Detecting the spatiotemporal characteristics of the supply-chain disruption and estimating its short term effects

Giannis Spiliopoulos*
Intelligent Transportation Systems Lab, University of the Aegean
Ermoupolis, Syros, Greece
gspiliopoulos@aegean.gr

César Ducruet*
French National Center for Scientific Research (CNRS), UMR 7235
EconomiX, University of Paris Nanterre, France
cesar.ducruet@economix.fr

Leonardo M. Millefiori*
NATO STO Centre for Maritime Research and Experimentation (CMRE)
La Spezia, Italy
leonardo.millefiori@cmre.nato.int

Paolo Braca*
NATO STO Centre for Maritime Research and Experimentation (CMRE)
La Spezia, Italy
paolo.braca@cmre.nato.int

Dimitris Zissis†*
Intelligent Transportation Systems Lab, University of the Aegean
Ermoupolis, Syros, Greece
dzissis@aegean.gr

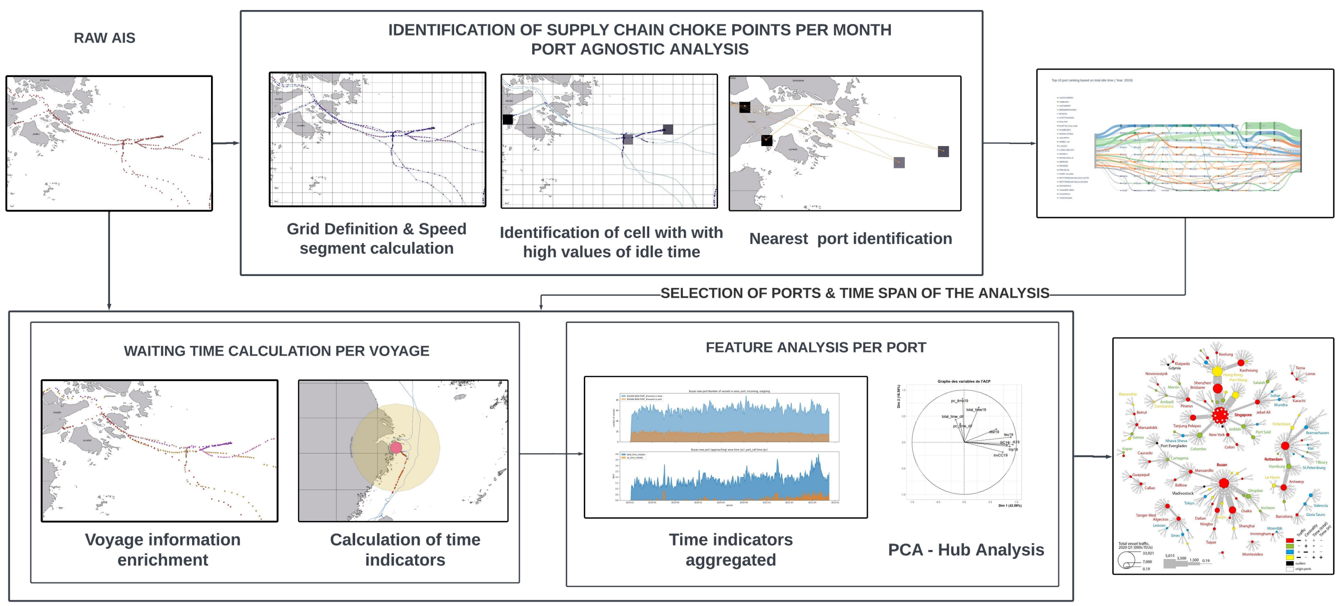


Figure 1: Visual representation of the data transformation flow.

ABSTRACT

Ships are often considered as the backbone of the global economy. A fundamental unresolved problem is how to best operate fleets, given a sudden increase in demand, such as that reported following the first months of the COVID19 pandemic. Advancing our knowledge of the supply chain's delicate equilibrium between demand and supply, requires analyzing huge amounts of ship-related positional data, thus revealing which areas should be avoided due to potential congestion buildup and where cargoes should be rerouted to. Herein, we analyze a large-scale high-resolution mobility data set of more than 7,000 container ships, collected over an extended period of 36 months, covering the entire globe, so as to measure quantitatively the effects of the pandemic post-hoc on the supply chain. To further understand these fine-grained mobility patterns, we introduce a mobility model for calculating ship presence times (or waiting times) at a global scale. We then reveal the congestion points, which strongly correlate with port waiting areas and anchorages. We analyse the data to reveal how times at port areas were affected by the rising number of ships waiting to load or unload cargo. Following this, we transform this data into a 'port to port' graph mapping the international flow of containerised trade. We

*These authors contributed equally to this work
†Corresponding author

 Published in the Proceedings of the Workshops of the EDBT/ICDT 2024 Joint Conference (March 25-28, 2024), Paestum, Italy.

apply methods of graph theory, complex networks, and multivariate statistics to unravel the hidden relationships between global maritime structure and ship time distribution. This analysis is novel in respect to the size of the data analyzed, the algorithmic approach and the impact of the results which reveal some affinities between pre-COVID and post-COVID shipping patterns.

1 INTRODUCTION

The supply chain is a complex network involving suppliers, trucks, distribution centers, warehouses, logistics centers and vectors, working together in concert to deliver goods to the customers. Today, this complex network is distributed all around the world. With more than 80% of global trade by volume and up to 70% of its value being carried by the shipping industry, vessels can be seen as the backbone of the global supply chain. Thus, shipping can be viewed as a barometer for the global economic climate, and any reduction in activity is expected to have a cascading effect on the global economy. On the other hand, ports are located at the nexus of the supply chain and connect global and regional actors.

