ISBN: 978-981-18-3959-7
2022 the 12th International Workshop on Computer Science and Engineering (WCSE 2022)
doi: 10.18178/wcse.2022.06.019

Research Landscape, Emerging Trends and New Developments in Data Warehouse: A Scientometric Analysis (1985 - 2021)

Changjun Fan [1], Li Zeng [2+], Kuihua Huang [1]

[1] College of Systems Engineering, National University of Defense Technology, Changsha, 410073 China

[2] School of International Business and Management, Sichuan International Studies University, Chongqing, 400000 China

Abstract. As the core component of bussiness intelligence, Data Warehouse (DW) has been studied by academia and industry for many years. In this paper, a scientometric review of DW over the past 30 years was conducted, with the aim to capture the landscapes, research hotspots and emerging trends in this field. The dataset was gathered from Web of Science between 1985 to 2021. Besides basic scientific outputs assessment based on statistical analysis and comparative analysis, scientometric softwares such as Citespace, VOSViewer and Bibliometrix were used to analysis the knowldege structure of Data Warehouse. Results showed that Data Warehouse research went up significantly in the past two decades, including a total of 2529 articles covering 93 countries/territories, and the top five most productive countries are USA, China, France, India, Germany, and Spain. There are 2028 research institutes involved in the field of DW and the top five most influential institutes are University of Alicante, University of Virginia, University of Georgia, University of Bologna and Stanford University. Besides, Keywords with strongest citation burst such as Spatial Data Warehouse, Data Mart, Etl Processe, Multidimensional Model, Business Intelligence, Cloud Computing, Dimensional Model, Big Data Warehouse, Nosql Database, Clinical Data Warehouse, Olap and Information Extraction, demonstrate the emerging trends of Data Warehouse. The results shown in this paper are expected to facilitate the research of Data Warehouse.

Keywords: scientometric, bibliometric, citespace, Data Warehouse.

1. Introduction

A Data Warehouse (DW) is a subject-oriented, integrated, time-invariant, non-updatable collection of data used to support management decision-making processes and business intelligence [1]. DW is widely used in Computer Science [2], Management [3], Business [4], Health [5], Eductaion [6] and so on.

Recent years, comprehensive reviews of the research related to Data Warehouse were conducted. Holmes JH et al. (2014) [7] reviewed the published, peer-reviewed literature on clinical research data warehouse governance in distributed research networks. Ramamurthy K et al. (2008) [8] examined the key organizational and innovation factors that influence the infusion (diffusion) of Data Warehouse within organizations and also examine if more extensive infusion leads to improved organizational outcomes. Wrembel R (2010) [9] surveyed challenges in designing, building, and managing data warehouses whose structure and content evolve in time. Triplet T et al. (2013) [10] provided a comprehensive and quantitative review of those genomic data warehousing frameworks in the context of large-scale systems biology. Sen A et al. (2007) [11] reviewed 30 commercial data warehousing methodologies and analyze the standard practices they have adopted with respect to data warehousing process. Moalla I et al. (2017) [12] provided a literature review on data warehouse design approaches from social media. Bogojevic P (2020) [13] conducted a systematic review about project management in data warehouse implementations with the aim to remove these gaps by conducting a systematic review of the literature.

Unlike those traditional methods, this paper gives a scientometric review of Data Warehouse research by investigating the scientific outputs, geographical distribution and international cooperation, distribution of institutions and journals with the aim to offer another perspective on the development of research in the field

+ Corresponding author. Tel.: + 18538882527.
E-mail address: crack521@163.com.

of Data Warehouse. Moreover, innovative methods such as co-citation analysis and burst detection were applied, which can vividly depict the landscape and trends from various aspects.

The sturcture of this paper are as follows. the data collection strategy and research methods are shown in Section 2. And Section 3 contains the results and discussion. At last, the conclusion is given in Section 4.

2. Data and Method

2.1. Data Collection

The data used for this paper were gathered from Web of Science (WoS) database on May 12, 2022, and the search strategy is as follows:

Title = “Data Warehous*”

Timespan = 1985-2021

Databases = SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

At last, a total of 2529 bibliographic records were downloaded for subsequent analysis.

2.2. Method

The primary objective of this study conducted a comprehensive scientometic review of Data Warehouse with the aim to uncover the landscape, research hotspots and emerging trends in this field. After the ETL operation (cleaning, conversion, deduplication) on the raw data, a basic analysis with regard to highly productive countries/territories and institutes, highly cited references and highly cited authors was conducted by Microsoft Excel. H-Index and other metrics were calculated by a self-developed scientometric software called Sciradar [14], geographic distribution of scholars was mapped by Bibliometrix [15] according to author affiliations, network analysis of different type entities such as countries/territories, institutes, categories and keywords was conducted by the scientometric software CiteSpace [16] and VOSViewer [17] with the aim to identify the intellectual structure, hotspots and emerging trends of the DW research. Besides, burst detection of keywords was conducted by the algorithm proposed by Kleinberg [18].

