

The Rise of Big Data in Communication Sciences: A Bibliometric Mapping of the Literature

İletişim Biliminde Büyük Verinin Yükselişi: Literatürün Bibliyometrik Haritalaması

Tuğba KARABOĞA¹, Hasan Aykut KARABOĞA², Yasin ŞEHİTOĞLU³



¹Ph.D. Candidate, Yıldız Technical University, Faculty of Economics and Administrative Sciences, Istanbul, Turkey

²Ph.D. Candidate, Yıldız Technical University, Faculty of Arts and Sciences, Statistics Department, Istanbul, Turkey

³Assoc. Prof. Dr., Yıldız Technical University, Faculty of Economics and Administrative Sciences, Istanbul, Turkey

ORCID: T.K. 0000-0003-3830-3536;
H.A.K. 0000-0001-8877-3267;
Y.Ş. 0000-0003-0074-6446

Sorumlu yazar/Corresponding author:

Hasan Aykut Karaboğa,
Yıldız Teknik Üniversitesi, Fen Edebiyat
Fakültesi, İstanbul, Türkiye
E-posta/E-mail: karaboga@yildiz.edu.tr

Geliş tarihi/Received: 24.12.2019

Revizyon talebi/Revision Requested:
10.02.2020

Son revizyon teslimi/Last revision
received: 06.04.2020

Kabul tarihi/Accepted: 12.05.2020

Atf/Citation: Karaboga, T., Karaboga, H.A., & Sehitoglu, Y. (2020). The rise of big data in communication sciences: a bibliometric mapping of the literature. *Connectist: Istanbul University Journal of Communication Sciences*, 58, 169-199.
<https://doi.org/10.26650/CONNECTIST2020-0083>

ABSTRACT

Today's digital world is characterized by advances in communication and information technologies. Internet technology provides a variety of communication channels like social media platforms, social network sites, search engines, blogs, forums, websites and e-mails. The users of these channels create digital traces which are the main source of big data in communication studies in social sciences. Big social data analytics in communication studies provides quantitative indicators to fully understand current situations rather than pre-defined cause and effect relationships. This study aims to investigate the studies in "big data and communication" in social sciences between the years 2014 and 2018. Web of Science Social Science Citation Index journals are selected to present systematic and quantitative analysis of the research field. Bibliometric analysis results provide insights about big data usage and expansion in the communication field not previously grasped by other reviews on this special topic. Bibliometric tools helped to identify research clusters, key research topics, and network and collaboration patterns in big data and communication studies in a social sciences context. This bibliometric mapping of the field visually illustrates the evolution of studies over time and identifies current research interests and future directions for the followers.

Keywords: Big data, communication, social media, bibliometric analysis, co-citation network

ÖZ

Günümüz dijital dünyası, iletişim ve bilgi teknolojilerindeki ilerlemelerle karakterize edilir. İnternet teknolojisi, sosyal medya platformları, sosyal ağ siteleri, arama motorları, bloglar, forumlar, web siteleri ve e-postalar gibi çeşitli iletişim kanalları sunmaktadır. Bu kanalların kullanıcıları, sosyal bilimlerdeki iletişim çalışmalarında büyük verinin ana kaynağı olan dijital izler yaratmaktadır.

İletişim çalışmalarındaki büyük sosyal veri analizi, önceden tanımlanmış neden sonuç ilişkilerinden ziyade mevcut durumları tam olarak anlamak için nicel göstergeler sağlamaktadır. Bu çalışma, 2014-2018 yılları arasında sosyal bilimlerde "büyük veri ve iletişim" konusundaki çalışmaları incelemeyi amaçlamaktadır. Web of Science Sosyal Bilimler Atf Dizini dergileri araştırma alanının sistematik ve kantitatif analizini sunmak için seçilmiştir. Bibliyometrik analiz sonuçları, daha önce bu özel konuyla ilgili diğer incelemelerde ele alınmayan iletişim alanında büyük

verinin kullanımı ve yayılımı hakkında bilgiler vermektedir. Bibliyometrik araçlar, sosyal bilimler kapsamında büyük veri ve iletişim çalışmalarındaki araştırma kümelerini, temel araştırma konularını, ağ ve işbirliği modellerini belirlemeye yardımcı olmuştur. Alanın bu bibliyometrik haritalaması, zaman içindeki çalışmaların evrimini görsel olarak gösterir ve takipçilere yönelik mevcut araştırma ilgi alanlarını ve gelecekteki yönelimleri tanımlar.