The COVID-19 pandemic and containment measures had a significant impact on global maritime traffic in the short term. Several studies have attempted to quantify these effects and assess the size of this reduction. However, after the initial shock of the pandemic, when customer demand rebounded, vessels stacked up high with containers could be seen anchored outside many of the world's major ports.

Starting from the idea that shipping traffic data can be used to assess the effects of restrictive measures on the global supply chain, in this work we focus on revealing the major delay points in the maritime global supply network, as well as attempting to answer the question if these correlate with major port locations. We also explore if congestion at ports is simply related to a sudden increase in ship traffic and evidently quantify the increase in delay times. Along these lines, we attempt to reveal how the pandemic has affected ports over an extended period of time and attempt to reveal the trajectory of recovery.

The main technical challenge is that of a big-data mining task of transforming huge amounts of geospatial data–as collected from vessels using the Automatic Identification System (AIS)–into a descriptive and compact data model, that can be used for identifying the underlying relationships or patterns. In our case, the patterns are those of normal port to port traffic connections. Our approach relies on data transformations and distributed raster-based analytics as a first step to reduce the size of the data, followed by graph analysis to reveal the hidden patterns in the data.

The main methodological contribution of this work is showing that mobility data can be processed to shed light on the temporal and spatial characteristics of the supply chain as a network at the global scale. To the best of our knowledge, no major study has attempted to analyze such a large dataset with this aim before.

1.1 Related work and contribution

Over the last years, there has been an exponential growth of scientific publications related to maritime traffic analysis involving big data analytics and/or novel Artificial Intelligence (AI) techniques. For instance, the analysis of vessel mobility data to understand the hidden patterns is intrinsic to trajectory data mining, and the seminal work in [28] covers the field in detail. This growth has been boosted also by the availability of data from large sensor networks, such as the AIS, which has provided researchers with enormous volumes of information for the study of maritime transportation and the maritime industry in general. According to the authors of [10], one of the main scientific topics discussed in the maritime literature is indeed the applications of big data techniques to AIS. Recently, mostly due to the COVID-19 pandemic, but also to other events, such as the Suez canal blockage in 2021 and the Red Sea crisis in 2024, the maritime transportation has been often disrupted [19, 23] and several studies were conducted to measure effects of this disruption [15, 17]. The strong academic interest to study large mobility data at scale combined with necessity for fast decision making during the pandemic has accelerated advancements in the field [29].

In this work, we use big data analytics and graph analysis to better understand the disruption in the supply chain. The utilisation of big data in AIS analysis is an area of research that has received a lot of attention [3, 24, 30] recently. For instance, in [22], authors evaluate the performance of clustering algorithms for route modelling on a full year global AIS dataset, and in [16], the KDE technique is adapted to map-reduce paradigm to compute seaports' extended areas of operations from AIS data. Other examples are [20], where an image analysis on density maps to detect traffic flows is introduced, as well as [27], in which authors introduce trip semantic objects and the use of density based clustering to identify clusters of way-points and stops. In a complementary to this work approach, authors in [4] build voyage graph feature time-series (VGT) to study their evolution from a time-series perspective. In this work, we quantify and allocate to ports the effects of each vessel slow-down in range by introducing the waiting and approaching time indicators. We study their evolution over time to understand which ports are affected the most.

Graph theory, and its widely known extension known as complex networks, can be employed to characterize (port) nodes by providing a hierarchy of centrality/accessibility in the container shipping network. Early applications provided global-network measures of connectivity [11, 12, 26], as well as a cartography of degree or betweenness centrality [9]. Research on the relationship between centrality and other port performance indicators remains relatively scarce in the literature [13, 14, 25], usually confirming the strong correlation between degree centrality (i.e., number of connections to other ports) and weighted degree (i.e., total traffic in twenty-foot equivalent units [TEUs]). Thus, in the present study, we innovate by applying a statistical analysis of traffic, centrality, and time indicators. It complements the work of Ducruet and Itoh [8] on the statistical relationships at stake between ship time, port centrality, and port traffic by focusing on a specific event and its supply chain consequences. A recent review of the field [7] also showed that within a corpus of 212 papers about shipping networks published between 2007 and 2022, nearly 20% concerned the topics of crisis and vulnerability, i.e., the second largest category after "network structure".

2 METHODOLOGY

2.1 Automatic Identification System

AIS was originally designed as a collision avoidance system for ships. Since 2002, the International Maritime Organisation (IMO) has made compulsory for all vessels with a tonnage including and above 300 gross tonnage to be equipped with a class-A AIS transceiver. At its core, each AIS transceiver sends and receives positional reports (i.e., types 1, 2, 3 and 18) every few seconds via VHF. The messages contain information about each vessel's identity, location, course and speed. Since 2006, the lower-power (and lower-cost) class-B transceiver was introduced, allowing also smaller vessels to use the AIS, even if with lower performance and priority than commercial fleets, which operate strictly on class-A transceivers [21]. The transmission rate of AIS ranges from 2 seconds, for fast moving vessels or maneuvering vessels equipped with a class-A transponder, up to 3 minutes for anchored or moored vessels.

For this study, we make use of AIS positional reports of container ships traveling across the globe for the years 2019, 2020 and 2021.

