Journal of Industrial and Management Optimization

doi:10.3934/jimo.2022190

A BIBLIOMETRIC AND SOCIAL NETWORK ANALYSIS OF DATA-DRIVEN HEURISTIC METHODS FOR LOGISTICS PROBLEMS

Nurcan Deniz[1] and Eren Ozceylan[2]*

[1]Department of Business Administration, Eskişehir Osmangazi University, Eskişehir, Turkey

[2]Department of Industrial Engineering, Gaziantep University, Gaziantep, Turkey

(Communicated by Kok Lay Teo)

Abstract. Transport and logistics systems include a range of activities that deal with all sorts of decisions and operations from material handling to vehicle routing. One of the main challenges for transport and logistics processes is to deal with large-scale and complex problems. However, with increasingly diverse sets of operational real-world data becoming available, data-driven heuristic approaches are promising to pave the path for solving the problems in the field of transport and logistics. Thus, a comprehensive review is needed to observe the reflections of this path in literature. To bridge this gap, a total of 40 papers on the topic of "data-driven heuristic approaches to logistics and transportation problems" are determined. Before the categorization and content analysis; descriptive, bibliometric and social network analysis are carried out to identify the current state of the literature. All the papers are systemically reviewed based on different perspectives, namely data-driven methodology, heuristics, sub-problems and etc. Based on the review, suggestions for future research are likewise provided. Subsequently, machine learning and deep learning methods are considered to be among the most promising data-driven methodologies. The review may be useful for academicians, researchers, and practitioners for a better understanding of data-driven heuristic approaches to transportation and logistics problems.

1. **Introduction.** Transportation and logistics systems include a range of activities. These activities deal with all sorts of decisions and operations from vehicle routing and scheduling to distribution of passengers and products in different regions. Decision makers in transportation and logistics systems deal with large-scale and complex problems at minimum cost while satisfying the needs of service users. In addition, the ever-changing global economy, fuel price instability, just-in-time requirement of customers, and the necessity of cost-effectiveness add to the complexity of the ecosystem. In this complex and dynamic ecosystem, along with the reliance on many moving parts at any point in the supply chain, logistics and transportation systems may be a perfect use case for big data.

Operators aim at solving these complex problems covered big data in a computationally efficient way. In recent years, with the increase of demand, this task has

2020 *Mathematics Subject Classification.* Primary: 00A64, 90-02; Secondary: 90B06, 90C59.

Key words and phrases. Data-driven, heuristic, systematic literature review, bibliometric analysis, social network analysis, logistics, transportation.

*Corresponding author: Eren Ozceylan.

become more challenging. However, with increasingly diverse sets of operational data becoming available in transportation and logistics systems, incorporating advanced technologies is promising to pave the path for improving the already existing models. One of these existing models is data-driven optimization methods that are motivated by available real-world data to solve the traditional problems.

Thanks to the developments digital and automation technologies, real time data can easily be collected from different sources. Data-driven optimization methods which combine traditional optimization methods and artificial intelligence prediction tools, became a fresh research area in the operations research field [8].Instead of abstracted models and simplified assumptions it can be possible to integrate real-time structured and unstructured information with data-driven models. Also data-driven decisions are delineated as remarkable, practical, and reasonable [47]. Control parameters can be set up and adjusted according to the historical data in the algorithms of the data-driven heuristics [55].

United Parcel Service (UPS) is a real-world case of big data logistics leading to big savings. UPS discovered that trucks turning left was causing a lot of delays, wasted fuel, and safety risk according to data analysis. These problems caused loss of money. In one of the posts titled "Why UPS drivers don't turn left and you probably shouldn't either", UPS stated that they used "10 m gallons less fuel, emits 20,000 tonnes less carbon dioxide and delivers 350,000 more packages every year" after switching to route optimization [41]. According to the report of "21st Annual Third Party Logistics Study", 98% of third party logistics providers (3PLs) said that improved data-driven decision making is "essential to the future success of supply chain activities and processes". Additionally, 81% of shippers and 86% of 3PLs expressed that using big data effectively will become "a core competency of their supply chain organizations" [15].

There are some review papers related to data-driven methodologies in some specific areas such as train dispatching [47] in the literature. However, to the best our knowledge there is not any review paper about data-driven heuristic methods focuses on transportation and logistic problems. The purpose of this paper is to close this gap and show the state-of-art in this area. The motivation behind this study is to shed the light for researchers. The researchers who want to make contributions in the data-driven heuristic methods in logistics context, will find a good starting point. We aimed to use both bibliometric and social network analyzes in addition to traditional systematic literature review and content analysis methodologies. Bibliographic records -author(s), address, title, cited references, keywords and research areas- of articles give chance to researchers to assess the scientific standing, identification of specialties, and the evolution of research traditions. Social network analysis enables to extract specialties and research traditions from citations. The main goals in social network analysis are detecting and interpreting patterns of social ties among actors [29]. Eighteen research questions were answered in the context of analyzes.

The rest of this paper consists of four main parts. First of all, the review methodology will be introduced. In the next section, readers can find the results of the detailed analyses of the literature. The paper ends with a conclusion section in which limitations and future directions are placed.

2. **Review methodology.** In addition to the descriptive and content analysis based traditional systematic literature review, bibliometric and social network analysis are used in this study. This review followed the guidelines proposed by [11] in order to ensure that it is systematic, replicable, thorough, and transparent in the traditional systematic review part. The steps suggested by [11] involve (i) the keywords, (ii) first refinement, (iii) inclusion and exclusion criteria, and (iv) second refinement.

We used networks in addition to table and graphics to detect and interpret the patterns of social ties among [29]. BibExcel developed by Olle Persson was used to make bibliometric analysis [35]. BibExcel can also produce input files for Pajek and MS Excel and prepare input file for Pajek. We used BibExcel version 2016.

Pajek 5.14 developed by De Nooy, Mrvar, and Batagelj was used to build co-occurrence networks [3]. MS Excel was also used to complement descriptive analysis and to make content analysis. Figure 1 depicts the flowchart of the review methodology.