3. Result and Discussion

3.1. Scientific Outputs of Data Warehouse Research

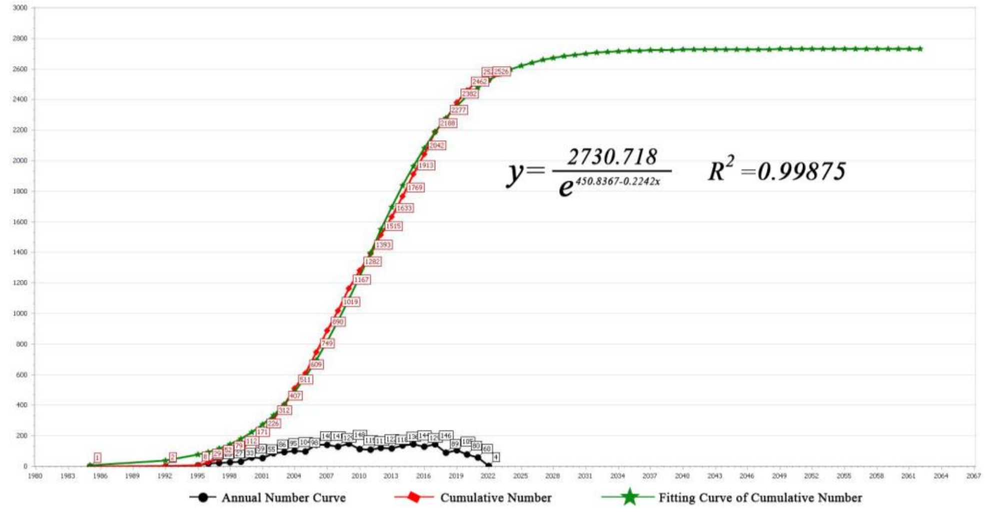


Fig. 1: Paper number of Data Warehouse.

Figure 1 displays the number of papers and maturity forecasting curve between 1985 and 2021. The black curve represents the annual paper number. From the curve, we found that a substantial interest in Data Warehouse research did not emerge until 1995, although a few articles were published previously. The year with the highest number was 2009, when 148 papers were published, accounting for 5.86% of the total number. The annual average number of papers was 68.3 per year. The red curve is the cumulative number of papers. According to the theory of technology maturity, the cumulative number of paper could be fitted by the Logistic Growth Model [19]. The least squares were used to get the parameters in the equation, where the green curve is the result which is described by formula (1).

$$y = 2730.718 / (1 + exp450.8367-0.2242x) \quad (1)$$

Here, x and y represent year and paper number respectively. According to formula (1), the development of Data Warehouse can be divided into four stages: infant period (before 2001), growth period (2002-2021), mature period (2022-2028) and stable period (after 2028). According to the above stage division, the research of Data Warehouse in 2021 was in the mature period with a maturity of 92.51%.

3.2. International Collaborations

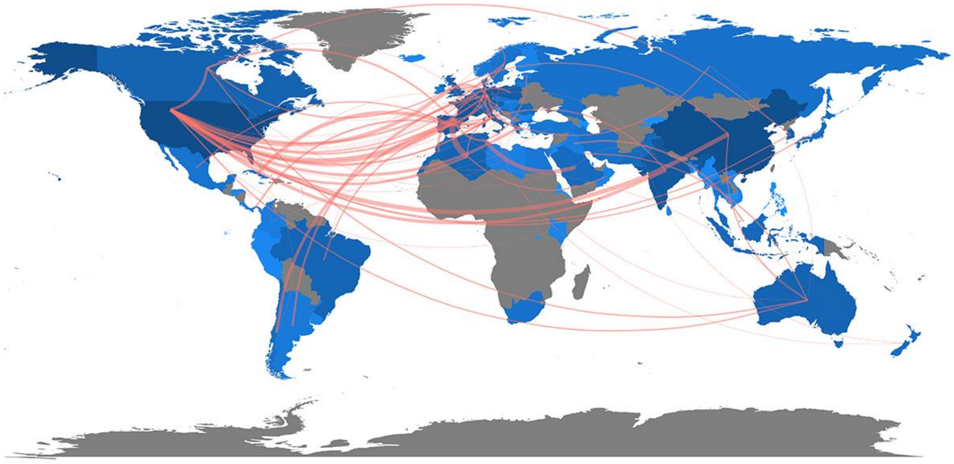


Fig. 2: Country/territories collaboration map in the filed of DW.

In order to vividly demonstrate the collaboration between countries/territories, a collaboration map was generated by the Bibliometrix [15] (Figure 2). The color of the place in the collaboration map represents the number of paper. In total, there are 93 countries/territories in the field of Data Warehouse. As can be seen, the major contribution of the total output mainly came from three countries, namely, USA, China, and France. "Burst Detection Algorithm" in CiteSpace was used to detect the surge in research interest within DW research, and top 3 countries with high frequent bursts are: USA (52.14), China (30.51) and France (12.75), suggesting that they have abrupt increases of interest in the research of Data Warehouse. Betweenness Centrality metrics provide a computational method for finding pivotal points between different specialties or tipping points in an evolving network [14]. Thus, high betweenness centrality nodes such as USA, Italy, Portugal and Poland indicates that these countries play an important role in this research filed.

Table 1: Top ten countries/ territories in DW.