2.2 Data transformation

To unravel the hidden information of global supply chain performance from raw AIS messages, we employ a multi-step sequential data mining process. Our main goal is understanding if supply chain disruption is measurable and correlated with port activities. Then, we also investigate what are the intrinsic characteristics of these ports, to understand if they can be possibly used in a predictive fashion as indicators of future disruptions, so that fleets can be rerouted suitably. In our case study, the effects of a disruption in global supply chain are not known beforehand. We introduce a process (Fig. 1) to infer a global supply chain network graph from AIS mobility data. Then, we apply advanced graph analytics to identify port typologies and changes over its connections to measure effects of disruptions over a three-year period.

Data cleaning & conditioning The first step in the process is a cleaning task to ensure that records comply with protocol standards and reject records with missing values. Then, we apply a geo-fencing technique to select records located within port areas, and exclude them from the identification of waiting areas part of the analysis. To facilitate the numerical calculations, all positional reports are re-projected into the Web Mercator (EPSG:3857) coordinate system.

Raster-based analysis. To identify waiting areas from AIS messages, we use a raster-based analysis. We first define the raster characteristics, such as the shape and size of its cells. For the analysis performed in this work, the raster consisted of square-shaped cells of a 9.7 km side length, each one of them covering an area of approximately 100 km^2 on average with respect to the projection systems' distortion. Then, we assign the AIS messages to the grid cells, by splitting each trajectory into segments that match the grid definition (i.e., each segment is allocated to exactly one cell and it is annotated with the cell's id). Each segment consists of either two consecutive AIS messages or a grid intersection point and an AIS message,where the location and timestamp of the intersection point are interpolated assuming constant speed. Then, we calculate the average speed required for a vessel to cover the distance of each segment. We annotate as idle any segments whose average speed is less than 2 knots, and finally we sum up the total time of idle segments for each cell and month. High values of idle times are indicators of choke points for shipping traffic, and the computation of idle time rasters on a monthly basis allows us to characterize how the distribution changes over the three-year period considered. Anticipating the results, we observe that high idle time cells typically appear near major container ports and canals.

Connecting waiting areas with ports. To further investigate this behavior and explicitly connect cells of high cumulative idle time with ports, we performed a nearest neighbor analysis [2] to assign each cell to its nearest port. Then, if the cell is located within a 100 km range of any top-50 ports [1](in terms of annual reported volume [1]) the cell is reassigned to its closest top-50 port. Again, anticipating a bit the results of our analysis, we notice an increase of cumulative idle time around major ports both in terms of cumulative values and number of cells where this happens.

Measuring in-port and approaching time. The previous steps leave us with the congestion epicenters near major ports, where vessels wait to enter the port. The epicenters are located, in most cases, near the ports. However, as congestion increases, the waiting areas expand vastly following different patterns with respect to topology and other local characteristics. It is also possible that a non-negligible number of vessels sheltered themselves in these areas and never entered the closest major port. To confirm or reject this hypothesis, we used accurate information about the end of each itinerary. The AIS protocol supports messages that include information about the destination port, but unfortunately it cannot be considered as a reliable source of information, as it is manually entered by the crew, without following a specific standard, making it thus extremely prone to errors. To tackle this problem, we performed a retrospective analysis on the data to identify the ports of origin and destination for each trip, and we calculated the exact time of approaching 300 km to destination, as well as the exact time of each vessel entering the port across all itineraries that reach any of the top-50 ports. Then, we calculate for each itinerary the total time spent within a 300 km radius and the time spent in the port. The 300 km radius ensures that we account for any intentional or unintentional delay that may occur for any vessel before it reaches its final destination. This radius is selected so that all waiting areas of the first part of our analysis are included. The time in port reflects the operational time of a vessel calling a port and it captures all time required to moor at berth and perform all kinds of loading and unloading operations and exit the port.

Defining the waiting time network. Maritime flows can be modeled as a graphical structure $\mathcal{G}$, where the ports (v) are the nodes (or vertices), which are connected by inter-port connections (e) as links (or edges), so that $\mathcal{G} = (v, e)$ [6]. The connections among ports are in general not known a priori and can change over time, but in principle they can be learned by inspection of the AIS data by looking at vessels navigating from one port to another. Each vertex v in the graph $\mathcal{G}$ stores static features and summary statistics of the port's traffic flow strength that corresponds to the graph weights and they are calculated on quarterly basis. The static features are the

[1]We consider the top-50 ports to be representative of the whole port system in their proximity.

port identifier and country each port belongs to, while the summary statistics measure the number of vessels calling the port, as well as their aggregated maximum capacity. Each link $e \in \mathcal{G}$ consists of the pair of ports identifiers it connects, as well the number of voyages on this connection and their cumulative maximum TEU capacity and aggregated time indicators. Those statistics are also calculated in correspondence to nodes on a quarterly basis.

Graph analysis To assess the level of disruption on port connections, we rely on quarterly created summary statistics for nodes (v) from the previous step, and we calculate the differences for the total and the in port time between the last quarter (Q4) of 2019 and the first quarter (Q1) of 2020. We apply linear transformations to summary statistics to define supply chain port characteristics such as the number of vessels calls (frequency) and total vessel traffic (frequency $\times$ vessel capacity in TEU). We complement our dataset with graph-theoretical indices calculated for all network nodes and both quarters, namely the degree centrality (number of shipping links), betweenness centrality (number of occurrences on shortest paths in the graph), and inverse clustering coefficient (local hub power). The average port time in Q4 2019 will also be used as a pre-existing characteristic. We apply a Principal Components Analysis (PCA) to all nodes to the (Q4) of 2019 quarter and quarterly calculated residuals to reveal the hidden trends at stake in the shipping traffic data. PCA is a statistical method serving to unravel a limited number of unobserved (latent) variables among a set of observed, correlated variables [5]. Such unobserved variables, often called principal components or "factors", constitute the basis of a clustering analysis that will distribute observations (port nodes) among distinct groups. The clusters, i.e. groups of ports of similar operational behaviour (see bottom right legend in Figure 4), are then confronted to initial variables to best describe their trends and characteristics. The next step is to illustrate the typology by means of a single linkage analysis. This method serves to highlight the main hubs and their "nodal regions" by keeping only the largest traffic flow link of each port in the graph. Finally, a multiple regression looks at the determinants of time evolution based on port characteristics.

