

Digital Transformation in the Shipping Industry: a Network-Based Bibliometric Analysis

Abstract

This paper presents a network-based bibliometric analysis of digital transformation in the shipping industry, a sector undergoing rapid change due to advancements in automation, AI, blockchain, and IoT. The study synthesizes existing knowledge to identify trends, challenges, and opportunities for industry stakeholders and researchers. Unlike previous literature reviews, this work adopts a graph theory approach applied to a large dataset of scientific publications, without predefined technological or industrial sub-domains.

Data was collected from EBSCO, ProQuest, and IEEE eXplore, then refined using OpenAlex to comprise 2293 scientific publications. The analysis includes descriptive statistics, co-authorship network analysis, co-citation network analysis, and thematic analysis.

The findings reveal a significant increase in publications since 2005, with exponential growth after 2015. A small percentage of authors and institutions account for a disproportionate share of publications, suggesting a skewed distribution of research efforts. The co-authorship network exhibits a heavy-tail distribution and interconnected communities, indicating extensive national and international collaboration. The co-citation analysis identifies key research areas such as fuel consumption optimization, safety and risk management, and smart port development. Thematic analysis highlights the growing importance of AI and cybersecurity.

Keywords: digital transformation, shipping industry, bibliometric analysis, network theory

1. Introduction

Digital transformation in the shipping and maritime industries is of strategic importance. It is shaped mainly by the rapid developments in digital technologies and the increasing demands for enhanced operational efficiency, sus-

tainability and competitiveness across the global maritime supply chain. The integration of innovative technologies in the maritime sector, including artificial intelligence (AI), machine learning (ML), blockchain, the Internet of Things (IoT), Big Data analytic, digital twins, autonomous vessels, and enhanced cybersecurity measures, has facilitated significant advancements in maritime logistics, port operations, and vessel management (Tijan et al., 2021; An, 2024). Maritime digital transformation aims to support decision-making in maritime operational processes, streamline fleet management tasks, enhance safety and environmental performance and reduce vessel environmental impact.

Despite the increasing adoption of innovative technologies, the maritime sector remains a complex and fragmented domain. There are various levels of digital maturity across different maritime sub-sectors. Previous research has examined diverse aspects of maritime digital transformation, ranging from industry-wide systematic reviews to bibliometric analyses that highlight specific technological trends or research gaps. However, prior studies primarily focus on predefined technological domains or specific industrial sub-sectors, which limits their findings' scope and generalizes their research conclusions. Additionally, the methodological frameworks employed across existing studies demonstrate substantial variation, encompassing systematic literature reviews (SLRs), conceptual papers, or bibliometric network analyses. Even though each contributes uniquely to the maritime sector research, sometimes it provides isolated findings to the broader transformation of the maritime landscape (Lau et al., 2024; Oğuz, 2024).

To bridge this research gap, the present study employs a network-based bibliometric analysis approach to provide a holistic perspective on the digital transformation of the shipping industry. Unlike many prior studies that categorize research into predefined technological areas, this study takes an unbiased, data-driven approach to map the broader research maritime landscape. Though a data-driven approach, the study uses graph-theoretical methods and natural language processing (NLP) to examine research trends systematically. Specifically, co-authorship and co-citation networks are analyzed using community detection algorithms and thematic clustering techniques to identify influential research clusters, collaboration patterns, and emerging topics in the field. Fi-

nally, analyzing an extensive dataset of scientific publications, this study identifies dominant research clusters, major contributing institutions, and key technological themes such as fuel efficiency, risk management, AI, and cybersecurity—providing a data-driven understanding of maritime digital transformation trends.

The study makes three key contributions. First, it conducts a detailed bibliometric analysis of the digital transformation in the maritime industry. Second, it applies network science to define structural patterns in research collaborations and citation networks. Finally, it employs machine learning-based thematic clustering to uncover emerging research directions, offering insights into future advancements in digital shipping technologies. Through this multi-dimensional approach, the study provides a robust foundation for academia, industry professionals, and policymakers to support data-driven decisions to navigate the evolving landscape of digital transformation in the maritime sector.

2. Literature Review

Previous work has investigated the state-of-the-art of digital transformation in shipping and maritime industries. Such research varied both in methodologies, domain of applications, and limitations. In this section, we offer an overview of the most relevant published reviews, compared to our study.

In 2019, (Sanchez-Gonzalez et al., 2019) published a systematic literature review to cover the advances in digitizing the maritime transportation. The authors started from a large set of articles (2900): however, being an SLRs they then applied stringent filters to reduce the initial set to a far smaller one (only 99 studies). It focuses on eight specific digital domains and three industry sub-sectors. Both domains and sub-sectors were defined as input to the analysis. The authors could identify interesting research trends. However, the selection of digital domains as a-prior potentially limits the extension of their results.

In (Poulis et al., 2020), the authors review the effects of digital transformation in a well-defined sub-domain of the maritime industry, namely that of unmanned vessels. The authors describe their work more as a conceptual paper rather than a proper SLR. In addition, their conclusions are limited to the specific sub-domain they have chosen for their analysis. Finally, covering a sub-

domain which is still being discovered, the research is by definition limited to qualitative analysis.

In the same year, a bibliometric study was published by (Munim et al., 2020). In their work, the authors applied network-based measures to investigate co-citations among 279 articles. Their bibliographic coupling analysis highlighted four key research clusters within the domains of big data and artificial intelligence. The article was published before the explosion of generative AI, hence their recommendations are mostly over legacy machine learning algorithms. This study is similar to ours when it comes to methodology. However, the graph analysis is applied to a limited set of pre-defined technologies.

Another interesting work is the SLR published by (Tijan et al., 2021). In their work, the authors set up to identify the major barriers, drivers, and successful factors for digital transformation in the maritime transport. The work is of particular interest because the authors are not focusing on pre-defined technological or industry domains. Instead, their effort remains open to any possible conclusion. Their analysis highlighted the influence of blockchain and autonomous shipping technologies, and paved the way to the development of effective strategies to support successful digital transformation in the field. Being a literature review, the authors curated a limited set of 139 articles from which they drew their findings and conclusions.

In (Jović et al., 2022) the authors present a bibliometric analysis on the digitization in both maritime transport and seaports. Being a bibliometric analysis, their methodology is similar to the one we applied in our work. Specifically, the authors apply thematic analysis to explore the state-of-the-art within the field of research, and identify key areas and gaps. Starting from over 8000 articles, the authors filtered them down to 280 for the subsequent analysis. Such filtering was done manually, hence leading to some extensive human effort. This, on one side, guarantees a curated final set of articles. However, it also potentially limits the extent of the results. In addition, as any other filtering technique, it may include biases in the process. Moreover, the authors recognized some of the limitations related to the tools they used for conceptualization of topics. The work was published before the impressive development of state-of-the-art transformer neural networks, which means the authors could not leverage modern tools such

as context-sensitive embedding systems.

Another bibliometric study was published by (Xiao et al., 2024). Similarly to what previously mentioned for other works, this work focuses on pre-defined technological domain, specifically on the application of artificial intelligence in shipping. The work is based on a relatively larger set of articles, 476, and the methodology is similar to the one applied in our work, including the identification of influential papers, countries, and institutions. The main difference with our work is in the limited technological domain being investigated.

A similar study is the bibliometric one by (Xiao et al., 2025). Once again, the methodology in place reflects the one we adopted as well. The number of examined publication was 201, and the focus was given to the digital transformation as a mean for decarbonizing the shipping industry.

To conclude, we refer to the work of (Filippopoulos et al., 2022). Despite not being an SLR nor a bibliometric study, this work is of interest because it presents a road-map for digital transformation as it is applied in a real case scenario. While building their road-map, the authors analyze the state-of-the-art of digital transformation in shipping industry, suggesting relevant strategies for its successful implementation.

In comparison to the works cited in this section, our work is rooted in graph theory, applied to a large cohort of scientific publications, and not limited by pre-defined technological or industrial sub-domains.

3. Research Methodology

In this section we describe the methodology we followed for the data collection and analysis. Figure 1 shows the overall methodology discussed in this section, while results and implications are discussed in the upcoming sections.

[Figure 1 about here.]

3.1. Keyword identification and data collection

We asked experts in the shipping industry to identify the most relevant keywords related to the industry itself and to digital technologies and digital transformation. Their analysis resulted in 35 keywords, listed in Table 1.

[Table 1 about here.]