No.	C/T	TP	IP	CP	TC	HI	TI	BC
1	USA	469	371	98	6475	40	132	0.62
2	CHINA	348	326	22	673	10	85	0.02
3	France	190	126	64	1093	16	142	0.09
4	India	149	132	17	384	9	79	0.00
5	Germany	122	86	36	1016	18	66	0.02
6	Spain	120	67	53	1761	23	148	0.16
7	Poland	84	74	10	278	9	80	0.18
8	Italy	92	62	30	1123	14	68	0.20
9	Portugal	57	51	6	302	9	72	0.21
10	South Korea	50	43	7	195	8	62	0.02

No., Rank By TP; C/T, Country/Territory; TP, Total papers; IP, independent papers; CP, internationally collaborated articles; TC, Total citations counts; HI, H Index; TI, Total Institutes numbers; BC, Betweenness centrality in the Cooperation Networks.

Table 1 lists the top ten most productive countries/territories in the field of Data Warehouse. Overall, USA is the first most productive, and first most influential country in this field, with a total number of 469 papers (371 independent papers, 98 internationally collaborated papers), 132 institutes and 6475 citations, and its H-Index is 30. China is the second most productive, but the sixth most influential country in this field, with a total number of 348 papers (326 independent papers, 22 internationally collaborated papers), 85 institutes and 673 citations, and its H-Index is 10. France is the third most productive and the fourth most influential country in this filed, with 190 papers (126 independent papers, 64 internationally collaborated papers), 142 institutes and 1093 citations and its H-Index is 16. Other countries/ territories such as India, Germany, Spain also make outstanding contributions in this field.

3.3. Institute Collaborations

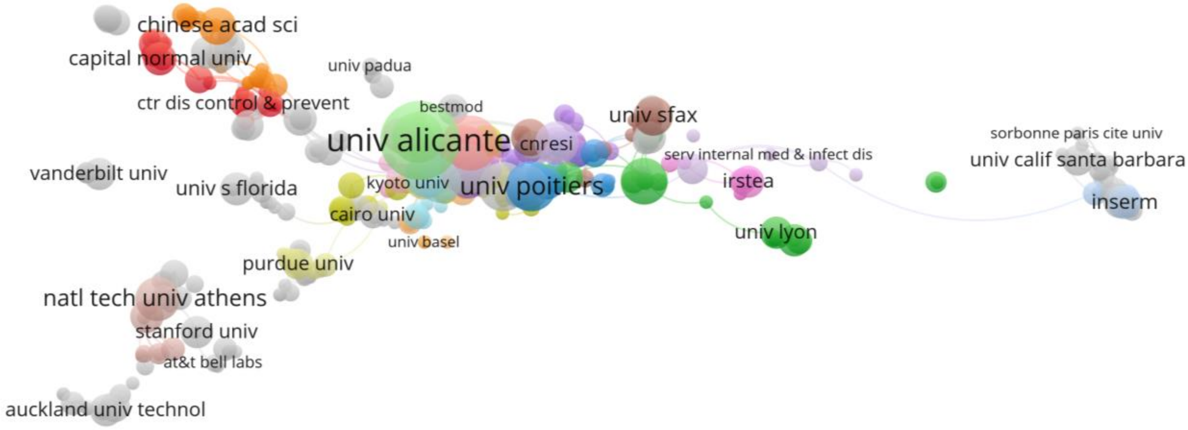


Fig. 3: Institute collaboration map in the filed of DW.

Figure 3 shows the institutes collaboration network of Data Warehouse. A total of 2028 institutes engaged in Data Warehouse research during the period 1985 to 2021. In order to show the core institutions of this field, we filter out the institutions with a small number of publications and get an institute co-occurring network with 546 nodes and 1261 links. Univ Alicante takes the first place with a 75 papers, 1346 citations. Silesian Tech Univ takes the second place with 31 papers, 126 citations. Univ Castilla La Mancha takes the third place with 25 papers, 371 citations. Other institutes such as Univ Coimbra, Univ Poitiers and Poznan Univ Tech also make great contributions to the research of Data Warehouse and the details are listed in Table 2.

Table 2: Top ten institutes in DW

No.	Name	Frequent	Citation Burst	Citations	Year
1	Univ Alicante	75	17.07	1346	2004
2	Silesian Tech Univ	31	9.39	126	2005
3	Univ Castilla La Mancha	25	8.79	371	2004
4	Univ Coimbra	25	3.85	93	2004
5	Univ Poitiers	24	5.68	174	2006
6	Poznan Univ Tech	23	5.10	63	2007
7	Univ Lyon 2	20	0	171	2005
8	Univ Minho	19	6.59	185	2015
9	Vienna Univ Technol	19	5.58	85	2000
10	Natl Tech Univ Athens	18	7.70	360	1998

3.4. Journal Distribution and Journal Co-citation Network Analysis

(A) Journal Cocitaion Network (Journal View)   (B) Journal Cocitaion Network (Clustering View)

Fig. 4: Journal cocitaion network of Data Warehouse.