Anahtar Kelimeler: Büyük veri, iletişim, sosyal medya, bibliyometrik analiz, ortak atf ağı

INTRODUCTION

In recent years, big data usage in communication studies has gained great interest both from academicians and business professionals (Schroeder, 2016). The widespread use of big communication data has affected many subfields of communication studies such as mass communication, political communication, health communication, digital media communication and other areas of communication (Parks, 2014). Digital technologies have increased the connectedness level of digital devices by modifying the world as an interconnected cyber world (Tsou, 2011). In this world, people get involved in online communication activities with or without being aware of so doing. Receiving or sending e-mails, visiting websites and communicating on social media channels such as Facebook, Twitter, WhatsApp, LinkedIn, Instagram, Pinterest and Youtube creates a large number of digital traces which is the main big data source of new era communication (Boyd, & Crawford, 2012; Lewis, Zamith, & Hermida, 2013).

The emergence of social media in the mid-2000s provided new opportunities for social science researchers because for the first time people's ideas, opinions, feelings, images, behaviors, communications and comments could be followed by researchers in their natural settings without any intervention (Monovich, 2011). The development of big social data led to the development of computational social sciences. Mass and interpersonal communication researches can benefit from the methods and findings of other disciplines with strong computational abilities (Cappella, 2017). In this new world of computational communication science, interpersonal data allow researchers to gain insights from unstructured, complex and huge data sets with big data analytics methods and algorithms.

The studies in the field of big data and communication in social sciences are typically gathered under two groups. The first group consists of computational communication studies that analyze large communication data from various media platforms, online information sources, social network sites and internet search engines by using different statistical analysis methods, revealing confidential and guiding information for all stakeholders of social life, business, economy, education and governments. The second group consists of studies that examine the impact of big data on communication studies at a more conceptual level and gives directive information for future research.

With the help of bibliometric analysis, this study aims to fill a research gap by conducting systematic and quantitative analysis of big data and communication studies in top social sciences journals in the Clarivate Analytics Web of Science Core Collection database. The study draws a big picture and tries to show the evolution of the research area in the years between 2014 and 2018. Bibliometric analysis is utilized to present the yearly increase in the number of articles, the most productive authors and their impacts, affiliation and country collaboration networks, keyword analysis and co-word network and finally citation analysis of the articles in the research field. With the help of these analyses, this study presents the comprehensive evolution of big data usage in communication studies by identifying clusters of research areas according to author keywords, journal network and country collaboration. From these results, current research interests and opportunities for future research are discussed.

The rest of the paper is structured as follows: firstly big data literature and the importance of big data usage in communication studies are given; next, the methodological design of the article is described by giving information about bibliometric analysis methods and tools, then, the results of bibliometric analysis are presented in different titles. Finally, the study is concluded with the discussion of findings, limitations of the study and suggestions for future research.

Big Data Definition

There is not a common definition of big data in the literature. Researchers have defined it from different perspectives. One of the definitions states big data as “A collection of huge volumes of diverse types of structured and unstructured data that cannot be handled by state-of-the-art data processing platforms” (Kaur, & Sood, 2017, p. 1). In another definition, big data is defined as “data that exceeds the processing

capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the structures of your database architectures. To gain value from this data, you must choose an alternative way to process it" (Dumbill, 2013, p. 1).

Previous studies also show that big data has three, four or five characteristics that distinguish it from other data sets. McAfee and Brynjolfsson (2012) emphasized 3 main characteristics that differentiate big data from traditional small data: volume, velocity and variety. Volume is the basic distinguishing dimension which indicates the huge amount of data gathering from various sources (O'Leary, 2013). Velocity refers to the speed of real time data collection speed (Russom, 2011). Variety is related to heterogeneity of data with various sources and formats, and a common conclusion can be drawn by using data from different sources in any relevant analysis (O'Leary, 2013; Russom, 2011). In the following studies Oracle (2012) added value dimension to refer to the business value of insights extracted from huge data sets as a result of various statistical analysis. Finally, White (2012) added veracity dimension to refer to reliability, consistency, accessibility and transparency of big data during the process of obtaining, storing and analyzing the data (Demchenko, Grosso, De Laat, & Membrey, 2013).

