Applied Data Science Master Thesis

Three Decades of Transport Corridor Research : A Literature Review Applying Text Mining.

Candidate : Theodosia Kosma [2879018]

Supervisors: Witte Patrick, Labib SM

Utrecht University

July 2023

Table of Contents

Abstract

Transport corridors are vital but face challenges like bottlenecks and disconnections. Academic discourse lacks cohesion and has diversified over time while practical research on corridor development lacks follow-up and scientific attention. This study aims to address the existing gap in the literature on corridor development by offering a unique and pioneering literature review analysis that employs computational methods. The objectives of the study are twofold: firstly, to examine the theoretical perspectives employed in corridor research and identify common areas of consensus, and secondly, to gain a better understanding of the scholarly influence within the field of corridor development. A comprehensive dataset of 79 papers is collected and further analyzed. The abstracts of each paper are first preprocessed and various analytical techniques are then applied, including frequency analysis, LDA topic modeling, and bibliometric analysis. The study finds that corridor development research is complex and requires interdisciplinary approaches due to the consideration of multiple variables and factors. The years influencing the most the corridor development are also identified while also the Netherlands, as a prominent contributor is discussed, emphasizing the country's expertise in infrastructure and it's beneficial geographical location.

1. Introduction

Transport corridors, serving as vital links connecting people, goods, cities, regions, countries, and even continents, have long been recognized as essential components of modern infrastructure networks (Priemus & Zonneveld, 2003). For several decades, they have been an influential academic and policy concept, involving a variety of multi-modal infrastructures. This idea has been thoroughly examined in academic literature on the development of transportation corridors, focusing on its integration, with urban planning, economic development, and environmental goals. More specifically, the European Commission has been actively promoting the development of transport corridors for several decades, starting with the trans-European transport networks (TEN-T) in the 1990s, and subsequently through the implementation of the Core Network Corridors as part of the Connecting Europe programs since the 2010s (Witte and Spit, 2016). Nevertheless, despite the considerable attention that transportation corridors have gotten over the past three decades, the reality of corridor development in Europe is plagued by bottlenecks and disconnections, rather than the intended synthesis and integration (Witte et al., 2012). These bottlenecks encompass various challenges, including inadequate or incomplete connectivity to transport infrastructure, conflicting political interests, lack of coordination among different transportation modes, and incompatible technical systems, among others.

The academic discourse surrounding transport corridor development has experienced an analogous lack of connectivity and cohesion. According to Reggiani et al. (1995), this subject has been the focus of scholarly debate since the 1990s, and in the early 2000s, Priemus and Zonneveld (2003) published a Special Issue on the governance of corridors aiming to provide a comprehensive understanding of corridors. However, since then, the evolution of their

development over time has resulted in a substantial expansion and diversification of approaches, with differing interpretations of what constitutes a transport corridor and how it should be planned and implemented. For instance, the United States has a vast rail network that successfully connects coastal ports to inland regions along an East-West freight transport corridor (Notteboom and Rodrigue, 2005). The infrastructure networks in Europe, in contrast, have been developed in a nationally oriented way, leading to fragmented methods. Also, Asia's primary focus lies in the Chinese Belt and Road Initiative (BRI), which aims to develop corridors that connect China to various regions, including Europe and Africa (Hoh, 2019). The main goal of this effort is to increase connectivity between the member nations while promoting and facilitating trade flows.

Given the practical implications and challenges associated with corridor development, there has been a noticeable increase in academic interest in this field that can be attributed to the finalization of TEN-T period and the implementation of Connecting Europe facility with the associated Core Network Corridors (European Commission, 2011). However, actual projects and investments are primarily motivating academic research on transportation corridors. This means that practical research frequently begins with the conception of a project idea, and concludes upon its accomplishment, eventually resulting in the publication of a scientific paper or project report. On the one hand, this approach enables researchers to thoroughly investigate and analyze corridor development within real-world contexts, providing valuable insights and expanding the existing knowledge base. However, on the other hand, it often results in a lack of follow-up research and scientific attention dedicated to that specific project. Consequently, there is a growing need to perform a literature review analysis on this topic, as there is an increased academic interest in corridor development, reinforced by prominent policy efforts, such as the Core Network Corridors in Europe in 2011, the Belt and Road Initiative (BRI) in 2013, the European structural and investment funds (ESIFs) in 2014, and the European Green Deal in 2019 among others. This review will delve into the complexities and implications of corridor development, thereby enriching the comprehension of this field.

This study encompasses a comprehensive analysis of a curated dataset consisting of 79 articles. This dataset serves as the foundation for the exploration of transport corridor development. Given the extensive body of research on corridors, conducting a systematic review can offer valuable insights into the existing literature landscape. The objectives of this study are threefold: firstly to investigate the theoretical perspectives and common areas of consensus utilized in corridor research, secondly to explore and better understand the scholarly influence within this field, and thirdly to analyse empirical insights and synthesize existing knowledge to enhance the understanding of corridor development. In pursuit of these objectives, computational methods are employed, recognized as a promising and time-efficient approaches by Boyer et al. (2017). Utilizing computational techniques in literature review analysis offers several advantages compared to traditional manual approaches, including enhanced reliability and efficiency by replacing human processing with computer processing (Asmussen & Moller, 2019). This research employs text mining techniques and bibliometrics as part of the literature review analysis. Text mining has already

demonstrated its effectiveness in various domains, such as risk management, fraud detection, business intelligence, and social media analysis (Ferreira‑Mello et al., 2019). It enables the discovery of hidden patterns, relationships, and trends within the literature corpus that may not be easily identifiable through traditional manual approaches. Additionally, bibliometrics provides a more objective and trustworthy analysis, compared to other qualitative and quantitative literature reviewing techniques (Aria & Cuccurullo, 2017). It facilitates a structured analysis of a vast amount of information, identifies influential scholars and institutions, and presents an overview of the existing research landscape. By adopting these advanced computational techniques, this study aims to surpass the limitations of traditional manual approaches, leading to a more thorough and insightful analysis of the literature.

To summarize this paper follows a structured approach to investigate corridor development. It starts with a data collection process, followed by a series of analytical techniques including frequency analysis, topic modeling, and bibliometric analysis. The results of each method are then discussed, providing valuable insights into the patterns within the dataset. Lastly, possible limitations are discussed and the potential areas for future research are identified.

2. Methodology

This section introduces the research framework in five phases: data collection, frequency analysis, topic modeling, topic analysis, and bibliometric analysis (Figure 1).

2.1. Data Collection

The dataset used in this study was obtained from a previous study, comprising a collection of 79 papers published between 1995 and 2021.Essential information were included such as the title of the papers, authors' names, and years of publication. Each paper's abstract was manually added to the dataset in order to facilitate text mining analysis, ensuring a comprehensive examination and extraction of insights from the textual content. This approach was adopted as the abstracts capture the essential information and disclose the larger content of an article (Karami et al., 2020).
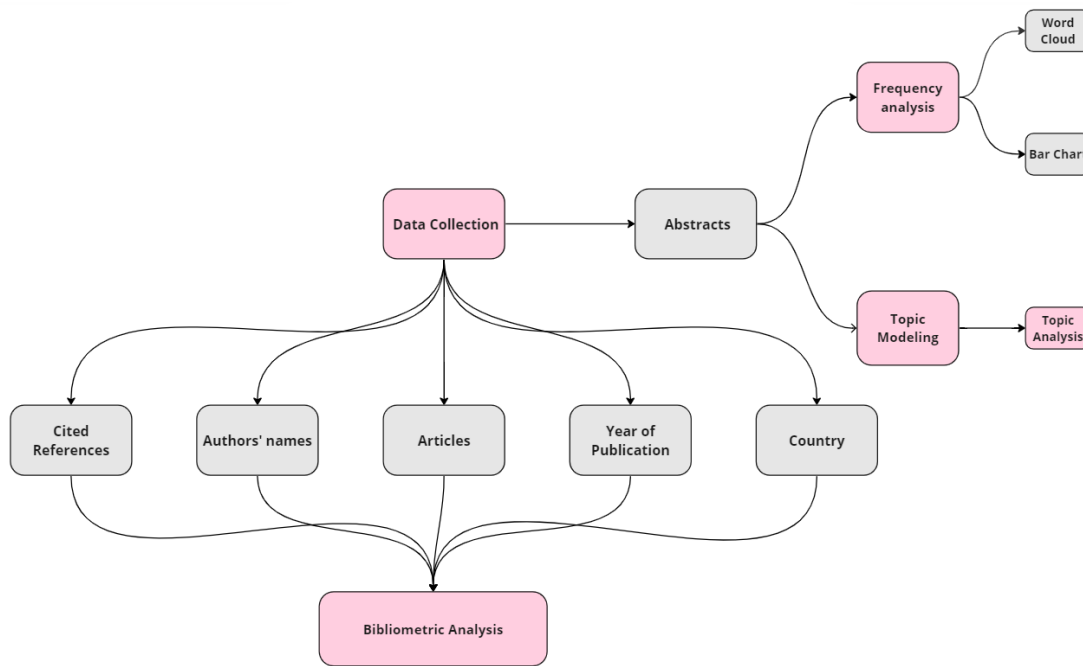
Figure 1. Research framework

2.2. Frequency analysis

Text mining involves converting unstructured text data into a structured format that can be processed by machines (Thakur & Kumar, 2022). Abstracts, in their original format, consist of unstructured textual data that requires interpretation. Employing text mining techniques enables the exploration of such data to identify underlying semantic patterns. In order to prepare the abstracts for analysis, a preprocessing step was performed using spaCy NLP pipeline (Honnibal et al., 2020). This pipeline included lowercasing all words to standardize the text, lemmatization to reduce words to their base form, and removal of stopwords. Furthermore, the stopwords list was extended with specific terms such as "corridor", "article" and "paper" among others, to ensure their exclusion from the analysis. Additionally, punctuation and digits were removed to further refine the dataset for topic modeling. This preprocessing step facilitated the transformation of the raw textual data into a cleaner and more manageable format for subsequent analysis. To gain a comprehensive understanding, an analysis of the top 10 and top 50 words was conducted, visualizing their frequency through bar charts and word clouds, respectively. Bar charts provide a visual representation of word frequency, where the height of each bar indicates the frequency of a specific word. A word cloud uses word size to reflect word frequency in a corpus, with larger size indicating higher frequency (van Altena et al., 2016).