Figure 4 shows the Journal co-cited network of Data Warehouse which was generated by Citespace. Node in the network represents a cited-journal and there exists a link between two cited-journals if they are cited by same paper. G-Index [20] was used to prune the whole network with the aim to display the core Journals and louvain algorithm [21] was used to extract the community structure of the network. Then a pruned network with 957 nodes, 1897 links and 26 communities were generated, and the metric of the network such as density, sihouette, modularity are calculated which are detailed in Table 3.

Table 3: Details of the journal co-cited network

Metric Name	Value
Node Number	957
Node number of Largest Component	947
Link Number	1897
Network Density	0.0041
Community Number	25
Weighted mean sihouette	0.9183
Modularity of community division	0.82
Top 5 Journal ranked by citations	LECT NOTES COMPUT SC, BUILDING DATA WAREHO, DATA WAREHOUSE TOOLK, SIGMOD Record, COMMUN ACM
Top 5 communites ranked by member size	#0 clinical data warehouse, #1 information quality, #2 schema versioning, #3 middleware, #4 mapreduce

Figure 4(A) and 5(B) show the node view of the network whose label represents cited journal and clustering view of the network whose label represents the community's name respectively. From Figure 4(A) and Table 3, we can find that the top ranked cited-journal by citation counts is LECT NOTES COMPUT SC (1998) in Cluster #8 with 672 citations. The second one is BUILDING DATA WAREHO (1997) in Cluster #16 with 533 citations. The third is DATA WAREHOUSE TOOLK (1997) in Cluster #16 with 444 citations. The 4th is SIGMOD Record (1999) in Cluster #17 with 329 citations. The 5th is COMMUN ACM (1997) in Cluster #8 with 312 citations. From Figures 4(B) and Table 3, we can find that the top 5 communites ranked by member size are Cluster #0 with 71 members whose label is clinical data warehouse, silhouette is 0.972 and mean year is 2015, Cluster #1 with 62 members whose label is information quality, silhouette is 0.884 and mean year is 2006, Cluster #2 with 61 members whose label is schema versioning, silhouette is 0.924 and mean year is 2007, Cluster #3 with 53 members whose label is middleware, silhouette is 0.757 and mean year is 2009, Cluster #4 with 51 members whose label is mapreduce, silhouette is 0.871 and mean year is 2009.

3.5. Character of Subject Categories

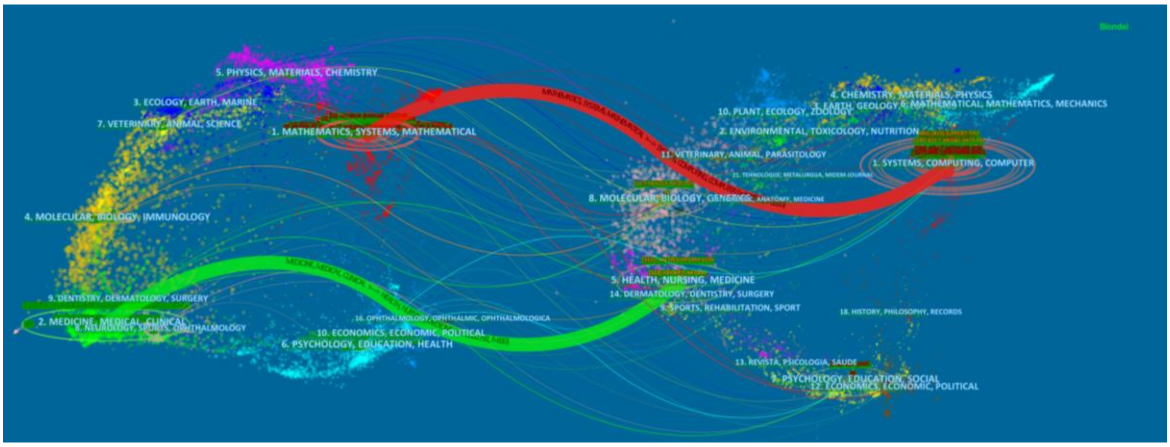


Fig. 5: Dual-map overlay of Data Warehouse (1985-2021). Wavelike curves portray the citation links which are colored by their source clusters.

The total of 2529 papers of Data Warehouse covered 89 subject categories and the top five categories were Computer Science (1913, 75.64%), Engineering (568, 22.46%), Business Economics (130, 5.14%), Telecommunications (127, 5.02%), Medical Informatics (126, 4.98%). In order to do portfolio analysis of subject category of Data Warehouse, the dual-map overlay technology was used and the result was exhibited in Figure 5. A dual-map overlay [22] of the science mapping represents the whole dataset in a global map of science created from over 10,000 journals indexed in the Web of Science Database and the Citation links in dual-map overlay are bundled by z-scores function in Citespace. In total there are 3107 Citation links in Figure 5. The left part are the source journals, while the right part are the target journals. The two major clusters of source journals are journals in medicine-medical-clinical (green), mathematics-systems-mathematical (red). We can see that the two major clusters in source journals are cited by the journals in health-nursing-medicine (830 links) and the journals in systems-computing-computer (890 links) which represents the movement of knowledge flow in the field of Data Warehouse.

