



BIBLIOMETRIC NETWORK TO IDENTIFY THE INTELLECTUAL STRUCTURE AND EVOLUTION OF THE BIG DATA RESEARCH FIELD

THE 19th INTERNATIONAL CONFERENCE ON INTELLIGENT DATA ENGINEERING
AND AUTOMATED LEARNING

J. R. LÓPEZ-ROBLES,, J. R. OTEGI-OLASO, I PORTO GOMEZ, N. K. GAMBOA-ROSALES, H. GAMBOA-ROSALES and H. ROBLES-BERUMEN

November 21-23, 2018 / Madrid (Spain)



TABLE OF CONTENTS

1. Introduction
2. Methodology
3. Dataset
4. Conceptual Analysis
5. Conclusions

TABLE OF CONTENTS

1. Introduction

2. Methodology

3. Dataset

4. Conceptual Analysis

5. Conclusions

1. INTRODUCTION

CONTEXT

Big Data has evolved from being an **emerging topic** to a growing research area in business, science and education fields.

The Big Data concept has a **multidimensional approach**, and it can be defined as a term describing the storage and analysis of large and complex data sets using a series of advanced techniques.

The professionals involved in this area of knowledge are seeking to uncover the **conceptual structure** of a research area of interest are worth and necessary.

OBJECTIVE

The main aim of this contribution is to develop a **bibliometric analysis to evaluate the performance and conceptual evolution** of the **Big Data** from **2012 to 2017**.

The analysis is developed using **SciMAT**.

TABLE OF CONTENTS

1. Introduction

2. Methodology

3. Dataset

4. Conceptual Analysis

5. Conclusions

2. METHODOLOGY

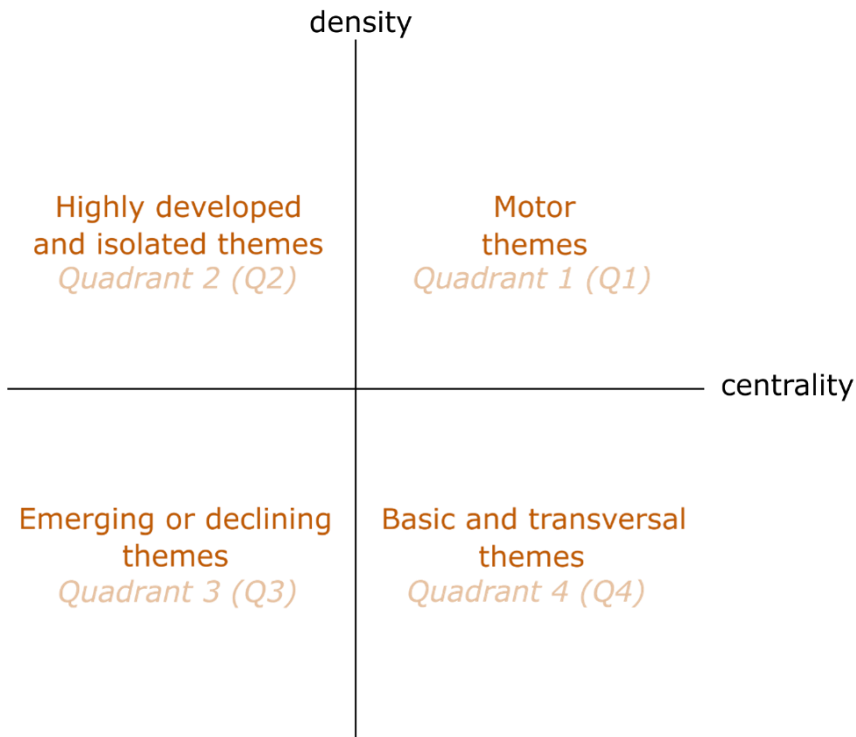
SOFTWARE TOOL

SciMAT was employed to develop a longitudinal conceptual science mapping analysis based on co-words bibliographic networks.