After preprocessing steps, a total of 2,136 unique words were identified in the abstracts. The frequency of these words ranged from 1 to 212 occurrences, with an average of 3,68. This indicated that the least frequent word appeared only once, while the most frequent word occurred 212 times. On average, each word appeared 3,68 times across the dataset. As can be seen in Figure 2 these findings align with Zipf's law, which describes an inverse relationship between the frequency of words and their frequency rank (Zipf, 1949). Zipf's law states that a

small number of words occur very frequently, while the majority of words occur infrequently. This inverse relationship can be visualized as a power-law distribution, where the frequency of a word is inversely proportional to its rank raised to a certain exponent. In Figure 2, a vertical line is used to indicate the position of the top-20 words among the more than 2,000 unique words identified.
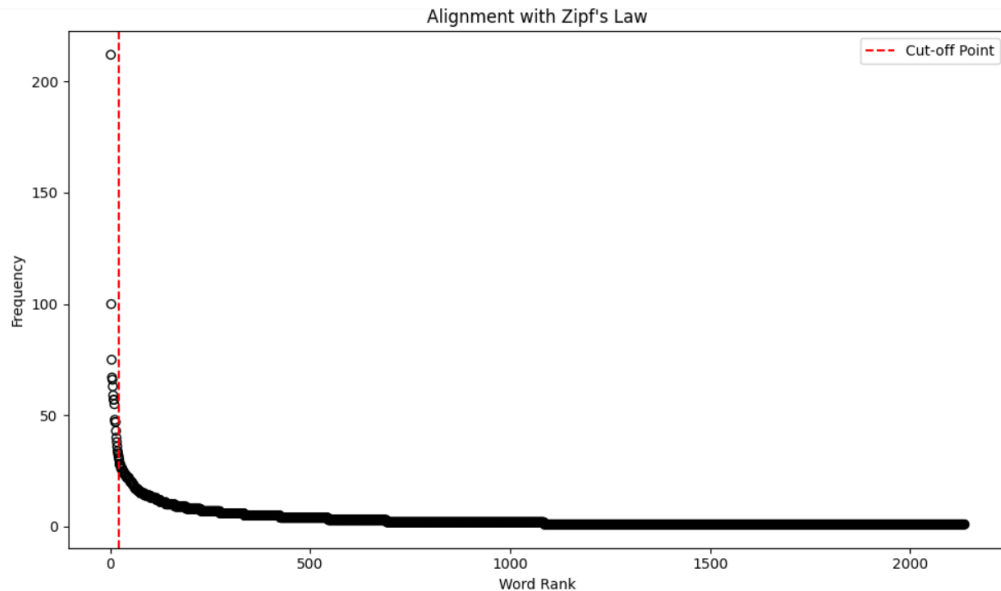

Figure 2 : Frequency of words. The vertical line shows the cut-off point for the top-20 words in the word cloud

2.3. Topic Modeling

Topic models are a type of generative model used in machine learning and natural language processing to analyze large collections of text data. This model has been widely used in different research areas, including digital humanities (Meeks & Weingart, 2012), bioinformatics (Liu et al., 2016), historical studies (Brauer & Fridlund, 2013), and political science (Wilkerson & Casas, 2017). The Latent Dirichlet Allocation (LDA) modeling technique was used in this study, to identify a set of key topics pertaining to corridor development, a widely employed method for topic modeling within the literature (Blei et al., 2003; Oehmer-Pedrazzi et al., 2022a). LDA aims to uncover the underlying themes or topics in a corpus of text data by using a generative process to model the data and make inferences about the topics. By analyzing the co-occurrence of words across documents, LDA identifies groups of words that frequently appear together, forming distinct topics. These topics are then interpreted by humans as meaningful "themes", providing valuable insights into the content and structure of the corpus (Karami et al., 2018). Whereas other models like LSI (Latent Semantic Indexing) use matrix factorization techniques to discover the topics (Rosario, 2000), LDA assumes that documents are generated from a mixture of topics, an explicit assumption made by the researchers (Ostrowski et al., 2015). In other words, LDA posits that each document in the corpus is a combination of multiple underlying topics, and these topics contribute to the generation of different parts of the document.

Finding the optimal number of topics, N, is a crucial step in implementing LDA as it directly influences the quality and specificity of the topics generated from the preprocessed abstracts. A higher value of N can lead to narrowly defined topics, while a lower value can result in overly generalized topics that fail to accurately capture the nuances of the text data. To determine the optimal N, a series of experiments in the form of an exhaustive grid-search was conducted. The coherence scores were evaluated by systematically varying N from 2 to 12, with steps of 1, no_below of 10,100 and 500, no_above of 0.2 and 0.5, alpha of 0.01, 0.3, 0.9 and beta 0.01, 0.3, 0.9. The coherence score was chosen as an evaluation metric since it is widely utilized in natural language processing and computation linguistics, to measure the semantic similarity and interrelation of generated topics (Röder et al., 2015).

The results revealed that a comparatively high coherence score was obtained when the number of topics ranged between 7 and 8. However, to enhance the determination of the optimal number of topics and to overcome the limitations of coherence score as a reliable metric, an intertopic distance mapping was employed. This method provided further information about optimal N by analyzing the overlap of topics and the sizes of topic clusters (Figure 3). Intertopic distance mapping is an interactive visualization tool used in topic modeling to visualize the distance between the identified topics (Sievert & Shirley, 2014). In the left panel of the visualization, topics are represented as circles and their positions are determined based on their distances between the other topics. The right channel presents a horizontal bar chart that displays the most significant terms for interpreting each selected topic.

Subsequently, the final LDA model was executed with a value of N=8, alpha=0.9, beta=0.3, no_below=5 and no_above=0.4. The results were saved, and the parameter for each topic was set to include the top 10 words with the highest probabilities. This process generated a list of 10 keywords for each identified topic.
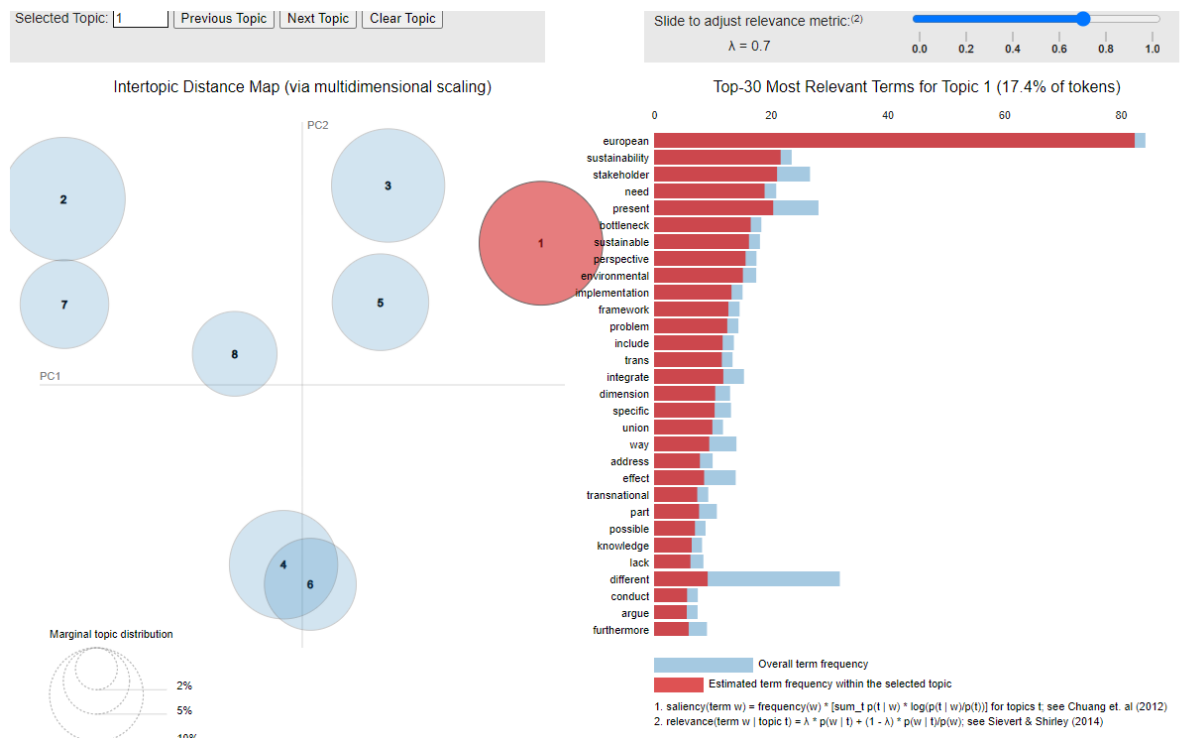
Figure 3: Intertopic Distance Map for tuned model, N=8. There are no major overlaps in topics, the relevance metric λ was set in this sample of topic 1 to 0.65, determining the weight given to the probability of each term within that specific topic. Higher values of λ emphasize more relevant terms, where lower values include a broader range of terms in the analysis.

2.4. Topic Analysis

In the process of topic analysis, it is important to note that the identified topics do not have meaningful interpretation without additional qualitative analysis. To decode the content of the topics, an exploration of the top 10 words associated with each topic was undertaken. Additionally, to provide a deeper understanding and interpretation of topics, the 5 most relevant abstracts related to each topic were also examined, revealing the underlying themes and content within the corpus. By performing this analysis, a comprehensive overview of the eight identified topics was achieved. Finally, each topic was assigned a label that effectively captures its content (see Results section).

2.5. Bibliometric Analysis

A bibliometric analysis is a quantitative method used to examine and evaluate scientific literature based on bibliographic data, that relies on the aggregation and reorganization of citation information from academic journals (Vogel, 2014). By analyzing citation patterns, researchers can evaluate the influence, visibility, and collaboration networks associated with academic articles from a particular field. In general, bibliometric analysis serves as a valuable tool for unraveling and visualizing the cumulative scientific knowledge and dynamic trends within established fields (Donthu et al., 2021). Consequently, well-executed bibliometric analyses can provide a comprehensive overview of the field, identify areas of knowledge gaps, and generate novel ideas for investigation. Co-authorship networks are an essential part

of bibliometric analysis, with a particular focus on investigating science and technology partnerships. These networks provide valuable insights into the patterns of scientific collaboration by representing authors as nodes and establishing connections between them when they collaborate on a paper (Newman, 2004). Despite ongoing discussions regarding its significance and interpretation (Beaver, 2001), co-authorship analysis remains extensively employed in understanding collaboration dynamics in scientific research (Sampaio et al., 2016; Liu et al., 2005).