Data was collected from three research engines: EBSCO (Vaughan, 2011), ProQuest (Cooke, 2017), and IEEE eXplore (Wilde, 2016). The search was performed on October the 22nd 2024. For each engine, we retrieved scientific articles containing any of the digital transformation related keywords and any of the shipping industry related keywords, in either their title or abstract. The exact query for each engine are available on request. We limited our results using the following criteria: *a* only English literature, and *b* only scientific contributions published in peer-reviewed journals. Table 2 shows the results.

[Table 2 about here.]

All search engines provided the digital object identifier for the articles. This allowed us to screen the resulting set and identify 2324 unique articles for the subsequent analysis. One challenge of using different data engines is the variety of attributes they return for each article. In order to have the same information for each article, we queried a fourth search engine for all the 2324 articles. We chose OpenAlex (Priem et al., 2022), which has been shown to be suitable for bibliometric analysis (Alperin et al., 2024). Our final result set comprised 2293 scientific publications.

3.2. Descriptive statistics

We started our analysis evaluating descriptive statistics across our article set. More specifically, we calculated:

1. the distribution of the number of publications per year;
2. the distribution of publications across authors, identifying the most prolific authors;
3. the distribution of publications across institutions, identifying the research centers with the highest number of publications;
4. the distribution of publications across countries.

3.3. Co-authorship network analysis

As a second step, we built and analyzed the network of co-authorship. Network analysis was performed in Python, using the NetworkX package (Hagberg et al., 2008). We identified 7723 distinct authors. We built the network using authors as nodes, and setting bi-directional links between them if there existed

at least one publication that they co-authored. For each link, we stored within the graph object information about the authors institutes and countries for further analysis.

To determine which distribution best fit the data, we run statistical tests comparing the likelihood of power-law distribution against the exponential distribution, the log-normal distribution, and the truncated power-law distribution.

Next, we focused on the largest connected component of the network, made of 883 authors and 2753 links between them. The choice of focusing on the largest component was dictated mostly by computational limitations.

Working on the largest component, we applied the Louvain community (Blondel et al., 2008) algorithm to identify the major communities of authors and investigated the distribution of institutions and countries across communities.

To conclude, we analyzed the network for small-world behavior. More specifically, we calculated both the clustering coefficient and the average path length and compared them to random networks of equivalent size.

3.4. Co-citation network analysis

We built a co-citation network of nodes (i.e., articles) and links (i.e. co-citation between two articles). The resulting graph had 1298 nodes. The degree distribution was tested for power-law characteristics against other plausible distributions (exponential, log-normal, and truncated power-law).

Next, we identified the most influential articles (i.e., the top 10 in terms of received citations). Our goal was to check if the most cited articles were literature reviews. As presented in the following section, this turned out not to be the case, allowing us to draw relevant considerations over the demand of SLRs at the conjunction of digital transformation and shipping industry.

We then moved our attention to the top 20% cited papers and analyzed their topics. To achieve this, we create a sub-network using only the top 20% cited papers and applied the Louvain community algorithm (Blondel et al., 2008). Next, for each community collected the titles and applied natural language processing (NLP) to model their topics (BERTTopic (Paul et al.)).

To conclude, we applied different centrality measures to the top 20% graph to identify the 5 most relevant articles. These were analyzed more in details in

terms of covered research area, as a preliminary trend analysis, further developed in our next and last analysis section.

3.5. Thematic analysis

Working on the entire set of articles (2290) we performed a thematic analysis to identify the major topic of research. We pre-processed the titles with the following steps:

1. lemmatization to transform words into their root forms;
2. removal of stop-words;
3. removal of non alpha-numeric text.

Next, we applied tokenization and embedded each title using BERT (Devlin et al., 2018). The resulting vectors were analyzed for unsupervised clustering. More specifically, we adopted two methods to identify the ideal number of clusters: the Calinski-Harabasz index (Caliński and Harabasz, 1974), and the Davies-Bouldin index (Davies and Bouldin, 1979). Having identified the best number of clusters, we applied the unsupervised K-means algorithm and calculated the centroid for each cluster. Next, we identified for each cluster the 10 articles closest to the corresponding centroid and applied BERTopic to extract the common themes.

We concluded our thematic analysis by building two word clouds. Using both titles and abstracts from all articles, we applied the TF-IDF algorithm to each word and use it as weight when building the clouds. The first cloud was built over the entire set of words in titles and abstracts, while the second cloud was built after removing all shipping related terms (hence focusing on the digital technologies only).

4. Results

In this section we present the results of our analysis. We then discuss them in the next section.

4.1. Descriptive statistics

Figure 2 shows the distribution of articles across years. Although the first publications are dates as back as the 1960s, only from the year 2005 we see an increasing interest in the effects of digital transformation within the shipping

and maritime industry. The number of publication increased minimally and not steadily between 2005 and 2015. From 2015 onward, we see an exponential increase in the number of publications. After reaching a peak in 2023, the trend seem to have stabilized. Considering that our data was collected at the end of October 2024, we can reasonably argue that the year 2024 has not witnessed a significant increase of publication, compared to the previous year.

[Figure 2 about here.]

Figures 3 show the top 20 authors, the top 20 institutes, and the top 20 countries in terms of number of publications. Considering the authors, we note how the 0.03% of all authors in our cohort (20 out of 7723) cover over 2.9% of the total publications, suggesting a skewed distribution of publications across authors. When looking at the top institutions, we see they cover over 21% of the total publications (see Table 3), while the top 5 countries cover up to 50% of total publications (see Table 4). Looking deeper into the top institute, one can notice how many of those Universities have strong historical bindings with the sea. Consider, as examples, the Dalian Maritime University, the Shanghai Maritime University, and the Delft Technical University. Similarly, looking at the most representative countries one can see they all have strong maritime industry and economy.

[Figure 3 about here.]

[Table 3 about here.]

[Table 4 about here.]

4.2. Co-authorship network analysis

The degree distribution of the co-authorship network seems to follow a power-law curve (see Fig. 4). However, several distributions may present similar curves. To establish which is the best fitting model we run statistical tests. We run statistical tests, calculating the log-likelihood and p-value between different pairs. The power-law distribution was significantly more accurate fit than the exponential one ($p < 0.01$). However, the comparison between power-law and truncated power-law distributions, as well as the one between power-law and log-normal distributions, did not lead to significantly different results ($p=0.32$ and

$p=0.39$ respectively). The results confirm the heavy-tail characteristic of the degree distribution (which holds true for both log-normal and power-law), but without further indicate the possible nature of such heavy tail (Mitzenmacher, 2004; Higaki et al., 2020; Liu et al., 2021; Smith, 2021).

[Figure 4 about here.]

As a second step in our co-authorship network analysis, we identified the largest component of the network (made of 2753 authors), and identified its main communities, using the Louvain algorithm (Blondel et al., 2008). We identified 28 communities (see Fig. 5), and map on them the distribution of institutions and countries linked to the authors. Results highlight a high level of international collaborations within each community, as well as a high level of national collaborations (within the same country). This can be seen in Figure 6, where we show the number of different countries and institutions per community. Furthermore, our network analysis does not show and closed cluster of collaborations. Communities are all well inter-connected, suggesting that the niche nature of this field (i.e., digital transformation in shipping) leads global actors to collaborate extensively in advancing research. In Figure 7 and Figure 8 we show the chord charts for both country and institution mapping on the co-authorship communities.

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

Lastly, we evaluated the small-world properties of the co-authorship network. To do so, we calculated both clustering coefficient and average path length, and compare them with equivalent random networks. Our results show a higher clustering coefficient (0.83 vs 0.007) and a higher average path (7.1 vs. 3.9). In a proper small-world topology, one would expect high clustering coefficient and small average path. Our results, instead, suggest that communities are strongly locally organized, but somehow lack efficiency in cross community communication.

4.3. Co-citation network analysis

The co-citation network was analyzed for its largest connected component (made of 1298 articles). The results on the degree distribution are similar to those we obtained for the co-authorship network. More specifically, the statistical comparison between degree distribution excluded an exponential distribution ($p < 0.05$), and did not favor a power-law distribution against log-normal or truncated power-law distribution ($p=0.06$ and $p=0.9$ respectively). Figure 9 shows the degree distribution.

[Figure 9 about here.]

Using the degree distribution, we identified the most influential articles (i.e. top 10 articles with the highest number of co-citation). Table 5 shows such influential works. One can see that amongh the most influential works we find SLRs and bibliometric studies, supporting the relevance of such publications within the industry.

[Table 5 about here.]

Next, we created a sub-graph considering only the 20% most cited articles (257 nodes). We identified the communities using the Louvain algorithm and perform a topic analysis on the titles of the articles per community. Table 6 reports the main topics for each of the 7 communities we identified, while Figure 10 show the color-mapped communities.

[Figure 10 about here.]

[Table 6 about here.]