Social big data is the main source of communication studies in today's digital world. Olshannikova and her colleagues (2017) express big social data as human-generated data and categorize them into three dimensions: digital-self representation data, technology-mediated communication data and digital relationships data. According to Ishikawa (2015), social big data has four characteristics: volume, variety, velocity and vagueness. The first three characteristics are discussed in many big data studies, but vagueness first appeared in his study as a special characteristic of social big data and it refers to inconsistency and deficiency in social data. Also, Ishikawa classifies social data sources into different categories such as blogging, micro blogging, social network service, sharing service, video communication, social search, social news, social gaming, crowd sourcing, and collaboration.

Big data analytics is a term used to express advanced analytical techniques applied on big data sets (Russom, 2011, p. 8). Our digitally connected daily lives create digital traces and these big digital records are valuable for companies and sciences studying human behaviors. At this point, big data analytics is applied to analyze huge data sets to gain valuable insights from the hidden information embedded in digital communication data (Boyd, & Crawford, 2012). Data-mining algorithms and data-driven decision-making

influences people, organizations and their communications. Analysis of social networks, data visualization, data aggregation, data mining, natural language processing, sentiment analysis, machine learning, deep learning, web mining and web analytics are some of the analytical methods used in analyzing big communication data (Park, 2014).

The Rise of Big Data in Communication Studies

Communication studies typically try to understand human communication processes and include many subjects such as media studies, media law and ethics, the Internet, social media, broadcasting, mobile communication, political communication, personal communication, international communication, journalism, journalism trends and education, human–computer interaction, advertising, public relations, cultural studies, audience studies, and film studies (Calhoun, 2011; Lee, Jung, & Song, 2016). In recent years, online communication tools have created social spaces for professionals and social interactions. Factors such as the speed of idea sharing, ease of use, cost efficiency, effective time management, elimination of distances between people, and being an important marketing, customer interaction and crisis communication tool are widening the usage of digital communication tools (Stieglitz, & Dang-Xuan, 2013).

Communication studies in the field of social sciences have recently been fed by big data sources. The widespread use of big communication data has contributed too many subfields of communication areas such as political communication, mobile communication, social network analysis and mass communication (Parks, 2014). Digital data flows from social media platforms, online information sources, social network sites, internet search engines and mobile/smart phone applications creating accessible data for researchers. Accessible big communication data sources have made social science more quantitative and statistical. Researchers and businesspeople obtain and analyze big data sets to take data-driven decisions which make social sciences more measurable and scientific (Schroeder, 2016).

Big data has created a computational social science era and many researchers have analyzed social behaviors by different statistical methods (e.g. Boyd, & Crawford, 2012; Kollanyi, Howard, & Woolley, 2016; Kreiss, & Jasinski, 2016; Parks, 2014; Shah, Cappella, & Neuman, 2015; Trilling, 2017). The analyses of communication data is fueled by three developments, which briefly are: digitally available huge data sets, advanced big data analytical methods and tools, and the decreasing costs of data processing, user friendly

computing infrastructure for data processing and crowdsourcing data platforms (van Atteveldt, & Peng, 2018).

Big digital data from social interactions and social media communications are mostly preferred by businesses and governments for practical purposes (Stieglitz, & Dang-Xuan, 2013). Corporations use big data analytics for marketing and advertising activities, new product development, investment plans, transportation decisions and many related business operations (Larson, & Watson, 2011). Therefore, big data is an indispensable part of communication and other business activities to create value for businesses. Also, governments use big social media data to understand public opinion and expectations about government policies and economics (Paris, & Wan 2011; Stieglitz, Brockmann, & Dang-Xuan, 2012). For example, a data-driven election campaign was conducted for president Obama in 2012 (Kreiss, & Jasinski, 2016) and political bots and automated accounts over Twitter were used in the U.S. Presidential debates in 2016 (Kollanyi, Howard, & Woolley, 2016).

AIM AND METHODOLOGY

Bibliometrics was first used by Pritchard (1969, p. 349) as “the application of mathematics and statistical methods to books and other media of communication”. Bibliometrics is generally combined with science mapping techniques and allows the examination of the body of literature in a specific research field to understand major themes (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011; Vogel, & Güttel, 2013). Bibliometric analysis is defined as a statistical approach which enables monitoring and evaluation of the developments and progresses in a specific discipline by analyzing the citations of the studies, relations between authors, keywords, and methods discussed in the study (Koseoglu, Rahimi, Okumus, & Liu, 2016; McBurney, & Novak, 2002). It provides more objective results by allowing larger literature search queries and providing clearer publication policies with a better understanding of author fields (Zupic, & Čater, 2015).