3.6. Research Hotspots and Emerging Trends of Data Warehouse

Table 4: Top 10 highly cited papers of Data Warehouse

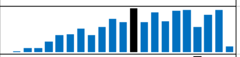
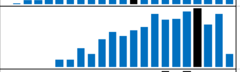
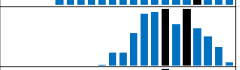
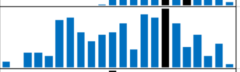
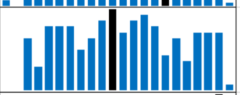
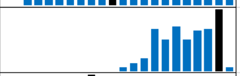
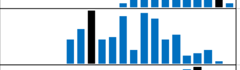
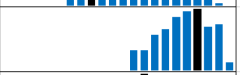
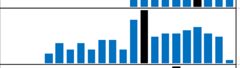
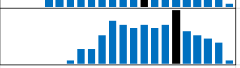
ID	Title	Frist Author	Publication Name	Year	DOI	Citation Count	Average Citation Count	Citation Distribution
1	An empirical investigation of the factors affecting data warehousing success	Wixom, BH	MIS QUARTERLY	2001	10.2307/3250957	643	29.23	
2	Antecedents of information and system quality: An empirical examination within the context of data warehousing	Nelson, RR	JOURNAL OF MANAGEMENT INFORMATION SYSTEMS	2005	10.1080/07421222.2005.11045823	401	22.28	
3	Hive - A Petabyte Scale Data Warehouse Using Hadoop	Thusoo, Ashish	26TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING ICDE 2010	2010	10.1109/ICDE.2010.5447738	352	27.08	
4	The dimensional fact model: A conceptual model for data warehouses	Golfarelli, M	INTERNATIONAL JOURNAL OF COOPERATIVE INFORMATION SYSTEMS	1998	10.1142/S0218843098000118	215	8.6	
5	Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing	Nemati, HR	DECISION SUPPORT SYSTEMS	2002	10.1016/S0167-9236(01)00141-5	185	8.81	
6	YeastMine-an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose toolkit	Balakrishnan, Rama	DATABASE-THE JOURNAL OF BIOLOGICAL DATABASES AND CURATION	2012	10.1093/database/bar062	153	13.91	
7	A UML profile for multidimensional modeling in data warehouses	Lujan-Mora, Sergio	DATA & KNOWLEDGE ENGINEERING	2006	10.1016/j.datak.2005.11.004	153	9	
8	InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data	Smith, Richard N	BIOINFORMATICS	2012	10.1093/bioinformatics/bts577	152	13.82	
9	Ligand Depot: a data warehouse for ligands bound to macromolecules	Feng, ZK; Chen, L	BIOINFORMATICS	2004	10.1093/bioinformatics/bth214	135	7.11	
10	Integrated decision support systems: A data warehousing perspective	March, Salvatore T.	DECISION SUPPORT SYSTEMS	2007	10.1016/j.dss.2005.05.029	128	8	

Table 4 displays the top 10 highly cited papers which represent the research hotspots of Data Warehouse and these papers' citation count, average citation count, citation distribution are listed in the table meanwhile. Wixom BH et al. (2001) [23] conducted a survey which involved data warehousing managers and data suppliers from 111 organizations by means of paired mail questionnaires on implementation factors and the success of the warehouse. Nelson RR et al. (2005) [24] done an empirical examination within the context of data warehousing. Thusoo A et al. (2010) [25] presented Hive, an open-source data warehousing solution built on top of Hadoop. Golfarelli M et al. (1998) [26] formalize a graphical conceptual model for data warehouses, called Dimensional Fact model, and propose a semi-automated methodology to build it from the pre-existing (conceptual or logical) schemes describing the enterprise relational database. Nemati HR et al. (2002) [27] proposed a knowledge warehouse (KW) architecture that will not only facilitate the capturing and coding of knowledge but also enhance the retrieval and sharing of knowledge across the organization to extend the data warehouse model. Balakrishnan R et al. (2012) [28] constructed a multifaceted search and retrieval environment called YeastMine which can provide access to diverse data types. Luján-Mora S et al. (2006) [29] present an extension of the Unified Modeling Language (UML) using a UML profile for multidimensional modelling in data warehouses. Smith RN et al. (2012) [30] developed an open-source data warehouse system InterMine which can facilitate the building of databases with complex data integration requirements and a need for a fast customizable query facility. Feng Z et al. (2004) [31] developed an integrated data resource called Ligand Depot for finding information about small molecules bound to proteins and nucleic acids. March ST et al. (2007) [32] provided both researchers and practitioners a clear view of the challenges and opportunities of applying data warehousing technology to support all levels of management decision-making.

