Incorporating an Unsupervised Text Mining Approach into Studying Logistics Risk Management: Insights from Corporate Annual Reports and Topic Modeling

David Olson 1 and Bongsug (Kevin) Chae 2,*

1 Department of Supply Chain Management and Analytics, University of Nebraska, Lincoln, NE 68588, USA; david.olson@unl.edu
2 Department of Management, Kansas State University, Manhattan, KS 66506, USA
* Correspondence: bchae@ksu.edu

Abstract: This study examined the Security and Exchange Commission (SEC) annual reports of selected logistics firms over the period from 2006 through 2021 for risk management terms. The purpose was to identify which risks are considered most important in supply chain logistics operations. Section 1A of the SEC reports includes risk factors. The COVID-19 pandemic has had a heavy impact on global supply chains. We also know that trucking firms have long had difficulties recruiting drivers. Fuel price has always been a major risk for airlines but also can impact shipping, trucking, and railroads. We were especially interested in pandemic, personnel, and fuel risks. We applied topic modeling, enabling us to identify some of the capabilities of unsupervised text mining as applied to SEC reports. We demonstrate the identification of terms, the time dimension, and correlation across topics by the topic model. Our analysis confirmed expectations about COVID-19’s impact, personnel shortages, and fuel. It also revealed common themes regarding the risks involved in international trade and perceived regulatory risks. We conclude with the supply chain management risks identified and discuss means of mitigation.

Keywords: unsupervised text mining; topic modeling; supply chains; risk management; logistics

1. Introduction

Supply chain risk management is defined as the identification and management of risks in a supply chain through coordinated efforts by its members to reduce vulnerability as a whole rather than a focus on a particular company within the supply chain [1–4]. Seyedghorbani et al. [5] applied bibliometric analysis to review trends in operations/supply chain management research. They found that the focus was still on production/manufacturing. Logistics is also a key element in supply chains. Dolgui et al. [6] focused on the ripple effect where disruption in one element of a supply chain impacts the rest of the chain. Supply chain risk management has been viewed as a generic process of collaboration to ensure profitability and continuity [7,8].

Supply chains in general involve many risks. These risks also impact the logistic component of supply chains. These risks include macro factors, such as natural disasters, war, and the political environment, as well as many micro factors, ranging from demand, supply, factors impacting production of goods or services, financial factors, and factors affecting logistics to transportation factors (fuel, skill availability) and information flow. Ho et al. [9] reviewed 14 studies through 2013 and categorized supply chain risks. Olson and Wu [10] reviewed seven studies (four overlapping with [9]) through 2014, doing the same. In both of these reviews, external (macro) factors involved natural disasters, political factors, and market competitors. Internal (micro) factors involved demand, manufacturing, supply, information systems, transportation, and finance. Nakano and Lau [11] included 173 studies ranging from 2001 to 2017 in their study.
In Section 2, we first review logistics risk management factors derived from many published studies of journal bibliometric analysis. This section includes the literature on supply chain risks and supply chain connectivity and bibliometric studies in logistic risk management. We also look at bibliometric analysis of corporate annual reports, a much more limited area of publication. In Section 3, we then describe an unsupervised text mining approach (topic modeling; e.g., [12,13]) as our research methodology, followed by an analysis of supply chain risk management focusing on transportation firms. This section describes the SEC filings data used and presents the structural topic modeling method used [14,15]. We examine risk reports of upstream inputs to logistic firms taken from the Securities and Exchange (SEC) annual reports of selected logistic firms over the period from 2006 through 2021. Section 4 gives the analysis and results, beginning with topic prevalence analysis and effects, topic correlation, and means used to detect community clusters. Section 5 discusses the results and their implications and ties the results to supply chain risk mitigation, and Section 6 gives conclusions.

2. Literature Review

2.1. Supply Chain Risks

Supply chain organizations need to worry about risks from every direction. In any business, opportunities arise from the ability of the organization to deal with risks. Most natural risks are dealt with either through diversification and redundancy or through insurance, both of which have inherent costs. As with any business decision, the organization needs to make a decision considering tradeoffs. Traditionally, this has involved the factors of costs and benefits. Society is moving more and more toward even more complex decision-making domains requiring consideration of ecological factors, as well as factors of social equity.

Dealing with other external risks involves more opportunities to control risk sources. Some supply chains in the past have had an influence on political systems. While most supply chain entities are not expected to be able to control political risks like wars and regulations, they do have the ability to create environments leading to labor unrest. On the other hand, disease and pandemics are beyond their control except for how they react. Choudhary et al. [16] extensively reviewed supply chain risk assessment articles, focusing on analytic approaches and multi-criteria decision-making techniques to assess supply chain risks including pandemics. Supply chain organizations have even greater expected influence over economic factors. While they are not expected to be able to control exchange rates, the benefit of monopolies or cartels is their ability to influence prices. Business organizations are also responsible for developing technologies providing competitive advantage and product portfolios in dynamic markets with product life cycles. The risks arise from never-ending competition.