Finally, we adopted several centrality measures to identify the most relevant articles. More specifically, we identified the 5 top articles for five different centrality measures: degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and page rank. We union the results and identified 10 relevant articles for further analysis. We then looked more in details to the theme covered in these articles to extrapolate relevant research areas the field. Results are shown in Table 7.

[Table 7 about here.]

4.4. *Thematic analysis*

For the thematic analysis, we considered all titles from the 2290 articles. After having pre-processed, tokenized, and vectorized all titles, we used clustering methods to identify the best number of clusters for the vectors. More specifically, we used the Calinski-Harabasz index and the Davies-Bouldin index. The curves are shown in Figure 11. All indexes pointed to 8 ideal clusters. For each cluster, we calculated the centroid and then selected the 10 closest vectors (i.e. articles) to each centroid. Focusing on their titles, we highlighted the main themes for each cluster. Results are shown in Table 8.

[Figure 11 about here.]

[Table 8 about here.]

Next, we built words cloud for our articles. In this case, we used both title and abstract words. A first word cloud was built using all words after pre-processing them. The second word cloud was built after removing shipping related key words. This allowed us to focus on technical key words for the second word cloud. Both word clouds were based on TF-IDF analysis and are shown in Figure 12 and Figure 13.

[Figure 12 about here.]

[Figure 13 about here.]

As a last analysis, we focused on the concept tags reported by OpenAlex for each paper. OpenAlex organized topics in a tree-like structure which is a modification of the one produced by (Shen et al., 2018). Specifically, there are 19 root-level concepts: engineering, computer science, business, economics, philosophy, mathematics, political science, medicine, psychology, environmental science, geology, geography, chemistry, physics, biology, sociology, art, history, and materials science. We have collected the root-level concepts related to the publications in our study and plot the number of corresponding papers over time (see Figure 14).

[Figure 14 about here.]

The predominant top-level concept over time are engineering, computer science, and business. We next focused on the second and third level of concepts of Open-Alex, limiting our search to those having as parents either engineering, computer science, or business. Among those, we selected the 10 most relevant and showed their evolution over time in a heat-map (see Figure 15).

[Figure 15 about here.]

5. Discussion

5.1. Descriptive analysis

Our work shows a significant increase in publications related to digital transformation in the shipping industry starting from 2005, with an exponential rise after 2015. Such growth indicates a growing recognition of the importance of digital technologies in the maritime sector. However, the data from the last year (October 2024) may suggest an inflection point, with the number of publication no longer increasing. Further analysis performed in the upcoming months may help validating this finding. If this was to be confirmed, then further investigation would be needed to understand the reason for such inflection: researchers could be moving to different sub-areas currently not fully identified yet, or it may as well be that financial funding for research in this sector is being reduced.

When analyzing the most prolific authors, we can found out that a very small percentage of all authors (approximately 0.03%) is responsible for over 2.9% of all publications. This finding suggests a skewed distribution of author contribution to the field, with the potential risk of limited research perspective and narrowing research venues. Ideally, one would prefer a more distributed contribution among different authors and teams. To achieve this, the industry and funding agencies should provide opportunities to less known authors to expand the horizon and outcome of their research.

The issue is reflected also at institutional and global levels. The top institutions account for over 21% of all publications and the top countries to over 50%, once again indicating a concentration of research efforts in specific locations. This is not completely surprising, since the shipping industry and its related research is particularly attracting for institutes and countries with a long-standing historical binding with the sea. However, new technologies allows

for relevant research to be done even when harbors and sea are not key elements of the country economy. As an example, one could think of the digital twin applications modeling complex systems: advances in such fields could be expected independently from the location of institutes and the geography of the country, allowing more researchers in the future to contribute to this field.

5.2. Co-authorship analysis

The degree distribution of the co-authorship network exhibits a heavy-tail behavior. This was confirmed by the statistical analysis comparing power-law versus exponential distributions. However, the statistical tests could not discriminate between power-law and log-normal distributions, preventing us to draw conclusions on the mechanism at the origins of such heavy-tail. The presence of authors with a high number of co-authors highlights once again the niche nature of such research field, where few co-authors have reached outstanding international recognition for their work and have participated in several research works, playing the role of hubs within the co-authorship network. As the field further evolves, one hope to see a less skewed distribution, with more authors acquiring leading roles and collaborating with each other.

Looking at communities, it appears clear that the network is highly cohesive. Clusters of co-authorship are present across the entire network, and such communities are strongly interconnected with each other. This can also be seen looking at the distribution of institutions and countries across communities of co-authors. Each community contains more than just one institute or one country, suggesting the research on digital transformation in shipping industry thrive on national and international collaborations. This could be the consequence of the niche nature of this research area, which brings different groups to join forces across borders in their research endeavors. Some institutes and some countries play a central role in holding the entire network connected. Important hubs are the Wuhan University of Technology, the Shanghai Maritime University, and the Dalian Maritime University; country-wise, the most important hubs are China, Singapore, and the United Kingdom. In order to mitigate this unbalanced situation, funding agencies should promote research at the crossroad of digital transformation and shipping in other areas, such as within the European Union.

Lastly, the analysis of small-world properties for the co-author network lead to interesting insights. The high clustering coefficient paired with a high average path suggests that, despite having tightly local structures, more work should be done to strengthen the collaboration among such clusters. Research groups should be incentivized to extend their collaborations beyond their usual collaborators, in order to favor knowledge transfer and advance this field of research.

5.3. Co-citation and thematic analysis

The co-citation network also exhibits heavy tail behavior, although in this case the evidence is not supported by statistical analysis. More data should be collected to further validate the heavy-tail nature of its degree distribution. When clustered in communities, the co-citation analysis allowed us to identify the main topics of research across the entire period under investigation. Important areas of research related to (a) the optimization and prediction of fuel consumption and, more in general, energy efficiency, with machine learning playing a key role; (b) safety and risk management, once again supported by machine learning based predictive maintenance and edge computing; (c) logistics, port operations and supply chain, supported by big data analysis, cloud computing, and blockchain technologies; (d) integration of Industry 4.0 technologies for smart port; (e) and cybersecurity, supported by IoT and machine learning.

When looking at themes over the year, the most predominant contributions came from engineering, computer science, and business science. The first two contributed mostly within sub-areas such as artificial intelligence, computer security, operating systems, mechanical and electrical engineering, telecommunications, and data mining. Business-wise, the main areas of research focused on marketing and operations research. One can see how artificial intelligence and computer security have become a very hot topic of research in the most recent years, while other fields, such as marketing and telecommunications seem to be in decline. Similarly, innovative technologies such as blockchain and cloud IoT do not emerge as predominant thematic trends. This could be interpreted as a potential gap to fill by research groups in the upcoming years.

5.4. *Related literature*

Others have published literature reviews related to digital transformation and shipping or maritime industry. In (Tijan et al., 2021), the authors have identified drivers, barriers, and successful factors for digital transformation specifically within the sub-field of maritime transportation. While reporting relevant insights on lack of awareness and strategies to facilitate digital transformation, their study remain focus on a specific area of the shipping industry. The authors also confirm that previous overviews of digital transformation in the maritime transport sector was scarce, further confirming then need of a more systematic and holistic literature review. The results are based on a classical SLRs, rather than adopting a network-based approach.

Another relevant review was published by Alahmadi et al. (2022). In their work, the authors focused specifically on blockchain and its application in the digital transformation of ports and shipping. The research is presented as a review article, rather than a proper SLR, and has particular focus on the supply chain. Again, an holistic view of digital transformation on the shipping industry was not the goal of this research and is therefore missing. Similarly, the work from (Kern, 2021) focuses mostly on digital transformation for logistics.

In (Poulis et al., 2020) the authors cover a rather innovative field of research, namely the intersection of digital transformation in unmanned vessels. Their review paper draws useful insights on the implications that unmanned vessels will have on the identity of the shipping industry. Once again, this review focuses on a specific sub-areas of the shipping industry, leaving space for a comprehensive review of digital transformation implications across the entire industry.

The two most relevant reviews to compare our work with are those by (Sanchez-Gonzalez et al., 2019) and (Jović et al., 2022). The first work by (Sanchez-Gonzalez et al., 2019) is a systematic literature review covering eight main technological domain and three industry sub-sectors. Starting with approximately 3000 research papers, the authors narrow the selection down to 191 by filtering the original set through titles and abstracts. In addition to the 191 articles, the authors analyzes all the referred papers of their subset, hence adding another 23 publications to their analysis. Although similar to our work

in their wide analysis of technological themes, the article is not based on network theory and does not provide holistic results across all the originally identified 3000 papers. Hence, their work is driven by the themes selected up-front, rather than identifying the most relevant themes as an output of the research. They conclude that maritime industry is moving towards digital transformation at different speeds within the different technological domains.