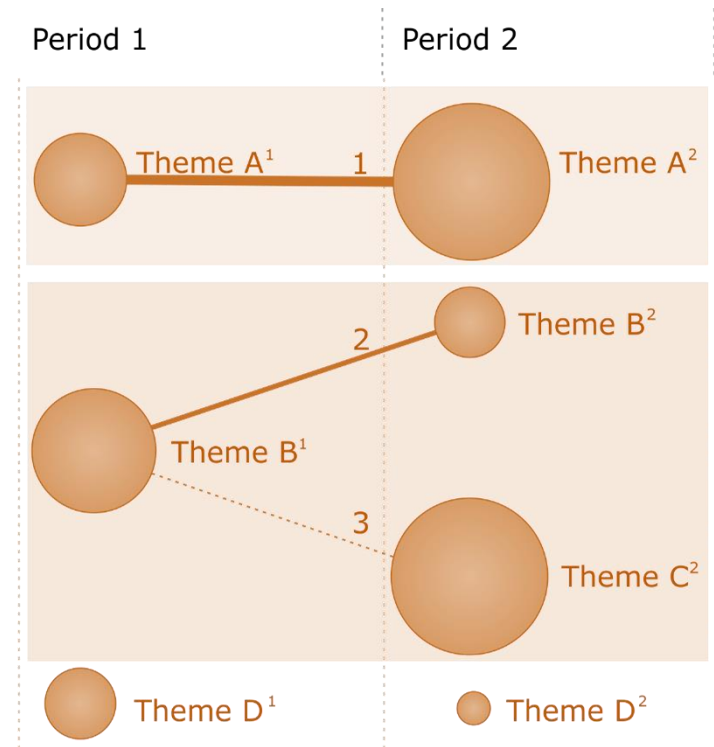
METHODOLOGY STAGES

- 1. Detection of the research themes.** **Co-word analysis**, followed by a **clustering** of keywords to topics/themes. The similarity between the keywords is assessed using the equivalence index.
- 2. Visualizing research themes and thematic network.** **Strategic diagram** and **thematic network** (centrality and density). Research themes mapped in a two-dimensional strategic diagram and classified into four groups (Figure 1): i) motor, ii) basic/transversal, iii) highly developed-isolated, and iv) emerging/declining
- 3. Performance analysis.** Relative contribution of the research themes to the whole research field: number of published documents, number of **citations**, and different types of bibliometric indices (**h-index**).

2. METHODOLOGY



(a) The strategic diagram



(b) Thematic evolution

TABLE OF CONTENTS

1. Introduction

2. Methodology

3. Dataset

4. Conceptual Analysis

5. Conclusions

3. DATASET

CORPUS AND DATABASE

Big Data research documents published in the WoS Core Collection.

QUERY

TS=("big data" OR "big-data") AND PY=2012-2017

TIME PERIOD

2012-2017 divided in three period: 2012-2013, 2014-2015 and 2016-2017.

CORPUS SIZE

- 25,658 documents (articles, proceedings, reviews...) and 140,809 keywords.
- Citations count up to 4th July 2018.
- 2012-2013: 1,643 documents and 7,109 keywords.
- 2014-2015: 8,552 documents and 34,953 keywords.
- 2016-2017: 15,463 documents and 98,747 keywords.