Internal risk management is more directly the responsibility of the supply chain organization and its participants. All business organizations are responsible for managing financial, production, and structural capacities. They are responsible for programs that provide adequate workplace safety, which has proven to be cost-beneficial to organizations, as well as fulfilling social responsibilities. Within supply chains, there is a need to coordinate activities with vendors and, to some degree, with customers (supported by data obtained through barcode cash register information providing instantaneous indications of demand). Information system technology provides effective tools to keep on top of supply chain information exchange. Another means of coping with risk across a supply chain is coordination using information technology. Zhou et al. [17] explored group purchasing organizations facilitated by shared information systems. Another factor of great importance is the responsibility of supply chain core organizations to manage the risks inherent in the tradeoff between the wider participation made possible through Internet connections (providing a larger set of potential suppliers, leading to lower costs) and the reliability provided by long-term relationships with a smaller set of suppliers that have proven to be reliable. Blockchain technology can improve supply chain communication security.
Etemadi et al. [18] applied bibliometric network analysis to examine how blockchains can be used to provide privacy and security, monitor counterfeiting, and trace food safety data. Subcontracting often provides a means to improve supply chain efficiency but can lead to problems of control. Caro et al. [19] empirically studied factors leading to unauthorized subcontracting that could involve liability for the core supply chain organization.


2.2. Supply Chain Connectivity

By their nature, supply chains are interconnected. Some supply chains are internally controlled, as was commonly the case in the 19th and early 20th centuries, where one organization controlled the entire supply chain (Standard Oil, US Steel, Alcoa Aluminum). However, it became apparent that a collection of specialist organizations focusing on their particular contributions to the supply chain was more efficient. This has even been the case in military organizations, where contracting has become the norm. This leads to interconnectivity without total control. When risks arise, the impact ripples throughout the organization (as in the beer game, see Forrester [26]). Berger et al. [27] presented a model accounting for the ripple effect of supply chain issues arising due to this interconnectivity of supply chain elements.

Dolgui et al. [6] proposed a control framework involving the four elements of resilience, redundancy, robustness, and flexibility:

- Resilience involves the ability to adapt to disruption events, including finding alternative supplies;
- Redundancy involves increasing product availability through building backup capacity or inventory;
- Robustness is the ability to survive in the face of challenge, which is enhanced by resilient actions and redundant systems;
- Flexibility is attained by being able to sense threats, react, and recover quickly, usually in the form of reallocation of inventory and capacity.

Nakano and Lau [11] studied supply chain risk management strategies. Redundancy can be gained through developing additional capacity, stocking more inventory, and multiple sourcing. Hedging was cited by Manuj and Mentzer [28] as having a globally dispersed portfolio of suppliers and facilities that enables at least some supply chain members to survive risk events. Flexible efforts can include collaboration, postponement, and development of suppliers. Agility is the ability to change policies in light of new threats or opportunities.

Fuel is of concern to trucking, shipping (bunker fuel), air transport (jet fuel), and rail transport (diesel). Trucking companies have long been short of operating personnel, and they have proactively sought to hire and train new drivers. Thus, key risks in logistics management include fuel prices and the need for operating personnel, such as truck drivers. Furthermore, beginning in 2020, COVID-19 critically strained supply chains. To independently look at key logistic risks, we examined data trends for fuel prices and COVID-19. Figure 1 displays the variation in Brent Crude, a key component in fuel. Anomalies in this series can be explained by the 2008 mortgage crisis, the onset of fracking in 2014, and COVID-19 in 2020. Volatilities were measured by the mean divided by the standard deviation—the coefficient of variation—for the period January 2001 through December 2022. The coefficient of variation for Brent Crude over this period was 2.29,
greater than that of gold (2.13) and the S&P 500 (2.04). Thus, we conclude that fuel has been a major risk that logistics firms have had to cope with.

![Plot of crude oil price by month for January 2001 through February 2023](data source: macrotrends.net)

In the early part of the 21st century, pandemics such as SARS and MERS had some impact on world trade. However, they were much less dramatic than the COVID-19 pandemic beginning in 2020. COVID-19 has had a more recent but possibly even more drastic impact on supply chain logistics. Thus, we begin with a sense of the risk impacts of personnel, fuel, and the pandemic. Logistics risk management involves the need for a systematic process to cope with these and other risks.

### 2.3. Bibliometric Analysis of Supply Chain Risk Management

Bibliometric analysis is a popular method for understanding risks in supply chain contexts, and such an approach applies tools to text-mine data, especially those found on Websites, for keyword analysis. For example, Nobanee et al. [29] applied it to reputational risk and sustainability. In the supply chain area, Kilubi [30] identified knowledge groups in the supply chain risk management field. Fagundes et al. [31,32] focused on decision-making models and support systems for supply chain risk management. Senna et al. [33] used bibliometrics to identify human factors impacting supply chain resilience. Mishra et al. [34] used bibliometrics to examine the ripple effects of risk events within supply chains.