The second work from (Jović et al., 2022) is the closest to ours. The authors do perform a bibliometric analysis leveraging some network theory concepts. Once again, their starting point is the recognition of a scarcity of reviews at the intersection of digital transformation and the maritime and shipping industry. Starting from over 8000 journal and conference articles, the authors manually filtered their initial set to exclude some categories such as ship building and design, and surveillance. The final number of papers reduced to 280 papers used for the downstream analysis. Their bibliometric analysis focused on countries and authors (i.e., most cited authors), in addition to thematic areas. It is particularly interesting how their analysis identified artificial intelligence as a field not strongly researched. Considering that their work was published in 2022, their results are in agreement with ours. However, our analysis covers the years up to 2024 and highlights how artificial intelligence has become a dominant research field within the industry. Moreover, the authors acknowledge the limitation of manual filtering, applied to reduce the total number of articles and guarantee high relevance of the publications. If on one hand we understand the importance of article relevance, in our work we recognize the added value of a large cohort in order to identify gaps in research and get statistically more robust results. To conclude, their work does mention the use of centrality measures, but without specify which centrality was adopted, hence not allowing us to directly compare our results.

6. Conclusion and future work

This study provides a comprehensive bibliometric analysis of digital transformation research in the shipping industry. The key findings indicate a significant increase in publications since 2005, with exponential growth after 2015. However, the data from the last year suggests a possible inflection point, which

requires further investigation to determine if research interest is shifting or if financial constraints are affecting the field. Additionally, the analysis of authorship patterns reveals a concentration of contributions among a small percentage of researchers, institutions, and countries, which may limit diversity in research perspectives. Encouraging broader participation through funding opportunities and institutional support could help mitigate this imbalance.

The co-authorship network analysis highlights the presence of a highly cohesive research community with strong national and international collaborations. While this structure has facilitated knowledge sharing, the dominance of a few key hubs suggests that fostering new collaborations among underrepresented researchers and institutions could further enhance the field. Additionally, the examination of co-citation networks has identified key thematic areas of research, including fuel consumption optimization, risk management, logistics and supply chain operations, smart ports, and cybersecurity. Notably, artificial intelligence has emerged as a dominant topic, while fields like telecommunications and marketing appear to be in decline.

A comparison with existing literature reviews confirms that previous studies have primarily focused on specific sub-fields such as maritime transportation, blockchain applications, or unmanned vessels. While these reviews provide valuable insights, they lack the holistic approach presented in this work. Compared to earlier bibliometric analyses, our findings indicate a growing emphasis on artificial intelligence, highlighting a shift in research priorities over recent years. Furthermore, our study benefits from a more comprehensive dataset, reducing the biases associated with manual filtering and enabling a broader understanding of research gaps.

Future research should focus on validating the observed inflection point in publication trends and investigating the underlying reasons behind potential shifts in research focus. Additionally, more efforts should be directed at fostering interdisciplinary collaborations and expanding participation beyond dominant institutions and countries. Further refinement of bibliometric methodologies, including more advanced statistical validation of network structures, would also enhance our understanding of the evolution of digital transformation in the shipping industry. By addressing these areas, future studies can provide even

deeper insights into the trajectory of this rapidly evolving field.

References

- Alahmadi, D.H., Baothman, F.A., Alrajhi, M.M., Alshahrani, F.S., Albalawi, H.Z., 2022. Comparative analysis of blockchain technology to support digital transformation in ports and shipping. *Journal of Intelligent Systems* 31, 55–69.
- Alperin, J.P., Portenoy, J., Demes, K., Larivière, V., Haustein, S., 2024. An analysis of the suitability of openalex for bibliometric analyses. *arXiv preprint arXiv:2404.17663* .
- An, J., 2024. Maritime logistics and digital transformation with big data: review and research trend. *Maritime Business Review* 9, 229–242.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, P10008.
- Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* 3, 1–27.
- Cooke, R., 2017. Proquest ebook central. *The Charleston Advisor* 19, 39–43.
- Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence* , 224–227.
- Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .
- Filippopoulos, I., Skiadas, C., Violaris, A., Tsaknaki, M., Kiouvrekis, Y., 2022. Road-map for digital transformation in shipping industry: a real use case, in: 2022 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), IEEE. pp. 1–6.
- Hagberg, A., Swart, P.J., Schult, D.A., 2008. Exploring network structure, dynamics, and function using NetworkX. Technical Report. Los Alamos National Laboratory (LANL), Los Alamos, NM (United States).

- Higaki, A., Uetani, T., Ikeda, S., Yamaguchi, O., 2020. Co-authorship network analysis in cardiovascular research utilizing machine learning (2009–2019). *International Journal of Medical Informatics* 143, 104274.
- Jović, M., Tijan, E., Brčić, D., Pucihar, A., 2022. Digitalization in maritime transport and seaports: bibliometric, content and thematic analysis. *Journal of marine science and engineering* 10, 486.
- Kern, J., 2021. The digital transformation of logistics: A review about technologies and their implementation status. *The digital transformation of logistics: Demystifying impacts of the fourth industrial revolution* , 361–403.
- Lau, Y.y., Chen, Q., Poo, M.C.P., Ng, A.K., Ying, C.C., 2024. Maritime transport resilience: A systematic literature review on the current state of the art, research agenda and future research directions. *Ocean & Coastal Management* 251, 107086.
- Liu, P., Jiang, L., Liu, H., 2021. Structural and evolutionary analysis of coauthorship network in scientometrics, in: *2021 International Conference on Education, Information Management and Service Science (EIMSS)*, IEEE. pp. 514–519.
- Mitzenmacher, M., 2004. A brief history of generative models for power law and lognormal distributions. *Internet mathematics* 1, 226–251.
- Munim, Z.H., Dushenko, M., Jimenez, V.J., Shakil, M.H., Imset, M., 2020. Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. *Maritime Policy & Management* 47, 577–597.
- Oğuz, S., 2024. Studies on maritime transport: A bibliometric analysis (2010–2023). *Mersin Üniversitesi Denizcilik ve Lojistik Araştırmaları Dergisi* 6, 17–29.
- Paul, P.C., Rahman, M., Begum, A., Ahmed, M.T., Chakraborty, D., Rahman, M.S., . Combining bert with lda: Improved topic modeling in bengali language .

- Poulis, K., Galanakis, G.C., Triantafillou, G.T., Poulis, E., 2020. Value migration: digitalization of shipping as a mechanism of industry dethronement. *Journal of Shipping and Trade* 5, 1–18.
- Priem, J., Piwowar, H., Orr, R., 2022. Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833* .
- Sanchez-Gonzalez, P.L., Díaz-Gutiérrez, D., Leo, T.J., Núñez-Rivas, L.R., 2019. Toward digitalization of maritime transport? *Sensors* 19, 926.
- Shen, Z., Ma, H., Wang, K., 2018. A web-scale system for scientific knowledge exploration. *arXiv preprint arXiv:1805.12216* .
- Smith, K.M., 2021. Explaining the emergence of complex networks through log-normal fitness in a euclidean node similarity space. *Scientific Reports* 11, 1976.
- Tijan, E., Jović, M., Aksentijević, S., Pucihar, A., 2021. Digital transformation in the maritime transport sector. *Technological Forecasting and Social Change* 170, 120879.
- Vaughan, J., 2011. Ebsco discovery services. *Library Technology Reports* 47, 30–38.
- Wilde, M., 2016. Ieee xplora digital library. *The Charleston Advisor* 17, 24–30.
- Xiao, G., Pan, L., Lai, F., 2025. Application, opportunities, and challenges of digital technologies in the decarbonizing shipping industry: a bibliometric analysis. *Frontiers in Marine Science* 12, 1523267.
- Xiao, G., Yang, D., Xu, L., Li, J., Jiang, Z., 2024. The application of artificial intelligence technology in shipping: A bibliometric review. *Journal of Marine Science and Engineering* 12, 624.