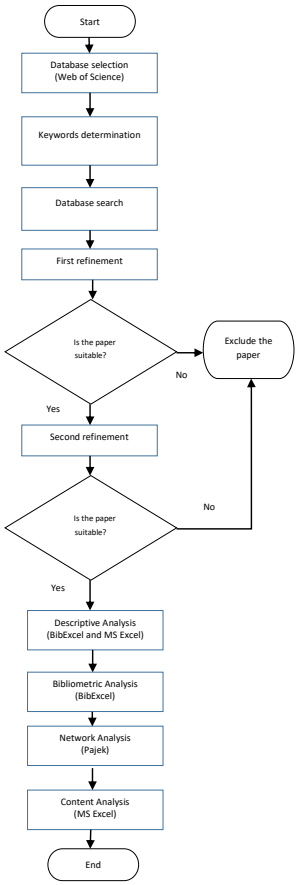


FIGURE 1. Flowchart of the Review Methodology

2.1. **Keywords.** Defining the keywords through an iterative process is the first step in a successful systematic review [34]. Group-based keywords [11] were defined based on the literature. Three groups (Group A, Group B, and Group C) of keywords are shown in Table 1. In order to generate the queries, the logical operator "AND" was used to create Boolean keywords combinations as '(keyword of group A) AND (keyword of group B) AND (keyword of group C). Totally 22 queries were made based on three-dimensional keyword structure and combination of keywords on WoS. The first dimension (Group A) consists of "data-driven", the second dimension (Group B) is "heuristic" part, and the third dimension consists of logistics and transportation related terms. The third dimension (Group C) of 16-22 queries was inspired by the work of [48].

TABLE 1. Database Search Information

Keywords			Query results	Included	Duplicated	Excluded
Group A	Group B	Group C				
Data driven	heuristic	logistic	12	4	8	
		transport	22	7	15	
	ad. large neighb.	logistic	1	1		
		transport	1			1
	heuristic	vehicle routing	13	5	2	6
		routing	26	6	7	13
	2-opt	logistic	0			
		transport	0			
	hill climbing	logistic	0			
		transport	0			
	lin karnighan	logistic	0			
		transport	0			
	greedy	logistic	5	1	3	1
		transport	6	1	4	1
	heuristic	supply chain	11	2	2	7
		material handling	3	2	1	
		storing	7	4	3	
		order picking	2		1	1
		warehouse	0			
		unitising	0			
		packaging	9	1	8	
		inventory	18	6	8	4
TOTAL			136	40	56+6=62	34

2.2. **First refinement.** Using the "all fields" search field of "Web of Science" database and limiting to documents published in English and Turkish, our keyword structure yields 136 papers published through January 2022. The search was not restricted according to research areas. [34] stated that peer-reviewed papers are important for systematic reviews. Hence, our search was refined to show only peer-reviewed papers. The marked list of WoS was transformed to a plain text containing all relevant information to be input for BibExcel and also a MS Excel file to make content analysis. To answer research questions, authors, publication years, journal names, keywords, titles, citations are generated from WoS database.

2.3. **Inclusion and exclusion criteria.** The exclusion criteria are determined as book chapter, proceeding, and unrelated disciplines such as medicine etc. Table 1 depicts the details about database research. 56 papers were excluded and 34 papers were determined as duplicated. As a result of first evaluation (title, abstracts and keywords), 46 papers remained which are non-duplicated.

2.4. **Second refinement.** All eligible papers analyzed in detail in the last phase called second refinement. In this step, we removed six papers which were seen relevant according to the abstract at first glance but out of scope after reading in detail. After this refinement, 40 papers were remained.

3. **Detailed analyzes of the literature.** The main research question is "What is the current status of the data-driven heuristics in logistics and transportation?" in this study. To make the results of this analysis more insightful, there are also other research questions (RQ) consistent with the research methodology. These eighteen questions are listed below under the analysis types in which they will be answered.

Descriptive analysis related research questions
RQ1. What is the chronological status of the area?
RQ2. In which journals most of the papers published?
RQ3. In which research areas the problems are studied?
Bibliometric analysis related research questions
RQ4. Who are working in this area?
RQ5. Who are working together?
RQ6. Which papers are mostly cited?
RQ7. Who are co-cited?
RQ8. Which keywords are mostly used?
RQ9. Which words are mostly stated in paper titles together?
RQ10. Which words are co-worded in titles?
Social network analysis related research questions
RQ11. How is the co-author network?
RQ12. How is the co-citation network?
RQ13. How is the co-word network?
Content analysis related research questions
RQ14. Which problems are studied?
RQ15. Which data-driven methodologies are used?
RQ16. Which heuristic types are used?
RQ17. What are the objectives (functions)?
RQ18. What are the future needs of research?

3.1. **Descriptive analysis.** Publication by year, publication per journals, and publication per research area are analyzed in the context of descriptive analysis to answer RQ1, RQ2, and RQ3 respectively.

3.1.1. *Publications by year.* The earliest study in the context of this review was published in 1998. In the following eighteen years, only six papers were published which means one paper per three years. However, only one of them was published in 2001 and the others were published between 2012-2016, this means one paper per each year. From Figure 2, it can be seen that the numbers of papers related to data-driven heuristic in transportation and logistics problems start to increase in 2017. This escalation indicates an increase in the interest of topic. The majority of the papers (77.5%) were published in the last three years. This increase can be related with the need for handling big data in the digital transformation context.

3.1.2. *Publications by journals.* Table 2 depicts the distribution of publications across journals. 40 papers were published in 32 different journals. It can be easily seen that from Table 2, "Production and Operations Management" is the prominent

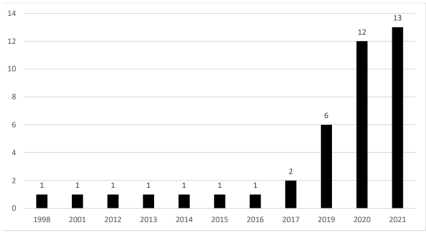


FIGURE 2. Paper numbers according to years (Early access papers are categorized in 2021)

journal in this area which published 10% (n=4) papers. "Operations Research" follows this journal with three papers published. "International Journal of Production Economics", "Transportation Science", and "Computers and Industrial Engineering" are the journals in which two papers were published. As a result, 32.5% (n= 13) of all published papers were published in these five journals. All of these five journals are also Q1 categorized journals which means highly impacted. To draw a conclusion, there is not a journal dominance and the variety of journals is very high.