Aim

This study aims to investigate the studies in the “big data and communication” field in social sciences between the years 2014 and 2018. Web of Science Social Science Citation Index journals and articles are selected as samples to present systematic and

quantitative analysis of the research field. Instead of selecting only communication journals, all journals in SSCI are selected because big communication data is used in many fields of social sciences such as company management, marketing, tourism management, health management, transportation, psychology, politics, and knowledge management. Also, the years between 2014 and 2018 are selected because it is seen that big data was studied intensively in 2014 and beyond in communication studies.

As a research methodology, bibliometric analysis tools are used to investigate and describe the quantity, characteristics, and productivity of “big data and communication studies” in a global context. Bibliometric approach enables monitoring and evaluation of the research field by analyzing the citations of the studies, relations between authors, keywords, and methods discussed in the studies. This is the first study to apply bibliometric analysis methods in exploring the evolution of the big data and communication field in social sciences.

In this context, this study aims to answer the following research questions:

- How does the use of big data in communication studies develop over time?
- Who are the most influential writers in this field?
- What are the most productive journals in this field?
- Which are the most dominant countries in the articles?
- How is the international collaboration structure of countries?
- What are the most commonly used keywords in the articles and how are they clustered?
- How is the trend topic evolution in the research field over time?
- What is the intellectual structure of the field like according to the co-citation network of articles and journals?

Methodology

Sampling and Data Collection

In this study, bibliometric analysis is used to investigate and describe the quantity, characteristics, and productivity of “big data and communication studies” in a global context. Big data development in information and communication technology has shifted the attention of researchers from different disciplines to the field of communication

and the big data-driven communication studies have rapidly started to gain great importance (e.g. Boyd, & Crawford, 2012; Lee, Jung, & Song, 2016; Lewis, Zamith, & Hermida, 2013; Manovich, 2011; Parks, 2017). Therefore, the number of articles in big data and communication studies has grown faster in the last decade and bibliometric analysis is considered to be the prominent way of reviewing related literature.

Clarivate Analytics Web of Science Core Collection database is the mostly preferred database for bibliometric studies thanks to its timeliness and high reliability (Falagas, Pitsouni, Malietzis, & Pappas, 2008). In order to decide which keywords to use, where the big data is used mostly in the subfields of communication studies was investigated and then the search query was written. After several trials to write a broad research query, the final query was created as below:

TS=(("big data*") AND (communication* OR "public relations*" OR journalism* OR advertis* OR "social media*"))

In the first step, Web of Science "TS" (Topic Search) operator searches selected terms, titles, abstracts, and keywords of the articles. In this search, English is selected as the language and a 5 year period, 2014-2018, is taken as the timespan because it is seen that big data was studied intensively in 2014 and beyond in communication studies. This initial search generated 4,294 articles. In the second step, the search is limited to articles and reviews as the document type and Social Science Citation Index (SSCI) is selected as the Web of Science Index category. SSCI is selected as the Web of Science Index to evaluate the effect of big data on communication studies in the field of social sciences. As a result, 803 entries were generated from this limited search query. In the third and final step, titles and abstracts of 803 academic articles were carefully reviewed for relevancy to "big data and communication" and unrelated articles were deleted. As a result, 538 articles remained in the final dataset to be used in the bibliometric analysis.

Data Processing

Researchers can conduct bibliometric analysis by using different software packages. In this study, the bibliometrix package developed in R statistical software program by Aria and Cuccurullo (2017) was used to make bibliometric analysis. R statistical software program is suitable for analysis thanks to its open source code and packages. This package is designed to analyze data using Scopus or WoS easily. Also, VOSviewer

package of BibExcel software program is used to visualize bibliometric networks. VOSviewer package can handle large data sets efficiently and innovatively visualize networks and collaborations.

FINDINGS

Initial Statistics

Figure 1 shows the number of articles published by years. While 40 studies were published in 2014 regarding the use of big data in communication studies, this number increased to 179 in 2018. Annual Percentage Growth Rate of the studies is 45.445%. It is also seen that the citations received from the first studies are quite high.