List of Figures

1	Network-based bibliometric analysis diagram	23
2	Distribution of publications across years	24
3	Top 20 leading authors (<i>top</i>), institutions (<i>middle</i>), and countries (<i>bottom</i>)	25
4	Co-authorship degree distribution (<i>top</i>), and corresponding log- log chart (<i>bottom</i>)	26
5	Co-authorship network with communities	27
6	Co-authorship distribution of institutions (<i>top</i>) and countries (<i>bot- tom</i>) across partitions	28
7	Chord diagrams for insitution mapping on co-authorship com- munities. The three diagrams show three relevant clusters: (<i>top</i>) Wuhan University of Technology, (<i>middle</i>) Shanghai Maritime University, and (<i>bottom</i>) Dalian Maritime University	29
8	Chord diagrams for countries mapping on co-authorship com- munities. The three diagrams show three relevant clusters:(<i>top</i>) China, (<i>middle</i>) Singapore, and (<i>bottom</i>) Great Britain	30
9	Co-citation degree distribution and log-log chart	31
10	Co-citation network with communities	32
11	Davies-Bouldin (<i>top</i>) and Calinski-Harabasz (<i>bottom</i>) indexes. . .	33
12	Word cloud based on TF-IDF	34
13	Word cloud focused on technological terms	35
14	Top-level OpenAlex concept distribution over time terms	36
15	Detail-level OpenAlex concept distribution over time for the 10 most relevant terms	37

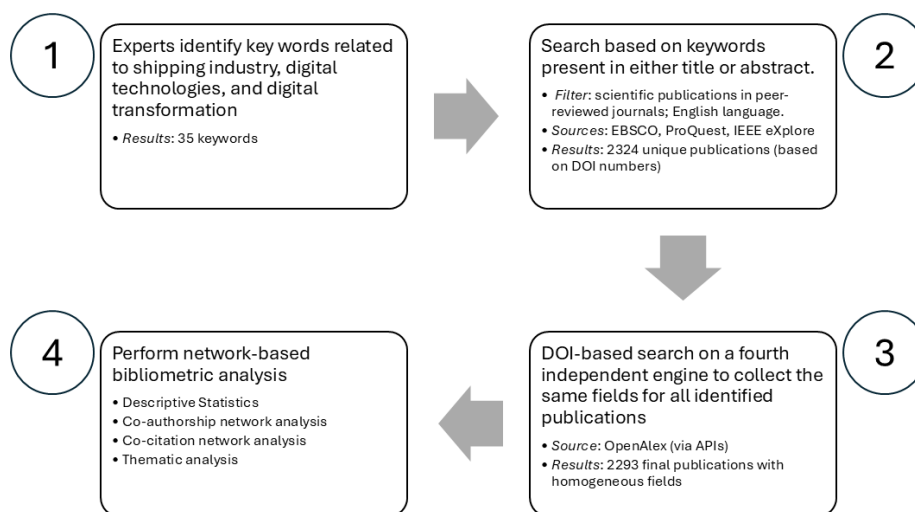


Figure 1: Network-based bibliometric analysis diagram

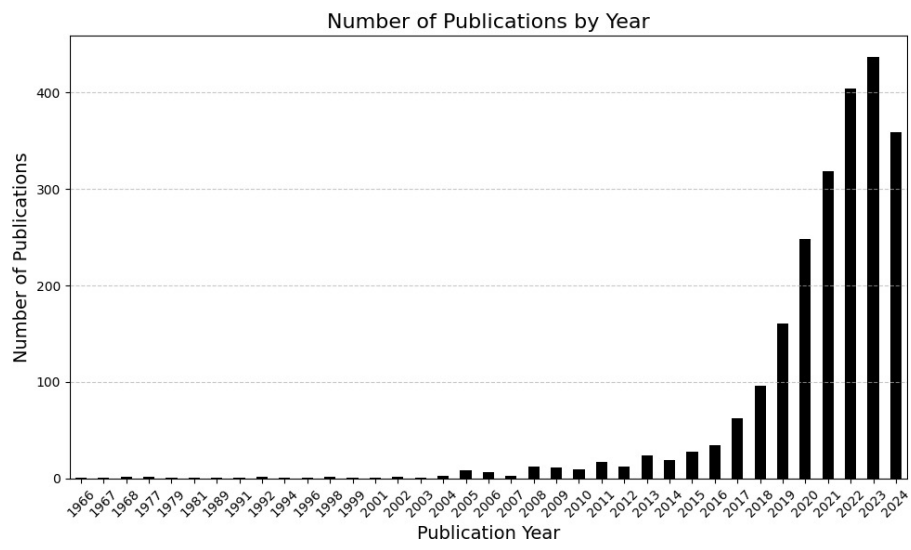


Figure 2: Distribution of publications across years

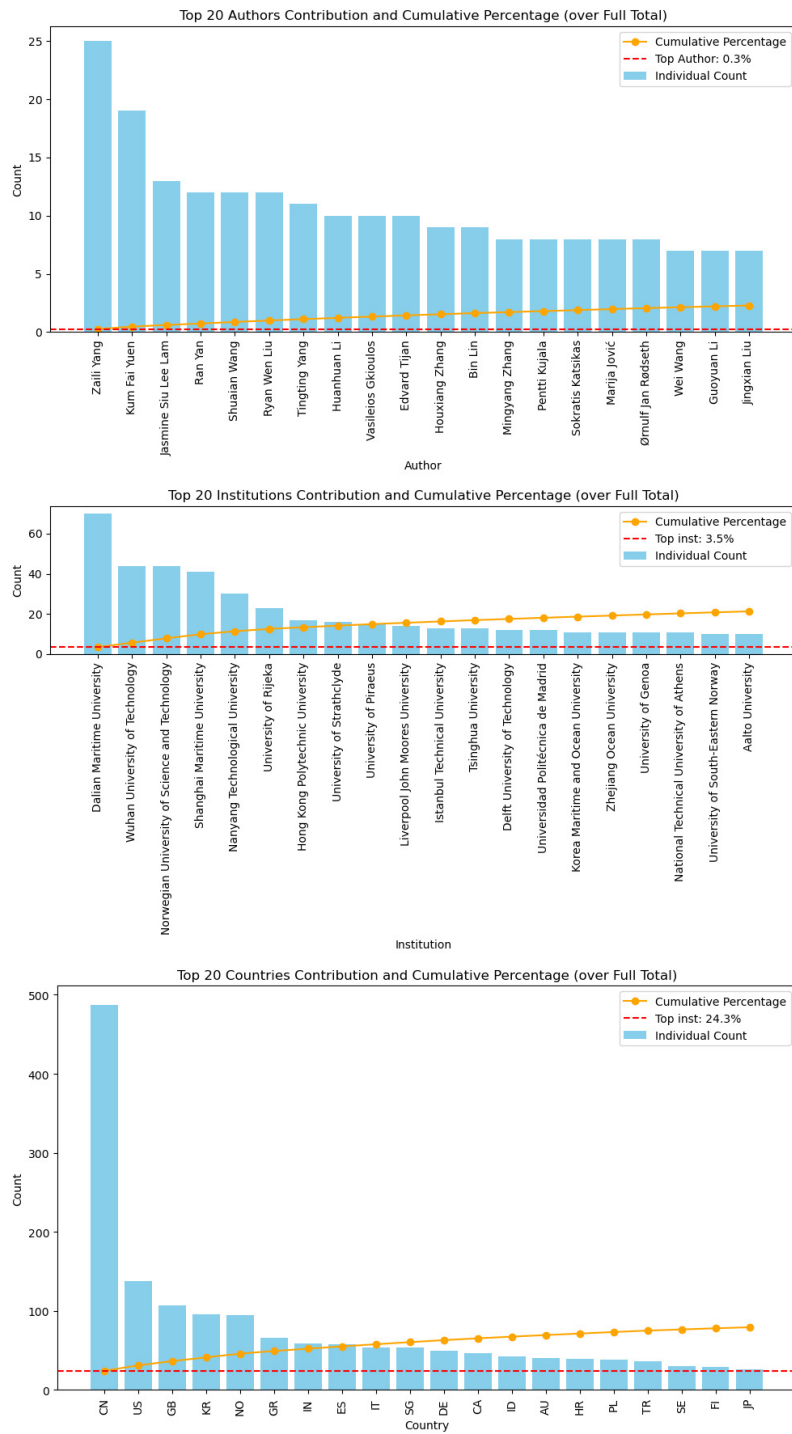


Figure 3: Top 20 leading authors (*top*), institutions (*middle*), and countries (*bottom*)

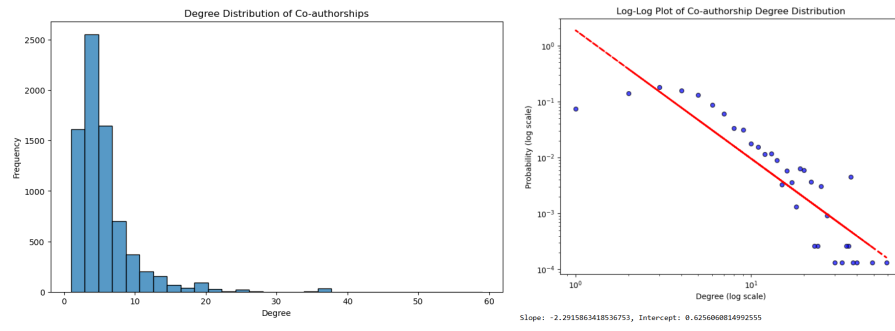


Figure 4: Co-authorship degree distribution (*top*), and corresponding log-log chart (*bottom*)