DOCUMENTS AND CITATIONS BY YEAR AND PERIOD

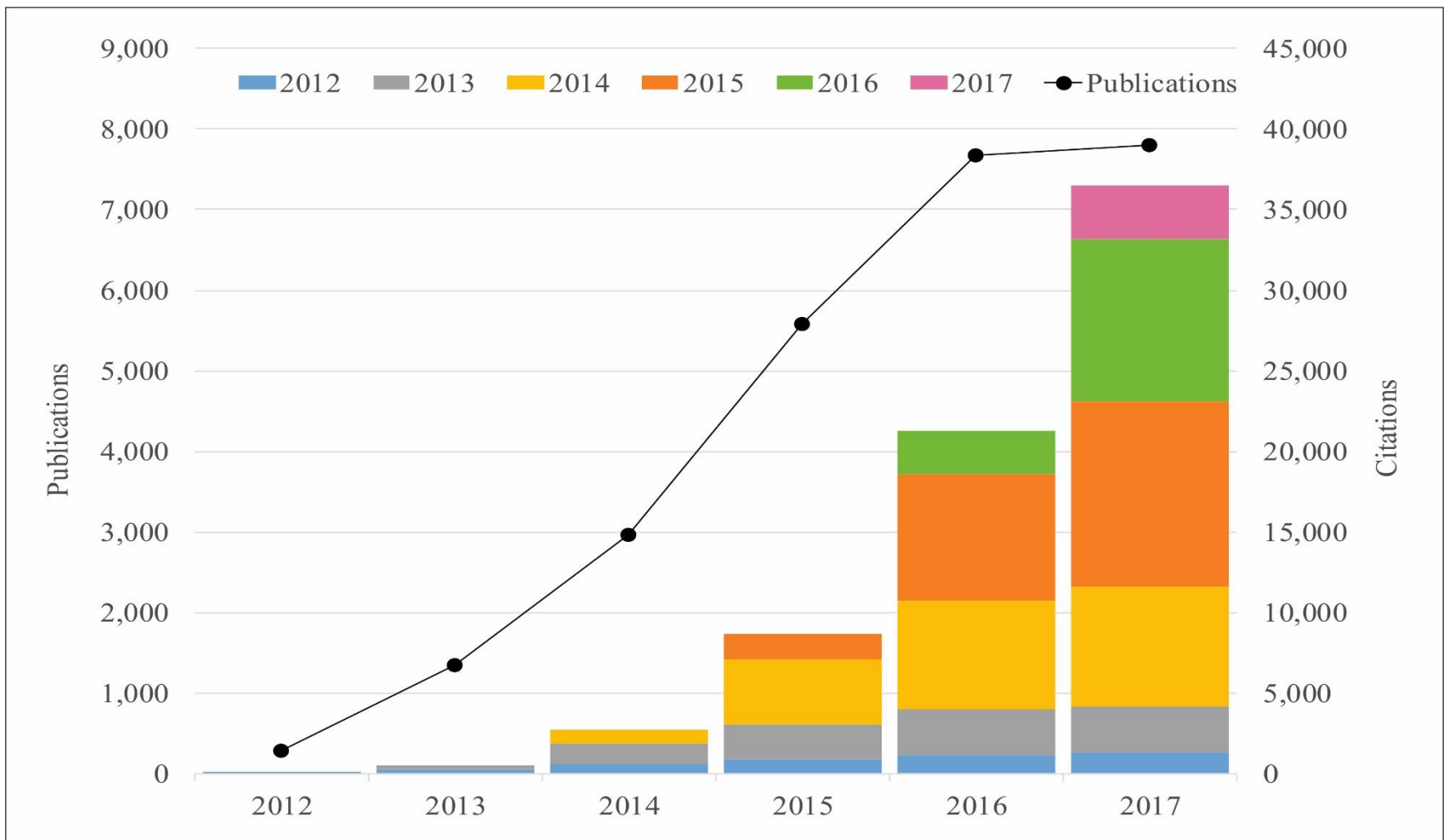


TABLE OF CONTENTS

1. Introduction

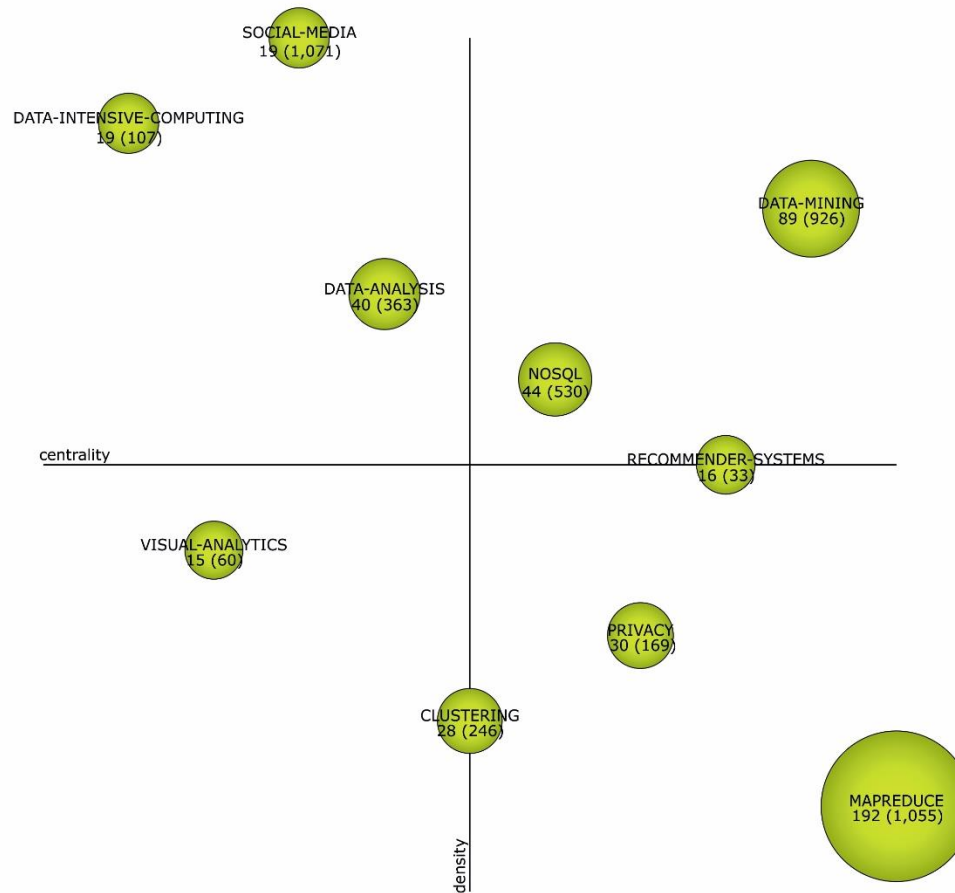
2. Methodology

3. Dataset

4. Conceptual Analysis

5. Conclusions

4. CONCEPTUAL ANALYSIS



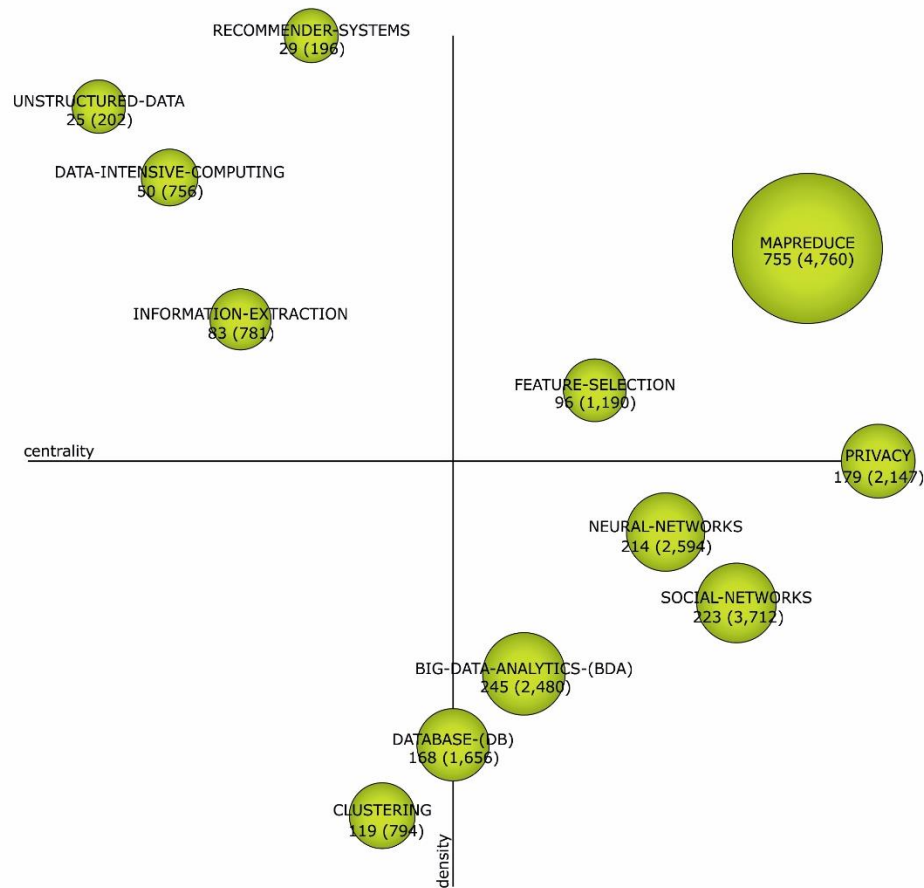
Strategic diagram 2012-2013

4. CONCEPTUAL ANALYSIS

Theme	Documents	Citations	h-index
MAPREDUCE	54	502	10
DATA-MINING	15	46	4
PRIVACY	9	80	3
DATA-ANALYSIS	9	25	2
SOCIAL-MEDIA	7	1,028	4
CLUSTERING	7	13	3
NOSQL	6	7	2
DATA-INTENSIVE-COMPUTING	6	10	1
RECOMMENDER-SYSTEMS	5	21	3
VISUAL-ANALYTICS	5	18	2