For the bibliometric analysis, R, an open-source programming language and environment, was utilized. The analysis involved the use of Biblioshiny, a package developed by Massimo Aria, which is an extension of the Bibliometrix package in R. The interactive web interface offered by Biblioshiny simplifies scientific measurement and visual analysis by providing a user-friendly platform, which successfully reduces the complexity of information input to a certain extent (Xu et al., 2022). To conduct the analysis, all the papers from the dataset were sought on Scopus to export them into BibTeX format with the required bibliographic details, including title, author, year of publication and citation references. However, four papers could not be found on Scopus, necessitating their manual addition to the dataset in BibTeX format without including the citation references.

Visualization is a crucial aspect of network analysis as it enhances the understanding and interpretation of the data, thereby complementing the analysis process. For the visualization of the co-citation network the Python package Networkx was employed, facilitating the identification of communities or groups of authors who have collaborated in the past. In network analysis, communities are groups of nodes that share common characteristics or similar properties (Dalmasso et al., 2013). Community detection is a widely used technique, which aims to identify a cohesive set of nodes that are more densely connected within the set than with nodes outside of it (Harenberg et al., 2014).

Analyzing social networks and understanding the role of their actors requires the use of centrality measures (Ni et al., 2011). Degree centrality is a commonly used measure that counts the number of connections a node has in the network. The most connected actors are the core actors in terms of degree centrality, showing a high level of activity and connectivity (Wasserman & Faust, 1994). Closeness centrality, on the other hand, assesses how close a node is to all the other nodes in the network (Wasserman & Faust, 1994), measuring the influence and accessibility of this node within the network. Lastly, betweenness centrality according to Borgatti (2005), examines the extent to which a node lies between other nodes, indicating its potential control over the flow of information.

In this study, centrality measures were initially computed for each node in the network. Subsequently, the Louvain community detection algorithm was applied to identify distinct groups or communities among the authors. By analyzing the connectivity patterns and interactions between the nodes, the Louvain algorithm effectively partitioned the network into communities, providing insights into the author groups within the network.

3. Results

The results section is structured to offer a comprehensive overview of the findings. Initially, a chronological analysis of the research field is presented, highlighting key developments and milestones. Subsequently, the frequency analysis's findings are then provided, offering insights into the occurrence of words within the literature corpus. Moving forward, the topic modeling analysis is explored, uncovering the underlying themes. Finally, the results of the bibliometric analysis are presented, examining quantitative measures of scholarly impact.

To initiate the analysis, the number of publications per year over time was explored, providing insights into the chronological progression of the research field (Figure 3). More specifically, initially, the scholarly attention towards corridors remained limited until the early 2000s. The cumulative number of papers focusing on corridor development during the period from 1995 to 2002 was only eight (as shown on the right axis of Figure 3), while these early papers primarily centered around corridors as a spatial planning concept. For example, Priemus (2001) explored corridors as a spatial planning tool in the Netherlands, Lambert (1998), and Williams et al. (2002) in the context of the UK and Ireland, while Kunzmann (2001) examined corridor development in Germany. Among these papers, Reggiani et al. (1995) played a pivotal role in capturing the diversity and complexity of corridors and identifying common characteristics, thus providing the initial framework for the academic discussion on corridor development.
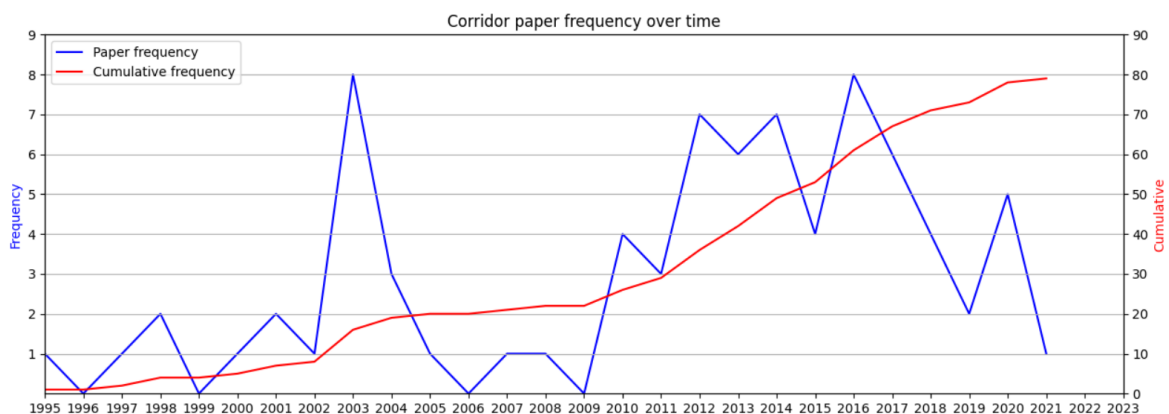

Figure 4: Publications per year over time

The publication of the Special Issue on corridor governance in the Journal of Transport Geography in 2003, guest edited by Priemus and Zonneveld (2003), marked a significant milestone in the academic exploration of corridors. This year a total of eight papers were written, delving into various aspects of corridor governance and focusing on corridors as transnational infrastructure axes, often referred to as "megacorridors". Chapman et al. (2003) was one of the key contributions of this Special Issue, which provided a comprehensive definition and understanding of corridors. Another notable aspect was the exploration of

corridor positioning in networks by Albrechts and Coppens (2003), and the governance challenges associated with the corridors. Specifically, Romein et al. (2003) and De Vries and Priemus (2003) delved into the complexities of multi-scalar and multi-level governance, highlighting the need for effective collaboration among various stakeholders, while Schönharting et al. (2003) examined the concept of multi-modality within the context of corridors. Following the Special Issue, the level of attention towards corridors diminished in the subsequent years up to 2010. However, throughout this period substantial studies were published that greatly improved the systematic and conceptual understanding of corridors, with a specific focus on the North-West European region. Especially the publications of Jean-Paul Rodrigue (Rodrigue, 2004; Hesse and Rodrigue, 2004; Notteboom and Rodrigue, 2005) emphasized the integrated nature of corridors as planning concepts.

From 2010 onwards, there has been a notable and consistent rise in the number of publications focused on corridor development, with this trend continuing until 2018. This increase can be attributed to several factors. Firstly, the completion of the TEN-T program and the subsequent establishment of the Connecting Europe Facility, which introduced core network corridors (European Commission, 2011), acted as a catalyst for renewed interest in the field. Secondly, during this period, specific authors emerged who focused their studies on the field of corridor development, among them Witte (2004). Furthermore, the involvement of some researchers in EU-funded projects related to corridor development (e.g., Lami, 2014; Drewello and Scholl, 2015), further fueled scholarly engagement in the topic. Thirdly, there has been a remarkable increase in the geographical diversity of corridor-related research. Although Europe and North America continued to receive significant attention, there has been a growing emphasis on exploring corridors in Asia (Regmi and Hanaoka, 2012; Wang and Ducruet, 2014), Pacific (Olaru et al., 2011; Rolfe et al., 2013), Middle East (Babalik-Sutcliffe, 2013) and Africa (Fraser and Notteboom, 2014).

In recent years, the study of corridors has witnessed ongoing trends, along with two significant developments. Firstly, the Chinese Belt and Road Initiative (BRI) has gained prominence, leading to a greater focus on its land bridge and international or even cross-continental transport routes (e.g., Yang et al., 2018). Secondly, there has been a growing emphasis on "green" or "sustainable" corridor development. Researchers have explored various aspects, such as reducing CO_2 emissions (Nocerra and Cavallaro, 2016), preserving biodiversity (Karlson et al., 2016), and incorporating sustainability principles (Maksin et al., 2017; Öberg et al., 2017; Tsigdinos et al., 2021). This trend builds upon the earlier emphasis on optimizing corridors and mitigating negative impacts since the early 2010s.

3.1. Frequency analysis

Subsequently, a frequency analysis was conducted to examine the occurrence of words within abstracts, providing insights into the prominence and prevalence of specific terms in literature. Figures 5 and 6 use a bar chart and a word cloud to display the top 20 and top 50 most frequently used words, respectively. Notably, the presence of the word "European" as the second most used word indicates that the literature has a particular focus on transport

corridors within Europe (e.g., Otsuka et al., 2017; Romein et al., 2003). Moreover, it is worth noting that a considerable number of papers delve into the policy frameworks, initiatives, and regulations implemented by the European Union to address challenges in transportation networks (e.g., Peters, 2003; Schönharting et al., 2003; Witte et al., 2012; Drewello & Scholl, 2015) and promote sustainable mobility (e.g., Janic & Vleugel, 2012; Panagakos et al., 2015; Öberg et al., 2017; Tsigdinos et al., 2021). Furthermore, the inclusion of the words "spatial", "planning", "economic", "policy" and "governance" suggests that the academic literature takes an interdisciplinary approach and incorporates concepts from various fields such as urban planning (Maksin et al., 2017), economics (e.g., Bruinsma et al., 1997; Wiegmans et al., 2020), policy-making (e.g., Williams et al., 2002), and governance (e.g., Priemus and Zonneveld, 2003; De Vries & Priemus, 2003; Ng et al., 2015). This interdisciplinary perspective allows for a holistic analysis and comprehensive planning of transport corridor development. Furthermore, the inclusion of the word "port" underscores the significance of this transportation node in facilitating the efficient and effective movement of goods and services within transport corridors and promoting international trade (e.g., Notteboom & Rodrigue, 2005; Wilmsmeier et al., 2011; Monios & Wilmsmeier, 2012). Also, the frequent occurrence of the word "rail" indicates a particular emphasis on rail infrastructure as one of the modalities within transport corridors. This suggests that a significant number of papers have examined various aspects related to rail corridors, such as planning, design, and policy considerations (e.g., Romein et al., 2003; Abramović et al., 2016; Karlson et al., 2016; Long & Zhang, 2019). Additionally, the prevalence of the word "freight" indicates that this type of corridor represents a substantial body of literature. Numerous studies are dedicated to examining the development and performance of freight corridors (e.g., Schönharting et al., 2003; Wilmsmeier et al., 2011; Wiegmans & Janic et al., 2019; Jiang et al., 2020 among others). Moreover, the term "infrastructure" reveals an investigation into the physical and operational aspects of transport corridors (Wilmsmeier et al., 2011; Regmi & Hanaoka, 2012), reflecting an understanding of the role of well-connected infrastructure networks in facilitating efficient transportation and enhancing overall corridor performance. Besides infrastructure, the occurrence of words such as "regional" and "urban" suggests that transport corridors are not solely focused on operational aspects but also encompass spatial and economic development considerations (e.g., Bruinsma et al., 1997; Williams et al., 2002; Priemus & Zonneveld, 2004). This indicates a broader perspective on the role of corridors in promoting regional connectivity and supporting urban growth.

In summary, the findings from these figures provide valuable insights into the academic literature on transport corridor development, indicating a clear acknowledgment of the complex nature of corridor development.