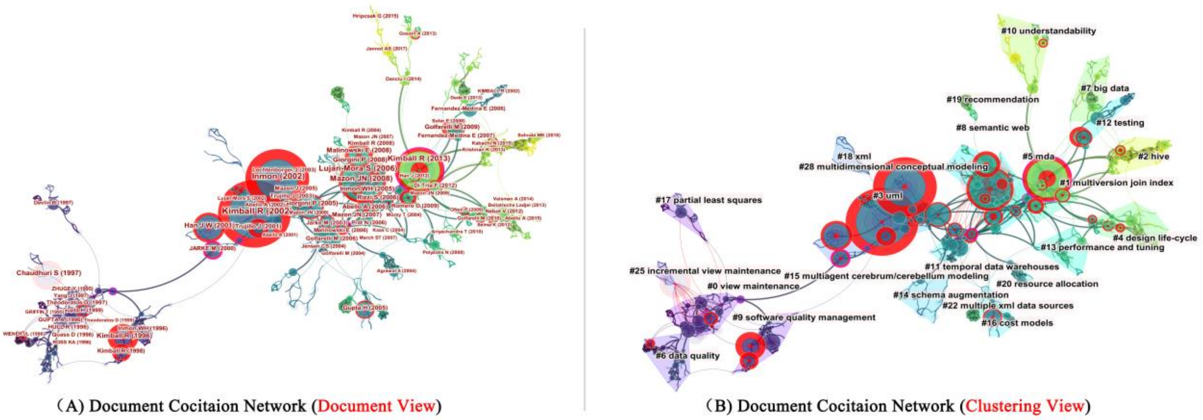


Fig.6: Document cocitation network of Data Warehouse.

In order to intensely distinguish the research frontier of Data Warehouse, document cocitation analysis was conducted. Document co-citation analysis scrutinizes the associations between papers citing the same references which was usually used to comprehend the intellectual infrastructure of the knowledge domain concerning periodic renovations [33].

In order to show the core references in the document co-citation network, G-Index [20] was used to prune the whole network and louvain algorithm [21] was used to extract the community structure of the network. Figure 6(A) and 6(B) show the node view of the network whose label represents cited reference and clustering view of the network whose label represents the community name respectively.

The pruned network consists of 1096 cited references and 2053 co-citation links. In total, there are 29 co-citation clusters in the network and the Modularity of community division is 0.9202. Table 5 lists the details of the top 5 largest clusters. The oldest and the largest cluster is cluster #0 with an average year 1996 and its labels are view maintenance, materialized views, view usability, data warehousing and view selection, and its representive papers are [34] [35] [36]. cluster #1 is the second largest one and its label are multiversion join index, multiversion data warehouse, requirements engineering, solap and join index, and its representive papers are [37] [38] [39]. cluster #2 is the third largest one and its label are hive, big data warehouse, agility, experimental evaluation, and geospatial data warehouse, and its representive papers are [40] [41] [42].

cluster #3 is the fourth largest one and its label are uml, multidimensional modelling, data warehouse metrics, unified process and uml extension, and its representive papers are [1] [37] [43]. cluster #4 is the fifth largest one and its label are design life-cycle, ontology-based approach, data warehousing, traceability and bitmap index, and its representive papers are [44] [45] [46].

Table 5: Top 5 largest clusters

#	Size	Silhouet	Mean Year	Labels	Representive Papers
0	61	0.897	1996	view maintenance; materialized views; view usability; data warehousing; view selection	[34] [35] [36]
1	53	0.904	2009	multiversion join index; multiversion data warehouse; requirements engineering; solap; join index	[37] [38] [39]
2	50	0.967	2015	hive; big data warehouse; agility; experimental evaluation; geospatial data warehouse	[40] [41] [42]
3	49	0.936	2002	uml; multidimensional modeling; data warehouse metrics; unified process; uml extension	[1] [37] [43]
4	48	0.955	2012	design life-cycle; ontology-based approach; data warehousing; traceability; bitmap index	[44] [45] [46]