The first period has lower number of publications than the other periods, we could identify ten themes related to the **Big Data** research field. In this regard, we could highlight five key themes (motor theme and basic and transversal themes) of the knowledge field: *DATA-MINING*, *MAPREDUCE*, *NOSQL*, *PRIVACY*, *RECOMMENDER-SYSTEMS*

4. CONCEPTUAL ANALYSIS



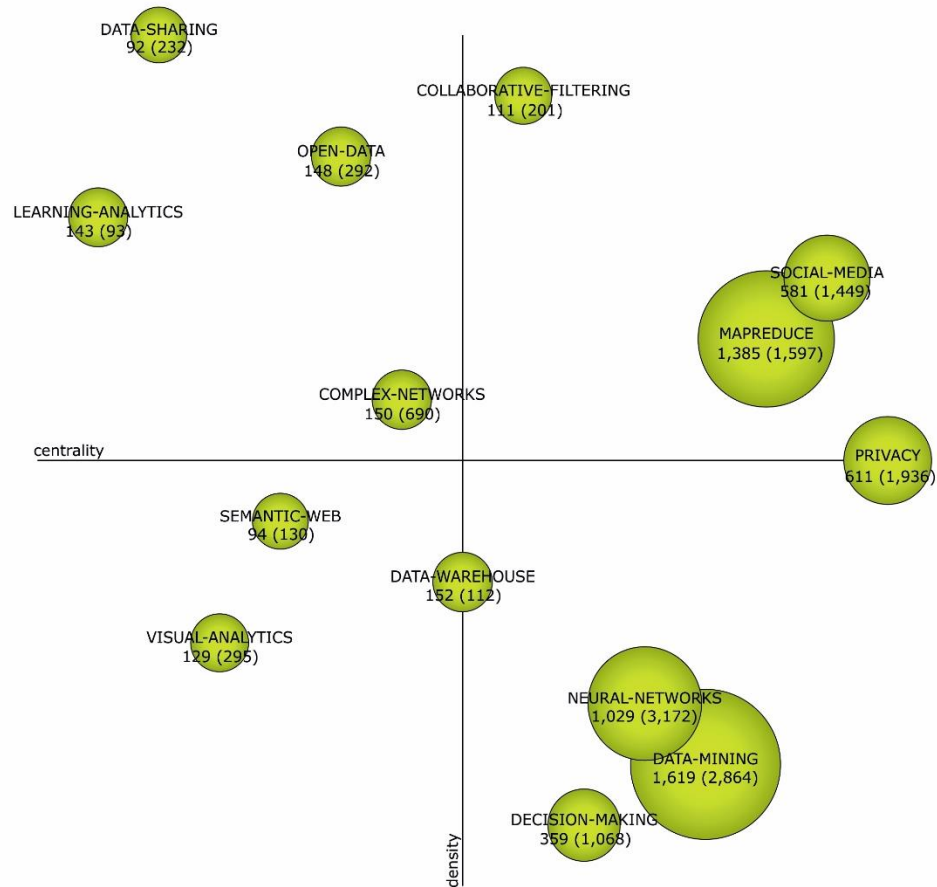
Strategic diagram 2014-2015

4. CONCEPTUAL ANALYSIS

Theme	Documents	Citations	h-index
MAPREDUCE	211	1,650	16
PRIVACY	45	490	11
SOCIAL-NETWORKS	32	423	10
BIG-DATA-ANALYTICS-(BDA)	28	486	7
NEURAL-NETWORKS	23	152	6
DATABASE-(DB)	17	104	6
FEATURE-SELECTION	11	90	5
CLUSTERING	9	59	3
RECOMMENDER-SYSTEMS	8	42	2
UNSTRUCTURED-DATA	6	5	1
INFORMATION-EXTRACTION	6	34	3
DATA-INTENSIVE-COMPUTING	5	50	3