We more closely examined bibliometric studies specific to supply chain risk management. All were focused on analysis of academic publications and co-citations of authors. Fahimnia et al. [35] screened 1108 academic papers, applying page rank analysis to identify relationships between authors in eight supply chain risk categories. Their focal interest was quantitative models, and they identified seven generative research areas for the eight risk categories. Perera et al. [36] focused on 137 papers by prolific authors, examining the role of human factors in supply chain forecasting. Pournader et al. [37] selected 119 core papers to generate 11 clusters of academic topics related to supply chain risk and resilience. In the supply chain risk cluster, assessment, mitigation, and sustainability issues were subclusters. Nakano and Lau [11] selected 173 publications for text analysis in their study of supply chain risk management mitigation strategies. Wicaksana et al. [38] selected 345 papers to analyze co-authorship in supply chain risk management. They identified 14 clusters of emerging topics. Choudhary et al. [16] applied bibliometric analysis to 136 multiple-criteria papers in the supply chain management field, providing a picture of how that type of analysis has been applied to supply chain risk management. Bibliometric analyses related to supply chain management also include the study by Chae and Olson [39,40].
Those studies scraped academic research publications. Other rich sources for bibliometric analysis are company annual reports. As of December 2005, the Securities and Exchange Commission (SEC) of the United States requires reporting companies to discuss in their annual reports the most significant factors that made their company risky [41]. These reports have been utilized by a number of studies. Janggu et al. [42] analyzed the importance of social risk management for sustainable development by applying content analysis to the 2013 and 2014 annual reports of plantation sector companies. Yang et al. [43] applied text mining to 10-K filings to extract financial, strategic, operational, and hazard risks. Azmi Shabestari and Romero [44] applied sentiment analysis to sections 1A (risk factors) and 7 (management discussion) of 10-K filings to obtain evidence of future performance expectations. Weber and Müsig [45] sampled annual reports of nonfinancial and nonutility firms from the European Economic Area over the period 2005–2017 for risk factors to analyze their risk strategies.

These report filings are also a rich source of data for supply chain risk management. Chiu et al. [46] applied risk factor disclosures from section 1A of annual reports to analyze the effects of downstream firm risk factor disclosures on upstream (supplier) firms in supply chains.

3. Research Methods and Data
3.1. Data
We analyzed risk factors reported in the SEC filings of companies focused on transportation. We examined these reports prior to the COVID-19 period and compared them with the reports during the pandemic period. SEC filings for logistic firms in the trucking, shipping, air, rail, and pipeline industries over the period 2006 through 2021 were processed for topic modeling analysis. There were 35 trucking firms, 35 shipping firms, 46 air transport firms, 11 railroad firms, and 22 pipeline firms, amounting to a total of 149 firms. Table 1 shows the number of reports by year and industry.

<table>
<thead>
<tr>
<th>Year</th>
<th>Air</th>
<th>Trucking</th>
<th>Rail</th>
<th>Shipping</th>
<th>Pipeline</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>12</td>
<td>2</td>
<td>38</td>
</tr>
<tr>
<td>2007</td>
<td>16</td>
<td>12</td>
<td>1</td>
<td>10</td>
<td>4</td>
<td>43</td>
</tr>
<tr>
<td>2008</td>
<td>13</td>
<td>7</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>2009</td>
<td>13</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>2010</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>2011</td>
<td>13</td>
<td>9</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>2012</td>
<td>11</td>
<td>23</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td>61</td>
</tr>
<tr>
<td>2013</td>
<td>21</td>
<td>23</td>
<td>5</td>
<td>12</td>
<td>16</td>
<td>77</td>
</tr>
<tr>
<td>2014</td>
<td>21</td>
<td>22</td>
<td>6</td>
<td>13</td>
<td>20</td>
<td>82</td>
</tr>
<tr>
<td>2015</td>
<td>22</td>
<td>23</td>
<td>6</td>
<td>15</td>
<td>19</td>
<td>85</td>
</tr>
<tr>
<td>2016</td>
<td>25</td>
<td>19</td>
<td>6</td>
<td>16</td>
<td>18</td>
<td>81</td>
</tr>
<tr>
<td>2017</td>
<td>20</td>
<td>17</td>
<td>6</td>
<td>15</td>
<td>17</td>
<td>75</td>
</tr>
<tr>
<td>2018</td>
<td>19</td>
<td>17</td>
<td>6</td>
<td>15</td>
<td>16</td>
<td>73</td>
</tr>
<tr>
<td>2019</td>
<td>21</td>
<td>18</td>
<td>6</td>
<td>13</td>
<td>14</td>
<td>72</td>
</tr>
</tbody>
</table>

3.2. Structural Topic Modeling (STM)
This study used topic modeling [12] to discover latent topics from section 1A of the annual reports. Topic modeling is an unsupervised machine learning approach that can be used to automatically detect latent structures in large texts and has been widely adopted in multiple academic disciplines [39,47,48]. There are different topic modeling methods in the
literature [49]. Our analysis adopted structural topic modeling (STM) [14,15]. STM is an innovative text analytics methodology incorporating variables such as time (e.g., year of the annual report) and company information as covariables in the topic modeling process. By default, STM utilizes the spectral algorithm explained in [14,50]. This approach was useful for our study because our focus was not just what type of supply chain management (SCM) risks are found in the annual reports but also how different risks are associated with each other and how the awareness of such risks has changed over the years. The STM framework used in this study is displayed in Figure 2.