3.1.3. *Publication per research area.* According to the Table 3, the half of the papers (50%) is categorized in "Operations Research & Management Science" research area as expected. "Transportation Science & Technology", "Engineering, Civil", "Engineering, Manufacturing", "Management", and "Transportation" are the hot research areas more than five papers were published.

3.2. **Bibliometric analysis.** The bibliographic records consist of author(s), title, cited references, keywords and research areas. The initial step is cleaning the bibliometric records from the different spelling of the same author, big/small letters to make accurate analysis. The bibliometric records were refined according to cited references (CD), authors (AU), titles (TI), keywords (DE), and research area (WC).

3.2.1. *Author analysis.* Initially, there are only four authors (Shen Z, Zhang W, Yang X, and Zhou C) wrote more than one paper in the context of RQ4. They wrote two papers in this field. Unfortunately, there is not any co-author in more than one paper. According to this result, the RQ5 can not be answered.

3.2.2. *Citation analysis.* According to the RQ6, Table 4, Clarke (1964) is the most cited author with five papers. Toth (2014) and Dantzig (1959) state at the second rank with four papers. The other authors listed in Table 5 are the authors who were cited in 3 papers. Although Dantzig G (1959) and Solomon M (1987) are the principal references in vehicle routing problems, the references such as Bao J (2017) needs to be highlighted for researchers.

After citation analysis, co-citation analysis was also conducted to answer RQ7. According to the results, there are only six co-cited papers as depicted in Table 5. Levi (2007), Gallego (1993), Godfrey (2001), Gallego (1993), and Scarf (1958) are

TABLE 2. Journals in which papers published

Journal Name	ISSN	Number of Papers
Production and Operations Management	1059-1478	4
Operations Research	0030-364X	3
International Journal of Production Economics	0925-5273	2
Transportation Science	0041-1655	2
Computers and Industrial Engineering	0360-8352	2
Networks	0028-3045	1
Manufacturing and Service Operations Management	1523-4614	1
Journal of Heuristics	1381-1231	1
Journal of Purchasing and Supply Management	1478-4092	1
Journal of the Operational Research Society	0160-5682	1
Ocean Engineering	0029-8018	1
Transportation Research Part D	1361-9209	1
Transportation Research Part E	1366-5545	1
Transportation Research Record	0361-1981	1
Pamukkale University Journal of Engineering Sciences	1300-7009	1
Royal Society Open Science	2054-5703	1
Transportation Research Part B	0191-2615	1
Complexity	1076-2787	1
Computers in Industry	0166-3615	1
Decision Support Systems	0167-9236	1
Complex & Intelligent Systems	2199-4536	1
ACM Transactions on Human-Robot Interaction	2573-9522	1
ASIA-Pacific Journal of Operational Research	0217-5959	1
Automation in Construction	0926-5805	1
Engineering Applications of Artificial Intelligence	0952-1976	1
International Journal of Advanced Computer Science and Applications	2158-107X	1
International Journal of Production Research	0020-7543	1
Journal of Advanced Transportation	0197-6729	1
IEEE Transactions on Knowledge and Data Engineering	1041-4347	1
European Journal of Operational Research	0377-2217	1
Geoinformatica	1384-6175	1
IEEE Transactions on Intelligent Transportation Systems	1524-9050	1
TOTAL		40

the co-cited authors in this Table. Among them, the oldest citation is Scarf (1958) and the latest citation is Levi (2007).

3.2.3. *Keywords analysis.* Papers were also analyzed in terms of keywords. Keywords are important elements to find the related journals in data bases. Table 6 depicts the highly used keywords in a descending order as an answer for RQ8. "Optimization" has the highest usage count in 25% (n=10) papers. "Chain" and "Routing" follow this keyword with 9 usages. The other keywords used more than 5 are "supply", "programming", "vehicle" and "planning" respectively. These results are consistent with the problem types and methodology but "data-driven" keyword is not placed at the top as expected. The details of these results can be found in "Section 3.4 Content Analysis" part.

3.2.4. *Title analysis.* Titles of the papers are also as important as keywords to find the related papers. Ten words used in paper titles more than two times are shown in Table 7. As expected "data-driven" states at the top of the list contrary to keyword results. This can be a result of an attempt to attract the readers for a hot topic. There is not a data-driven methodology name in the title. "Problem", "routing", "vehicle", and "approach" are the other words used at least five times in the title. These results are the answer of RQ9.

TABLE 3. Research areas in which papers published

Research Area	**Number of Papers**
Operations Research & Management Science	20
Transportation Science & Technology	8
Engineering, Civil	7
Engineering, Manufacturing	7
Management	7
Transportation	6
Engineering, Industrial	5
Computer Science, Artificial Intelligence	5
Computer Science, Information Systems	3
Engineering, Electrical & Electronic	3
Computer Science, Interdisciplinary Applications	3
Multidisciplinary Sciences	2
Computer Science, Theory & Methods	2
Engineering, Multidisciplinary	2
Economics	2
Mathematics, Interdisciplinary Applications	1
Oceanography	1
Robotics	1
Construction & Building Technology	1
Computer Science, Hardware & Architecture	1
Automation & Control Systems	1
Engineering, Marine	1
Geography, Physical	1
Environmental Studies	1
Engineering, Ocean	1
TOTAL	40

In addition to co-author and co-citation, bibliometric analysis gives also a chance to make co-word analysis. In the context of RQ10, co-words occurred in the title more than two times are listed in Table 8. "Data-driven" and "routing" couples states in the first rank in this list with five papers. The consecutive couples are "data-driven" and "vehicle"; "data-driven" and "problem"; "vehicle" and "routing". This result is important as being a messenger of the problem type analyzed in content analysis (Table 9).