Figure 5: Frequency of the top-20 high-frequency words

Figure 6: Word cloud of the top-50 high frequency words

3.2. Topic Analysis

The following section presents a comprehensive exploration of the emergent themes from the LDA, providing valuable insights into the research landscape and highlighting the aspects that researchers have concentrated on in their studies of corridors. The resulting labels for the 8 topics obtained from the LDA analysis can be found in Table 1.

1) *Environmental sustainability:* This topic addresses a crucial aspect of corridor development. The aim is to encourage sustainable practices and minimize the negative effects of transportation systems on the environment. By recognizing the importance of balancing economic growth with environmental concerns, corridor development has undergone a significant transformation in recent years. Initially, the primary focus was on the economic potential of transport corridors. However, a notable trend has emerged since 2014, giving rise to the concept of "green transport corridors". The studies reviewed in this topic are primarily centered on Europe and highlight key areas of focus within corridor development, including reducing CO_2 emissions and minimizing costs in the freight transport sector (Blinge, 2014), analyzing the impact of sustainable practices on supply chains (Panagakos et al., 2015), predicting the effects on sustainable territorial development in waterway corridors (Maksin et al., 2017), and transforming urban car traffic arterials into multimodal corridors that promote sustainable mobility (Tsigdinos et al., 2021).

2) *Railway corridors:* A railway corridor can be defined as a transportation route consisting of a railway line that operates within an urban or regional level and encompasses the surrounding areas, including the stations located along the line (Chorus & Bertolini, 2016). Railway corridors play a crucial role in corridor development, with a focus on optimizing sustainability and integrating public transport and land-use developments. The analysis of transport and land-use development dynamics within railway corridors reveals a strong relationship between the two, underscoring the importance of careful planning to enhance transport operations' efficiency. Several papers on this theme address the concept of transit-oriented development in passenger transportation to enhance connectivity and accessibility and emphasize the need for more sustainable transportation by proposing the replacement of road connections with rail (e.g., Bertolini & Dijst, 2003; Curtis, 2008; Chorus & Bertolini, 2016; Olaru et al., 2011; Zimny-Schmitt & Goetz, 2020).

3) *Governance of corridors:* This theme refers to the processes and mechanisms through which various stakeholders, including transport authorities (local, regional, national, and transnational), non-governmental organizations, private companies, and other relevant actors, collaborate and coordinate their efforts to plan, develop and manage corridors (De Vries & Priemus, 2003; Öberg, 2014). The literature on this topic provides insights into corridor governance, with a particular focus on transnational axes, commonly referred to as megacorridors (Chapman et al., 2003). These studies also highlight the impact of socio-economic and spatial changes on corridor development and the formulation of government policies. Additionally, they emphasize the necessity of an enhanced governance framework that encompasses

aspects such as infrastructure, urbanization, and economic development (De Vries & Premius, 2003). Therefore, it is worth mentioning that this topic intersects with other areas of research. It is closely related to Topic 1, which explores the promotion of environmentally friendly practices within corridors (Marshall, 2014; Blinge, 2014), as well as Topic 4, which examines the maximization of the economic benefits derived from corridor development (Witte et al., 2014). Additionally, corridor governance plays a significant role in the context of urban planning (Topic 6), ensuring that transport infrastructure aligns with urban development objectives (Williams et al., 2002). By considering the interconnected nature of corridor governance with other topics, researchers gain a comprehensive understanding of how governance practices influence various fields within transport corridors.

4) *Economic impact of corridors:* The economic impacts of corridors have been a subject of extensive research, with numerous studies attempting to assess the relationship between corridor development and regional economic development. Articles in the literature delve into this aspect by examining the economic impacts of specific corridors, including examples such as corridors in the Netherlands (Bruinsma et al., 1997; Priemus, 2001), Rockhampton in Australia (Rolfe et al., 2013), the main transalpine corridors (Nocera & Cavallaro, 2016), the Rhine- Danube Core Network Corridor (Maksin et al., 2017), and the China-Pakistan economic corridor (Ahmed, 2018) among others. By examining these impacts of corridors on various aspects, such as corporate investments, employment levels, accessibility, travel time, and accuracy in delivery times, researchers aim to contribute to a better understanding of corridors' potential as tools for economic growth and inform decision-making processes. It is through such assessments that the significance of corridors in driving economic development can be fully comprehended, leading to the formulation of effective policies and strategies to maximize their positive impacts.

5) *Inland and dry ports:* This topic focuses on the analysis and exploration of port and inland integration, with a specific emphasis on the institutional challenges and strategies associated with inland ports, including dry ports. Inland ports play a crucial role in facilitating the accessibility of ports to the inland distribution system, acting as key nodes along major transportation corridors. As the functions of port terminals evolve, the development of inland terminals becomes essential to accommodate the increasing port-inland linkages. The research investigates the formation of transport and supply chain functions within inland ports, examining how different actors contribute to their planning and establishment operation. By employing effective challenges, the studies on this topic (e.g., Notteboom & Rodrigue, 2005; Rodrigue et al., 2010; Wilmsmeier et al., 2011) provide valuable insights into enhancing the performance and integration of inland ports within broader transportation systems.

6) *Urban planning:* Urban planning within the context of corridor development has been recognized as an important topic in the literature. Studies have highlighted the significance of urban planning in designing and shaping cities and urban areas to optimize connectivity, accessibility, and livability. Researchers in this discipline have discovered that transportation networks, including corridors, can be effectively organized to optimize connectivity, accessibility, and livability for both residents and

businesses (e.g., Marshall, 2014; Long & Zhang, 2019; Zimny-Schmitt & Goetz, 2020). Additionally, by incorporating spatial planning elements, corridors can be designed to preserve natural and cultural elements (Kunzmann, 2001), resulting in harmonious urban environments. In this way, corridors emerge as dynamic and interconnected components that greatly enhance the overall quality of urban life.

7) *Corridor development in China:* This theme has gathered particular attention in the literature, with a particular focus on the post-2012 period, coinciding with the initiation of the BRI. China has been actively investing in and developing transportation corridors both domestically and internationally as part of its Belt and Road Initiative (BRI). The BRI aims to enhance connectivity and foster economic cooperation across six major economic corridors that span China and various regions, including Mongolia and Rusia, Eurasian countries, Central and West Asia, Pakistan, other countries of the Indian sub-continent, and Indochina (OECD, 2018). These corridors serve as crucial pathways for trade, infrastructure development, and economic integration. Scholars have increasingly focused on the land bridge function of BRI, exploring the anticipated transport routes from a global perspective (Yang et al., 2018). As attention shifts beyond mainland China, the academic discourse on the BRI has expanded to cover the international implications and potential benefits of these corridors. This topic encompasses studies by Regmi & Hanaoka (2012), Wang & Ducruet (2014), Ahmed (2018), and Long & Zhang (2019), among others.

8) *Capacity in transport corridors:* This topic refers to the ability of a transportation system to handle the volume of traffic and goods efficiently. It encompasses various factors such as infrastructure, resources, and operational capabilities. In the context of corridor development, enhancing capacity becomes crucial to meet the demands associated with transportation networks. This involves not only physical improvements, such as expanding and upgrading existing infrastructure but also institutional capacity building, which refers to the overall quality of the relational networks among existing governments and non-governmental organizations involved in corridor development (e.g., De Vries & Priemus, 2003; De Borger et al., 2007). Other papers that contribute to the understanding of this topic include Lambert, (1998), Schönharting et al, (2003), Monios & Lambert, (2013), Witte et al., (2014) among others.

Table 1: Topic Names and top 10 words for corridor development related topics

ID	Topic Names	Topic keywords
T1	Environmental Sustainability	European sustainability stakeholder present need bottleneck sustainable perspective environmental implementation
T2	Railway corridors	rail method use propose land time service model base link
T3	Governance of corridors	planning regional project infrastructure national impact governance plan future region

T4	Economic impact of corridors	freight route growth economic provide measure global context social emerge
T5	Inland and dry ports	port inland institutional strategy challenge integration finding require private function
T6	Urban Planning	urban city current core accessibility indicate activity relationship concern connectivity
T7	Corridor development in China	performance china road economic country trade assess datum international cost
T8	Capacity in transport corridors	system capacity structure increase rhine number investment build export traffic

3.3. Bibliometric Analysis

The following section presents a comprehensive bibliometric analysis, which encompasses various aspects, including an examination of the main researchers in the field, an analysis of the most cited papers, an exploration of the co-authorship network, and an analysis of the distribution characteristics of major research countries. Each of these analyses sheds light on different dimensions on corridor development research, offering a deeper understanding of the scholarly contributions and collaborations within this domain.

3.3.1 Analysis of Main Researchers

The paper data set on corridor development involved a total of 155 authors. Among them, the top three authors based on the number of publications were Witte, Wiegmans, and Spit, with 9, 7, and 7 publications, respectively (as shown in Table 2). Witte, in particular, had the highest number of publications, with an h-index of 6, a g-index of 9, and a total number of citations reaching 136. The h-index, proposed by Hirsch (2005), is a metric used to assess the impact and significance of a researcher's body of work. It considers both the number of publications and the number of citations those publications have received. For instance, an h-index of 10 indicates that an author has published at least 10 papers that have received 10 or more citations. The g-index, introduced by Egghe (2006), is a metric that gives more weight to highly cited articles. It is defined as the largest value g for which the top g articles, ranked in descending order based on the number of citations, received at least g2 citations.

According to the information presented in Figure 7, the most frequently cited paper by Witte, published in 2014, has received a total of 57 citations. Interestingly, this paper was co-authored by Spit, Wiegmans, and Van Oort, making it the most cited paper among these authors as well. This article entitled "Governing inland ports - A multi-dimensional approach to addressing inland port-city challenges in European transport corridors", explores the application of the port-city concept to inland ports. The results of the study reveal a balance between positive and negative externalities associated with these ports and highlight that

proactive governance strategies are the most effective in managing conflicting land uses and accommodating the challenges arising from the interaction between the port and the surrounding city (Witte et al., 2014).

According to Hirsch (2007), the h-index serves as a valuable tool not only for evaluating the importance, significance, and broad impact of authors' cumulative research conditions but also for assessing their current publication output and predicting their future performance. However, relying solely on the h-index for comparing the influence of authors can be misleading, as it disregards the citation information (Zhang, 2009). To address this limitation, the m-index can be introduced as m=h/n, which is the ratio of the h-index to the age of the author's publication in the specific field of study (Roldan-Valadez et al., 2019).

Witte and Wiegmans achieved the highest m-index, which was calculated as 0.5. Both actors have been actively publishing papers in the field of corridor development since 2012. Interestingly, their most cited paper emerged two years after they started their research in the field.