Figure 2. Structural topic modeling (STM) process used (the kye information is highlighted).

The STM process involves text cleaning and processing, the development of many topic models for model selection, and selection of the optimal model. The first step involved a series of text preprocessing tasks, including tokenizing the texts in the reports, removing stopwords and nonalphanumeric characters, and lemmatizing and lowercasing each word. Then, the text data were converted into a document term matrix with term frequency, and this process used the tm package in R [51]. The second step is called “model selection”, which is necessary because topic modeling is unsupervised machine learning where the number of topics, $k$, is required as the input to the modeling process. In the literature, there are different methods or metrics for evaluating the quality of topic models [49]. The model fit was estimated by comparing the residuals [52] of the topic models, with the number of topics ranging from 10 to 150. The results are displayed in Figure 3. The results indicated that the optimal number of topics was between 40 and 100 and, specifically, the residual was the lowest when $k$ was 48, 52, and 57. The authors reviewed the results of three models ($k = 48, 52, 57$) in terms of topic interpretability, and 52 was chosen as the optimal $k$ value.
There are three broad types of analysis that use topic modeling results [14,15]. They are topic prevalence analysis, topic evolution analysis, and topic correlation analysis.

3.3. Analysis of STM Topic Model

In general, a topic model generates two outputs: the document-topic distribution (DTD) and the topic-term distribution (TTD). The DTD referred to the topic probability for each annual report; section 1A (risk factors) of each SEC filing was considered to be composed of 52 topics with different degrees. Thus, using DTD, it was possible to figure out which topics were more prevalent than others in the corpus (953 annual reports in this study). This is called topic prevalence analysis. In a topic model, a series of associated words with different probabilities represents a topic. The TTD offers these words and their probabilities per each topic so researchers can develop proper labels for the topics [12]. In addition, a key feature of STM is its capability to consider annual reports’ metadata, such as industry type (e.g., air, railroad), as covariates so they can influence topic prevalence. This is called estimation with the topical prevalence parameter [15].

Topics (or SCM risks in this study) are not static. They evolve over time in terms of their prevalence (or popularity) [53]. For example, pandemic-related risks may have never been a dominant topic in annual reports prior to 2020. As explained earlier, STM supports the incorporation of metadata, and time-related information, such as publication year, can be entered as covariates. This supports the analysis of topic evolution [39,54] using regression, and this type of regression analysis can offer meaningful information about the trends for different SCM risks over the years and whether such trends (e.g., upward, downward) are significant or not.

Topics or SCM risks are not standalone either. Instead, they are correlated with each other and, thus, the topic probabilities are related [13]. For example, China-related risks may be more related to specific industries (e.g., air rather than pipelines) and to other topics, such as currency and exports. The association of topics is estimated from the DTD using correlation coefficients. Network analysis can be useful to visualize topic correlations, and such methods as community detection can discover clusters of associated topics [55,56].

4. Analysis and Results

4.1. Topic Prevalence Analysis

This analysis focused on labeling topics (or risks), figuring out probabilistic proportions per topic, and estimating the effect of metadata; specifically, the type of topic prevalence for each industry. The STM model generated a series of associated words for each of the 52 topics. The process of labeling topics involved the authors recursively reviewing the most exclusive and frequent words by topic, as well as by industry. Figure 3 shows the relative density of these 52 topics labeled by industry and selected terms. The
most common topic was topic 23, which was most associated with trucking, with the most common terms being those such as driver, diesel, engine, emission, surcharge, truck, CSA, MFCSA, and center.

The authors were looking for evidence of terms involving COVID-19, fuel, and personnel, as these have been common risk elements in supply chains in recent times. They also noted evidence of regulatory issues; catastrophes, such as weather, war, and other risks; and international elements. Note that section 1A of the SEC filings can be as long as 34 pages with many thousands of words per report, so the STM model is useful for focusing on the most dense terms.

Thus, the authors further reviewed all 52 topics and selected 12 topics closely related to SCM risks. The most associated words for each of the 12 topics and the topic prevalence are summarized in Table 2. Among these 12 topics, the popular topics were topics 23 (T23TruckPersonnelDieselEmission), 40 (T40TruckPersonnelRegulation), and 39 (T39RailWarCleanupCoalPersonnel). Among the least prevalent topics were topics 6 (T6PipeAirPandemicPersonnelRegulation) and 37 (T37ShipRailPirateSanctionChina).

As mentioned in Section 3, STM is an innovative technique enabling the incorporation of variables as covariates in the topic modeling process. The topic model with 52 topics was built with the years of the annual reports (e.g., 2010, 2011) and the types of logistics industries (e.g., air, railroad). Figure 5 gives the topic probabilities by industry. This shows the effects of logistic industries on topic prevalence. For example, topic 6 was strongly related to the pipeline industry, topics 12 and 13 to the ship industry, topics 14
and 39 to the railroad industry, topic 18 to the air industry, and topics 23, 40, and 52 to the trucking industry.

Table 2. Selected SCM risk topics and proportions.

