the intersection of a row (i) and a column (j) is a set of OD. The odMS framework aims to exploit the computational efficiency of OD matrices by combining it with the flexibility of sets. It suits qualitative and quantitative analysis of entities described. The case study is a European port network defined by several characteristics forming an odMS. Then, on an operational point of view, odMS method addresses the need for port authorities to benchmark and compare themselves with each other, with an aim to increase their competitiveness and to identify and characterise their competitors or possible partners in port alliances (estimate complementarities and compatibility within such alliances).

Key-words: Ports, Europe, similarities, OD matrices, matrices of sets, Jaccard index.

Résumé : L'analyse des similarités et des différences entre les ports au niveau mondial peut aider les autorités en charge du secteur maritime de différentes manières (planification des échanges commerciaux, de l'exploitation et de la réglementation). L'intérêt de notre approche des similarités a été précédemment démontrée sur un cas particulier, le trafic de "short sea shipping" observé en un port. Aucune limite structurelle à sa portée et à son adaptation n'ayant été introduite, la nouvelle version que nous proposons s'appuie sur un cadre récent de matrices d'ensembles, où l'efficacité de calcul des matrices est combinée avec la polyvalence et la flexibilité des ensembles, pour capturer des informations quantitatives et qualitatives caractéristiques des ports. Les Matrices d'Ensembles Origine Destination (OD) (odMS) sont des tableaux croisés où la valeur Mij placée à l'intersection d'une ligne (i) et d'une colonne (j) est un ensemble d'OD. Le cadre odMS vise à exploiter l'efficacité de calcul des matrices OD en la combinant avec la flexibilité des ensembles. Il convient donc à l'analyse qualitative et quantitative des entités décrites. L'étude de cas est présentée est un réseau portuaire européen défini par plusieurs caractéristiques formant un odMS. Ensuite, d'un point de vue opérationnel, la méthode odMS répond au besoin des autorités portuaires de se comparer entre elles, dans le but d'accroître leur compétitivité, d'identifier et de caractériser leurs concurrents ou leurs partenaires éventuels dans le cadre d'alliances portuaires (estimer les complémentarités et la compatibilité au sein de ces alliances).

Mots-clés : Ports, Europe, similarités, matrices OD, sets de matrices, indice Jaccard.

1. Introduction

The analysis of the differences between places (here port entities) at the global level is generally carried out either from the point of view of the analysis of the networks of actors as Social Network Analysis (SNA), within the framework of the theory of graphs, or from the point of view of the comparison of indicators of specialization of places within the framework of descriptive statistical analysis of the locations (Mandják T. et al, 2018; Rodrigue J-P, 2020; Ducruet et al., 2011/2022).

Our proposal falls in this second theoretical framework. The general idea is to propose a summary of a set of data available on the same couples of geographical entities. To do this, we select from a set of matrices only their margins, in order to focus our analysis from the point of view of the entities, thus of the places (and no longer of their relations). The aim is to analyze the spatial differentiations of places that result from their interrelationships.

Our analysis mobilizes statistical data on port traffic which will also be apprehended from the point of view of their positional parameters (quartile, etc.), at different dates. For this purpose, we will consider the different positional parameters of these time series and consider initially only what generalizes the notion of amplitude (A) of the series, considering the MIN and the MAX taking into account a "minimum encompassing Rectangle". We thus use A: [MAX; MIN], i.e. the smallest interval or "encompassing Rectangle" in which the series fits.

The comparison and benchmarking of Ports may have multiple motivations and objectives. One of them is the logistics of European short sea shipping, on which we will focus as a starting use case.

We will proceed with an overview of public data made available for the European Region by Eurostat, and we will continue with modelling tools of interest for comparing maritime routes, sea basins, countries, and harbours. We will recall the recent framework of Matrices of Sets, which are tables with line index i and column index j, in which the element at position (i,j) is taken to be $M_{i,j}$, a set. Such a matrix of sets can be seen as formed of column vectors of sets. After giving a self-contained description of possible use and applications of this framework, we will come more specifically to the case where sets are intervals of the set of real numbers (or Cartesian products of such intervals). We will recall definitions and properties of Jaccard similarity measures on sets, and a simplification of such a similarity measure will be developed for a computationally efficient comparison of maritime countries and harbours, with detailed hands-on results for the case of short-sea shipping.