3.3. **Social network analysis.** In addition to traditional literature reviews based on content analysis, recently, social network analyses are also used. The powerful side of this analysis is to give opportunity to make co-occurrence analysis. Mutual appearances of pairs of units over a consecutive number of bibliographic records can be determined with co-occurrence analysis [36]. Co-author, co-citation and co-word are types of co-occurrence analyses. In an attempt to answer RQ11, RQ12, and RQ13 all of these analyses were made respectively.

3.3.1. *Co-author analysis.* The academic cooperation among the active authors in the research area can be visualized with co-author analysis [53]. In the Figure 3 in which social networks are visualized, it can be easily seen that there is a disconnected network and there is not big components in network. This network reveals that the

TABLE 4. Cited papers more than two times

Cited Author	Number of Papers
Clarke G, 1964	5
Toth P, 2014	4
Dantzig G, 1959	4
Solomon M, 1987	3
Scarf H, 1958	3
Ropke S, 2006	3
Bao J, 2017	3
Liyanage L, 2005	3
Gallego G, 1993	3
Pisinger D, 2007	3
Godfrey G, 2001	3
Levi R, 2007	3
Breiman L, 2001	3
Bertsimas D, 2006	3

TABLE 5. Co-cited papers more than two times

Citation 1	Citation 2	Co-citation
Gallego G, 1993	Levi R, 2007	3
Levi R, 2007	Scarf H, 1958	3
Gallego G, 1993	Godfrey G, 2001	3
Godfrey G, 2001	Scarf H, 1958	3
Godfrey G, 2001	Levi R, 2007	3
Gallego G, 1993	Scarf H, 1958	3

area is immature. There are only four researchers (Yang X, Zhang W, Zhou C and Shen Z) who are in the author list of papers more than one time. This four initial phase components highlighted with circles around them. It is expected that these components will be enlarged with new scholars in this area.

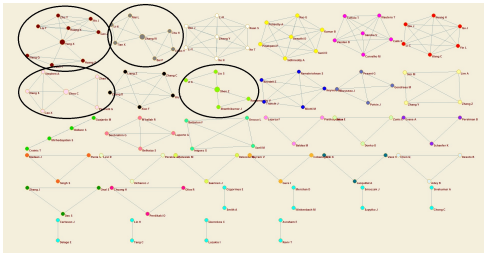


FIGURE 3. Co-author network

3.3.2. *Co-citation analysis.* The second co-occurrence analysis called co-citation analysis. Co-citation occurs when two papers share one or more same references. Co-citation analysis also gives us chance to show the authors who could but did not collaborate actively [12]. The size of each node is proportional with the number of citation [23]. As it can be seen from Figure 4, Clarke G is the biggest node

TABLE 6. Keywords more than three times

Keyword	**Usage Count**
Optimization	10
Chain	9
Routing	9
Supply	8
Programming	7
Vehicle	6
Planning	6
Prediction	5
Algorithm	5
Data	5
Data-Driven	5
Location	4
Logistics	4
Scheduling	4
Heuristic	4
Problem	4
Time	4
Search	4
Model	4

TABLE 7. Words in the title more than two times

Words in the Title	**Usage Count**
Data-Driven	15
Problem	8
Routing	7
Vehicle	5
Approach	5
Time	4
Chain	4
Optimization	4
Urban	3
Based	3
Model	3
System	3
Control	3
Bike	3
Supply	3
Windows	3
Planning	3

according to the results depicted in Table 4. Clark G wrote this mostly cited article -Scheduling of Vehicle Routing Problem from a Central Depot to a Number of Delivery Points- with Wright JR. It is an expected result taken into consideration of Vehicle Routing Problem's importance in logistics area. In addition, Clarke and

TABLE 8. Co-words in the title more than two times

Word 1	Word 2	Times
Data-Driven	Routing	5
Data-Driven	Vehicle	4
Data-Driven	Problem	4
Vehicle	Routing	4
Time	Windows	3
Routing	Problem	3
Bike	Trajectories	3
Supply	Chain	3

Wright saving algorithm -which is one of the most known algorithm in VRP – was proposed in this article.

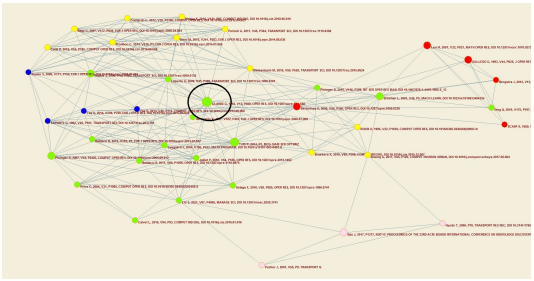


FIGURE 4. Co-citation network

3.3.3. *Co-word analysis.* The last but not least, it is possible to make co-word analysis based on titles of papers, to detect which word couples are mostly used in the titles. By the same token, the size of each node is proportional with the number. Figure 5 visualizes the co-word analysis results. Taken into consideration with Table 8, the biggest node in the network is data-driven. In addition, "data driven" is in the first three rank in Table 9 with being "routing", "vehicle", "problem". This results are also the messenger of highly studied problem in this area. This result also shows the importance of determining the title which focus on the methodology.

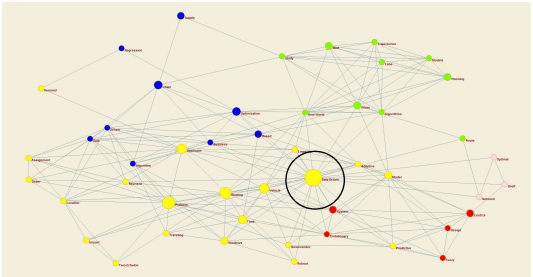


FIGURE 5. Co-word network

3.4. **Content analysis.** Content analysis is defined by Krippendorff (2004) as "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" it helps researchers to describe and quantify specific phenomena [11]. In this section, we divide the papers based on problem type (RQ14), data-driven methodology (RQ15), and heuristic type (RQ16). In addition the objectives (functions) (RQ17) of the papers are also analyzed.