<table>
<thead>
<tr>
<th>Topic #</th>
<th>Terms</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Loan, pandemic, treasury, restricting, treasury</td>
<td>0.009958</td>
</tr>
<tr>
<td>12</td>
<td>Cruise, ship, port, guest, treaty</td>
<td>0.034844</td>
</tr>
<tr>
<td>13</td>
<td>Tanker, charter, pool, detention, voyage</td>
<td>0.014237</td>
</tr>
<tr>
<td>14</td>
<td>Railroad, concession, Mexico, traffic, crime</td>
<td>0.015758</td>
</tr>
<tr>
<td>18</td>
<td>Helicopter, gas, production, exploration, Brazil</td>
<td>0.021635</td>
</tr>
<tr>
<td>23</td>
<td>Driver, tractor, diesel, engine, emission</td>
<td>0.046674</td>
</tr>
<tr>
<td>37</td>
<td>Sanction, mining, bulk, China, Libor, export</td>
<td>0.005057</td>
</tr>
<tr>
<td>39</td>
<td>Rail, railroad, coal, clean-up, concerning</td>
<td>0.036162</td>
</tr>
<tr>
<td>40</td>
<td>Driver, contractor, lawsuit, turnover, legislator</td>
<td>0.040267</td>
</tr>
<tr>
<td>44</td>
<td>Libor, brand, cybersecurity, pandemic, coronavirus</td>
<td>0.023404</td>
</tr>
<tr>
<td>50</td>
<td>Dealer, China, currency, exchange, freight</td>
<td>0.010799</td>
</tr>
<tr>
<td>52</td>
<td>Driver, contractor, anti-terrorism, disrupt, closure</td>
<td>0.018439</td>
</tr>
</tbody>
</table>

![Graphs showing topic prevalence over time](image)

Figure 5. Estimation of the effects of industry type on topic prevalence.

4.2. Topic Evolution Analysis

Figure 6 graphically displays the relative proportions of topics over the time frame 2006 through 2021. This figure shows that topic prevalence or popularity changes following certain evolutionary patterns, as discussed in the relevant literature [53,57]. The results of the nonparametric Mann–Kendal test for trend analysis are shown in each chart to indicate whether or not the trend is statistically significant. Two popular trends are upwards (or hot) and downwards (or cold). Topics such as 6, 12, 14, 39, 40, 44, and 52 increased in prevalence over the years. For example, two topics (6 and 44) were related to the pandemic and both
topics displayed upward trends, especially since 2020. The opposite trend was observed for such topics as 50.

![Graph showing topic densities over time](image)

**Figure 6.** Topic densities over time.

### 4.3. Topic Correlation Analysis

One annual report contains multiple risks and, thus, there are correlations among the topics or risks. As a probabilistic topic model, the STM model was used to produce the document–topic matrix showing the proportion of each topic or risk in each document. Thus, it was possible to analyze the correlations among the topics. A network graph can be used to show the correlation between topics. The edges in a network mean positive correlations and the line thickness shows the strength of the correlations. The network given in Figure 7 includes 41 topics and their associations (54 edges). Specifically, it represents how the 12 topics were correlated with each other and with the other 29 topics. Topic 6 was associated with 11 other topics and its subnetwork had the largest number of edges.

Building on Figure 5 (estimation of the effects of industry type on topic prevalence), Figure 8 reports first-degree connections for the 12 focus topics. Topic 6 was found in two industries (pipeline, air) and thus was strongly associated with topics 24 (T24PipeGreenhousePetroleumScientistDeception) and 51 (T51AirPersonnelRegulation). Topic 14 (RailNAFTACrimeViolenceMexico) referred to a risk in the railroad industry and was found to be strongly associated with another railroad topic (or risk), T39RailWarCleanupCoalPersonnel. Topic 23 (T23TruckPersonnelDieselEmission) was a trucking-related risk highly correlated with T27TruckWeatherPersonnelRegulation.

Further analysis focused on finding clusters of correlated topics (or risks). Community structure detection from networks is well-suited for this purpose [56]. Specifically, the Louvain modularity algorithm [55] was used to detect clusters of correlated topics in the network, and six clusters were discovered. The results are summarized in Table 3. Three large clusters were #1, #2, and #3. Several topics in cluster #1 were related to international issues or risks involving foreign countries, such as China. Cluster #2 largely consisted of topics related to regulation. A common theme among the topics in Cluster #3 was personnel. In the remaining section, we review each of the selected topics in greater detail.
Figure 7. Network visualization of positively correlated topics.

Table 3. Cluster elements with focus topics highlighted.