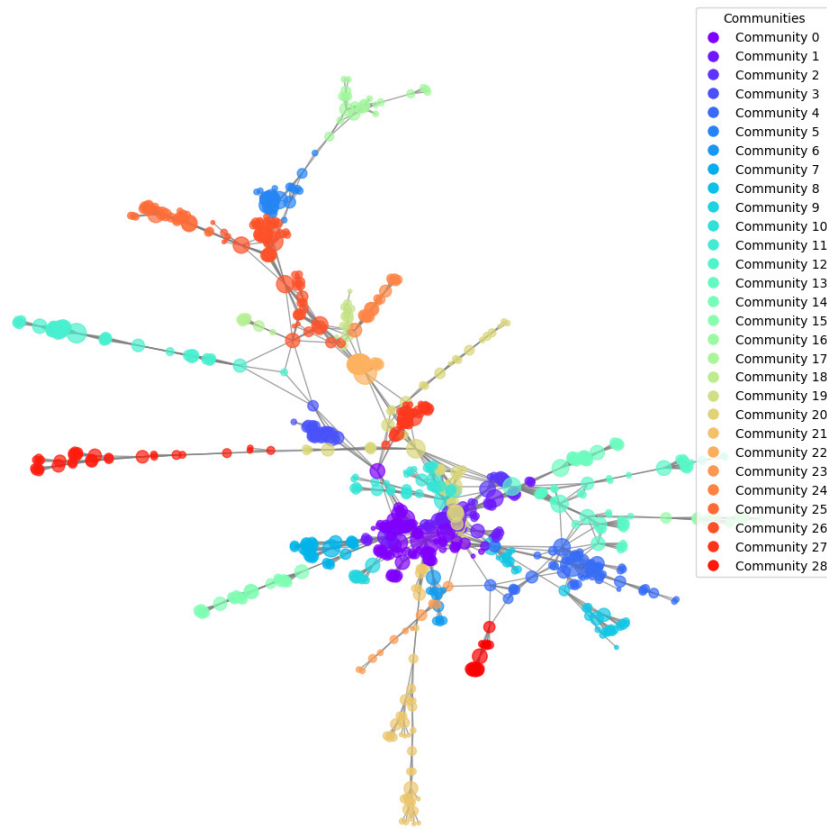


Figure 5: Co-authorship network with communities

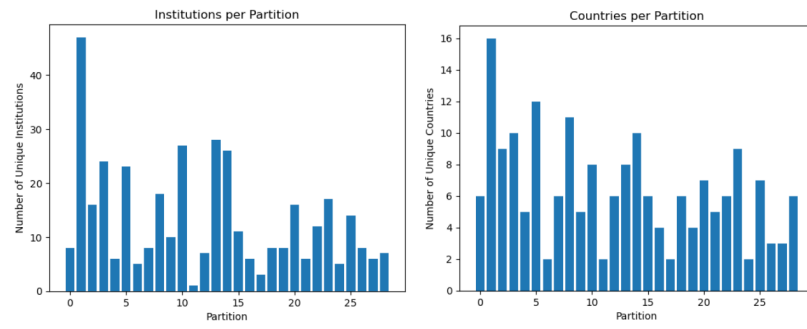


Figure 6: Co-authorship distribution of institutions (*top*) and countries (*bottom*) across partitions

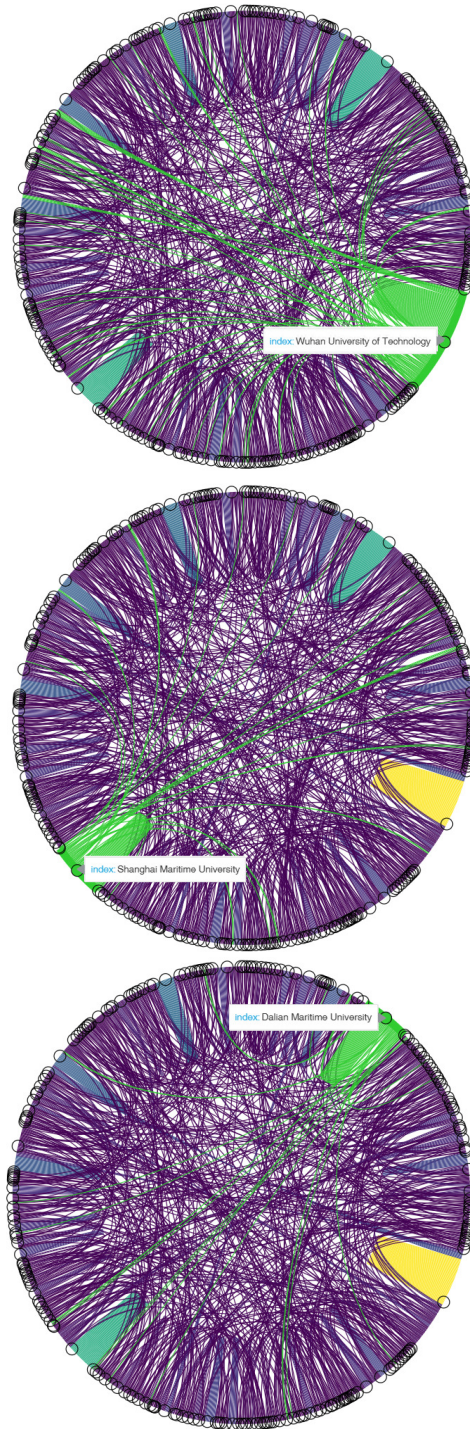


Figure 7: Chord diagrams for insitution mapping on co-authorship communities. The three diagrams show three relevant clusters: (*top*) Wuhan University of Technology, (*middle*) Shanghai Maritime University, and (*bottom*) Dalian Maritime University

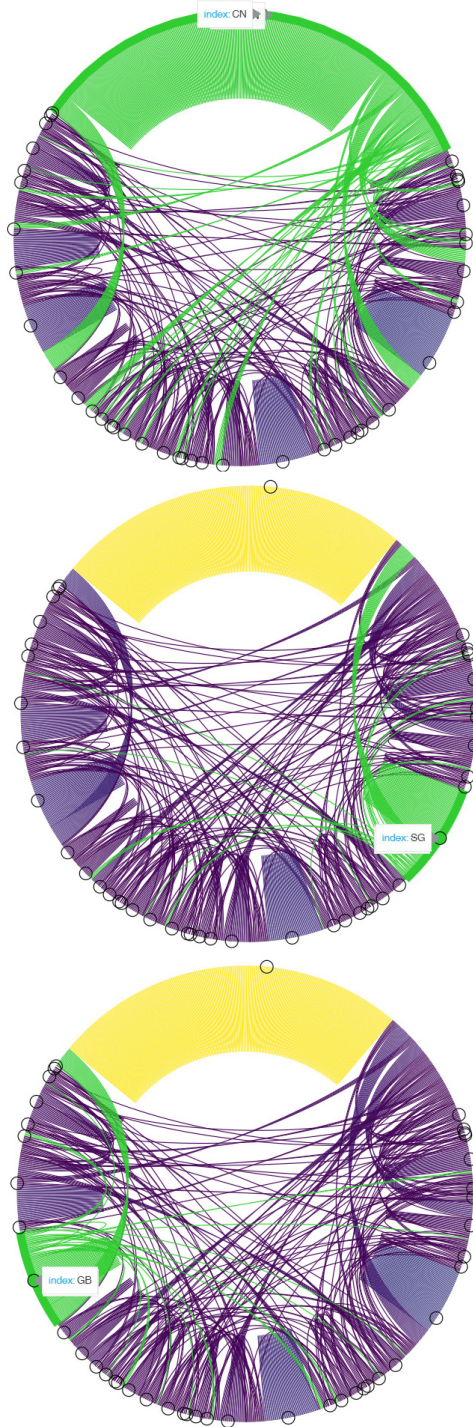


Figure 8: Chord diagrams for countries mapping on co-authorship communities. The three diagrams show three relevant clusters:(*top*) China, (*middle*) Singapore, and (*bottom*) Great Britain