The Jaccard index, initially introduced to compare discrete sets, is the division of the cardinal (number of elements) of the intersection of two sets, by the cardinal of the union of these two sets. We adapt this notion slightly for intervals and Cartesian products of intervals, as the length of the intersection divided by the length of the union, with demonstrated efficiency when comparing cities using the monthly temperature [min,max] interval (Di Francesco, 2022)

The case study is the network of European ports on which a simplified version of Jaccard's similarity measure will be applied. We will show the efficiency of this measure on short distance sea shipping. The comparison will be carried out at different geographical scales: harbours, countries, sea basins. The response will be tested for robustness against the so-called Modifiable areal unit problem (MAUP) of Openshaw (1984) distinguishing ports, port hinterlands (or basins) and European maritime countries.

2. Short-sea-shipping: an illustrative use case

Let us start by introducing short-sea shipping, where and why it may matter. Maritime transport usually distinguishes between deep-sea shipping, typically on intercontinental routes, and short sea shipping, for the maritime transport of goods over shorter distances, with inclusion of feeder services when they are used.

Short sea shipping serves and connects deep sea hubs, extending the capillarity of the maritime reach. Further on multimodal transport extends through inland waterways, railways, and road transport.

The well-known harbour of Rotterdam connects to a short sea network of more than 200 destinations from Russia to the Mediterranean, mainly from Waal and Eemhaven and soon from the former ECT City Terminal. Such localized grouping allows for optimized transhipment, with multimodal connection to extend the reach and penetration. Feeder services ensure transhipment from scheduled intercontinental freight transport to most locations in Europe from Rotterdam, with high granularity in time and size through short-sea-shipping. Italy is at the top of the league for short-sea-shipping in the European Union. In other regions, one can name, as another example (Konstantinus et al, 2019), short-sea shipping within the Southern African Development Community of 16 countries, providing sustainable economic growth to the area. Let us also refer to a detailed study of short-sea shipping for Cyprus and the Eastern Mediterranean (Michaelides et al, 2019) which has looked at ship type, port of origin, and shipping agent, to calculate Key Performance Indicators (KPIs) relating to predictability of timings and resource allocation. Another study uses data analytics to track liquid gas utilization in short-sea shipping (González-Cancelas et al, 2019).

Overall, the examples above make a strong case for trying to develop a robust and efficient methodology for comparing basins, countries, harbours under different aspects of short-sea-shipping.

3. Setting the stage for the framework

3.1. Origin Destination Matrices of Sets (odMS)

Origin Destination (OD) Matrices of Sets (odMS) are an ensemble of cross tables where the value M_{ij} placed at the intersection of a row (i) and a column (j) is a set of OD. We look at the ports from the point of view of the margins of the matrix: on the total incoming (Dj) and outgoing (Oi). We are interested in the gross flows or volume (Vi) per port, defined as follows: (Vi)=Oi+Dj

These gross flows correspond to the "Gross weight of seaborne freight transported to/from main ports by type of shipping" during 2006-2020 in (million tonnes), from Eurostats datasets.

Such OD Matrices have proven computational efficiency and can help organize and manage data access. Sets are flexible and can aggregate and describe diversely acquired data, overcoming the barrier of heterogeneity, and the need to regularize and normalize required by the classical vector processing, and matrix-based linear models.

3.2. Matrices of Sets (reminder of key definitions and properties)

Matrices of sets (Di Francesco et al., 2021) combine the computational efficiency and programmability of matrices with the set flexibility to track events and features at the early stage of their exploration, typically.

A matrix of set is a data arrangement, similar to classical matrices in using rows and columns to index the objects it stores and encapsulates at row i column j, but different since the object is not a number $m_{i,j}$ but a set $M_{i,j}$.