3 RESULTS

Areas where ships "wait" evidently depict problematic spots in the supply chain. In order to reveal the spatiotemporal characteristics of delay areas globally, we first define and quantify areas where ships are idle for long periods, as these can be an indicator of disruptions in the supply and demand balance [17]. Our analysis focused on first understanding if these areas overlapped with specific port areas and then to further understand if these had specific characteristics.

As a first result, which may have been expected, we confirmed that waiting areas are close to port locations. All top-30 locations are within 80 nautical miles range from ports, and we can assume reasonably that ships stationing in these areas are waiting to enter the port. This result is in line with reports and papers reporting the increase turn around in port areas [18].

In Fig.2, we illustrate a Sankey Diagram ranking ports according to the measured total idle time over the last time period of the analysis, where ports with longer waiting times are located at the top of the diagram. Interestingly, we can see the fluctuation in the positions, with ports switching their ranking throughout the year.

To understand the characteristics of the ports and areas of delay, we move onto the second part of the analysis, which makes use of graph theory and PCA.

3.1 Single and multivariate linkage analysis

We apply PCA to two distinct datasets: 1) based on static port characteristics in Q4 2019; and 2) based on absolute changes of these characteristics between Q4 2019 and Q1 2020. Both datasets include time evolution as a means of checking its affinity with other variables, which are its potential determinants. The two PCAs provided interesting results, with 72.1% of variance contained in the three first components for static variables (with eigenvalues > 1), and 76.0% for the first four components for dynamic variables (with eigenvalues > 1). Figure 3 represents the distribution of variables along the two first components for each dataset (left, static; right, dynamic). Interestingly, worsening time is opposed to traffic and centrality level/growth in the two figures. This is even truer for dynamics, where calls (trip_dif) and traffic (teu_dif) are more directly opposed to port time evolution. Another difference between the two PCAs is the opposition between connectivity changes and traffic changes along the second component (vertical axis) for dynamics. It means that although growing ports in general witnessed reduced port time, those with growing connectivity tended to increase port time, contrary to ports increasing traffic. Lastly, the static analysis shows that ports with longer times in Q4 2019 were also the ones increasing port time in Q1 2020 (vertical axis).

The situation of each port in the observed trends is revealed by means of a hierarchical clustering analysis, which is applied to the main components of each PCA, to produce a typology. This is combined with a single linkage analysis, to test whether the obtained types have a specific position in the network's backbone.

The dynamics-based typology provides the picture of world ports reported in Fig. 4. It considers absolute changes of port characteristics as for the second PCA, and the single linkage analysis is based on Q1 2020. The most impacted category (yellow) is marked by drastic traffic decline, slight reduction of centrality, and the strongest increase of total and in-port time. It includes a vast majority of gateway ports (Le Havre, Constanta, Koper, Alexandria, Liverpool, Felixstowe, Zeebrugge, Fos, Los Angeles, Long Beach, Tianjin, Lianyungang, Kobe, Ho Chi Minh, Manila) as well as Hong Kong and Port Klang. Except from the latter two ports, these gateways have, in general, a limited role in the architecture of nodal regions, due to their specialization in import/export cargoes. Another category has lost similar amounts of traffic on average (red) but such ports slightly increased their centrality. Several of them are large hub ports polarizing their respective nodal region, the largest being, like in the previous figure, Singapore, Busan, and Rotterdam. Like for the other categories (green, blue), these ports experienced a slight increase of total port time and small decrease of in-port time. While they also lost traffic, the secondary hubs (green) gained enormous centrality in Q1 2020, contrary to what we can call second-tier hubs (blue), which have the opposite profile. There is no apparent geographic or functional logic in those two categories, which are disseminated across regions and contain both