3.4.1. *Problem type.* "Vehicle routing problem" (VRP), which is the generalization of the travelling salesman problem (TSP), is one of the most studied optimization problems in the literature [3]. Similarly, in our study VRP is the most studied problem. It is shown in Table 9 that 15% papers (n=6) deals with VRP. "Newsvendor Problem" follows VRP with 3 papers. "Bike lane planning", "location routing problem (LRP)", "material handling", and "travelling salesman problem" (TSP) are the other problem types studies in at least two papers.

TABLE 9. Problem type based content analysis results

Problem type	**Frequency**
Vehicle routing problem (VRP)	6
Newsvendor Problem	3
Bike lane planning	2
Freight distribution/Location routing problem (LRP)	2
Material handling	2
Travelling salesman problem (TSP)	2
Credit term determination	1
Eco-driving	1
Facility layout/ Shelf space allocation	1
Fastest route recommendation	1
Flight timetabling and fleet assignment	1
Inventory Shrinkage Problem/phantom stockout	1
Labor-planning	1
Last-Mile Problem/ the joint order assignment and routing	1
Multi-Channel Switching	1
Multimodal Logistics	1
Network growth	1
Optimal route selection	1
Order picking/ joint order batching and generalized assignment model	1
Production–distribution problem	1
Route planning	1
Shelf space selection	1
Storage Location Assignment Problem (SLAP)	1
Supply chain coordination	1
Supply chain production and distribution planning	1
Train speed profile optimization	1
Unit Loading Device handling sequence, workstation assignment, and vehicle schedule	1
Urban transport/vehicle scheduling	1
Vehicle sequence	1

3.4.2. *Data-driven methodologies of papers.* According to the classification of [47], statistical methods (SM), graphical models (GM), and machine learning (ML) are the main types of data-driven methodologies. The data-driven mechanism in [37] paper exploits pruning rules developed in operations research by using them not only a priori but also dynamically during the search. After three years from this paper discrete event simulation was used by [42]. Linear correction and support vector regression approaches [40], robust optimization framework in which partitioning

the region with respect to the worst-case distribution that satisfies first and second-order moments information [5] and Informative Probabilistic Bound [20] are the other data driven approaches in the former literature.

In addition, predictive/forecasting models based on fuzzy logic [43], ARIMA [9], multilevel modeling [25], and linear projection [50] were also reported.

In the last three years according to the digital transformation and big-data, machine learning [4, 17, 26, 14, 55, 8, 16, 31] based data driven methodologies are highly used.

In detail [17] used Random Forest Regression (RFR) algorithm and Support Vector Machine Regression (SVR) algorithm in the machine learning methodology. SVM was also used by [55] in addition to Generalized Linear Models (GLM). On the other hand, [26] used Gaussian Mixture Model (GMM) with K-mixture components using an Expectation-Maximization (EM) algorithm, Principal Component Analysis (PCA), Ordinary Least Squares (OLS) regression. K-Means Clustering [10], Spatial Clustering-Based Initialization [16], Cluster Analysis [33, 52], association rule mining [31] are the other machine learning methodologies.

As a superset of machine learning, [38] used artificial intelligence. Beside, as a subset of machine learning, there are also some papers related with deep learning. Artificial Neural Network (ANN) [14], deep learning-based demand estimation [39], and adaptive graph convolutional recurrent networks [44] were used in the papers in this context.

The stochasticity was modeled using infinite horizon Markov chain model by [1] and a set of scenarios based on historical data by [2]. The paper of [23] differs from the other studies using game-theoretic framework (Stackelberg game). Dynamic programming is another methodology that used by [45] and [6].

[27] preferred to use simulation in their papers. Regression and extrapolation guided optimization method (REGO) [49], Generalized Attraction Model (GAM) [46], distributionally robust optimization [51], multinomial logit model [24], and an information-based approach with a multi-criteria index function [54] are the other data-driven methodologies stated in the literature.

3.4.3. *Heuristic types of papers.* Heuristics are defined as a subset of strategies to find the near optimal solutions [32]. In terms of heuristic types used in the papers we see the constraint logic programming (CLP) and exact branch-and-bound optimization algorithms firstly in [37] paper. Branch-and-bound algorithms were also used by [5] and [2]. Similarly, exact branch-and-cut approach was used by [51]. [51] not only used branch-and-cut approach but also they used variable neighborhood search (VNS) meta-heuristic algorithm. This study is not the only paper used VNS. VNS was combined with the first-free approach and used as Adaptive Large Neighborhood Search (ALNS) algorithm by [25]. Additionally, [2] used specialized ALNS heuristic with branch-and-bound algorithm. On the other hand, [26] used large neighborhood search heuristic. In addition to VNS, tabu search [55, 31], enhanced quantum genetic algorithm [22] and simulated annealing (SA) [8] are the other metaheuristics used in this context.

Greedy heuristics [17, 16, 10, 38, 30] are highly used heuristic types in this context. As an example [30] used two types of greedy heuristics called Largest-to-Second (L2S) and Largest-to-Closest (L2C) to solve the network growth problem in urban transportation.

[42] used dispatching rules, [40] used operational statistics based heuristics, [43] used the evolutionary algorithm (EA), and [20] used sample average approximation

(SAA) approach where [9] used a data-driven staffing heuristic that incorporates the prediction loss function and uses past traffic to predict optimal labor. On the other hand, [14] preferred to use modified version of Dijkstra's algorithm.

Multi-start data-driven evolutionary heuristics [4], Linear Moving Average Trend (LMAT) heuristic [50], fixed credit term heuristic [23], spatial mental models [33], Fleet-Fixing Approach and Symmetry-Inducing Approach [46], k-means algorithm [52], Bernoulli shrinkage process [6], a modification of the heuristic Clarke–Wright algorithm [55], Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [7], temperature-bounding heuristic [19], decomposition-based heuristic [28], Lagrangian Relaxation heuristic [24], Multi-Objective Convergent Optimisation via Most Promising Area Stochastic Search (MO-COMPASS) algorithm [54] and A* algorithm [44] are the other heuristics types found in literature.

3.4.4. *Objectives of the papers.* In this part of the content analysis we aimed to analyze the objectives of the papers. First of all, the objective functions of the papers are listed in which mathematical programming are used. In the other papers, the general objectives of the papers are deduced.