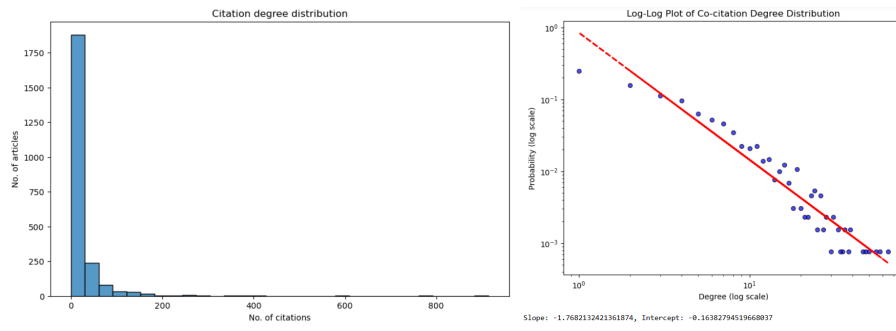


Figure 9: Co-citation degree distribution and log-log chart

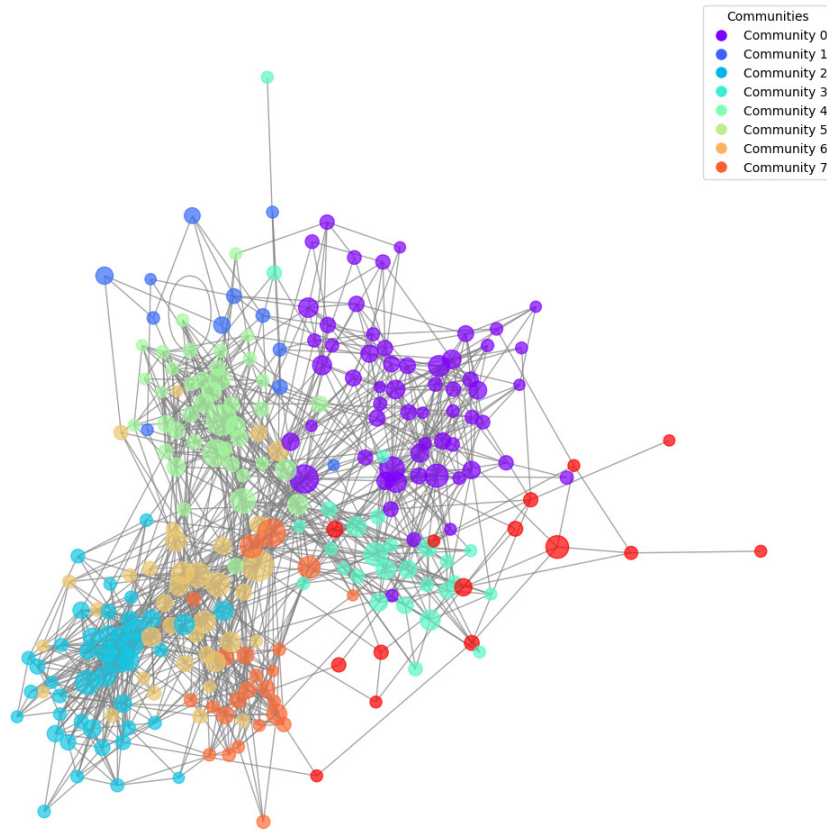


Figure 10: Co-citation network with communities

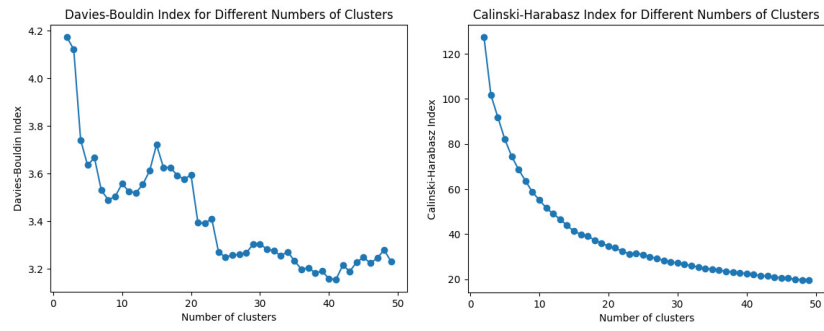


Figure 11: Davies-Bouldin (top) and Calinski-Harabasz (*bottom*) indexes.

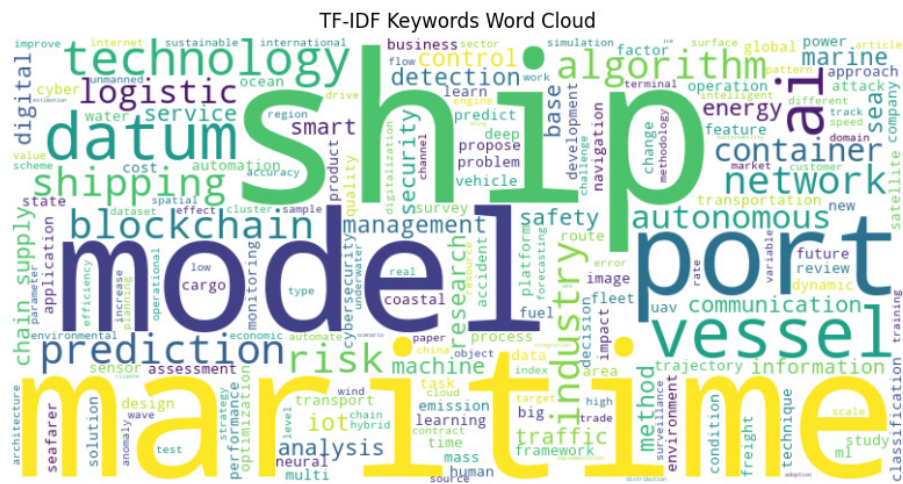


Figure 12: Word cloud based on TF-IDF

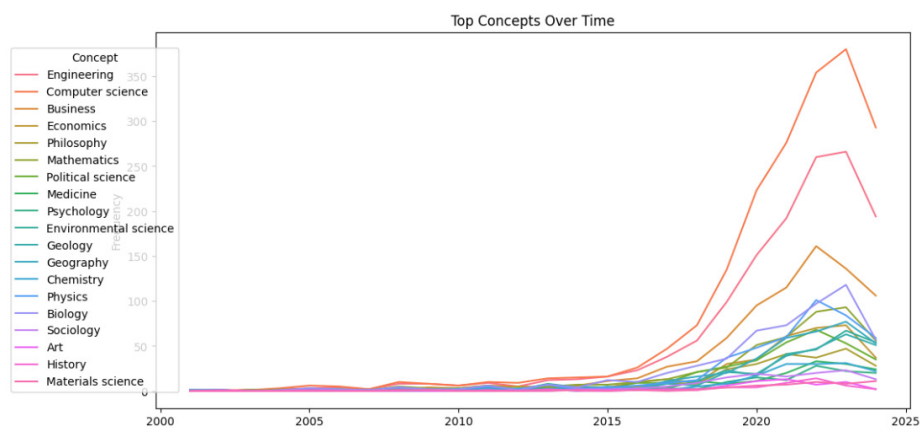


Figure 14: Top-level OpenAlex concept distribution over time terms

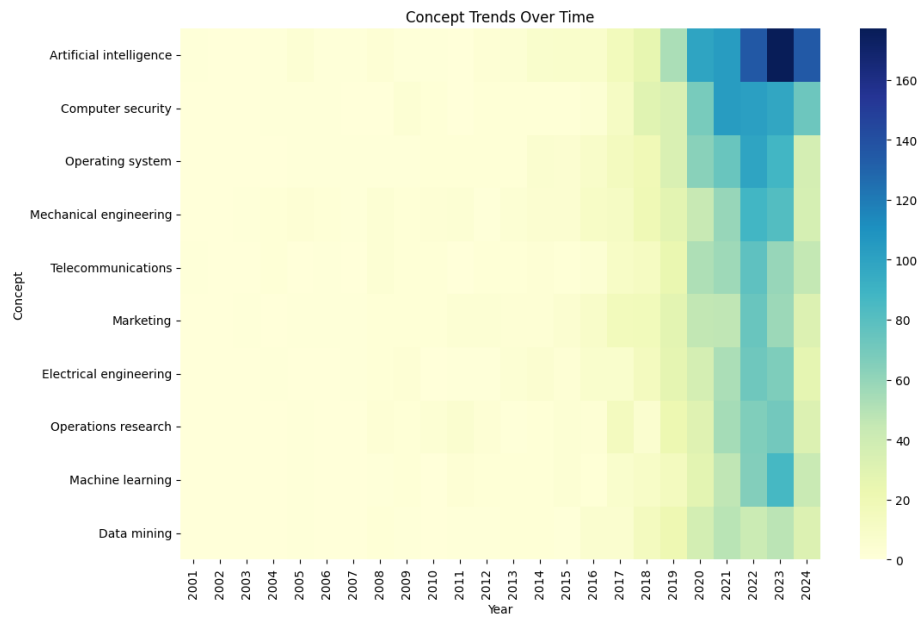


Figure 15: Detail-level OpenAlex concept distribution over time for the 10 most relevant terms

List of Tables

1	List of keywords identified by experts.	39
2	Number of retrieved articles per research engine.	40
3	Distribution of publications across institutions.	41
4	Distribution of publications across countries.	42
5	Top 10 influential papers based on citations.	43
6	Community-based topic analysis	44
7	Centrality-based topic analysis.	45
8	Results of thematic analysis.	46

Table 1: List of keywords identified by experts.