<table>
<thead>
<tr>
<th>#</th>
<th>Topic Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>T7TruckShipRailAirPersonnel, T12ShipPersonnelRegulationIncident, T23TruckPersonnelDieselEmission, T10TruckPersonnelRegulation, T16AirTruckShipRailPipeRegulation, T40TruckPersonnelRegulation, T27TruckWeatherPersonnelRegulation, T31TruckRailFinance (eight topics)</td>
</tr>
<tr>
<td>4</td>
<td>T22TruckPersonnel, T44TruckShipViolenceCybersecurityPersonnelRegulationPandemic, T35ShipPollutionPersonnel, T42AirPassenger (four topics)</td>
</tr>
<tr>
<td>5</td>
<td>T18AirGasBrazil, T28ShipGasHydrocarbonImmigration, T29AirTourismGasRegulation (three topics)</td>
</tr>
<tr>
<td>6</td>
<td>T14RailNAFTACrimeViolenceMexico, T39RailWarCleanupCoalPersonnel (two topics)</td>
</tr>
</tbody>
</table>
Figure 8. First-degree connections for the 12 focus topics.

4.4. Logistics Risk Analysis

4.4.1. Personnel and Fuel

Four topics were related primarily to fuel and personnel risks. The topic T23TruckPersonnelDieselEmission had the highest score. The topic presence started low in 2006, rose to a peak in 2011–2012, and then declined. It was in the second largest cluster with other trucking topics (one shipping topic with personnel and incident issues was also found in this cluster). The implication is that personnel, fuel, and risks from CO₂ emissions are important.
The topic T40TruckPersonnelRegulation had the third highest score. The topic started to rise in 2012 and steadily decreased thereafter. It was in the second largest cluster with other trucking and linked industry topics. Salient issues were personnel and regulation. The topic T39RailWarCleanupCoalPersonnel had the fourth highest score. This topic steadily rose in proportion, with a jump in 2021. It was in the sixth cluster with T14. The salient issue for T39 was the rail industry. The topic T12ShipPersonnelRegulationIncident had the fifth highest score. This topic started low in proportion and quickly rose around 2013. It was in the second largest cluster with other trucking and linked industry topics. The predominant issues were personnel and regulation.

4.4.2. Pandemic

The topic T44TruckShipViolenceCybersecurityPersonnelRegulationPandemic had the 14th highest score. Its proportion rose steadily after 2011, with COVID-19 making a clear impact in 2020. It was in the largest cluster with air and truck topics. Salient issues were the pandemic (COVID-19) and cybersecurity. The topic T6PipeAirPandemicPersonnelRegulation had the 47th highest score. This topic steadily rose in proportion, with a jump in 2021. This topic started low, rising to a peak in 2011–2012 and then declining. It was in the fourth largest cluster with pipeline, air, and ship industries. Regulatory and emission issues were present. The pandemic element was not COVID-19 but presumably some other respiratory pandemic, such as swine flu or MERS.

4.4.3. International

The topic T18AirGasBrazil had the 15th highest score. This topic’s proportion started to rise in 2012, peaked in 2016, and then declined. It was in the fifth largest cluster with air and shipping topics. The salient feature was Brazil, which had economic issues in 2016. The topic T52TruckAntiterrorismPersonnelDieselMexico had the 23rd highest score. This topic’s proportion rose steadily after 2012, when crime in Mexico began to rise. It was in the third largest cluster with ship, truck, and rail topics. Salient issues were antiterrorism and Mexico. The topic T14RailNAFTACrimeViolenceMexico had the 28th highest score. This topic’s proportion started to rise in 2012 and reached a high plateau from 2012 to 2021, again reflecting the rise in crime in Mexico. It was in the sixth cluster with one other rail topic—NAFTA and crime were salient. The topic T13ShipPersonnelLeakageAfrica had the 32nd highest score. Its proportion generally declined from 2006, with a small rise in 2016–2018 and a fall thereafter. It was in the third largest cluster with ship and rail topics (as well as some truck topics), with international terms involving Africa (leakage), Puerto Rico (vice), China (piracy), Australia, and Mexico (antiterrorism). The topic T50TruckFraudChina had the 45th highest score. This topic was relatively high from 2006 through 2012 and then declined. It was in the second largest cluster with other truck and linked industry topics. Salient were China and fraud. The topic T37ShipPiracySanctionChina had the 52nd highest score. This topic’s proportion tended to rise slowly, with a high anomaly in 2009. It was in the third largest cluster with ship and rail topics (as well as some truck topics). China and piracy were salient issues.

5. Discussion: Mitigation Strategies

We reviewed terms within each cluster by dominant industry, and classified terms related to seven supply chain risk categories. These are displayed in Table 4.

Xu et al. [58] reviewed the literature on supply chain disruption, applying bibliometric analysis to identify keywords found that would be of interest to academic researchers. They analyzed the prevalence of these keywords over time from 1996 through 2019. Initially, disruption management and supply chain management had the strongest presence through 2010. Around 2005, coordination became a key focus, followed quickly by demand disruption. Around 2007, the focus included inventory systems and coordination, often with information systems. Interruption became a key focus beginning in 2008, along with increased interest in retailers. In 2010, EOQ models received notable attention. In 2012, there was notable attention given to policy, followed by a focus on decision making in 2015.
Our study of the period 2006 through 2020 found that the pandemic of 2020 was important, as well as regulatory restrictions. For logistic firms specifically, personnel shortages, fuel prices, weather, and international trade exposure continued to be issues. There were also one-of-a-kind risks, such as war, riots, piracy, and fraud.

Table 4. Terms by industry.