During the second period we could identify twelve themes related to the **Big Data** research field. Consistent with the last period, seven themes are considered keys in the knowledge field: *BIG-DATA-ANALYTICS*, *DATABASE*, *FEATURESELECTION*, *MAPREDUCE*, *NEURAL-NETWORKS*, *PRIVACY*, *SOCIAL-NETWORKS*

4. CONCEPTUAL ANALYSIS



Strategic diagram 2016-2017

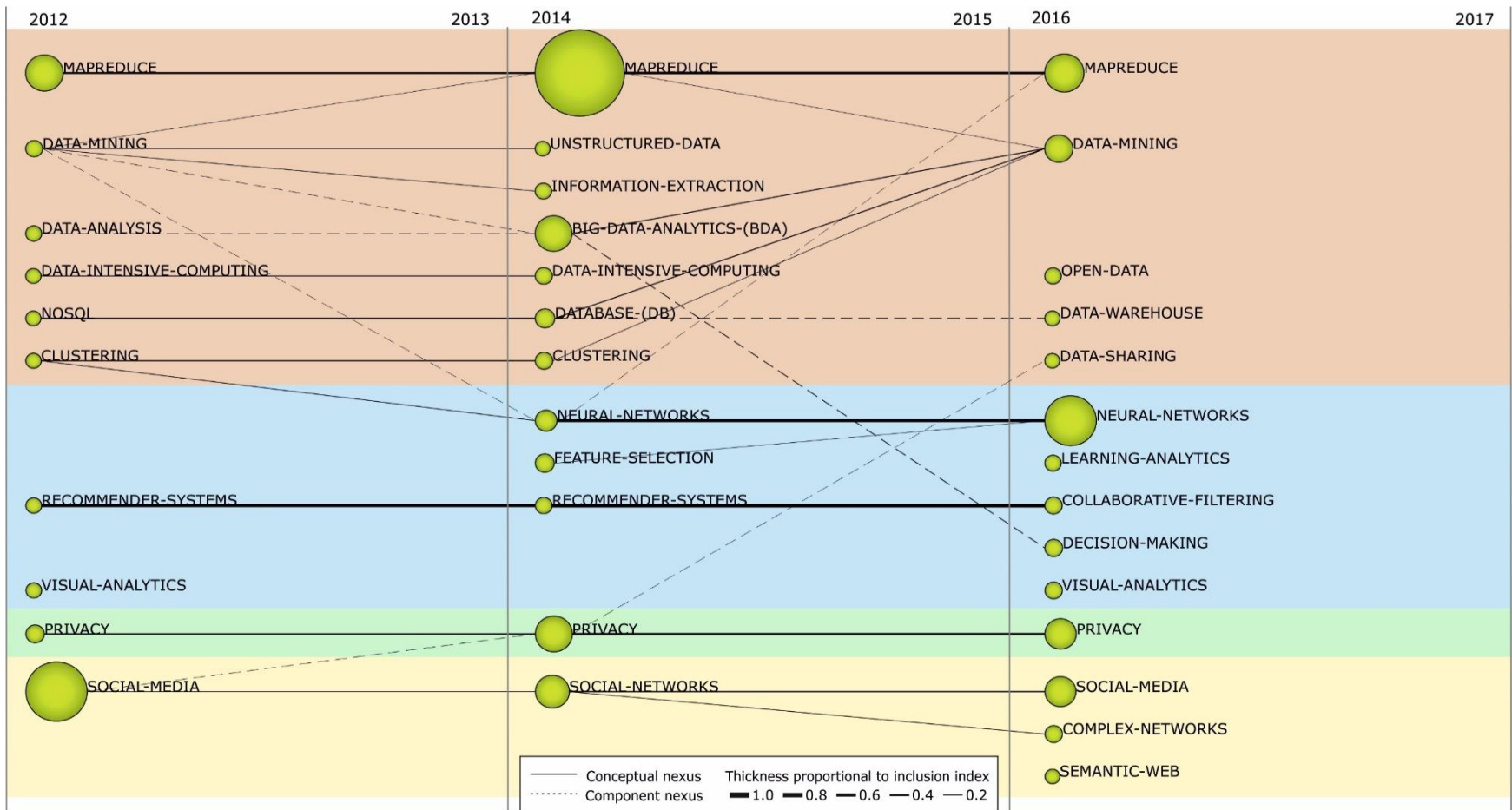
4. CONCEPTUAL ANALYSIS

Theme	Documents	Citations	h-index
MAPREDUCE	419	544	9
DATA-MINING	185	297	9
NEURAL-NETWORKS	157	814	14
PRIVACY	143	367	10
SOCIAL-MEDIA	116	360	10
LEARNING-ANALYTICS	31	35	3
COLLABORATIVE-FILTERING	19	55	3
COMPLEX-NETWORKS	17	59	5
OPEN-DATA	15	39	4
DATA-SHARING	13	4	1
DECISION-MAKING	12	64	5
SEMANTIC-WEB	12	3	1
VISUAL-ANALYTICS	12	56	2
DATA-WAREHOUSE	12	11	2

The third period is the most productive and hosts fourteen themes. In this regard, eight of the total themes are considered key: *COLLABORATIVE-FILTERING*, *DATA MINING*, *DATAWAREHOUSE*, *DECISION-MAKING*, *MAPREDUCE*, *NEURAL-NETWORKS*, *PRIVACY*, *SOCIAL-MEDIA*

CONCEPTUAL EVOLUTION MAP

BIG DATA RESEARCH FIELD FROM 2012 TO 2017



Conceptual evolution map 2012-2017

CONCEPTUAL EVOLUTION MAP

BIG DATA COMPONENTS AND THEMATIC AREAS

In the Big Data evolution map we can identify four kinds of topics: *Data Management*, *Decision Support*, *Privacy* and *WEB & Social Networks*. Accordingly, *MAPREDUCE* is the most representative research theme in the period evaluated followed by *SOCIAL-MEDIA* and *NEURAL-NETWORKS*.