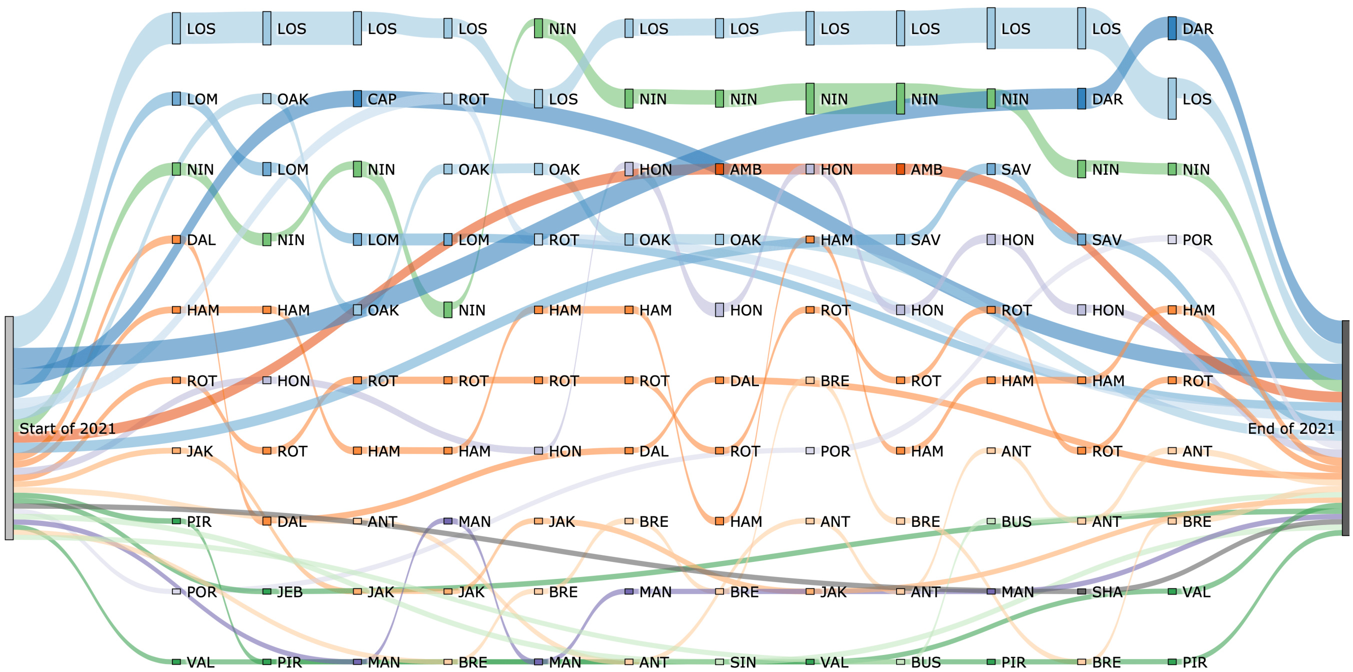
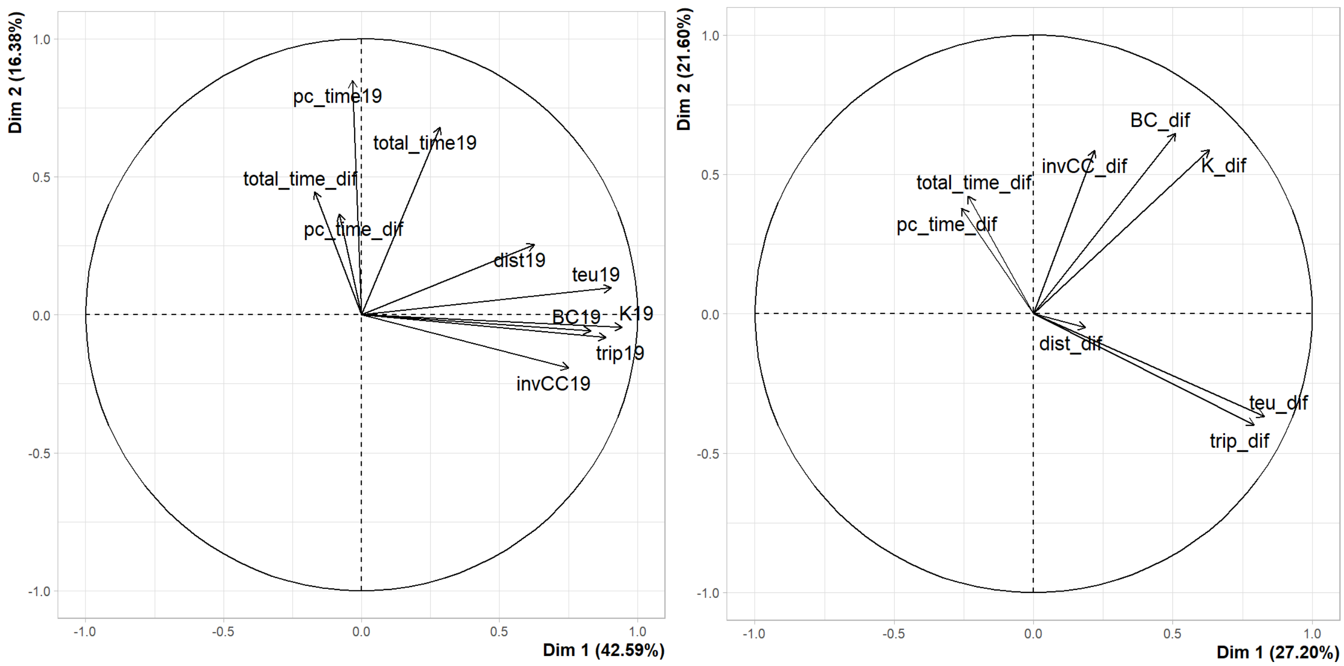


Figure 2: Visual representation in terms of total idle time for the top-10 ports of each month across 2021 year. Line width represents the value of total idle time. AMB:Ambarli, ANT:Antwerp, BRE:BremerHaven, BUS:Busan, CAP:Cap Town, DAL:Dalian, DAR:Dar ES Salaam, HAM:Hamburg, HON:Hong Kong, JAK:Jakarta, JEB:JEBEL ALI, LOM:LOME, LOS:Los Angeles, MAN: Manzanillo, NIG: Nigbo, OAK:Oakland, PIR: Piraeus, POR: Port Lang, Rot: Rotterdam (blue: Waalhaven, orange:Maasvlakte), SAV: Savannah, SHA: Shanghai, SIN: Singapore, VAL: Valencia.

gateway ports and transshipment ports. The loss of centrality (blue) is, still, relatively common to European ports while the increase of centrality (green) is better found in Asia.