Table 2: Top 10 influential authors in the field of corridor development

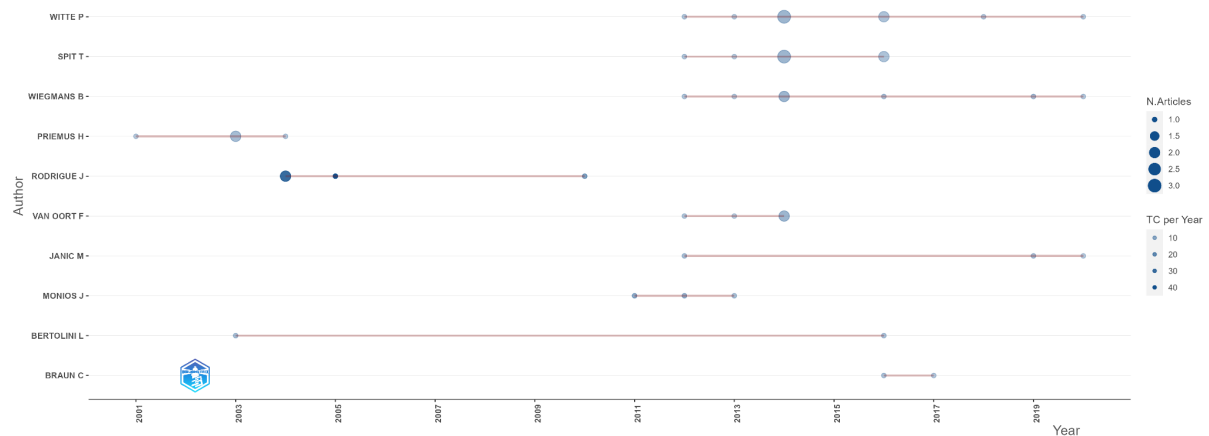| Author	h-Index	g-index	m-Index	TC	NP	PY-Start
Witte	6	9	0.500	136	9	2012
Wiegmans	6	7	0.500	147	7	2012
Spit	5	7	0.417	124	7	2012
Priemus	4	4	0.174	132	4	2001
Rodrigue	4	4	0.200	1559	4	2004
Van Oort	4	4	0.333	102	4	2012
Janic	3	3	0.250	72	3	2012
Monios	3	3	0.231	269	3	2011
Bertolini	2	2	0.095	123	2	2003
Braun	2	2	0.250	30	2	2016

Figure 7: Authors' production over time in the field of corridor development

3.3.2. Analysis of most cited papers

The most cited paper in the field of corridor development is "Port regionalization: towards a new phase in port development", authored by Notteboom & Rodrigue, (2005), as it can be seen in Figure 8. While not explicitly a "corridor paper", it has played a crucial role in shaping the discourse around corridor development. This article introduces the concept of port regionalization as a progressive stage in the evolution of port systems. It expands on existing theoretical frameworks, such as the Anyport model by Bird, by incorporating the concept of hinterland connectivity and considering the broader regional impacts of port development. The results of this paper indicate that regionalization, driven by logistic decisions and actions of shippers and third-party logistics providers, plays a crucial role in port development, emphasizing the importance of inland distribution, efficient hinterland access, and active engagement of port authorities in addressing port- related challenges.

The second most cited paper in the field of corridor development, written by Hesse & Rodrigue (2004), titled "The transport geography of logistics and freight distribution", provides an overview of the evolving perspective on transportation as an integrated demand influenced by logistical requirements. The paper explores the fundamental aspects of transport geography, including flows, nodes, and networks, while introducing the concept of logistical friction. The findings highlight that transportation cannot be solely regarded as a derived demand, but rather as an interconnected component of integrated demand that incorporates both physical distribution and materials management.

Indeed, the fact that three out of the four most cited papers in the field of corridor development were written (or co-written) by Rodrigue highlights the significant influence of this author in the domain. Additionally, the years 2003, 2004, and 2005 played a crucial role in shaping the evolution of corridor development, as shown by the high impact and citation rates of papers published during that period.
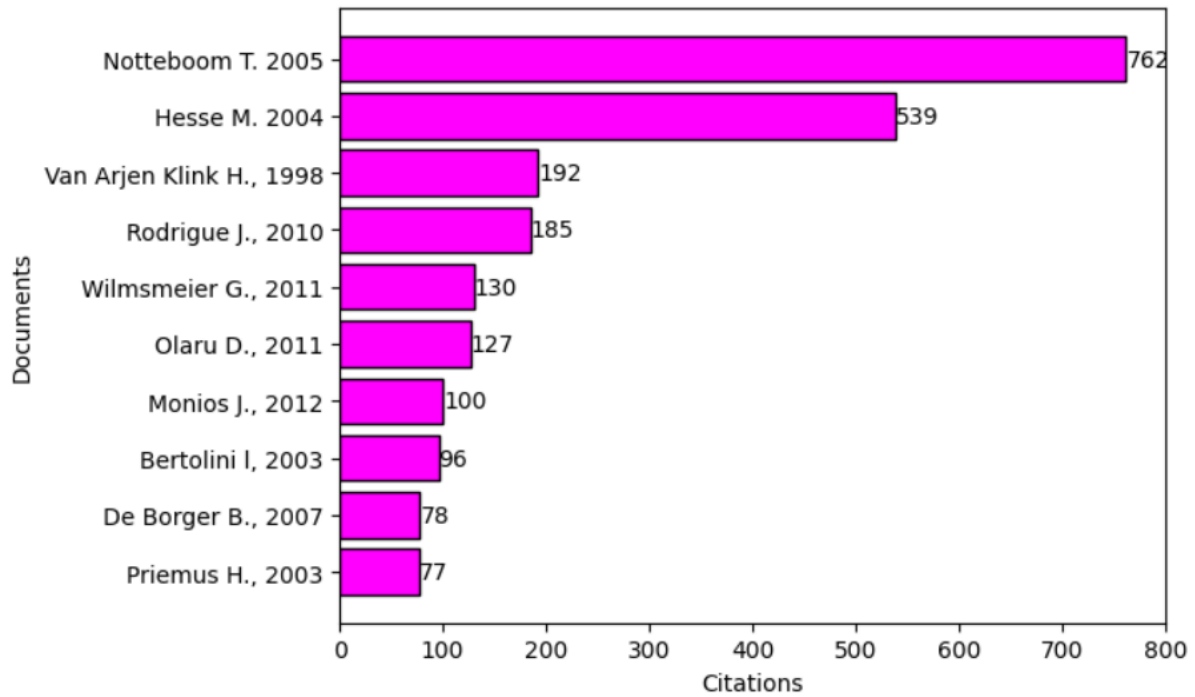
Figure 8: Most cited papers in the field of corridor development

3.3.3. Analysis of co-authorship network

Table 3 presents the top 5 authors ranked based on degree, betweenness, and closeness centrality measures. Degree centrality captures the extent of an author's collaboration scope, betweenness centrality reflects an author's influence over the spread of information in the network, measuring the extent to which they act as a bridge between other authors, and closeness centrality assesses an author's position and virtual distance to others in the field, indicating how easily they can access and disseminate information (Umadevi, 2013).

Based on the degree centrality measure, Witte emerges as the most well-connected author in the co-authorship network (Figure 9, top image). Similarly, Braun and Witte rank highest in terms of betweenness and closeness centrality, respectively (Figure 9, middle and bottom images). Notably, there is a considerable overlap among the top-ranked authors across these centrality measures. This suggests that these authors (Witte, Wiegmans, Braun, Rodrigue) have a significant presence in the network, combined with strong influence and active engagement in shaping the flow of information.

Figure 10 provides a visual representation of the communities identified through the community detection algorithm. Each community is represented by a different color, aiding in the identification and differentiation of distinct groups of authors within the network. Among these communities, the green community on the right side of the figure stands out as the largest one, consisting of authors such as Witte, Wiegmans, van Oort, Spit, Braun, Straalen, Janic, Jong, Otsuka, Günther, Tosoni, and Vleugel. It is worth noting that the papers associated with this green community were written between 2012-2020. Furthermore, a significant number of these papers are linked to the INTERREG IV-B funded project "Code

24-Corridor Development Rotterdam-Genoa" (2010-2015), which focused on the collaborative and integrated development of the TEN-T core network Rhine-Alpine corridor. This project aimed to integrate economic growth with spatial, transport, and ecological considerations, addressing various challenges related to capacity, sustainability, and quality of life along the corridor (EGTC, n.d.).

Table 3: Top five ranked authors based on centrality measures. Centrality values were normalized in accordance with the size of the network

Author Name	Degree Centrality	Author Name	Betweenness Centrality	Author Name	Closenness Centrality
Witte	0.052	Braun	0.0027	Witte	0.06
Wiegmans	0.045	Witte	0.0023	Wiegmans	0.056
Braun	0.045	Ng	0.0013	Braun	0.053
Rodrigue	0.037	Rodrigue	0.0012	Spit	0.048
Rooney	0.030	Wiegmans	0.0012	Janic	0.043

Degree
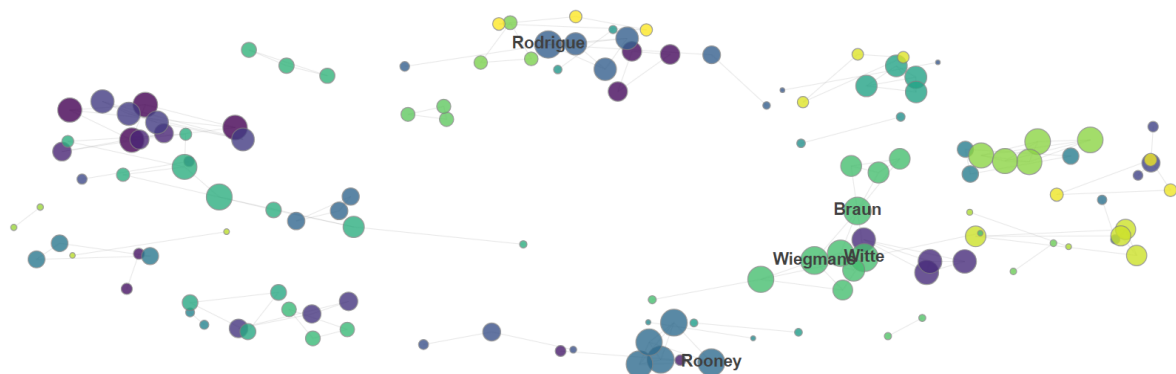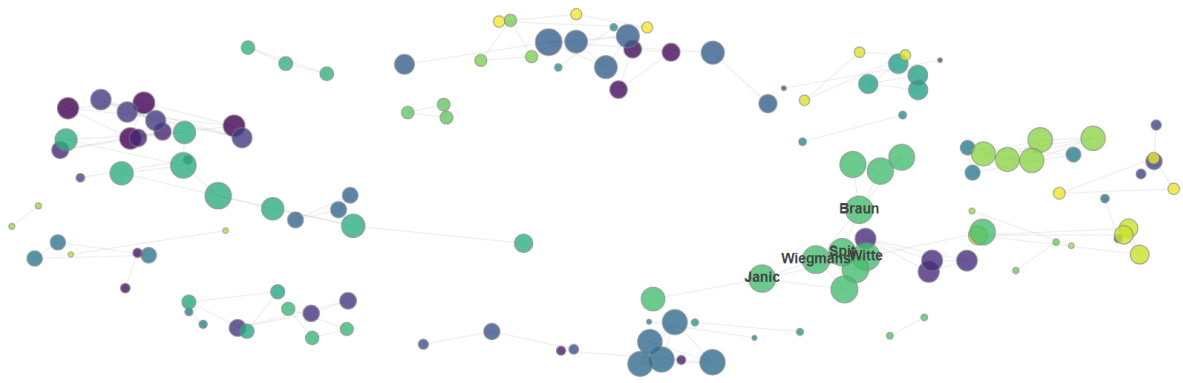