Period 1: 2012-2013 Big Data components	Period 2: 2014-2015 Big Data components	Period 3: 2016-2017 Big Data components
MAPREDUCE, DATA-MINING, PRIVACY, DATA-ANALYSIS, SOCIAL-MEDIA, CLUSTERING, NOSQL, DATA-INTENSIVE-COMPUTING, RECOMMENDER-SYSTEMS and VISUAL-ANALYTICS	MAPREDUCE, PRIVACY, SOCIAL-NETWORKS, BIG-DATA-ANALYTICS-(BDA), NEURAL-NETWORKS, DATABASE-(DB), FEATURE-SELECTION, CLUSTERING, RECOMMENDER-SYSTEMS, UNSTRUCTURED-DATA, INFORMATION-EXTRACTION and DATA-INTENSIVE-COMPUTING	MAPREDUCE, DATA-MINING, NEURAL-NETWORKS, PRIVACY, SOCIAL-MEDIA, LEARNING-ANALYTICS, COLLABORATIVE-FILTERING, COMPLEX-NETWORKS, OPEN-DATA, DATA-SHARING, DECISION-MAKING, SEMANTIC-WEB, VISUAL-ANALYTICS and DATA-WAREHOUSE

TABLE OF CONTENTS

1. Introduction

2. Methodology

3. Dataset

4. Conceptual Analysis

5. Conclusions

CONCLUSIONS

SUMMMARY

- An amount of 25,658 documents (articles, proceedings and reviews) were retrieved from the Web of Science Core Collection.
- The corpus was divided in three period: 2012-2013, 2014-2015 and 2016-2017
 - 2012-2013: 1,643 documents and 7,109 keywords.
 - 2014-2015: 8.552 documents and 34,953 keywords.
 - 2016-2017: 15,463 documents and 98,747 keywords.
- The impact achieved is summarized in the following indicators:
 - Average citations per publication: 3.51
 - Sum of Times Cited (without self-citations): 90,164 (85,584)
 - Citing articles (without self-citations): 21,954 (66,009)