The framework has been developed using products of matrix form on such matrices of sets M and M′, with a generic form of:

$M'' = M \circ M'$

defined by its term

$M''_{i,j} = \bigvee_k M_{i,k}$ ž $M'_{k,j}$

where large operator $\bigvee$ emulates the sum operator used in classical matrices, and ž emulates the product of numbers in classical matrices.

The case where $\bigvee$ is the set reunion $\bigcup$ and ž is the Cartesian product of sets x has been studied in some detail (Di Francesco, Dec. 2021), establishing interesting spectral properties (generalized eigenvectors and eigenvalues) for these objects, in the absence of the classical and comforting linear space structure.

In a fruitful collaboration aimed at applications in the short food supply chains, ž was taken as the set intersection operator $\cap$ leading to a new comparison method generalizing the Hamming distance. $\bigvee$ can be taken for such objective as the set reunion $\bigcup$ or as the disjoint set union $\sqcup$ which brings it closer to the Hamming distance method: every occurrence found of the same element is added as if it were different (tagged for difference), equivalently one could define

$\bigsqcup_k A_k = \bigcup_k A_k \times \{k\}$.

This method for aggregating similarities into sets instead of just measuring them can be compared with the Jaccard index to compare sets A and B of which one at least is not empty, by computing the quotient of the cardinal of their intersection divided by the cardinal of their reunion.

3.3. Particularly convenient Matrices of sets: matrices of intervals of the real line

Being aware that Cantor had developed set theory to address problems in number theory, a venture into the Master's territory at the crossroads of numbers and sets seemed natural. Hence the case where sets $M_{i,j}$ are intervals of the set of real numbers was considered. In this instance, vectors of sets were formed with real intervals as sets. A first application proved successful, because open data was available for testing and computing. The yearly records (or average of these) of monthly minimum and maximum temperature in cities of the world were used (Di Francesco, 2022).

Data was of the form [T_{min}(month, city), T_{max}(month, city)], therefore allowing to compare month by month two cities City_1 and City_2 using the following generalized Jaccard similarity measure:

J(month, City_1, City_2) = (Length ([T_{min}(month, City_1), T_{max}(month, City_1)] Ç [T_{min}(month, City_2), T_{max}(month, City_2)])) / (Length ([T_{min}(month, City_1), T_{max}(month, City_1)] È[T_{min}(month, City_2), T_{max}(month, City_2)]))

In short, we divide the length of the intersection by the length of the reunion (the latter being supposedly non-zero).

This corresponds to the Jaccard index, in the case of uniformly distributed data.

One could modify this similarity measure by imposing a non-linear weight on the interval.

Now, in computational terms the above intervals intersect if and only if the smallest of their maximums is not inferior to the largest of their minimums.

Even a very rough programming resource such as a spreadsheet can handle the computation:

-let us simplify the notations with the two intervals to compare being

I= [m, M]

I'= [m', M']

We then obtain

if min (M,M') < max(m,m')

then

J(I,I') = 0

else

J(I,I') = (min(M,M') - max(m,m'))/(max(M,M') - min(m,m'))

This comparison tool is well suited to filter all cases where the two objects compared have "no common operation mode" (empty intersection) and focuses on close objects.

3.4. Comparing vectors of sets: the interval similarity proposed and the original Jaccard similarity

Let us consider the case where observations are subsets of the intervals I and I', with I resp. minimal interval containing the observations.

The actual observation sets are denoted Ω and Ω', respectively included in I and I'.

Then

Ω ∩ Ω' is included in I ∩ I'

Suppose that we have N observations in each Ω and Ω', then comparing the two is reduced to computing Card (Ω ∩ Ω') ≤ Card (I ∩ I').

We see that the original index and the interval similarity measure index are different notions. However, the interval similarity builds on the acquisition of minimum and maximum values, followed by computationally light processing. In short, the interval as a minimum container of data observed (in the family of intervals) captures summary information on the data, with reduced computational complexity.