Keyword	Type (Digit. Trans. or Shipping)
Digital transformation	Digit. Trans.
Digital innovation	Digit. Trans.
Digital ecosystems	Digit. Trans.
Digitization	Digit. Trans.
Digitalization	Digit. Trans.
Digital platforms	Digit. Trans.
Industry 4.0	Digit. Trans.
Smart technologies	Digit. Trans.
Data-driven transformation	Digit. Trans.
Automation	Digit. Trans.
Internet of Things	Digit. Trans.
Blockchain	Digit. Trans.
Data analysis	Digit. Trans.
Artificial intelligence	Digit. Trans.
Machine learning	Digit. Trans.
Big data	Digit. Trans.
Cloud computing	Digit. Trans.
Cyber-physical systems	Digit. Trans.
Digital twins	Digit. Trans.
Edge computing	Digit. Trans.
5G networks	Digit. Trans.
Predictive analytics	Digit. Trans.
Cybersecurity	Digit. Trans.
Supply chain integration	Digit. Trans.
shipping	Shipping
maritime	Shipping
Sea freight	Shipping
Smart ports	Shipping
Autonomous ships	Shipping
Fleet management	Shipping
Cargo tracking	Shipping
Digital shipyards	Shipping
Port digitalization	Shipping
Port automation	Shipping
Vessel performance	Shipping

Table 2: Number of retrieved articles per research engine.

Engine	No. of scientific articles
EBSCO	1904
ProQuest	2011
IEEE eXplore	300

Table 3: Distribution of publications across institutions.

Institution	No. of publi- cations	Cumulative %	% Of total
Dalian Maritime Univer- sity	70	3.48	3.48
Wuhan University of Technology	44	5.67	2.19
Norwegian University of Science and Technology	44	7.86	2.19
Shanghai Maritime Uni- versity	41	9.90	2.04
Nanyang Technological University	30	11.39	1.50
University of Rijeka	23	12.53	1.14
Hong Kong Polytechnic University	17	13.38	0.85
University of Strathclyde	16	14.17	0.80
University of Piraeus	15	14.92	0.75
Liverpool John Moores University	14	15.61	0.70
Istanbul Technical Univer- sity	13	16.26	0.65
Tsinghua University	13	16.91	0.65
Delft University of Tech- nology	12	17.50	0.60
Universidad Politécnica de Madrid	12	18.10	0.60
Korea Maritime and Ocean University	11	18.65	0.55
Zhejiang Ocean Univer- sity	11	19.19	0.55
University of Genoa	11	19.74	0.55
National Technical Uni- versity of Athens	11	20.29	0.55
University of South- Eastern Norway	10	20.79	0.50
Aalto University	10	21.28	0.50

Table 4: Distribution of publications across countries.

Country ISO	No. of publica- tions	Cumulative %	% Of total
CN	487	24.29	24.29
US	138	31.17	6.88
GB	107	36.51	5.34
KR	96	41.30	4.79
NO	95	46.03	4.73
GR	66	49.33	3.29
IN	59	52.27	2.94
ES	58	55.16	2.89
IT	54	57.86	2.69
SG	54	60.55	2.69
DE	50	63.04	2.49
CA	47	65.39	2.34
ID	42	67.48	2.09
AU	40	69.48	2.00
HR	39	71.42	1.95
PL	38	73.32	1.90
TR	36	75.11	1.80
SE	30	76.61	1.50
FI	29	78.05	1.45
JP	26	79.35	1.30

Table 5: Top 10 influential papers based on citations.

DOI	Title	Publication Year
https://doi.org/10.1080/03088839.2020.1788731	Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions	2020
https://doi.org/10.3390/jmse12060919	Comprehensive Analysis of Maritime Cybersecurity Landscape Based on the NIST CSF v2. 0	2024
https://doi.org/10.1016/j.tre.2019.09.020	Maritime shipping digitalization: Blockchain-based technology applications, future improvements, and intention to use	2019
https://doi.org/10.1080/01441647.2019.1649315	How big data enriches maritime research—a critical review of Automatic Identification System (AIS) data applications	2019
https://doi.org/10.3390/app14145994	Harnessing AI for sustainable shipping and green ports: Challenges and opportunities	2024
https://doi.org/10.1016/j.ijcip.2022.100571	Developments and research directions in maritime cybersecurity: A systematic literature review and bibliometric analysis	2022
https://doi.org/10.1109/tits.2019.2908191	Traffic pattern mining and forecasting technologies in maritime traffic service networks: A comprehensive survey	2019
https://doi.org/10.3390/s19040926	Toward Digitalization of Maritime Transport?	2019
https://doi.org/10.3390/info13010022	Cyber security in the maritime industry: A systematic survey of recent advances and future trends	2022
https://doi.org/10.3390/jmse10040486	Digitalization in maritime transport and seaports: bibliometric, content and thematic analysis	2022

Table 6: Community-based topic analysis

Community	Main theme
1	Optimization and prediction of fuel consumption, energy efficiency, and environmental impact in the maritime industry, with a focus on machine learning, big data, and modeling techniques.
2	Maritime safety, risk management, and the application of machine learning techniques to predict, analyze, and mitigate accidents and hazards in maritime operations.
3	Machine learning, artificial intelligence, and big data applications in the maritime domain.
4	Integration of Internet of Things (IoT), mobile edge computing, communication networks, and security within the maritime industry, with a particular focus on autonomous ships, data offloading, latency minimization, and communication technologies for maritime transportation systems.
5	Digital transformation and technological advancements within the maritime sector, particularly in relation to maritime logistics, container shipping, port operations, and the broader shipping supply chain.
6	Development and optimization of smart ports, focusing on the integration of Industry 4.0 technologies.
7	Maritime cybersecurity, with an emphasis on cyber risks associated with the digital transformation of the maritime industry, particularly as it pertains to the rise of autonomous vessels, smart shipping technologies, and the IoT-enabled maritime environment.
8	Adoption and application of blockchain technology in the maritime industry, specifically in areas like shipping, supply chains, and port management.

Table 7: Centrality-based topic analysis.

OpenAlex ID	Top 5 in No. of centrality measures	Main topic
W3041382323	5	Big data and artificial intelligence
W4400493457	4	Artificial intelligence, sustainable shipping, and green ports
W2964482263	4	Big data
W2978644098	3	Blockchains
W4386245296	2	Data and IoT
W4225993858	2	Digitalization
W4399283331	2	Cybersecurity
W3090216936	1	Blockchain conceptual framework
W4205557186	1	Cybersecurity
W3213918042	1	Blockchain

Table 8: Results of thematic analysis.

Topic 1: Advancements in AI and Machine Learning for Maritime Applications. This topic explores the integration of advanced AI techniques in the maritime sector, focusing on automation, situational awareness, and safety.
Topic 2: Spatial and Environmental Analysis in Maritime and In-land Waters. This topic focuses on the intersection of spatial data analysis, environmental monitoring, and maritime operations.
Topic 3: Predictive Modeling and Risk Assessment in Maritime Operations. This topic delves into predictive models and analytics for enhancing safety and efficiency in maritime activities.
Topic 4: Data-Driven Optimization in Maritime Engineering and Traffic Management. This topic addresses the use of data-driven and AI approaches in ship design, traffic prediction, and environmental forecasting.
Topic 5: Blockchain and Edge Computing in Maritime Operations. This topic investigates the application of blockchain and edge computing technologies in maritime logistics and communication.
Topic 6: Innovations and Reviews in Autonomous and Smart Shipping. This topic reviews the state of autonomous and smart shipping technologies, exploring their implications for design, safety, and logistics.
Topic 7: Digital Transformation and Sustainability in Maritime Logistics. This topic focuses on the digitalization of maritime operations and its impact on sustainability and efficiency.
Topic 8: Maritime Industry's Adaptation to Digital and Sustainable Practices. This topic explores the maritime industry's response to emerging technologies and sustainability challenges.