Betweenness
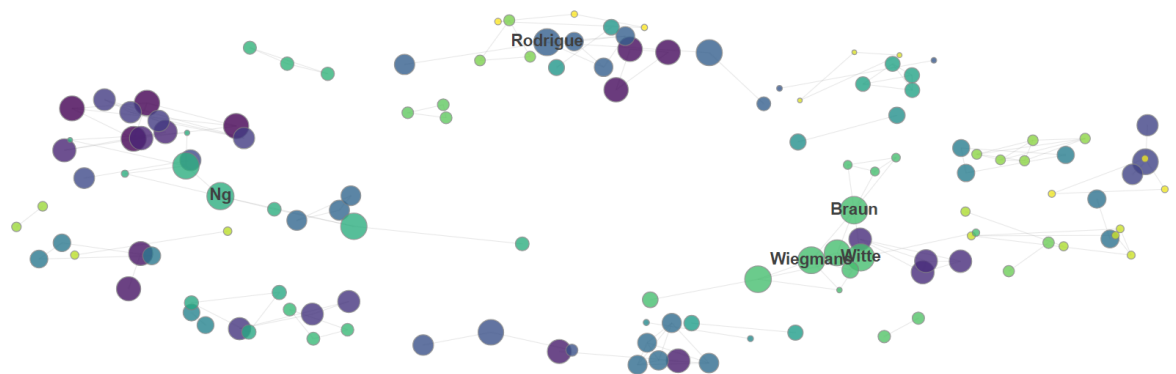
Closeness

Figure 9: The top three authors with highest centrality, according to each metric. The size of the nodes indicates their degree centrality (top image), betweenness centrality (middle image) and closeness centrality (bottom image)
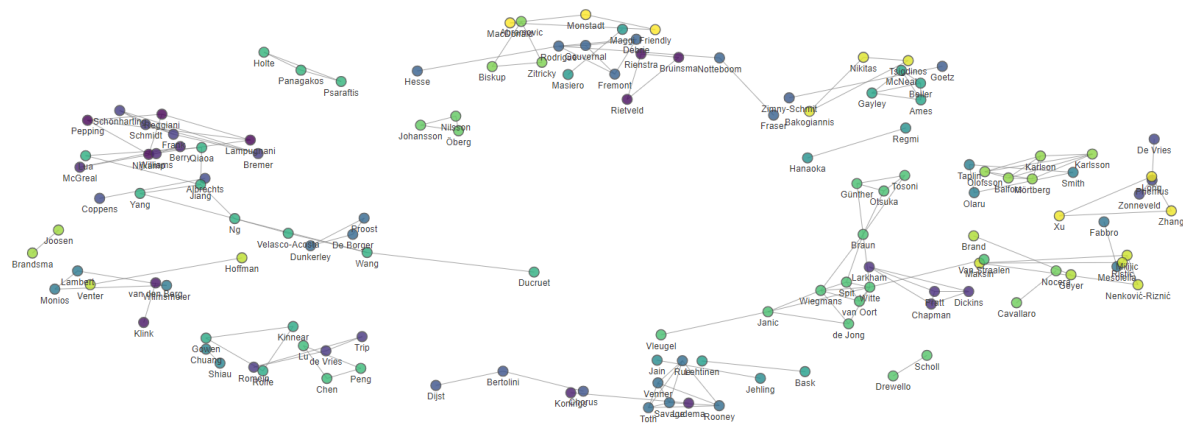

Figure 10: Community detection results

3.3.4. Analysis of Distribution Characteristics of Major Research Countries

The distribution characteristics of major research countries/regions in the field of corridor development reflect the influence and contributions of each country in this domain. In the dataset analyzed for this article, publications were identified from 25 countries or regions. Among the top 10 publishes papers, the distribution was as follows: four European countries (Netherlands, Belgium, United Kingdom, and Germany), one American country (United States), and one Oceania country (Australia).

Figure 11 illustrates the publication distribution, where darker colors indicate higher paper counts. It can be observed that papers on corridor development were predominantly published in the Netherlands, the United States, and China. Notably, the Netherlands account for approximately 45% of the total production, with 35 articles published there. The high production of papers in the Netherlands can be attributed to several reasons. Firstly, the Netherlands benefits from its geographical location and well-developed transport infrastructure. Situated in a strategic position within Europe's transportation network, the country has efficient ports, airports, and an extensive road and rail network. This

infrastructure facilitates corridor development activities, leading to a higher number of scientific publications originating from the Netherlands. Also, a significant number of papers were written as part of the INTERREG IV-B funded project 'Code 24', which involved four countries, including the Netherlands. This project likely contributed to a considerable number of publications coming from the Netherlands. Additionally, the Dutch government's commitment to sustainable transportation and regional development has played a crucial role (en Waterstaat, 2013). The government has implemented policies and provided funding support that encourages research and innovation in the field of corridor development. Lastly, the successful lobbying efforts of the Dutch government to have their corridors included in (TEN-T) corridor network has had a profound impact on research in this field within the Netherlands. These factors have contributed to the Netherlands emerging as the country with the highest number of papers in the dataset, indicating its prominent role in the field of corridor development.
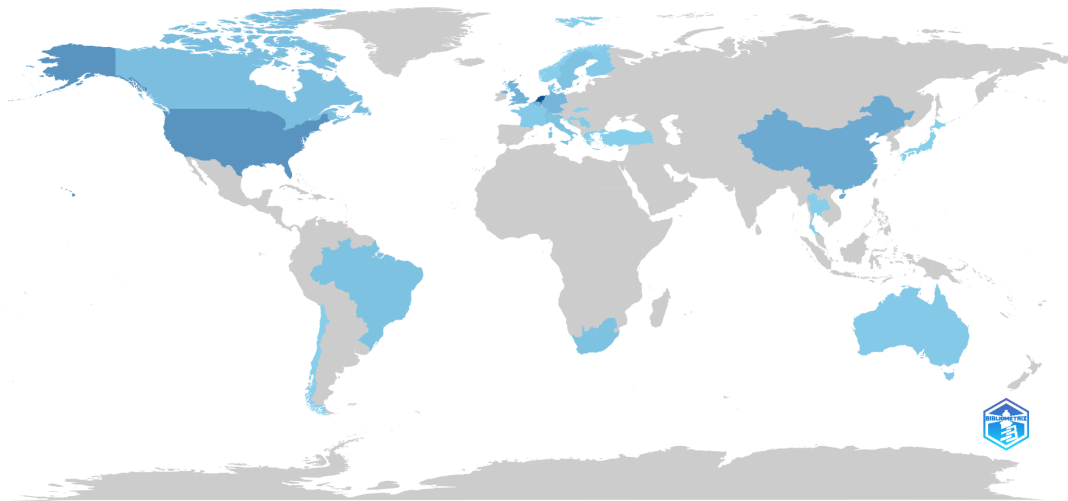

Figure 11: Scientific production distribution in the field of corridor development

4. Discussion

This section encompasses the significant conclusions drawn from the analysis and the interpretation of results while providing a critical reflection on potential limitations, which can serve as a valuable reference for future research.

4.1. Conclusions

The literature review analysis conducted in this study aimed to investigate the theoretical perspectives employed in corridor research, to identify the common areas of consensus, and to better understand the scholarly influence on the field of corridor development. The topic modeling analysis revealed the complex and interconnected nature of transport corridors, emphasizing the need to consider multiple dimensions when investigating them. By uncovering the relationships and dependencies among these factors, the analysis provided

valuable insights for decision-making, policy formulation, and strategic planning. These insights include a deeper understanding of the interconnections between different aspects of corridor development, the potential environmental implications, the economic opportunities, the importance of effective governance, and the role of various infrastructure components. Furthermore, the findings shed light on the evolving trends and interests in transport corridors, with a significant focus on Europe and a growing shift towards Asia influenced by the Belt and Road Initiative (BRI). The research also acknowledged the substantial contribution of the Netherlands, demonstrating its expertise and influence in shaping the discourse on transport corridors. Moreover, the identification of pivotal years, specifically 2003-2005, highlighted the foundation role played by papers published during that period, which subsequently paved the way for further exploration and research in corridor development. Lastly, the co-authorship network analysis revealed isolated research efforts, suggesting that collaborations have often been limited to individual papers without sustained partnerships and interconnections.

The present study offers a unique contribution to the existing literature on corridor development, as there is currently a noticeable dearth of research exploring this specific area. It stands out as the only literature review analysis in this field that utilizes computational methods for analysis, providing a novel approach to understanding and examining the subject matter. By employing computational techniques, this study brings a fresh perspective that has not been previously explored in the existing literature.

4.2. Limitations

Although the results of this research provided valuable insights, several limitations should also be considered. Firstly, the dataset's timeframe only includes papers published until 2021, which may exclude relevant research published afterwards. Additionally, the aforementioned papers were written in English, excluding non-English publications. Possible limitations concerning topic modeling should also be taken into consideration. To begin with, the selection of an appropriate number of topics is not a straightforward task and lacks a definitive method. Consequently, this can lead to the emergence of vague and ambiguous topics, diminishing the quality and interpretability of the results. The preprocessing steps, such as stemming and stopword removal may also have influenced the findings of this study. It is also well known that topic modeling only considers word co-occurrence patterns and does not take into account the context in which words appear, leading to limitations in the contextual information that is captured. However, to minimize the effect of such limitations, the validation of individual abstracts and interpretation was considered carefully. Finally, in this paper, other models beyond LDA were not explored, however, extensions and other models of interest exist. Consequently, it is crucial to consider cautiously the presented results and proceed with further research to overcome the aforementioned limitations.