Objective functions are categorized as minimization and maximization based. As expected, "minimization of costs" was the objective function mostly used. In 37.5% (n=15) papers [8, 37, 20, 43, 4, 14, 52, 38, 39, 6, 49, 7, 19, 28, 13] it was aimed to minimize costs. Thanks to the increasing importance of green issues and sustainability, minimizing energy consumption [17] started to find place in the objective functions. Minimizing the sum of total order processing time [25], minimizing the expected route duration plus penalties for late arrivals [2], minimizing the space [10], minimizing the waiting time cost of passengers [22], and minimizing the delays [51] are the other minimization based objective functions in the literature.

Minimizing energy consumption can be expressed as to maximize energy savings [27] also. Maximizing the profit follows minimizing costs in maximization based objective functions interms of frequency. In three papers [40, 46, 21] it was aimed to maximize profits. [50] and [31] differs from these studies with aiming to maximize total revenue. Both of minimizing costs and maximizing profits objectives shows the dominance of economic side of sustainability. To maximize overall usage [16], and to maximize riders' utility [24] are the other maximization based objectives.

According to the type of objective functions, there are also bi-objective papers in which two objectives are optimized simultaneously in a multi-objective way. As an example [45] aimed to optimize the vehicle sequence and fuel consumption simultaneously. Optimizing fuel consumption is not only important for minimizing costs bu also important for reducing the carbon emissions in sustainability concept.

To visit a set of cities in the shortest possible path [55] is the general objective of VRP. Thanks to the data-driven methodology, the objective can be expressed as to find the real-time fastest route [44]. In terms of general objective of the papers, to reduce factory cycle time and distribution [42], to find an optimal partition [5], to improve staffing decisions and store performance [9], to control expiration of perishable inventory [1], to predict near-optimal route distances [26], to improve the connectedness and directness of network with potentially crucial implications for sustainable transport policy planning [30], improving the speed and quality of routes produced by human-intelligent agent (H-IA) teams [33], to reduce sorting operation duration (SOD) and the Route Fulfilment Time (RFT) [54] can be listed.

4. **Conclusion and research implications.** Data-driven methodologies gain importance in last years according to the digital transformation and big-data. They are used with heuristics in a diversity of reseach area. One of the important areas is logistics and transportation. According to this importance the articles published in this area were analyzed in detail to answer "What is the current status of the data-driven heuristics in logistics and transportation?" In addition 18 research questions were also determined to analyze the area in detail. These research questions are related with descriptive (year, journal, research area), bibliometric (authors, citations, keywords and words in title), social network (co-author, co-citation and co-word) and content (problem type, data-driven methodology, heuristic type, objective (function) and future studies) analysis.

One of the contributions of this research is to apply a systematic literature review based on data-driven heuristics in logistics and transportation area. Best of our knowledge, this paper is the first attempt in the literature. This paper differs from the traditional review papers with using bibliometric and social network analysis tools. In order to achieve this, MS Excel, BibExcel and Pajek based analysis were made. It can be resulted that, this topic has acceleration in the last three years. According to the development in big-data and easy access to the real-time data, data-driven methodologies are expected gain more importance.

The main findings of the paper conducted are as follows in accordance with research questions:

RQ1. The paper numbers increased in the last three years.

RQ2. Despite the "Production and Operations Management" is the journal in which highest paper have been published, there is richness in the journal type.

RQ3. "Operations Research & Management Science" research area is in the first rank.

RQ4. Shen Z, Zhang W, Yang X, and Zhou C wrote more than one paper.

RQ5. Can not be answered due to lack of any co-author in more than one paper.

RQ6. Clarke G (1964) is the mostly cited paper.

RQ7. There are only six co-cited papers.

RQ8. "Optimization" is the mostly used keyword.

RQ9. "Data-driven"is the mostly stated word in paper titles.

RQ10. "Data-driven" and "routing" are the mostly co-worded in titles.

RQ11. Co-author network is disconnected.

RQ12. "Clarke G" is the biggest node in co-citation network.

RQ13. "Data-driven" is the biggest node in co-word network.

RQ14. Vehicle Routing Problem is the highly studied problem type.

RQ15. Machine learning is generally used in the data-driven methodologies.

RQ16. Greedy heuristics are mainly used in this context.

RQ17. Cost minimization is the mostly used objective function.

This study has several limitations. The first limitation is related with the search process. As the process starts with a set of keywords, this brings a risk not to capture all relevant papers. The other limitation is using only Web of Science database. To review only "articles" and the language (English and Turkish) of the papers are also the other limitations. Also studying with relatively small number of papers can be seen as a limitation but the immaturity of the research area is the reason of this. The two-dimensional visualization of social network analysis is another limitation related with the methodology.

In this section, possible future researches are explored to answer final research question (RQ18). This paper not only aims to show the state-of-the art but also reveal the future directions. As the authors of this paper, our motivation is being an inspired starting point for researchers in this area. These directions are listed below in the categories:

- As this is the first attempt to discover the research area, in the future studies the other databes such as Scopus or Google Scholar can be used to expand the papers.
- To gain more insight conference proceedings, books, and book chapters can also be reviewed in the future studies. At this point researchers who has an intention to add conference proceedings needs to be careful in accessibility limitations, being peer- reviewed and potential duplication of proceedings with published papers [34].
- In the next literature reviews on the same issues, the social network will be more connected and it is expected to see big components in the structure. This means researchers can find the hubs and authorities in the network.
- As special issues are important in generating interest amongst researchers [34] it will be a good idea to publish special issues (SI) in leading journals for dissemination and elaboration.
- In the latest research it is highlighted that the development of predictive analytics models based on the huge amount of data available from sensors with data-driven models and methods is an important avenue for researchers.
- A contemporary trend is combining data-driven methodologies and traditional mathematical models [47]. In this context, the researchers can establish such hybrid models.
- Other uncertainties that arise in practice can be analyzed [19].
- The data handling capacity can be enhanced [22].
- Behavioral variables of workers can be included [25]. These variables can be related with customers also [28].
- Alternative prediction methods can be explored to enhance flexibility and to improve predictive performance [26].
- Hybrid demand estimation methods can be used [39].
- Faster and more effective algorithms can be applied [8].
- Objectives related with environmental and social dimension of sustainability can be added to model.
- Longitudinal studies can be conducted [30].
- Interdisciplinary studies will be helpful to get a holistic insight. As an example market basket data can be integrated with the path of the customer and emotions in a sociological way [31].
- The results of the data-driven heuristic methodology for transportation and logistics related problems can shed valuable insight for different types of problems in different contexts.
- Data can be collected from integrated digital platforms to support joint decision-making activities [6].
- Robust optimization approach and data-driven techniques can be combined [18].