Table 1 shows that a total of 538 articles in the study come from 285 sources. 511 of these studies are research articles and 27 of them are review articles. It is observed that a total of 1491 authors took part in the studies. It was found that each author published an average of 0.361 articles and 2.77 authors in each article. However, a total of 96 authors published 100 single-name studies. Considering that the cooperation index is 3.18, it is seen that the studies are generally written in cooperation. It should be emphasized that this field includes highly collaborative studies.

Table 2 shows the number of articles published by journals. Accordingly, 1 journal has published 21 articles, 1 journal has 16 articles, 1 journal has 12 articles and 2 journals have published 11 articles. Of the 285 journals, 196 journals published one article and 33 journals published two articles. Accordingly, 80.3% of the journals published very few articles and 5 journals are dominant in the field. The most productive journals are given in Table 3.

According to Table 3, it is understood that five most productive journals are *Digital Journalism*, *International Journal of Communication*, *Sustainability*, *Information Communication & Society* and *New Media & Society* respectively. The first publications of *Sustainability* and *New Media & Society* journals were published in 2016, while the first publications of the others were given in 2014, which is the starting year of our study. Accordingly, it is thought that these journals will continue to dominate “big data and communication studies” by increasing the number of publications and citations.

Author Influence

Information about the authors' production is given in Table 4. According to the table, Chen, Y. seems to be the most productive author in terms of the total number of articles contributed by the author. However, when compared to the number of authors in the articles, Vargo, C. J. has passed the first three writers. As the number of authors in the article decreases, even if the total number of articles of the author decreases, the fractionalized effect of the author can increase. For example, Zhang Y., who ranked 10th according to the total number of articles contributed, raised to the second rank in the latter ranking. Bail, C. A., Carlson, M. and Couldry, N. were not among the top 20 authors, but they were at the top of the list. Accordingly, it is observed that the authors of the list have a high tendency to cooperate.

In Figure 2, the production of the authors is examined according to time. It is understood that the most productive authors have published their articles related to big data and communication in the last two years. According to the graph, Poorthuis, A. is one of the most prolific writers among authors. Table 5 confirms this by the author's influence. In figure 2, Vargo, C. J. is the most remarkable writer according to M-index which is the median number of cited publications (Bornmann, Mutz, & Daniel, 2008). Although the author only started to produce in 2016, his effectiveness is increasing rapidly. Ye, X. is also ranked first in the list in terms of G-index which is more sensitive to high-cited works and refers to the top g publications with g^2 or more citations (Egghe, 2006). Considering that the number of articles is 7 and the year of commencement of production is 2016, it is determined that this author also has effective publications, and the probability of being more effective is increasing as time goes on.

Affiliation and Country Collaboration Network

Table 6 shows the total number of publications produced by 41 countries. Accordingly, USA is the most productive country with 196 publications. 36.431% of the analyzed publications were produced with the contribution of the authors in USA.

Figure 3 shows whether the publications of the first 20 countries are published in collaboration with authors from different countries or with the citizens of their own countries. In general, the publications of Single Country Papers (SCP) are more common than Multi Country Papers (MCP).

In addition to identifying the most productive countries, country collaboration network shows the dominant countries in the field and cooperation between countries. The size of the bubbles in Figure 4 indicates the dominance of countries in big data and communication studies and the thickness of the lines indicates the strength of the collaboration between countries. The color of the bubbles indicates the clusters, which countries cooperate mostly with each other. According to the given network, USA, China, UK and Netherlands are the most productive countries. In addition to producing the highest number of publications, these blue cluster countries are also the ones that create strength collaboration networks between each other and with other countries. Australia, Italy, Spain and Korea dominated the second largest cluster, green cluster, in the collaboration network. The third cluster, red cluster, is composed of the dominance of Canada, Germany, Switzerland and Saudi Arabia. Also, the green and red clusters have high collaboration network with the blue cluster. The fourth cluster is quite small and the distance of this cluster to the other clusters is great. Accordingly, South Africa and Finland conduct studies on big data and communication by collaborating with each other and with other countries.

Keyword Analysis and Co-word Network

Table 7 shows the number of times certain keywords occur in the articles. According to this table, big data and social media are the most frequently used keywords in big data and communication studies. In addition to these words, social media applications, Twitter and Facebook, are also included in the table as mostly frequently used keywords in the articles. This is not surprising, because in the new internet era, the big communication data is generally formed on social media applications. Today, most of the studies in the field of communication are fed from this data. As the data obtained from these sources are analyzed in the articles, the words sentiment analysis, text mining, content analysis, social media analytics, machine learning, natural language processing, data mining and other related analytical words are used extensively as keywords in the articles.