4. Short sea shipping: similarity quantification at several levels

4.1. Eurostats datasets

For Europe, a free and easily accessible data repository is available from Eurostat. The gathering of it had been initially motivated by the need to orient Maritime and Multimodal Transport at Policy Level, following the Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions – "The Development of Short Sea

Shipping in Europe: A dynamic alternative in a sustainable transport chain - Second two-yearly progress report", COM/99/0317 final.

We mobilize two sets of digital data, statistical and geographical. Statistical dataset is from Eurostat (2022) which provides a substantial data set on short sea shipping available, and easy to use. This data set covers ports situated in geographical Europe, on the Mediterranean and Black Seas on the other hand, with ports in: EU maritime countries; EEA maritime countries (Iceland and Norway); candidate countries; the Baltic Sea area (Russia) ; the Mediterranean Sea area (Algeria, Bosnia and Herzegovina, Egypt, Israel, Lebanon, Libya, Morocco, Occupied Palestinian territory, Syria, and Tunisia) ; the Black Sea area (Georgia, Moldova, Russia, and Ukraine). Geographical dataset is from Eurostat (2022) two, especially from GISCO, the Geographic Information System of the commission (GIS). See Figure 1.

Figure 1. Location of European ports and those of interest

In this statistical and geographical data collection, a dataset of practical interest is the gross weight of goods transported yearly, for years ranging from 2005 to 2020. Such time series is given for countries, harbours, and sea basins. We extract from the time series a summary formed of intervals of minimum and maximum value for traffic over the time ranging from 2005 to 2020.

4.2. Compare the similarities with the Jaccard index

The Jaccard index allows for the comparison of port types by evaluating their degree of similarity based on different characteristics. This index, which varies between 0 and 1, makes it possible to estimate the diversity of port infrastructures location observed in an area (here in the neighbourhood of Europe). An index close to 0 means that there are few common features between the ports being compared, that is to say there are few ports of a given type. Conversely when the Jaccard index is close to 1, a significant number of identical features are present in the two ports being compared, suggesting that these couple of port are similar.

The method thus makes it possible to create a typology of ports (from the characteristics contained in the odMS) that are more or less similar.

The Jaccard similarity can be easily computed, when comparing location A with [min, max] interval denoted by I(A), and location B with associated interval I(B):

J(I(A),I(B)) = 0 if I(A) and I(B) have an empty intersection.

Else, one can easily verify that

J(I(A),I(B)) = (min(max(I(A)),max(I(B))) – max(min(I(A)),min(I(B)))) / (max(max(I(A)),max(I(B))) - min(min(I(A),min(I(B)))

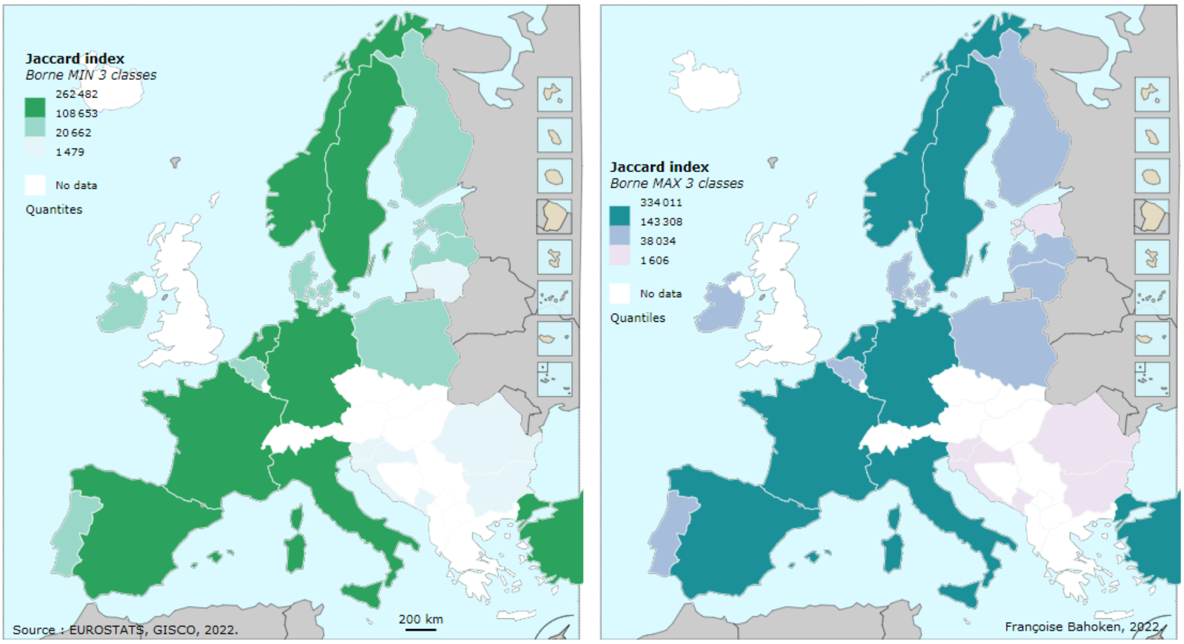
Details can be found above in the section describing “matrices of intervals of the real line”.