4.3. Future work

Building upon the identified limitations, this section presents innovative solutions and future directions aimed at addressing these challenges and advancing the field. To start with, future research could employ a more comprehensive approach for retrieving papers by combining multiple academic search engines such as Scopus, Google Scholar, and Web of Science. This broader search strategy would allow for a broader coverage of relevant topics, including research published beyond the cutoff date of 2021. Furthermore, incorporating non-English publications into the search can provide a more diverse and inclusive perspective on corridor development. Additionally, beyond LDA several other alternative models should be explored that have the potential to provide additional insights while also overcoming the limitations of the chosen model. One such alternative model is the Hierarchical Dirichlet Process (HDP), which allows for an automatic determination of the number of topics and hierarchical relationships among them (Wang et al., 2011). By considering the HDP model, a more flexible and nuanced representation of the underlying topics could have been potentially obtained. Another alternative worth exploring is the Non-Negative Matrix Factorization (NMF) model, which has been successfully applied in topic modeling tasks. A notable advantage of NMF is its ability to provide consistent results across multiple runs on the same set of documents, enhancing the reliability and robustness of its findings (Choo et al., 2013). To overcome the word co-occurrence patterns because of topic modeling, efforts should be made to incorporate contextual information to provide a more nuanced understanding of the topics. To do so, more advanced natural processing techniques should be utilized in the future. One such technique is BERTopic, which utilizes Bidirectional Encoder Representations from Transformers (BERT) embeddings to capture the contextual meaning of words by considering their surrounding words (Grootendorst, 2022). This approach improves the representation of topic relationships and themes, leading to a more accurate analysis and understanding of textual data.

In the context of corridors, one key observation that should be considered is that the academic debate on them appears to be fragmented, and research efforts often occur in isolation. This is evident from the co-authorship network analysis (Figure 10), which indicated that collaborations primarily revolve around individual papers rather than establishing sustained partnerships or interconnections among researchers. To address this, future research should adopt an interdisciplinary approach, exploring how transport corridors intersect with fields such as economics, environmental studies, urban planning, or political science. This integrated perspective will provide a more holistic and comprehensive understanding of the multifaceted aspects of corridor development and its impacts across various domains. Furthermore, promoting collaboration and knowledge exchange among researchers from different continents is essential. Embracing a global perspective will enable cross-cultural learning and facilitate the sharing of best practices. Additionally, it is recommended to conduct an integrated policy analysis of empirical outcomes from different corridor projects utilizing data science techniques, such as data mining, machine learning, and statistical modeling. This analysis will allow researchers to analyze large and diverse datasets, identify correlations, and gain deeper insights into the impacts and effectiveness of various policy

interventions. By adopting this data-driven approach, policymakers can receive evidence-based recommendations, make informed decisions, and enhance policy formulation in the field of corridor development. Lastly, it is important to address the challenge of limited data availability in corridor research. To overcome this obstacle, it is crucial to establish data-sharing mechanisms among researchers. Data-sharing platforms will provide researchers with tools to upload, access, and retrieve shared data, facilitating collaboration and enhancing the availability of data for corridor research. Additionally, data science techniques, such as cleaning and preprocessing, can be used to integrate data for different sources and formats. By promoting collaboration and data pooling, a more robust foundation for analysis and decision-making at a global level can be achieved.

References

Abramović, B., Zitricky, V., & Biškup, V. (2016). Organisation of railway freight transport: case study CIM/SMGS between Slovakia and Ukraine. European transport research review, 8(4), 1-13.

Ahmed, M. A. N. Z. O. O. R. (2018). The economics and politics of China Pakistan Economic Corridor and Balochistan. Regional Studies, 36(4), 71-111.

Albrechts, L. & T. Coppens (2003), Megacorridors: Striking a balance between the space of flows and the space of places. Journal of Transport Geography 11 (3), pp. 215-224.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of informetrics, 11(4), 959-975.

Asmussen, C. B., & Møller, C. (2019). Smart literature review: a practical topic modelling approach to exploratory literature review. Journal of Big Data, 6(1), 1-18.

Babalik-Sutcliffe, E. (2013). Urban form and sustainable transport: Lessons from the Ankara case. International Journal of Sustainable Transportation, 7(5), 416-430.

Beaver, D. (2001). Reflections on scientific collaboration (and its study): past, present, and future. Scientometrics, 52(3), 365-377.

Bertolini, L., & Dijst, M. (2003). Mobility environments and network cities. Journal of urban design, 8(1), 27-43.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Blinge, M. (2014). Policy measures to realise green corridors—A stakeholder perspective. Research in Transportation Business & Management, 12, 55-62.

Borgatti, S. P. (2005). Centrality and network flow. Social Networks, 27(1), 55-71

Boyer, R. C., Scherer, W. T., & Smith, M. C. (2017). Trends over two decades of transportation research: a machine learning approach. Transportation research record, 2614(1), 1-9.

Brauer, R., & Fridlund, M. (2013, October). Historicizing topic models, a distant reading of topic modeling texts within historical studies. In International Conference on Cultural Research in the context of "Digital Humanities", St. Petersburg: Russian State Herzen University.

Bruinsma, F. R., Rienstra, S. A., & Rietveld, P. (1997). Economic Impacts of the Construction of a Transport Corridor: A Multi-level and Multiapproach Case Study for the Construction of the A1 Highway in the Netherlands. Regional studies, 31(4), 391-402.

Chapman, D., Pratt, D., Larkham, P., Dickins, I., 2003. Concepts and definitions of corridors: Evidence from England's Midlands. Journal of Transport Geography 11 (3), 179-191

Choo, J., Lee, C., Reddy, C. K., & Park, H. (2013). Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization. IEEE transactions on visualization and computer graphics, 19(12), 1992-2001.

Chorus, P., & Bertolini, L. (2016). Developing transit-oriented corridors: Insights from Tokyo. International Journal of Sustainable Transportation, 10(2), 86-95.

Curtis, C. (2008). Evolution of the transit-oriented development model for low-density cities: a case study of Perth's new railway corridor. Planning, Practice & Research, 23(3), 285-302.

Dalmasso, I., Datta, S. K., Bonnet, C., & Nikaein, N. (2013, July). Survey, comparison and evaluation of cross platform mobile application development tools. In 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC) (pp. 323-328). IEEE.

De Borger, B., Dunkerley, F., & Proost, S. (2007). Strategic investment and pricing decisions in a congested transport corridor. Journal of Urban Economics, 62(2), 294-316.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285-296.

Drewello, H., & Scholl, B. (Eds.). (2015). Integrated spatial and transport infrastructure development: the case of the European North-South corridor Rotterdam-Genoa. Springer.

Egghe, L. (2006). Theory and practise of the g-index. Scientometrics, 69(1), 131-152.

EGTC. (n.d.). https://www.egtc-rhine-alpine.eu/projects/code-24/

Ferreira‑Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(6), e1332.

Fraser, D., & Notteboom, T. (2014). A strategic appraisal of the attractiveness of seaport-based transport corridors: the Southern African case. Journal of Transport Geography, 36, 53-68.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). " I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. arXiv preprint arXiv:2212.05856.

Harenberg, S., Bello, G., Gjeltema, L., Ranshous, S., Harlalka, J., Seay, R., ... & Samatova, N. (2014). Community detection in large‑scale networks: a survey and empirical evaluation. Wiley Interdisciplinary Reviews: Computational Statistics, 6(6), 426-439.

Hesse, M. & J.P. Rodrigue (2004), The transport geography of logistics and freight distribution. Journal of Transport Geography 12 (3), pp. 171-184.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National academy of Sciences, 102(46), 16569-16572.

Hirsch, J. E. (2007). Does the h index have predictive power?. Proceedings of the National Academy of Sciences, 104(49), 19193-19198.

Hoh, A. (2019). China's belt and road initiative in Central Asia and the Middle East. Digest of Middle East Studies, 28(2), 241-276.

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength natural language processing in python.

Janic, M., & Vleugel, J. (2012). Estimating potential reductions in externalities from rail–road substitution in Trans-European freight transport corridors. Transportation Research Part D: Transport and Environment, 17(2), 154-160.

Jiang, Y., Qiao, G., & Lu, J. (2020). Impacts of the new international land–sea trade corridor on the freight transport structure in China, central Asia, the ASEAN countries and the EU. Research in Transportation Business & Management, 35, 100419.

Karami, A., Gangopadhyay, A., Zhou, B., & Kharrazi, H. (2018). Fuzzy approach topic discovery in health and medical corpora. International Journal of Fuzzy Systems, 20, 1334-1345.

Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and research: A systematic literature review through text mining. *IEEE access*, *8*, 67698-67717.

Karlson, M., Karlsson, C.S.J., Mörtberg, U., Olofsson, B., & Balfors, B. (2016). Design and evaluation of railway corridors based on spatial ecological and geological criteria. Transportation Research Part D: Transport and Environment, 46, 207-228.

Kunzmann, K.R. (2001). State planning: A German success story?. International Planning Studies, 6(2), 153-166.

Lambert, C. (1998). Planning for major urban expansion: A Case study in the UK's M4 growth corridor. Planning Practice & Research, 13(4), 371-388.

Lami, I.M. (Ed.). (2014). Analytical decision-making methods for evaluating sustainable transport in European Corridors (Vol. 11). Springer.

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. Information processing & management, 41(6), 1462-1480.

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. SpringerPlus, 5(1), 1-22.

Long, Y., & Zhang, W. (2019). Prioritizing future funding and construction of the planned high-speed rail corridors of China–According to regional structure and urban land development potential indices. Transport Policy, 81, 381-395.

Maksin, M., Nenković-Riznić, M., Milijić, S., & Ristić, V. (2017). The impacts of spatial planning on the sustainable territorial development of the Rhine-Danube Trans-European Transport Corridor through Serbia. European Planning Studies, 25(2), 278-297.

Marshall, T. (2014). The European Union and major infrastructure policies: The reforms of the trans-European networks programmes and the implications for spatial planning. European Planning Studies, 22(7), 1484-1506.

Meeks, E., & Weingart, S. B. (2012). The digital humanities contribution to topic modeling. Journal of Digital Humanities, 2(1), 1-6.