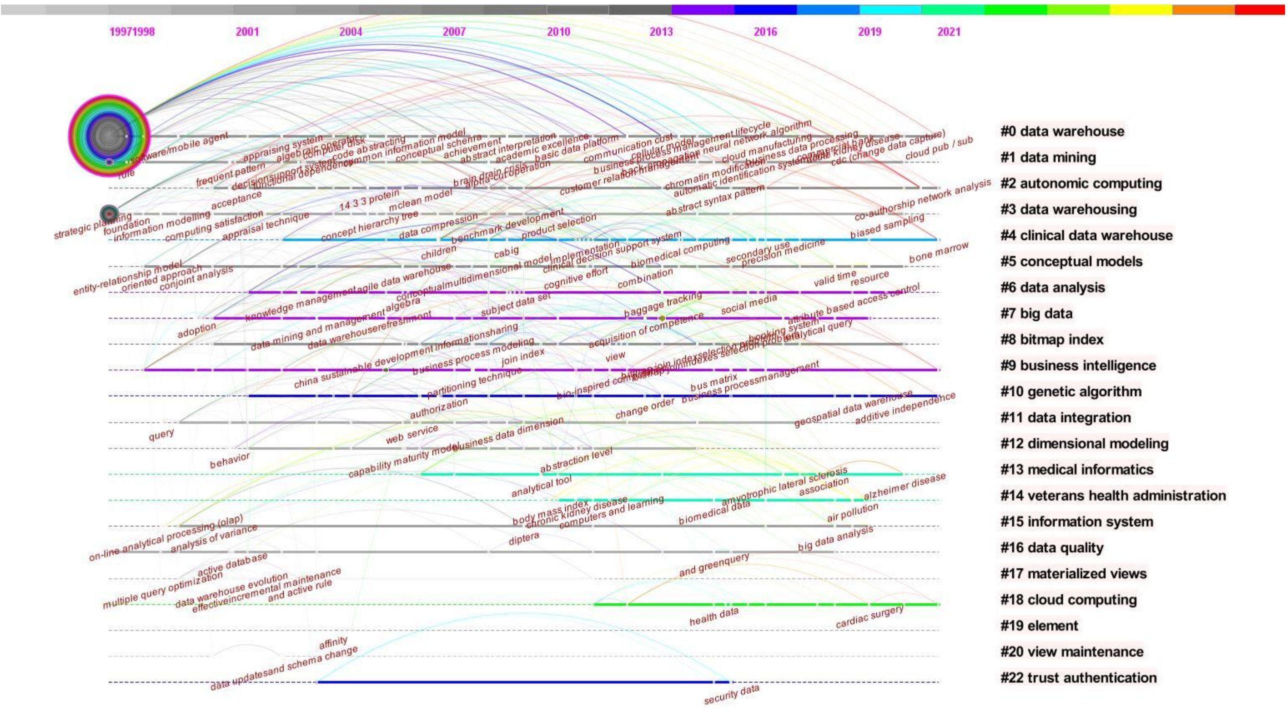


Fig. 7: Timeline graph of the keyword cooccurence network of Data Warehouse.

In order to find the research landscape about Data Warehouse in detail, the keyword occurrence network was generated by Citespace and the result was shown as a timeline graph. In total, there are 805 nodes, 1160 links and 23 communities in the network.

Figure 7 shows the temporal graph of burst keywords detected by CiteSpace, which can be seen as the micro research front of Data Warehouse research. According to the order of this emergence of the research front, Keywords such as Spatial Data Warehouse, Data Mart, Etl Processe, Multidimensional Model, Business Intelligence, Cloud Computing, Dimensional Model, Big Data Warehouse, Nosql Database, Clinical Data Warehouse, Olap and Information Extraction can be used to denote the emerging trends of Data Warehouse.

4. Conclusion

This paper conducts a quantitative assessement the landscape, research hotspots and emerging trends of Data Warehouse a broad assessment of publication data in the Knowledge Management domain based on a comprehensive scientometric analysis using the related literatures from the Web of Science database between 1985 and 2021. Analysis about Data Warehouse were concentrated on scientific outputs, geographic distribution, institutions, journals, and subject categories. Moreover, innovative methods such as co-citation analysis and burst detection were applied, the conclusion this paper can brilliantly uncover the research landscape and emerging trend of Data Warehouse from various perspectives.

5. References

[1] Inmon WH. Building the Data Warehouse. *John Wiley & Sons*; 2005.

[2] Nelson RR, Todd PA, Wixom BH. Antecedents of Information and System Quality: An Empirical Examination Within the Context of Data Warehousing. *Journal of Management Information Systems*. 2005, 21(4):199-235.

[3] Cooper BL, Watson HJ, Wixom BH, Goodhue DL. Data warehousing supports corporate strategy at First American Corporation. *MIS Quarterly*. 2000, 24(4):547.

[4] Sen A, Sinha AP, Ramamurthy K. Data Warehousing Process Maturity: An Exploratory Study of Factors Influencing User Perceptions. *IEEE Transactions on Engineering Management*. 2006, 53(3):440-455.

[5] Gorina Y, Kramarow EA. Identifying Chronic Conditions in Medicare Claims Data: Evaluating the Chronic Condition Data Warehouse Algorithm. *Health Services Research*. 2011, 46(5):1610-1627.

[6] John Thomas L. Data Warehousing and Data Mining Techniques to Improve Student Retention, within International Universities in the U.A.E. *SSRN Electronic Journal*. Published online 2017.

[7] Holmes JH, Elliott TE, Brown JS, et al. Clinical research data warehouse governance for distributed research networks in the USA: a systematic review of the literature. *Journal of the American Medical Informatics Association*. 2014, 21(4):730-736.

[8] Ramamurthy K, Sen A, Sinha AP. Data Warehousing Infusion and Organizational Effectiveness. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*. 2008, 38(4):976-994.

[9] Wrembel R. A Survey of Managing the Evolution of Data Warehouses. In: *Business Information Systems. IGI Global*; 2010:894-928.

[10] Triplet T, Butler G. A review of genomic data warehousing systems. *Briefings in Bioinformatics*. 2013, 15(4):471-483.

[11] Sen A, Sinha AP. Toward Developing Data Warehousing Process Standards: An Ontology-Based Review of Existing Methodologies. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*. 2007, 37(1):17-31.

[12] Moalla I, Nabli A, Bouzguenda L, Hammami M. Data warehouse design approaches from social media: review and comparison. Social Network Analysis and Mining. 2017, 7(1). Bogojevic P. Project Management in Data Warehouse Implementations: A Literature Review. *IEEE Access*. 2020, 8:225902-225934.

[13] Bogojevic P. Project Management in Data Warehouse Implementations: A Literature Review. *IEEE Access*. 2020, 8:225902-225934.

[14] Li Z, Li Z, Zhao Z, et al. Landscapes and Emerging Trends of Virtual Reality in Recent 30 Years: A Bibliometric Analysis[C]// 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, 2018.

[15] Aria M, Cuccurullo C. bibliometric: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*. 2017, 11(4):959-975.