<table>
<thead>
<tr>
<th>Risk Categories</th>
<th>Trucking</th>
<th>Shipping</th>
<th>Air</th>
<th>Rail</th>
<th>Pipeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personnel</td>
<td>Driver retention, operator, pension, recruitment, unionization</td>
<td>Crew, hire, Manning, captain</td>
<td>Pilot, furlough, attendant, dispatcher, union</td>
<td>Injury</td>
<td>Scientist, pension, payroll</td>
</tr>
<tr>
<td>Fuel</td>
<td>Diesel</td>
<td>Gas, diesel, hydrocarbon</td>
<td>Gas</td>
<td>Coal</td>
<td>Energy, gas, petroleum, crude, distillate</td>
</tr>
<tr>
<td>Weather</td>
<td>Snow, winter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risks</td>
<td>Terrorism, fraud, underground, emission, destruction</td>
<td>Incident, leakage, immigration, pollution, vice</td>
<td>Catastrophe, war, emission, noise</td>
<td>Crime, clean-up, war, violence</td>
<td>Conflict, emission, riot, greenhouse, deception</td>
</tr>
<tr>
<td>International</td>
<td>China, Mexico</td>
<td>Africa</td>
<td>Brazil</td>
<td>Mexico, Australia, Canada</td>
<td></td>
</tr>
<tr>
<td>Pandemic</td>
<td>Health</td>
<td>Coronavirus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulatory</td>
<td>License, ordinance, hours-of-service, taxation, legislature</td>
<td>Inspection, arrest, treaty, warrant, certificate</td>
<td>FAA, federalization, legislature, auditor, warrant</td>
<td>NAFTA</td>
<td>IRS, shutdown, fracturing, restricting</td>
</tr>
</tbody>
</table>

There are many means available to control risks within supply chains. One strategy would be to try to do a great job in relation to the fundamental supply chain performance measures of consistent fulfillment of orders, delivery dependability, and customer satisfaction. That would basically amount to doing a good job in what the company does. Of course, many effective organizations have failed when faced with changing markets or the catastrophic risks outlined in Section 5 as external risks. Some strategies proposed for supply chains are reviewed in Table 5.

Table 5. Risk mitigation strategies.

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>Mitigation (Zhao and Huchzermeier [59])</th>
<th>Elements (Nakano and Lau [11])</th>
<th>Robust Strategy (Tang [8])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply</td>
<td>Multiple suppliers (backup production/suppliers), quantity-flexible contracts, invest in supplier improvement, strategic stock</td>
<td>Redundancy</td>
<td>Strategic stock, economic supply incentives</td>
</tr>
<tr>
<td>Processing</td>
<td>Production flexibility, modularization, timing of product introduction</td>
<td>Flexibility</td>
<td>Postponement, make-or-buy, transportation flexibility</td>
</tr>
<tr>
<td>Demand</td>
<td>Shift demand over time, markets, products, customize service</td>
<td>Agility</td>
<td>Influence customer selection, dynamic assortment planning</td>
</tr>
<tr>
<td>Network</td>
<td>Coordinate decisions (contracts, information sharing)</td>
<td>Collaboration</td>
<td></td>
</tr>
</tbody>
</table>
Zhao and Huchzermeier [59] provided a framework considering the integration of the operational and financial risk aspects of supply chain risk management. They concluded that if operational and financial flows are partial substitutes, centralization should be applied; however, if there is no interaction, decentralization is expected.

Logistic firms have been dealing with specific forms of risk. Personnel shortages were common across trucking, shipping, and air. In cluster one, personnel shortages were found for trucking firms and the interaction of trucking with shipping in 6 of the 14 topics. In cluster two, 5 of the 10 topics mentioned personnel, predominately for air firms. In cluster three, studies on trucking and shipping firms mentioned personnel in six of eight topics, while studies on the same combination in cluster four mentioned personnel in four of six topics. In cluster six, personnel were mentioned in one topic (involving rail). Rail firms began to downsize crews about 50 years ago, relying more and more on automation. Trucking firms continue to advertise for truckers. Automation is another means of dealing with this shortage, although to date the perceived risk of driverless trucks has stopped trucking firms from taking that approach.

All logistic firms have always had to cope with fuel prices. In cluster one, topic 34 (involving rail) included mentions of risks for coal, topic 45 for ship fuel, and topic 52 for diesel for trucking firms. In cluster two, gas was mentioned in topic 19 (pipeline) and topic 33 (shipping). In cluster three, diesel was mentioned in topic 23 related to trucking firms. In cluster five, studies on airlines mentioned gas in two topics and a shipping-related topic. In cluster six, coal was mentioned for the rail-related topic 39. Storing fuel via hedging and stocking when prices are low is one method used. However, that approach only works for a short period, as stored fuel is usually exhausted before a high-price cycle is over. Furthermore, holding inventory is an expensive way to manage fuel risk. Alternative energy sources are obviously to be considered, although the development of reliable delivery systems and the cost of generating the extra electricity need to be considered.