4.3. Short sea shipping models visualised

The following maps show the similarities (Jaccard Index) at country level and then at main port level. They represent respectively the Jaccard similarity MIN and MAX at both scales, discretized into three classes according to the tertiles.

Figure 2 presents the MIN, MAX at a national level and shows how they are identical in terms of ranking (not on value).

Figure 2. Jaccard similarity at country level (MIN MAX)



As the MIN and MAX maps are identical in terms of ranking (no value), we finally present a representation of the center of this class [MIN ; MAX].

Figure 3. Jaccard similarity at country level (CENTER)

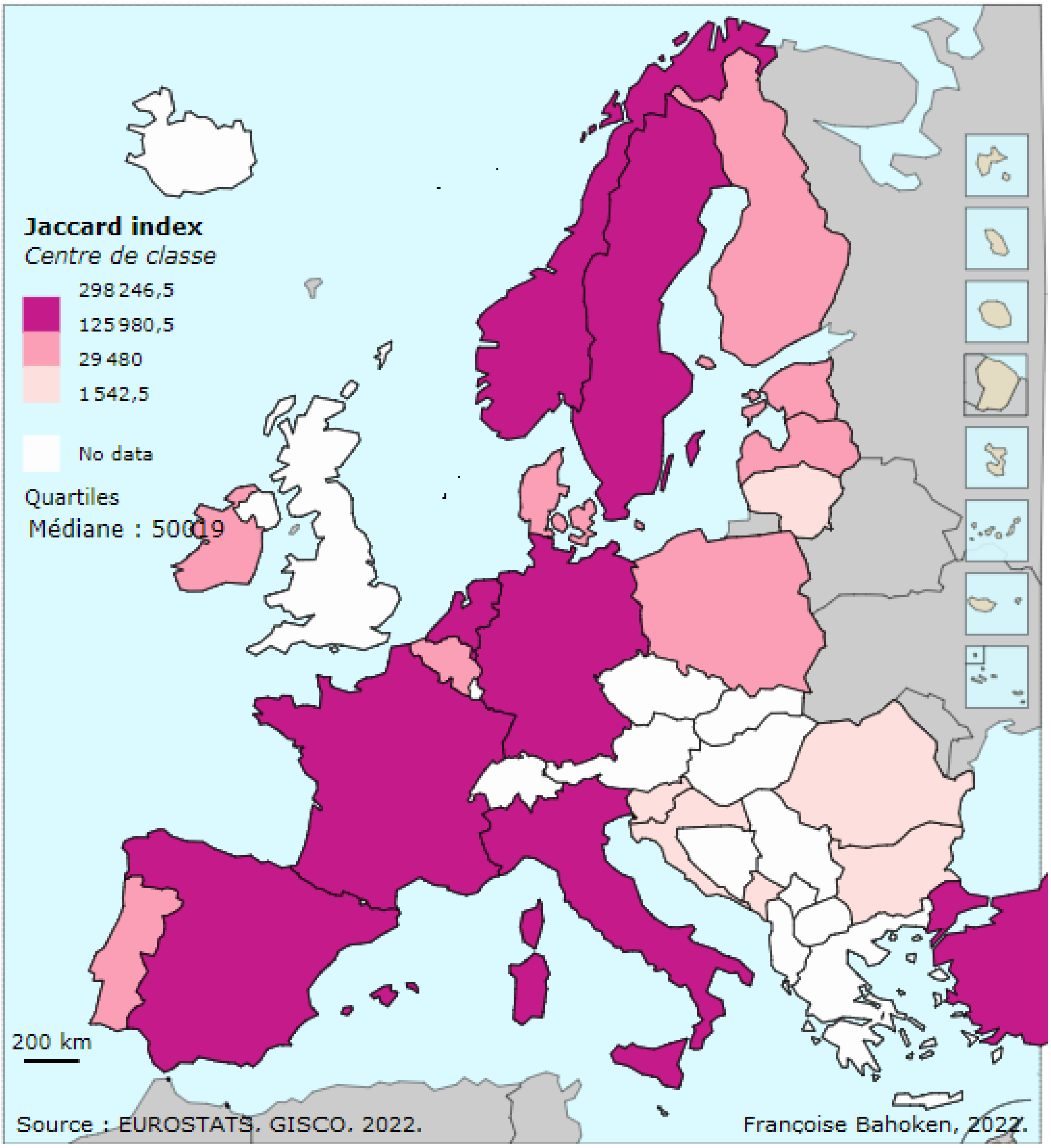


Figure 3 describes the gross weight of good transport from/to main ports at country level. It clearly shows the old countries in the heart of Western Europe, which are grouped together in the upper class. The less similar countries are located at the margins of Europe and in its neighbourhood.