- "Responsible Artificial intelligence" is started to gain importance in the sustainability context. One of the main element is "explainability" to be responsible. Data-driven heuristics need to be designed as easily explained for stakeholders.

REFERENCES

[1] A. Akkas, Shelf space selection to control product expiration, *Prod. Oper. Manag.*, **28** (2019), 2184-2201.

[2] E. Avraham and T. Raviv, The data-driven time-dependent traveling salesperson problem, *Transp. Res. Part B*, **134** (2020), 25-40.

[3] V. Batagelj, *Networks/Pajek Program for Large Network Analysis*, 2018. Available from: `http://vlado.fmf.uni-lj.si/pub/networks/pajek/`.

[4] S. Belhaiza, R. M'Hallah, G. Ben Brahim and G. Laporte, Three multi-start data-driven evolutionary heuristics for the vehicle routing problem with multiple time windows, *J. Heuristics*, **25** (2019), 485-515.

[5] J. G. Carlsson and E. Delage, Robust partitioning for stochastic multivehicle routing, *Oper. Res.*, **61** (2013), 727-744.

[6] L. Chen, Fixing phantom stockouts: Optimal data-driven shelf inspection policies, *Prod. Oper. Manag.*, **30** (2021), 689-702.

[7] Q. Chen, B. G. De Soto and B. T. Adey, Supplier-contractor coordination approach to managing demand fluctuations of ready-mix concrete, *Autom. Constr.*, **121** (2021), 103423.

[8] H. Chu, W. Zhang, P. Bai and Y. Chen, Data-driven optimization for last-mile delivery, *Complex. Intell. Syste.*, (2021).

[9] H. H.-C. Chuang, R. Oliva and O. Perdikaki, Traffic-based labor planning in retail stores, *Prod. Oper. Manag.*, **25** (2016), 96-113.

[10] I. Cobanoglu, I. Gure and V. Bayram, Data driven storage location assignment problem considering order picking frequencies: A heuristic approach, *Pamukkale University J. Eng. Sci.*, **27** (2021), 520-531.

[11] A. Dolgui, F. Sgarbossa and M. Simonetto, Design and management of assembly systems 4.0: Systematic literature review and research agenda, *Int. J. Prod. Res.*, **60** (2021), 184-210.

[12] Y. Feng, Q. Zhu and K.-H. Lai, Corporate social responsibility for supply chain management: A literature review and bibliometric analysis, *J. Clean. Prod.*, **158** (2017), 296-307.

[13] V. M. S. Gandra, H. Calik, T. Wauters, T. A. M. Toffolo, M. A. M. Carvalho and G. Vanden Berghe, The impact of loading restrictions on the two-echelon location routing problem, *Comput. Ind. Eng.*, **160** (2021), 107609.

[14] C. Gkerekos and I. Lazakis, A novel, data-driven heuristic framework for vessel weather routing, *Ocean Eng.*, **197** (2020), 106887.

[15] F. Gong, D. S. Kung and T. Zeng, The impact of different contract structures on IT investment in logistics outsourcing, *Int. J. Prod. Eco.*, **195** (2018), 158-167.

[16] T. He, J. Bao, S. Ruan, R. Li, Y. Li, H. He and Y. Zheng, Interactive bike lane planning using sharing bikes' trajectories, *IEEE Trans. Knowl. Data. Eng.*, **32** (2020), 1529-1542.

[17] K. Huang, J. Wu, X. Yang, Z. Gao, F. Liu and Y. Zhu, Discrete train speed profile optimization for urban rail transit: A data-driven model and integrated algorithms based on machine learning, *J. Adv. Transp.*, **2019** (2019), 7258986.

[18] Y. Kou and Z. Wan, A new data-driven robust optimization approach to multi-item newsboy problems, *J. Ind. Manag. Optim.*, **19** (2023), 197-223.

[19] F. Lejarza, I. Pistikopoulos and M. Baldea, Scalable real-time solution strategy for supply chain management of fresh produce: A Mexico-to-United States cross border study, *Int. J. Prod. Econ.*, **240** (2021), 108212.

[20] R. Levi, G. Perakis and J. Uichancoi, The data-driven newsvendor problem: New bounds and insights, *Oper. Res.*, **63** (2015), 1294-1306.

[21] H. Li, L. Mai, W. Zhang and X. Tian, Optimizing the credit term decisions in supply chain finance, *J. Purch. Supply. Manag.*, **25** (2019), 146-156.

[22] H. Lin and C. Tang, Intelligent bus operation optimization by integrating cases and data driven based on business chain and enhanced quantum genetic algorithm, *Trans. Intell. Transp. Syst.*, **23** (2022), 9869-9882.

[23] L. Liu and S. Mei, Visualising the GVC research: A co-occurrence network based bibliometric analysis, *Scientometrics*, **109** (2016), 953-977.