Ministerie van Infrastructuur en Waterstaat. (2017, November 6). Summary National Policy Strategy for Infrastructure and Spatial Planning. Publication | Government.nl. https://www.government.nl/documents/publications/2013/07/24/summary-national-policy-str ategy-for-infrastructure-and-spatial-planning

Monios, J., & Lambert, B. (2013). The Heartland Intermodal Corridor: public private partnerships and the transformation of institutional settings. Journal of Transport Geography, 27, 36-45.

Monios, J., & Wilmsmeier, G. (2012). Giving a direction to port regionalisation. Transportation Research Part A: Policy and Practice, 46(10), 1551-1561.

Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. Proceedings of the national academy of sciences, 101(suppl_1), 5200-5205.

Ng, A. K., Velasco-Acosta, A. E., & Wang, T. (2015). Institutions and the governance of transport infrastructure projects: Some insight from the planning and construction of the CentrePort Canada Way. Research in Transportation Business & Management, 14, 25-33.

Ni, C., Sugimoto, C., & Jiang, J. (2011, July). Degree, Closeness, and Betweenness: Application of group centrality measurements to explore macro-disciplinary evolution diachronically. In Proceedings of ISSI (Vol. 1, pp. 1-13).

Nocera, S., & Cavallaro, F. (2016). Economic valuation of Well-To-Wheel CO_2 emissions from freight transport along the main transalpine corridors. Transportation Research Part D: Transport and Environment, 47, 222-236.

Notteboom, T.E. & J.P. Rodrigue (2005), Port regionalization: towards a new phase in port development. Maritime Policy & Management 32 (3), pp. 297-313.

Öberg, M. (2014). Governance structure for transport corridors (Doctoral dissertation, Luleå tekniska universitet).

Öberg, M., Nilsson, K.L., & Johansson, C. (2017). Major transport corridors: the concept of sustainability in EU documents. Transportation research procedia, 25, 3694-3702.

OECD (2018), "The Belt and Road Initiative in the global trade, investment and finance landscape", in OECD Business and Finance Outlook 2018, OECD Publishing, Paris

Oehmer-Pedrazzi, F., Kessler, S. H., Humprecht, E., Sommer, K., & Castro, L. (2022a). Automated Content Analysis. Standardized Content Analysis in Communication Research: A Handbook. Springer Publishing

Olaru, D., Smith, B., & Taplin, J.H. (2011). Residential location and transit-oriented development in a new rail corridor. Transportation Research Part A: Policy and Practice, 45(3), 219-237.

Ostrowski, D. A. (2015, February). Using latent dirichlet allocation for topic modelling in twitter. In Proceedings of the 2015 IEEE 9th international conference on semantic computing (IEEE ICSC 2015) (pp. 493-497). IEEE.

Otsuka, N., Günther, F. C., Tosoni, I., & Braun, C. (2017). Developing trans-European Railway corridors: lessons from the rhine-alpine corridor. Case Studies on Transport Policy, 5(4), 527-536.

Panagakos, G., Psaraftis, H. N., & Holte, E. A. (2015). Green corridors and their possible impact on the European supply chain. Handbook of Ocean Container Transport Logistics: Making Global Supply Chains Effective, 521-550.

Peters, D. (2003). Cohesion, polycentricity, missing links and bottlenecks: Conflicting spatial storylines for Pan-European transport investments. European Planning Studies, 11(3), 317-339.

Priemus, H. (2001), Corridors in The Netherlands: Apple of discord in spatial planning. Tijdschrift voor Economische en Sociale Geografie 92 (1), pp. 100-107.

Priemus, H. & W. Zonneveld (2003), What are corridors and what are the issues. Introduction to special issue: the governance of corridors. Journal of Transport Geography 11 (3), pp. 167-177.

Priemus, H., & Zonneveld, W. (2004). Regional and transnational spatial planning: problems today, perspectives for the future: Guest Editorial. European Planning Studies, 12(3), 283-297.

Reggiani, A., Lampugnani, G., Nijkamp, P., & Pepping, G. (1995). Towards a typology of European inter-urban transport corridors for advanced transport telematics applications. Journal of Transport Geography, 3(1), 53-67.

Regmi, M.B., & Hanaoka, S. (2012). Assessment of intermodal transport corridors: Cases from North-East and Central Asia. Research in Transportation Business & Management, 5, 27-37.

Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In Proceedings of the eighth ACM international conference on Web search and data mining (pp. 399-408).

Rodrigue, J.P. (2004), Freight, gateways and mega-urban regions: The logistical integration of the Bostwash corridor. Tijdschrift voor Economische en Sociale Geografie 95 (2), pp. 147-161.

Rodrigue, J. P., Debrie, J., Fremont, A., & Gouvernal, E. (2010). Functions and actors of inland ports: European and North American dynamics. Journal of transport geography, 18(4), 519-529.

Roldan-Valadez, E., Salazar-Ruiz, S. Y., Ibarra-Contreras, R., & Rios, C. (2019). Current concepts on bibliometrics: a brief review about impact factor, Eigenfactor score, CiteScore, SCImago Journal Rank, Source-Normalised Impact per Paper, H-index, and alternative metrics. Irish Journal of Medical Science (1971-), 188, 939-951.

Rolfe, J., Kinnear, S., & Gowen, R. (2013). Simplified assessment of the regional economic impacts of interruption to transport corridors with application to the 2011 Queensland floods. Australasian Journal of Regional Studies 19(2), 215-238

Romein, A., Trip, J.J. & J. de Vries (2003), The multi-scalar complexity of infrastructure planning: Evidence from the Dutch-Flemish megacorridor. Journal of Transport Geography, 11 (3), pp. 205-213.

Rosario, B. (2000). Latent semantic indexing: An overview. Techn. rep. INFOSYS, 240, 1-16.

Sampaio, R. B., Fonseca, M. V. D. A., & Zicker, F. (2016). Co-authorship network analysis in health research: method and potential use. Health research policy and systems, 14(1), 1-10.

Schönharting, J., Schmidt, A., Frank, A. & S. Bremer (2003), Towards the multimodal transport of people and freight: interconnective networks in the RheinRuhr Metropolis. Journal of Transport Geography 11 (3), pp. 193-203.

Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63-70).

Thakur, K., & Kumar, V. (2022). Application of text mining techniques on scholarly research articles: Methods and tools. New Review of Academic Librarianship, 28(3), 279-302.

Tsigdinos, S., Nikitas, A., & Bakogiannis, E. (2021). Multimodal corridor development as a way of supporting sustainable mobility in Athens. Case Studies on Transport Policy, 9(1), 137-148.

Umadevi, V. (2013). Case study–centrality measure analysis on co-authorship network. Journal of Global Research in Computer Science, 4(1), 67-70.

van Altena, A. J., Moerland, P. D., Zwinderman, A. H., & Olabarriaga, S. D. (2016). Understanding big data themes from scientific biomedical literature through topic modeling. *Journal of Big Data*, *3*(1), 1-21.

Vogel, R. (2014). What happened to the public organization? A bibliometric analysis of public administration and organization studies. The American Review of Public Administration, 44(4), 383-408

Vries, J. de & H. Priemus (2003), Megacorridors in north-west Europe: Issues for transnational spatial governance. Journal of Transport Geography 11 (3), pp. 225-233.

Wang, C., & Ducruet, C. (2014). Transport corridors and regional balance in China: the case of coal trade and logistics. Journal of Transport Geography, 40, 3-16.

Wang, C., Paisley, J., & Blei, D. M. (2011, June). Online variational inference for the hierarchical Dirichlet process. In Proceedings of the fourteenth international conference on artificial intelligence and statistics (pp. 752-760). JMLR Workshop and Conference Proceedings.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications.

Wiegmans, B., & Janic, M. (2019). Analysis, modeling, and assessing performances of supply chains served by long-distance freight transport corridors. International Journal of Sustainable Transportation, 13(4), 278-293.

Wiegmans, B., Witte, P., Janic, M., & de Jong, T. (2020). Big data of the past: Analysis of historical freight shipping corridor data in the period 1662–1855. Research in Transportation Business & Management, 34, 100459.

Wilkerson, J., & Casas, A. (2017). Large-scale computerized text analysis in political science: Opportunities and challenges. Annual Review of Political Science, 20, 529-544.

Williams, B., Berry, J., & McGreal, S. (2002). The East coast corridor: spatial development strategies for the Dublin-Belfast metropolitan regions. Journal of Irish Urban Studies, 1(2), 19-31.

Wilmsmeier, G., Monios, J., & Lambert, B. (2011). The directional development of intermodal freight corridors in relation to inland terminals. Journal of Transport Geography, 19(6), 1379-1386.

Witte, P. (2014). The Corridor Chronicles. Integrated perspectives on European transport corridor development. Eburon Academic Publishers.

Witte, P., & Spit, T. (2016). Challenges for corridors: Future perspectives on European corridor development. In Integrated Spatial and Transport Infrastructure Development (pp. 99-114). Springer, Cham.

Witte, P., Van Oort, F., Wiegmans, B., & Spit, T. (2014). European corridors as carriers of dynamic agglomeration externalities?. European Planning Studies, 22(11), 2326-2350.

Witte, P., Wiegmans, B., Oort, F. van & T. Spit (2012), Chokepoints in corridors: Perspectives on bottlenecks in the European transport network. Research in Transportation Business & Management 5, pp. 57-66.

Witte, P. A., Wiegmans, B. W., van Oort, F. G., & Spit, T. J. (2012). Chokepoints in corridors: Perspectives on bottlenecks in the European transport network. Research in Transportation Business & Management, 5, 57-66.

Witte, P., Wiegmans, B., van Oort, F., & Spit, T. (2014). Governing inland ports: a multi-dimensional approach to addressing inland port–city challenges in European transport corridors. Journal of Transport Geography, 36, 42-52.

Xu, X., Chen, Q., & Zhu, Z. (2022). Evolutionary overview of land consolidation based on bibliometric analysis in web of science from 2000 to 2020. International Journal of Environmental Research and Public Health, 19(6), 3218.

Yang, D., Jiang, L., & Ng, A.K.Y. (2018). One Belt one Road, but several routes: A case study of new emerging trade corridors connecting the Far East to Europe. Transportation Research Part A: Policy and Practice, 117, 190-204.

Zhang, C. T. (2009). The e-index, complementing the h-index for excess citations. PLoS One, 4(5), e5429.

Zimny-Schmitt, D., & Goetz, A. R. (2020). An investigation of the performance of urban rail transit systems on the corridor level: A comparative analysis in the American west. Journal of Transport Geography, 88, 102848.

Zipf G. K., Human Behavior and the Principle of Least Effort. Boston, MA, USA: Addison-Wesley, 1949