[16] Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*. 2006, 57(3):359-377.

[17] van Eck NJ, Waltman L. Vosviewer: A Computer Program for Bibliometric Mapping. In: Advances in Data

Analysis: *Proceedings of the 30th Annual Conference of the German Classification Society*. Springer; 2007:299-306.

[18] Kleinberg J. Bursty and hierarchical structure in streams. In: *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining- KDD '02. ACM Press*; 2002.

[19] Rogosa D, Brandt D, Zimowski M. A growth curve approach to the measurement of change. *Psychological Bulletin*. 1982, 92(3):726-748.

[20] Egghe L. Theory and practise of the g-index. *Scientometrics*. 2006, 69:131-152.

[21] Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008;2008(10): P10008.

[22] Chen C, Leydesdorff L. Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis. *Journal of the Association for Information Science and Technology*. 2013,65(2):334-351.

[23] Wixom BH, Watson HJ. An Empirical Investigation of the Factors Affecting Data Warehousing Success. *MIS Quarterly*. 2001,25(1):17. *Wixom BH* et al. (2001) [23] *Nelson RR* et al. (2005) [24] *Nelson RR* et al. (2005) [24]

[24] Nelson RR, Todd PA, Wixom BH. Antecedents of Information and System Quality: An Empirical Examination Within the Context of Data Warehousing. *Journal of Management Information Systems*. 2005,21(4):199-235.

[25] Thusoo A, Sarma JS, Jain N, et al. Hive - a petabyte scale data warehouse using Hadoop. In: *2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)*. IEEE; 2010.

[26] Golfarelli M, Maio D, Rizzi S. THE DIMENSIONAL FACT MODEL: A CONCEPTUAL MODEL FOR DATA WAREHOUSES. *International Journal of Cooperative Information Systems*. 1998;07(02n03):215-247.

[27] Nemati HR, Steiger DM, Iyer LS, Herschel RT. Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence, and data warehousing. *Decision Support Systems*. 2002,33(2):143-161.

[28] Balakrishnan R, Park J, Karra K, et al. YeastMine—an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. *Database*. 2012.

[29] Luján-Mora S, Trujillo J, Song I-Y. A UML profile for multidimensional modeling in data warehouses. *Data & Knowledge Engineering*. 2006,59(3):725-769.

[30] Smith RN, Aleksic J, Butano D, et al. InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics*. 2012,28(23):3163-3165.

[31] Feng Z, Chen L, Maddula H, et al. Ligand Depot: a data warehouse for ligands bound to macromolecules. *Bioinformatics*. 2004,20(13):2153-2155.

[32] March ST, Hevner AR. Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*. 2007,43(3):1031-1043.

[33] Small H. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*. 1973,24(4):265-269.

[34] Theodoratos D, Sellis TK. Data Warehouse Configuration. In: *Proceedings of the 23rd International Conference on Very Large Data Bases. Morgan Kaufmann Publishers Inc*.; 1997:126-135.

[35] Hull R, Zhou G. A framework for supporting data integration using the materialized and virtual approaches. *ACM SIGMOD Record*. 1996,25(2):481-492.

[36] H. Gupta, I. S. Mumick. Selection of views to materialize in a data warehouse. *IEEE Transactions on Knowledge and Data Engineering*. 17(1):24-43.

[37] Kimball R, Ross M. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. John Wiley & Sons, 2011.

[38] Di Tria F, Lefons E, Tangorra F. Hybrid methodology for data warehouse conceptual design by UML schemas. *Information and Software Technology*. 2012,54(4):360-379.

[39] Mazón J-N, Pardillo J, Trujillo J. Applying Transformations to Model Driven Data Warehouses. In: Data Warehousing and Knowledge Discovery. *Springer Berlin Heidelberg*; 2006:13-22.

[40] Krishnan K. Data Warehousing in the Age of Big Data. *Newnes*; 2013.

[41] Arres B, Kabachi N, Boussaïd O, Bentayeb F. A Data Mining-based Blocks Placement Optimization for Distributed Data Warehouses. *Proceedings of the 19th International Database Engineering & Applications Symposium*. Published online 2015.

[42] Sohrabi MK, Ghods V. Understanding the Impact of the Interconnection Network Performance of Multi-core Cluster Architectures. *Journal of Computers*. 2016,11(2):132-139.

[43] Giorgini P, Rizzi S, Garzetti M. Goal-oriented requirement analysis for data warehouse design. In: *Proceedings of the 8th ACM International Workshop on Data Warehousing and OLAP - DOLAP*. *ACM Press*; 2005.

[44] Nebot V, Berlanga R. Building data warehouses with semantic web data. *Decision Support Systems*. 2012,52(4):853-868.

[45] Abello A, Romero O, Pedersen TB, et al. Using Semantic Web Technologies for Exploratory OLAP: A Survey. *IEEE Transactions on Knowledge and Data Engineering*. 2015,27(2):571-588.

[46] Selma K, Ilyès B, Ladjel B, Eric S, Stéphane J, Michael B. Ontology-based structured web data warehouses for sustainable interoperability: requirement modeling, design methodology and tool. *Computers in Industry*. 2012,63(8):799-812.