The weather seems to have more variance in the minds of those stressing global warming. Whether that is so or not, keeping close track of short-term weather seems to be improved with better short-term weather forecasting. Weather was mentioned in cluster three in a topic related to trucking. Ships need to worry about hurricanes, which are closely tracked, enabling rerouting. Trucking firms can, at least to some extent, reroute to avoid blizzards. Airplane firms, on the other hand, are more negatively impacted by short-term weather events, as those who fly regularly know full well.

The COVID-19 pandemic has had a severe impact on logistic firms. A great deal of this impact is due to heavy reliance on Chinese manufacturing. The pandemic was mentioned in cluster two (an air-related topic) and in cluster four in a topic related to interaction between shipping and trucking. An obvious mitigation is to obtain alternative sources, although that is very difficult to do in an economy manufacturing complex products with thousands of components, many of which have very limited options due to essentially monopolistic sourcing. It is simply too expensive to eat high fixed costs for production facilities. However, in the long run, supply chains need to consider risk more in their creation of linked sources of components.

6. Conclusions

Redundancy obviously would be beneficial in terms of obtaining operating personnel versus automation and redundant fuel supplies, and alternative manufacturing sites would have been useful during the COVID-19 pandemic. However, generating such redundancy takes time, violates the focus on minimizing operating costs due to higher inventories, and involves its own risk. Redundancy also provides more tools for flexibility and agility. This study considered text as data and adopted an innovative topic modeling method to discover supply chain management (SCM)-related risks from annual reports. In topic modeling applied in the literature, methods have been successfully adopted to discover valuable insights from different types of text data, including news articles, social media
posts, and journal abstracts [49,60]. A method to select the number of topics was applied, as shown in Figure 3, based on minimizing residuals.

6.1. Implications

Topic modeling revealed terms obtained from the SEC filings of logistic firms over the period 2006–2021. Table 4 recapitulates the dominant risk terms by industry. All industries faced regulatory issues. Most have had problems hiring people in recent times. Fuel has been a serious issue for the air industry. COVID-19 impacted trucking, shipping, air, and rail industries from 2020 on. We categorized these in seven groups: personnel (hiring shortages), fuel (price fluctuations), weather, risks (adverse events), international events, the pandemic, and regulatory issues. The trucking industry has had major issues in finding drivers but also faces risks involving fuel, winter weather, and hiring issues. The air industry has been severely threatened by fuel price variations while also facing personnel shortages. The rail industry seems to have faced fewer threats, although specific regulatory issues were found. The pipeline industry did not have as many personnel shortages or fuel issues but had more regulatory issues.

6.2. Contributions

Structural topic modeling (STM), an unsupervised text analytics approach, was demonstrated to be a means to identify key logistics risks in supply chains, their cluster relationships, and their trends over time. Supply chain managers and analysts can utilize this text analytics technique to learn about logistics risks in their industries and supply chains and take appropriate measures in advance. We concluded with the identification of supply chain risks and means of mitigation. These include designing systems with more options and developing linked communication networks to more closely monitor events. STM also offers value to academic research. A review of previous studies in Section 2 indicated that bibliometric analysis is a popular method for studying supply chain risks in the literature. Our study shows that STM in particular—and topic modeling in general—can play an important role as a tool for academic research on supply chain risk management.

6.3. Limitations and Future Research Directions

While this computational approach is needed for processing and analyzing large volumes of text data, such an approach is not without limitations and, in response, various methods have been developed for model selection and greater model accuracy. However, several challenges remain as diverse digital texts with different characteristics are introduced as data for topic modeling. As mentioned earlier in the paper, topic modeling requires the number of topics from researchers/analysts and, according to the residual method for model selection, high k values between 40 and 100 appear to indicate the optimal number of topics. This challenge remains strong as topic modeling is applied to long texts like annual reports and full journal texts [61].

We make two suggestions for future research. The first suggestion is regarding the utilization of different text preprocessing approaches. For example, the challenge of long texts can be addressed by splitting each annual report into multiple paragraphs [62]. In this case, one paragraph or a set of paragraphs can be treated as a single document during the topic modeling process, and then it may even be assumed that each document contains only one topic rather than multiple topics. The second suggestion is to utilize state-of-the-art algorithms for topic modeling. Topic modeling as a method for text analysis has drawn much attention and there are emerging algorithms combining topic modeling with word embedding techniques in the field of natural language processing (NLP). For example, Top2Vec combines document clustering with word embedding and automatically detects the number of topics in texts [63]. Similarly, BERTopic discovers topics through the clustering of word embeddings and offers a hierarchy of topics, allowing researchers to choose the optimal number of topics [64]. These new algorithms address the challenges
from model selection and can potentially generate more semantically coherent topics than traditional algorithms.

**Author Contributions:** Conceptualization, D.O.; methodology, B.C.; data curation, B.C.; writing—original draft preparation, D.O.; writing—review and editing, B.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data used in this study can be accessed through SEC API (https://sec-api.io/, accessed on 6 July 2023).

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


18. Etemadi, N.; Borbon-Galvez, Y.; Strozzì, F.; Etemadi, T. Supply chain disruption risk management with blockchain: A dynamic literature review. *Information* 2021, 12, 70. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.