3.1.1 What determined port time changes in 2020? A multiple regression analysis is applied in two steps, each being a model focusing on a distinct independent variable: total port time difference (model 1) and in-port time difference (model 2), as shown in Table 1. As a matter of fact, among the selected dependent variables, only two have a statistically significant effect. It is the case of in-port time in model 1 (0.05 significant), which increased total port time difference between Q4 2019 and Q1 2020. The other case is the regional dummy Africa in model 2, which increased in-port time difference. Despite the low significance of other dependent variables, some of them may be discussed according to the direction of their effect on time evolution. Among the ones that deserve attention, inverse clustering coefficient stands out as it has the same, negative effect on time evolution, and is near-to-significant in both models. It means that ports ensuring stronger hub functions before the crisis have witnessed lesser congestion and, even, more fluid cargo transfers. Such a result is in line with the single linkage analyses, showing that pivotal hubs ensure and maintain their domination towards other ports, often within a certain geographic radius (nodal regions). It also confirms the work of Ducruet and Itoh [8] about the determinants of ship turnaround times on a long period. Although this measure is very much correlated with port traffic and other centrality indicators (cf. Principal Component Analysis (PCAs)), it expresses a specific dimension of port connectivity, namely the ability to polarize neighboring, or adjacently connected, ports.

A counter-intuitive result is the negative effect of city size (population) on time evolution in both models, as the inclusion of this variable was meant to test the role of potential congestion played by cities on port operations, in terms of lack of space and density. This can be explained by our focus on the top of the port hierarchy, where most ports are in fact major metropolitan areas. Another commonality between the two models is the negative influence of total port time and the positive influence of in-port time. In-port time is a component of total port time, but it better represents the core activity of the port, as it is the closest to the length taken by terminal operations. This crucial component of the whole transport chain, if prolonged, will inevitably have strong consequences on the rest of the chain, as seen with its positive impact on total time difference (slowdown, queuing), which includes the water vicinity of the port (e.g. port entrance, access channel). Thus, ports with an already high in-time (turnaround time) have witnessed worsening operations (prolonged times) during the COVID-19 crisis. At the contrary, total port time had a negative effect on time evolution in both models. While such a result may seem to contradict the former, it should be understood in the light of other port variables in each model. In model 1, the negative effect of total port time on total port time difference goes along with a negative effect of port size (calls, TEUs), meaning that large, busy ports in Q4 2019 (but



(a) (b)

Figure 3: Principal component analyses

Table 1: Determinants of port time evolution.

	Model 1 *total time difference*				**Model 2** *in-port time difference*			
	Estimate	Std. err.	t-value	$\Pr(>\|t\|)$	Estimate	Std. err.	t-value	$\Pr(>\|t\|)$
(Intercept)	1.409 00	1.201 00	1.173	0.2440	−0.720 11	0.524 60	−1.373	0.1735
Population	−0.024 69	0.035 61	−0.694	0.4899	−0.008 62	0.015 55	−0.554	0.5809
Calls	−0.093 89	0.232 51	−0.404	0.6874	0.164 42	0.101 56	1.619	0.1092
TEUs	−0.127 66	0.141 70	−0.901	0.3702	0.024 27	0.061 89	0.392	0.6960
Distance	−0.037 87	0.088 34	−0.429	0.6692	0.038 61	0.038 59	1.000	0.3200
Betweenness Centrality	0.023 93	0.079 31	0.302	0.7636	0.010 52	0.034 64	0.304	0.7621
Degree Centrality	0.460 94	0.332 03	1.388	0.1687	−0.238 50	0.145 03	−1.645	0.1038
I.C.C.†	−0.048 19	0.031 99	−1.506	0.1357	−0.016 58	0.013 97	−1.186	0.2388
Total time	−0.167 66	0.135 24	−1.240	0.2185	−0.083 76	0.059 07	−1.418	0.1599
In-port time	0.581 33	0.263 06	2.210	0.0298*	0.084 71	0.114 91	0.737	0.4630
Africa	−0.025 91	0.699 50	−0.037	0.9705	0.525 12	0.305 55	1.719	0.0893
Americas	0.180 97	0.664 12	0.273	0.7859	0.147 99	0.290 09	0.510	0.6113
Asia	0.405 55	0.665 50	0.609	0.5439	0.194 31	0.290 69	0.668	0.5057
Europe	0.087 54	0.662 63	0.132	0.8952	0.264 30	0.289 44	0.913	0.3638

† Inverse clustering coefficient

not necessarily the most central, as seen with the positive influence of betweenness and degree) had more chance to perform better in times of crisis. In model 2, the same negative effect is associated with a positive influence of port size (worsening in-port time), and a negative influence of degree centrality (numerous connections). This means that hub ports with a relatively lower size managed to improve their core operations in the advent of the crisis.