[24] S. Liu, Z. J. M. Shen and X. Ji, Urban bike lane planning with bike trajectories: Models, algorithms, and a real-world case study, to appear, *Manuf. Serv. Oper. Manag.*
[25] M. Matusiak, R. de Koster and J. Saarinen, Utilizing individual picker skills to improve order batching in a warehouse, *Eur. J. Oper. Res.*, **263** (2017), 888-899.
[26] D. Merchan and M. Winkenbach, An empirical validation and data-driven extension of continuum approximation approaches for urban route distances, *Networks*, **73** (2019), 418-433
[27] M. Miotti, Z. A. Needell, S. Ramakrishnan, J. Heywood and J. E. Trancik, Quantifying the impact of driving style changes on light-duty vehicle fuel consumption, *Transp. Res. D. Transp. Environ.*, **98** (2021), 102918.
[28] S. M. Mirhedayatian, T. G. Crainic, M. Guajardo and S. W. Wallace, A two-echelon location-routing problem with synchronisation, *J. Oper. Res. Soc.*, **72** (2021), 145-160.
[29] W. de Nooy, A. Mrvar and V. Batagelj, *Explarotary Social Network Analysis with Pajek* , 2^{nd} edition, Cambridge University Press, New York, 2011.
[30] L. G. N. Orozco, F. Battiston, G. Iniguez and M. Szell, Data-driven strategies for optimal bicycle network growth, *R. Soc. Open. Sci.*, **7** (2020), 201130.
[31] E. Ozgormus and A. E. Smith, A data-driven approach to grocery store block layout, *Comput. Ind. Eng.*, **139** (2020), 105562.
[32] S. K. Paul, R. Sarker and D. Essam, Managing risk and disruption in production-inventory and supply chain systems: A review, *J. Ind. Manag. Optim.*, **12** (2016), 1009-1029.
[33] B. S. Perelman, A. W. Evans and K. E. Schaefer, Where do you think you're going? Characterizing spatial mental models from planned routes, *ACM Trans. Hum.*, **9** (2020), 1-55.
[34] H. N. Perera, B. Fahimnia and T. Tokar, Inventory and ordering decisions: A systematic review on research driven through behavioral experiments, *Int. J. Oper. Prod.*, **40** (2020), 997-1039.
[35] O. Persson, *Bib-Excel*, 2018. Available from: `http://homepage.univie.ac.at/juan.gorraiz/bibexcel/`.
[36] O. Persson, R. Danell and J. W. Schneider, How to use Bibexcel for various types of bibliometric analysis. In: Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at His 60th Birthday, International Society for Scientometrics and Informetrics, Sweden, 2009.
[37] G. Pesant, M. Gendreau, J. Y. Potvin and J. M. Rousseau, An exact constraint logic programming algorithm for the traveling salesman problem with time windows, *Transp. Sci.*, **32** (1998), 12-29.
[38] P. Pinakpani, A. Polisetty, G. B. N. Rao, D. H. Sunil, B. M. Kumar, D. Deepthi and A. Sidhireddy, An algorithmic approach for maritime transportation, *Int. J. Adv. Comput. Sci. Appl.*, **11** (2020), 764-775.
[39] S. Punia, S. P. Singh and J. K. Madaan, From predictive to prescriptive analytics: A data-driven multi-item newsvendor model, *Decis. Support Syst.*, **136** (2020), 113340.
[40] V. Ramamurthy, J. G. Shanthikumar and Z. J. M. Shen, Inventory policy with parametric demand: Operational statistics, linear correction, and regression, *Prod. Oper. Manag.*, **21** (2012), 291-308.
[41] S. Raman, N. Patwa, I. Niranjan, U. Ranjan, K. Moorthy and A. Mehta, Impact of big data on supply chain management, *Int. J. Logist. Res. Appl.*, **21** (2018), 579-596.
[42] A. I. Sivakumar and C. S. Chong, A simulation based analysis of cycle time distribution, and throughput in semiconductor backend manufacturing, *Comput. Ind.*, **45** (2001), 59-78.
[43] J. Smoczek and J. Szpytko, Evolutionary algorithm-based design of a fuzzy TBF predictive model and TSK fuzzy anti-sway crane control system, *Eng. Appl. Artif. Intell.*, **28** (2014), 190-200.
[44] C. Wang, C. Li, H. Huang, J. Qiu, J. Qu and L. Yin, ASNN-FRR: A traffic-aware neural network for fastest route recommendation, to appear, *Geoinformatica*, (2021).
[45] Q. Wang, X. Yang, Z. Huang and Y. Yuan, Multi-vehicle trajectory design during cooperative adaptive cruise control platoon formation, *Transp. Res. Rec.*, **2674** (2020), 30-41.
[46] K. J. Wei, V. Vaze and A. Jacquillat, Airline timetable development and fleet assignment incorporating passenger choice, *Transp. Sci.*, **54** (2020), 139-163.
[47] C. Wen, P. Huang, Z. Li, J. Lessan, L. Fu, C. Jiang and X. Xu, Train dispatching management with data- driven approaches: A comprehensive review and appraisal, *IEEE Access*, **7** (2019), 114547-114571.
[48] S. Winkelhaus and E. H. Grosse, Logistics 4.0: A systematic review towards a new logistics system, *Int. J. Prod. Res.*, **58** (2020), 18-43.

[49] T. Wu, F. Xiao, C. Zhang, D. Zhang and Z. Liang, Regression and extrapolation guided optimization for production-distribution with ship-buy-exchange options, *Transport. Res. E Log.*, **129** (2019), 15-37.
[50] J. Zhang, S. Onal and S. Das, Price differentiated channel switching in a fixed period fast fashion supply chain, *Int. J. Prod. Econ.*, **193** (2017), 31-39
[51] Y. Zhang, Z. Zhang, A. Lim and M. Sim, Robust Data-driven vehicle routing with time windows, *Oper. Res.*, **69** (2021), 469-485.
[52] Q. Zhao, C. Zhou and G. Pedrielli, Decision support system for data-driven driver-experience augmented vehicle routing problem, *Asia Pac. J. Oper. Res.*, **37** (2020), 2050018.
[53] S. Zhong, Y. Geng, W. Liu and W. Chen, A bibliometric review on natural resource accounting during 1995-2014, *J. Clean. Prod.*, **139** (2016), 122-132.
[54] C. Zhou, A. Stephen, X. Cao and S. Wang, A data-driven business intelligence system for large-scale semi-automated logistics facilities, *Int. J. Prod. Res.*, **59** (2021), 2250-2268.
[55] E. Zunic, D. Donko and E. Buza, An adaptive data-driven approach to solve real-world vehicle routing problems in logistics, *Complex.*, **2020** (2020), 7386701.

Received June 2022; revised September 2022; early access September 2022.