Trend topic analysis according to author and keywords of the articles between the years 2014 and 2018 shows which keywords come to the forefront over time (Figure 5). In 2014, when studies in the field of usage of big data communication first started, it was understood that there were more methodological discussions about how to use big data in communication studies. Then, in 2015 and 2016, as indicated in Figure 3,

studies were formed around concepts such as algorithms, computational social sciences, computational journalism, big data analytics and social media analysis. In 2017 and 2018, the concepts of social media, Twitter, Facebook, decision making, sentiment analysis, text mining, machine learning, artificial intelligence, deep learning, and natural language processing came to the forefront. Trend topic analysis is a good visualization of how the relationship between big data and communication is rapidly shaped over a five year period. This visualization also helps researchers to formulate ideas about how the studies in this field will be shaped in the future. Based on this figure, it becomes clear that the use of big data in the field of communication is not only realized simply by big data analysis, but also the language and images in big communication data can be analyzed by various artificial intelligence methods. In other words, large unstructured communication data can also be analyzed with artificial intelligence methods. Future researchers will extensively use artificial intelligence methods for the analysis of communication data.

Using VOSviewer, 538 studies were analyzed according to the keywords used (Figure 6). Co-word analysis creates clusters with different colors according to the author keywords used in the title, abstract, and keywords of the studies. This classification indicates the frequency of a keyword appearing with other keywords. In Figure 6, social media and Twitter are two main clusters/keywords used in connection with big data research in communication studies. That means social media and twitter are basic communication sources in big data studies. Also, Figure 6 indicates the most frequently used keywords separately. For example; big data (red color cluster) is mostly used with the keywords of methodology, networks, social networks, social network analysis, content analysis, decision making, urban planning, user-generated content, surveillance and online news, whereas social media (blue color cluster) is frequently linked to Facebook, social media analytics, customer relationship management, social networks, social big data, computational social science and topic modeling.

Citation Analysis and Co-citation Network

Co-citation networks of the references are visualized in Figure 7 (created with the VOSviewer). Co-citation network is a good visual representation to understand interconnections of references (Sainaghi, Phillips, Baggio, & Mauri, 2018). Co-citation network is used to investigate the intellectual structure of a research field (Leung,

Sun, & Bai, 2017; Rauchfleisch, 2017). The size of the circle represents the number of citations received by articles, the thickness of the lines indicates the strength of the co-citation ties and, finally, connection and proximity between two references indicates the co-citation relationship between them. The color of the circles and ties indicates clusters of the references. As seen in Figure 7, the co-citation network consists of five clusters. The yellow cluster is called “big data transformation,” which focused on the evolutionary effects of big data on social lives, communication habits, culture, knowledge creation and decision making. The second cluster, colored red, defined “computational methods in social sciences,” which emphasized the role and characteristics of computational methods and algorithms in social big data researches. The third cluster, in green, is related to “big data analytics,” which is used to identify and analyze social media data and other communication data. The fourth cluster, which is blue, identifies “social media analytical methods,” which include text mining, user-generated content analysis, opinion mining, sentiment analysis, natural language processing, data visualization and other related methods to analyze online communication data. The last cluster, the purple one, is related to “geographical data analysis of social media traces,” which includes geographical mapping, geographic intelligence, tourist route propositions and smart spaces. Co-citation of references indicates that big data usage in communication studies have diverse theoretical foundations from different areas.

Co-citation networks of journals are shown in Figure 8. The size of the circles shows the normalized number of citations received by journals and the darkness of the lines shows the strength of ties. The connection and closeness between two journals define the co-citation relationship between them. The color of the circles indicates clusters. Journals are classified into 3 clusters and the publications cited in these clusters are considered to have common characteristics. The green cluster on the left is composed of journals which are thought to be dominant in big data, communication and media. The red colored cluster on the right includes business and technology-oriented journals which have studies about marketing, human behavior, decision-making, tourism and management. The blue cluster on the top left consists of journals such as Science, Plos One and Nature which investigate the technical characteristics of big data and communication studies. In this respect, it can be concluded that the studies carried out in the field of big data and communication are grouped under 3 main headings: business implementation; communication, media and journalism; technological infrastructure and methodological investigation.