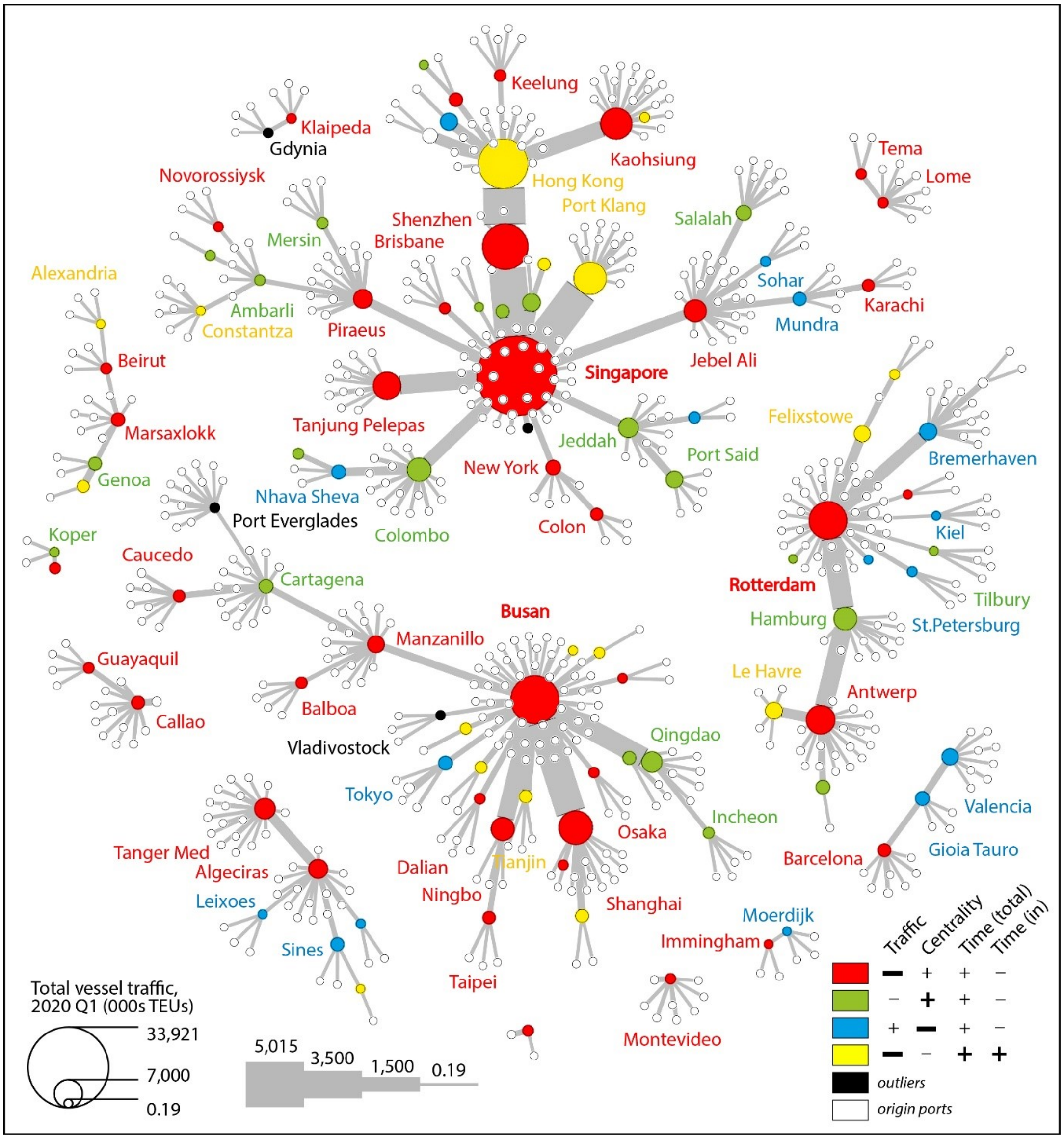


Figure 4: Single linkage analysis and port typology – Q1 2020 and port/time evolution

4 CONCLUSION

Based on geospatial data mining approaches, in this paper we presented a methodology to determine the misbalances in the global supply and demand equilibrium as captured through ship movements. The presented approach is capable, firstly, of detecting and defining the areas of potential delays, which in most cases overlap with the main port areas, and secondly to determine the specific characteristics of these ports. Our approach relies on methods of raster based analysis, graph theory and complex networks analysis. Future work will be focused on applying additional methods from the field of graph analysis and complex networks to similar datasets.

ACKNOWLEDGMENTS

This work was partially supported by MarineTraffic. This research is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 101092749, project Critical Action Planning over Extreme-Scale Data (CREXDATA). The work of L. M. Millefiori and P. Braca is supported by NATO Allied Command Transformation (ACT) via project "Data Knowledge Operational Effectiveness" (DKOE). The contribution of César Ducruet is supported by the French National Research Agency (ANR) through the research project No. ANR-22-CE22-0002 "Maritime Globalization, Network Externalities, and Transport Impacs on Cities" (MAGNETICS).

REFERENCES

[1] World Shipping Council. 2020. The Top 50 Container Ports. https://www.worldshipping.org/top-50-ports Accessed: 19-Jan-2024.

[2] Jon Louis Bentley. 1975. Multidimensional Binary Search Trees Used for Associative Searching. *Commun. ACM* 18, 9 (sep 1975), 509–517.

[3] Samuele Capobianco, Leonardo M. Millefiori, Nicola Forti, Paolo Braca, and Peter Willett. 2021. Deep Learning Methods for Vessel Trajectory Prediction Based on Recurrent Neural Networks. *IEEE Trans. Aerospace Electron. Systems* 57, 6 (2021), 4329–4346.

[4] Emanuele Carlini, Vinicius Monteiro de Lira, Amilcar Soares, Mohammad Etemad, Bruno Brandoli, and Stan Matwin. 2021. Understanding evolution of maritime networks from automatic identification system data. *GeoInformatica* 26 (2021), 1–25.

[5] Dennis Child. 2006. *The essentials of factor analysis*. A&C Black.

[6] Pasquale Coscia, Paolo Braca, Leonardo M. Millefiori, Francesco A. N. Palmieri, and Peter Willett. 2018. Multiple Ornstein–Uhlenbeck Processes for Maritime Traffic Graph Representation. *IEEE Trans. Aerospace Electron. Systems* 54, 5 (2018), 2158–2170.

[7] César Ducruet. 2023. *Shipping network analysis: State-of-the-art and application to the global financial crisis*. Routledge. 300 – 333 pages.