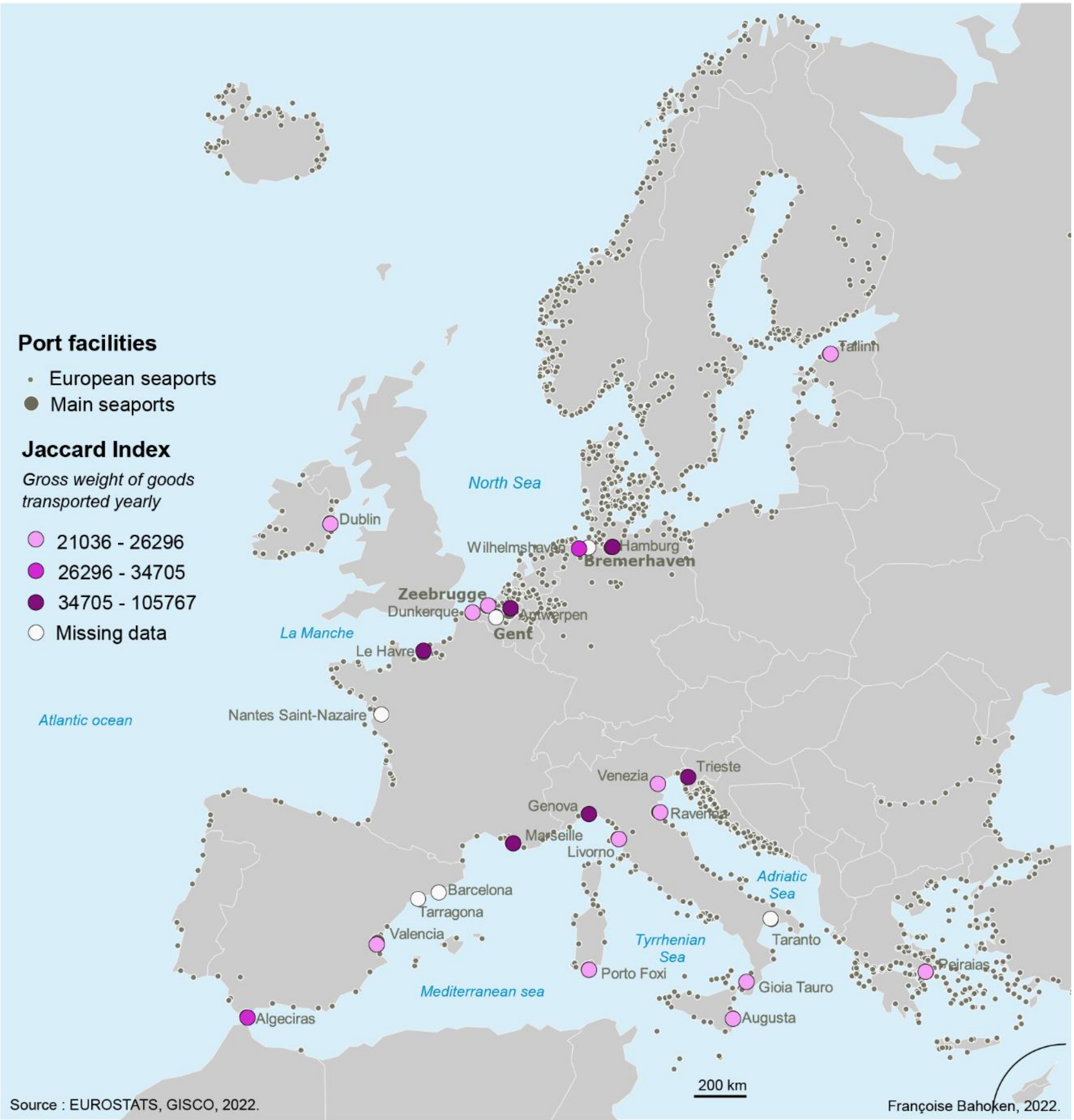
The map thus shows that the similarity between ports decreases along a gradient from the heart of Western Europe to its periphery. It should be noted that these results are identical with a discretization containing a larger number of classes.

They are however likely to change at the scale of the main ports (see Figure 3).

If the maps are thus comparable two by two scale (in terms of country/port gross weight positions) they are also logically submitted to the Modified Areal Unit problem (MAUP) for areal entities and to the ecological error for punctual and area entities, within the framework of a cross-scalar analysis perspective. That is because each of the countries' or the harbours' scales corresponds to a real geographical analysis level.

As previously, Figure 3 describe the gross weight of good transport from/to main Europeans ports.

Figure 4. Jaccard similarity applied to individual harbours



5. Conclusion and Future Opportunities

The analysis presented was based on a subset of the port matrix series corresponding to the amplitude-based minimum congestion rectangle with A: [MAX; MIN].

It is important to note that since the analysis is based on the positional parameters of the series, other intervals could also have been considered.

On the one hand, other subsets or discretised subspaces containing the values of the series of matrices such as: [Min, Q2], [Q2, Max] for example, or even the interquartile range: [Q3-Q1].

On the other hand, by considering all the series and not only the subsets discretised according to the quartiles, e.g.: [(Mi,Q1), (Q1,Q2),(Q3,3), (Q3, Max)].

Finally, as the analysis presented on temporal port data is reproducible, it could be interesting to reproduce it on EUROSTAT airport or even multimodal data (air, rail, road, river and sea), to be able to compare the similarity of transport modes on the European territory.