CONCLUSIONS

MAIN CONCLUSION

- The size of literature related to **Big Data** research field showed a noticeable increase in the past decade (2012-2017). Given the large volume of publications and citations received in this field, it is expected that the use of these will be seen as part of other knowledge fields.
- The main themes used in the **Big Data** literature are: *MAPREDUCE, PRIVACY, NEURAL-NETWORKS, CLUSTERING, RECOMMENDER-SYSTEMS, DATA-INTENSIVE-COMPUTING, SOCIAL-MEDIA, DATA-MINING and VISUAL-ANALYTICS.*

FUTURE WORKS

- Evaluate the evolution of the research themes across the consecutive time periods.
- Study the relationship to other disciplines as **Business Intelligence** and **Competitive Intelligence.**



THANK YOU

THE 19th INTERNATIONAL CONFERENCE ON INTELLIGENT DATA ENGINEERING
AND AUTOMATED LEARNING

Acknowledgments: The authors J. R. López-Robles, N. K. Gamboa-Rosales, H. Gamboa-Rosales and H. Robles-Berumen acknowledge the support by the CONACYT-Consejo Nacional de Ciencia y Tecnología (Mexico) and DGRI-Dirección General de Relaciones Exteriores (México) to carry out this study.

REFERENCES

- Liebowitz, J.: Strategic intelligence: business intelligence, competitive intelligence, and knowledge management. Auerbach Publications (2006)
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data: The next frontier for innovation, competition, and productivity. McKinsey & Company, (2011)
- López-Robles, J.R., Otegi-Olaso, J.R., Arcos, R., Gamboa-Rosales, N.K., Gamboa-Rosales, H.: Mapping the structure and evolution of JISIB: A bibliometric analysis of articles published in the Journal of Intelligence Studies in Business between 2011 and 2017. J. Intell. Stud. Bus. 8, (2018)
- Gandomi, A., Haider, M.: Beyond the hype: Big data concepts, methods, and analytics. Int J Inf Manage 35, 137-144 (2015)
- Sagiroglu, S., Sinanc, D.: Big data: A review. Collaboration Technologies and Systems (CTS), 2013 International Conference on, pp. 42-47. IEEE (2013)
- López-Robles, J.R., Otegi-Olaso, J.R., Porto-Gómez, I.: Bibliometric analysis of worldwide scientific literature in Project Management Techniques and Tools over the past 50 years: 1967-2017. Research and Education in Project Management (Bilbao, 2018) 49 (2018)
- Glänzel, W.: Bibliometric Methods for Detecting and Analysing Emerging Research Topics. Prof. Inf. 21, 194-201 (2012)
- López-Robles, J.R., Otegi-Olaso, J., R., Porto-Gómez, I., Gamboa-Rosales, H., Gamboa-Rosales, N.K.: Intelligence: origin, evolution and trends. VISIO 2018 Conference, (2018)
- Glänzel, W.: The role of core documents in bibliometric network analysis and their relation with h-type indices. Scientometrics 93, 113-123 (2012)
- Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V.: Critical analysis of Big Data challenges and analytical methods. J. Bus. Res. 70, 263-286 (2017)
- López-Robles, J.R., Otegi-Olaso, J.R., Gamboa-Rosales, N.K., Gamboa-Rosales, H., Cobo, M.J.: 60 Years of Business Intelligence: A Bibliometric Review from 1958 to 2017. In: New Trends in Intelligent Software Methodologies, Tools and Techniques: Proceedings of the 17th International Conference SoMeT_18, pp. 395. IOS Press, (2018)
- Van Raan, A.F.: The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. Technology Assessment-Theory and Practice 1, 20-29 (2003)
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: SciMAT: A new science mapping analysis software tool. J. Am. Soc. Inf. Sci. Technol. 63, 1609-1630 (2012)
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. Journal of Informetrics 5, 146-166 (2011)
- Zupic, I., Čater, T.: Bibliometric methods in management and organization. Organizational Research Methods 18, 429-472 (2015)
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools. J. Am. Soc. Inf. Sci. Technol. 62, 1382-1402 (2011)
- Gutiérrez-Salcedo, M., Martínez, M.A., Moral-Munoz, J.A., Herrera-Viedma, E., Cobo, M.J.: Some bibliometric procedures for analyzing and evaluating research fields. Appl Intell 48, 1275-1287 (2018)