[8] César Ducruet and Hidekazu Itoh. 2022. Spatial network analysis of container port operations: the case of ship turnaround times. *Networks and Spatial Economics* (July 2022).

[9] César Ducruet and Theo Notteboom. 2012. The worldwide maritime network of container shipping: Spatial structure and regional dynamics. *Global Networks* 12, 3 (2012), 395 – 423.

[10] Ziaul Haque Munim et al. 2020. Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. *Maritime Policy & Management* 47, 5 (2020), 577–597.

[11] Yihong Hu and Daoli Zhu. 2009. Empirical analysis of the worldwide maritime transportation network. *Physica A: Statistical Mechanics and its Applications* 388, 10 (2009), 2061 – 2071.

[12] Pablo Kaluza, Andrea Kölzsch, Michael T. Gastner, and Bernd Blasius. 2010. The complex network of global cargo ship movements. *Journal of the Royal Society Interface* 7, 48 (2010), 1093 – 1103.

[13] Dong-Joon Kang and Su-Han Woo. 2017. Liner shipping networks, port characteristics and the impact on port performance. *Maritime Economics & Logistics* 19 (2017), 274–295.

[14] Arom Kim and Jing Lu. 2015. A study on the effects of network centrality and efficiency on the throughput of Korean and Chinese container ports. In *ICTE 2015*. 760–769.

[15] David March, Kristian Metcalfe, Joaquin Tintoré, and Brendan J Godley. 2021. Tracking the global reduction of marine traffic during the COVID-19 pandemic. *Nature communications* 12, 1 (2021), 2415.

[16] L. Millefiori, D. Zissis, L. Cazzanti, and G. Arcieri. 2016. A distributed approach to estimating sea port operational regions from lots of AIS data. In *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 1627–1632.

[17] Leonardo M. Millefiori, Paolo Braca, Dimitris Zissis, Giannis Spiliopoulos, Stefano Marano, Peter K. Willett, and Sandro Carniel. 2021. COVID-19 impact on global maritime mobility. *Scientific Reports* 11 (sep 2021), 18039.

[18] Theo Notteboom and Thanos Pallis. 2020. *IAPH-WPSP Port Economic Impact Barometer*. Technical Report. International Association of Ports and Harbors (IAPH).

[19] Richard Partington. 2024. What is the Red Sea crisis, and what does it mean for global trade? The Guardian, Jan 3rd 2024 https://www.theguardian.com/world/2024/jan/03/ what-is-the-red-sea-crisis-and-what-does-it-mean -for-global-trade,. Accessed: 19-Jan-2024.

[20] Roeland Scheepens, Christophe Hurter, Huub Van De Wetering, and Jarke J Van Wijk. 2015. Visualization, selection, and analysis of traffic flows. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 379–388.

[21] Solas. 2002. Solas Chapter V. Regulation 19 - Carriage requirements for shipborne navigational systems and equipment. https://assets.publishing.service.gov.uk/media/5a7f0081ed915d74e33f3c6e/solas_v_on_safety_of_navigation.pdf. [Accessed: 19-Jan-2024].

[22] Giannis Spiliopoulos, Konstantinos Chatzikokolakis, Dimitrios Zissis, Evmorfia Biliri, Dimitrios Papaspyros, Giannis Tsapelas, and Spyros Mouzakitis. 2017. Knowledge extraction from maritime spatiotemporal data: An evaluation of clustering algorithms on Big Data. In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 1682–1687.

[23] United Nations Conference on Trade and Development (UNCTAD). 2023. Review of Maritime Transport. https://unctad.org/publication/review-maritime-transport-2023.

[24] Xiaoguang Wang, Xuan Liu, Bo Liu, Erico N de Souza, and Stan Matwin. 2014. Vessel route anomaly detection with Hadoop MapReduce. In *2014 IEEE international conference on big data (big data)*. IEEE, 25–30.

[25] Yuhong Wang and Kevin Cullinane. 2016. Determinants of port centrality in maritime container transportation. *Transportation Research Part E: Logistics and Transportation Review* 95 (2016), 326–340.

[26] Deng Wei-Bing, Guo Long, Li Wei, and Cai Xu. 2009. Worldwide Marine Transportation Network: Efficiency and Container Throughput. *Chinese Physics Letters* 26, 11 (nov 2009), 118901.

[27] Zhaojin Yan, Yijia Xiao, Liang Cheng, Rong He, Xiaoguang Ruan, Xiao Zhou, Manchun Li, and Ran Bin. 2020. Exploring AIS data for intelligent maritime routes extraction. *Applied Ocean Research* 101 (2020), 102271.

[28] Yu Zheng. 2015. Trajectory Data Mining: An Overview. *ACM Trans. Intell. Syst. Technol.* 6, 3, Article 29 (may 2015), 41 pages.

[29] Chenghu Zhou, Fenzhen Su, Tao Pei, An Zhang, Yunyan Du, Bin Luo, Zhidong Cao, Juanle Wang, Wen Yuan, Yunqiang Zhu, Ci Song, Jie Chen, Jun Xu, Fujia Li, Ting Ma, Lili Jiang, Fengqin Yan, Jiawei Yi, Yunfeng Hu, Yilan Liao, and Han Xiao. 2020. COVID-19: Challenges to GIS with Big Data. *Geography and Sustainability* 1, 1 (2020), 77 – 87.

[30] D. Zissis, K. Chatzikokolakis, G. Spiliopoulos, and M. Vodas. 2020. A Distributed Spatial Method for Modeling Maritime Routes. *IEEE Access* 8 (2020), 47556–47568.