References

Mandják Tibor, Lavissiere Alexandre, Hofmann Julian, Bouchery Yann, Lavissière (Dillon) Mary Catherine, Faury Olivier, Sohier Romain (2018), Port marketing from a multidisciplinary perspective: A systematic literature review and lexicometric analysis, Transport Policy, 2018/12/05, Vol. 84, https://doi.org/10.1016/j.tranpol.2018.11.011

Rodrigues Jean-Paul (2020), The Geography of Transport Systems, FIFTH EDITION, New York: Routledge, 456 pages. ISBN 978-0-367-36463-2. https://doi.org/10.4324/9780429346323

César Ducruet, Sung-Woo Lee and Adolf Koi Yu Ng, « Port Competition and Network Polarization in the East Asian Maritime Corridor », Territoire en mouvement Revue de géographie et aménagement [Online], 10 | 2011, Online since 01 June 2013, connection on 30 June 2022]. http://journals.openedition.org/tem/1327

DOI https://doi.org/10.4000/tem.1327

Di Francesco R., Boukachour J., El Yaagoubi A., Charhbili M. IAME-2021-Containers transport and logistics models with Matrices of Sets, Conference: IAME 2021, Rotterdam, July 2021: researchgate access

Di Francesco R., Di Francesco-Isart P., Maritime Economics Computable Models using Matrices of Sets: study cases of 1) economics of routing for multimodal transport, 2) expression of preference across heterogeneous dynamic baskets Applications illustrating the efficiency of economic modelling with Matrices of Sets Conference: International Association of Maritime Economists (IAME) 2021 Conference, Rotterdam, July 2021: researchgate-access

Di Francesco, R., Scalable Computations on Large Numbers of Heterogeneous Sets -supporting use categories of graph structured data, transactional features for economic agents, logical propositions combined, with Matrices of Sets

December 2021. https://doi.org/10.13140/RG.2.2.12495.20643/1

Di Francesco, R., Geographic matrices of set: -comparing cities based on their monthly temperature intervals

January 2022. https://doi.org/10.13140/RG.2.2.13512.65283

Ducruet, C. 2006, Port-city relationships in Europe and Asia, Journal of International Logistics and Trade 4(2), p. 13-35.

Konstantinus, A.; Zuidgeest, M.; Christodoulou, A.; Raza, Z.; Woxenius, J. Barriers and Enablers for Short Sea Shipping in the Southern African Development Community. Sustainability 2019, 11, 1532. https://doi.org/10.3390/su11061532

Merk, O. (ed.), The Competitiveness of Global Port-Cities: Synthesis Report, OECD Port-Cities Programme,

Michaelides, M.P.; Herodotou, H.; Lind, M.; Watson, R.T. Port-2-Port Communication Enhancing Short Sea Shipping Performance: The Case Study of Cyprus and the Eastern Mediterranean. Sustainability 2019, 11, 1912. https://doi.org/10.3390/su11071912

Nicoleta González-Cancelas, Beatriz Molina Serrano, Francisco Soler-Flores; Seaport Sustainable: Use of Artificial Intelligence to Evaluate Liquid Natural Gas Utilization in Short Sea Shipping. Transportation Journal 1 July 2019; 58 (3): 197–221. doi: https://doi.org/10.5325/transportationj.58.3.0197

Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions - The Development of Short Sea Shipping in Europe: A dynamic alternative in a sustainable transport chain - Second two-yearly progress report

/* COM/99/0317 final */

https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A51999DC0317

Openshaw S., 1984, The modifiable areal unit problem, Norwick: Geo Books

Openshaw S. and Taylor P.J., 1979, A Million or so Correlation Coefficient: Three Experiments on the Modifiable Areal Unit Problem, In: Wrigley N. (ed.), *Statistical Applications in the Spatial Sciences*, 127-144, London, Pion.

Tobler, W. 1989. Frame Independent Spatial Analysis, In: Goodchild M. F. and Gopal S. (eds). *Accuracy of Spatial Databases*, 115–122. London: Taylor and Francis.

Eurostat data repository for short-sea shipping:

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Maritime_transport_statistics_-_short_sea_shipping_of_goods

Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions - The Development of Short Sea Shipping in Europe: A dynamic alternative in a sustainable transport chain - Second two-yearly progress report /* COM/99/0317 final */

https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A51999DC0317

Eurostat data repository for short-sea shipping:

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Maritime_transport_statistics_-_short_sea_shipping_of_goods