DISCUSSION AND CONCLUSION

This study presents a systematic and quantitative analysis of big data usage in communication studies in the Web of Science Social Science Citation Index journals. Various bibliometric analysis methods (citation analysis, co-citation network, keyword analysis, co-word network and collaboration network) are applied to show the evolution in the research area in the years between 2014 and 2018.

Descriptive statistics of the 538 articles indicated that the number of articles in the research field increased by about 45 percent annually between the years 2014 and 2018. The presence of a total of 285 journals indicated that big communication data can be used in many different social science disciplines. Author effectiveness is measured with the productivity of authors. Chen, Y., Wang, Z., Ye, X., Vargo, C.J. and Eichstaedt, J.C. are the most productive five authors in this area (Table 4). USA (196 articles), China (62 articles), United Kingdom (61 articles), Australia (23 articles) and Germany (23 articles) are the most productive five countries (Table 6). Very strong collaboration network ties have been established between USA and China, USA and UK, USA and Netherlands and USA and Australia (Figure 4). It is clearly evident that USA is the primary collaborative partner country in the research field.

Keyword analysis results showed that big data, social media, twitter, sentiment analysis and machine learning are the most frequently used keywords in big data and communication studies (Table 7). This was an expected result because big communication data is generally gathered from social media applications especially from twitter. Sentiment analysis and machine learning is used to analyze the data obtained from various communication channels. Co-word analysis (Figure 6) indicated the frequency of a keyword appearing with other keywords. This plot was a good visual representation of keyword clusters according to their interrelated usage in the articles. Big data, social media and twitter were the three main clusters. There is a strong interrelation between big data and social media; big data and twitter; social media and twitter keywords. That means these pairs of keywords were used together in the articles.

The study also analyzed the evolution of trend topics in big data and communication studies over time (Figure 5). In 2014, the first studies in the field started with methodological discussions on the usage of big data in the communication field. In the following three years, studies were shaped around the concepts such as algorithms,

computational social sciences, computational journalism, big data analytics, social media analysis, sentiment analysis and machine learning. This three-year period was an implementation period of big data analytical methods on social media data and other online communication data. In the last year, in 2018, it was understood that there were variations in both analytical methods and in the research area. More artificial intelligence methods came to the forefront. The visualization of keywords over time is helpful to formulate ideas about how the studies in this field will be shaped in the future. Based on the trend topic analysis, it becomes clear that the use of big data in the field of communication is not only realized by simple big data analysis, but also the language and images in big communication data can be analyzed by various artificial intelligence methods. In other words, large unstructured communication data can also be analyzed with artificial intelligence methods. Artificial intelligence methods will be extensively used by future researchers to analyze both structured and unstructured of communication data.

The results of the citation analysis showed that the *Journal of Communication* (431 citations), *Digital Journalism* (364 citations), *Information Communication & Society* (240 citations), the *International Journal of Communication* (232 citation) and *Information & Management* (152 citation) were the most frequently cited journals in the research field (Table 3). The journal co-citation network showed the way in which journals were interconnected and considered together. Figure 8 classified journals in three clusters. One cluster was dominant in big data, communication and media studies. The second one is related to business and technology-oriented journals which had studies about marketing, human behavior, decision-making, tourism and management. The journals in the final cluster mainly investigated the technical characteristics of big data and communication studies. These journal clusters gave insights on which sub-areas journals make publications.

The article citation analyses (Table 8) showed that the five studies which were most frequently cited in the research query results were Kramer, Guillory and Hancock, (2014); Bello-Orgaz, Jung, and Camacho (2016); Youyou, Kosinski and Stillwell (2015); Xiang et al. (2015) and Colleoni, Rozza and Arvidsson (2014). The co-citation network of articles (Figure 7) visualized the interconnections of references in the articles. This was a good visualization to investigate the intellectual structure of a research field. Five clusters were identified to show the theoretical foundations in the research area. These five clusters were identified as big data transformation, computational methods in social

sciences, big data analytics, social media analytical methods, and geographical data analysis of social media trace. The co-citation network of references indicated that big data usage in communication studies have diverse theoretical foundations from different research areas.

Although bibliometric analysis has already been separately used in the analyses of communication studies (e.g. Borgman, & Furner, 2002; Borgman, & Rice, 1992; Feeley, 2008) social media studies (e.g. Gan, & Wang, 2015; Leung et al., 2017) and big data studies (e.g. Kalantari et al., 2017; Mishra et al., 2018), no bibliometric study has been conducted to examine big data and communication studies in social sciences. This is the first study to apply bibliometric analysis methods in exploring the evolution of the big data and communication field in social sciences. Combined usage of different bibliometric tools helped us to understand details of the research field and the state of art of the big data and communication studies.

Although our study showed the spread of the use of big data in the field of communication through bibliometric mapping, it has some limitations. Primarily, the data source was limited to the Web of Science Social Sciences Citation Index journals. The results of the analyses in the study may not be generalized to all big data and communication studies in other indexes such as Scopus. Behind, the study only considered social sciences to understand the big data usage in communication studies. Other research areas could also use big communication data from various sources to conduct research about their field with different analytical methods. Future researchers should consider larger research areas to generalize the evolution of big data usage in communication studies. Also, due to the limitations of bibliometric analysis, cluster definitions as theoretical foundations may be biased. Future research could align bibliometric analysis results with systematic reviews of the articles in the research query to better identify current research trends and future directions.

Beside the limitations of the study, bibliometric findings brought to light certain future research interests related to big data usage in communication studies which are under-researched in the literature. Firstly, more studies could be conducted with artificial intelligence methods to analyze big communication data from various online sources. Trend topic analysis results indicated that the usage of artificial intelligence methods in big data analysis is becoming widespread. Therefore, artificial intelligent applications such as machine learning, deep learning, natural language processing, text mining,

sentiment analysis and data mining can be useful to analyze complex big communication data for decision making, disaster management, marketing management, health management, journalism, crisis management, public relations and other areas of communication. Second, collaboration networks indicated that top researchers had many coauthored papers. Strong collaborations between communication scientists in social sciences and data scientists or computer scientists can be helpful for the analysis of big communication data in social sciences. Because, compared to social scientists, data scientists and computer scientists are very successful in analyzing big data sets. Finally, more effort is needed to study theoretical foundations in the context of communication theories. According to the results, social media is the most important theoretical foundation and researchers should pay more attention to analyzing social media data from various methods and to provide underlying theories to strengthen the research field.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors has no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

REFERENCES

- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, 28, 45-59. <https://doi.org/10.1016/j.inffus.2015.08.005>
- Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual review of information science and technology*, 36(1), 2-72. Retrieved from <https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/aris.1440360102>
- Borgman, C. L., & Rice, R. E. (1992). The convergence of information science and communication: A bibliometric analysis. *Journal of the American Society for Information Science*, 43(6), 397-411. Retrieved from <https://www.dhi.ac.uk/san/waysofbeing/data/health-jones-borgman-1992.pdf>
- Bornmann, L., Mutz, R., & Daniel, H. D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and technology*, 59(5), 830-837. <https://doi.org/10.1002/asi.20806>
- Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>

- Calhoun, C. (2011). Plenary| Communication as Social Science (and More). *International Journal of Communication*, 5, 18. <https://doi.org/10.1590/S1809-58442012000100014>
- Cappella, J. N. (2017). Vectors into the future of mass and interpersonal communication research: Big data, social media, and computational social science. *Human Communication Research*, 43(4), 545-558. <https://doi.org/10.1111/hcre.12114>
- Chae, B. K. (2015). Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, 165, 247-259. <https://doi.org/10.1016/j.ijpe.2014.12.037>
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, 63, 67-80. <https://doi.org/10.1016/j.dss.2013.08.008>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). 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(1), 146-166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Coddington, M. (2015). Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), 331-348. <https://doi.org/10.1080/21670811.2014.976400>
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of communication*, 64(2), 317-332. <https://doi.org/10.1111/jcom.12084>
- Conway, B. A., Kenski, K., & Wang, D. (2015). The rise of Twitter in the political campaign: Searching for intermedia agenda-setting effects in the presidential primary. *Journal of Computer-Mediated Communication*, 20(4), 363-380. <https://doi.org/10.1111/jcc4.12124>
- Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013, May). Addressing big data issues in scientific data infrastructure. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (pp. 48-55). IEEE. <https://doi.org/10.1109/CTS.2013.6567203>
- Dumbill, E. (2013). Making sense of big data. <https://doi.org/10.1089/big.2012.1503>
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131-152. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.395.9064&rep=rep1&type=pdf>
- Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., ... & Weeg, C. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science*, 26(2), 159-169. <https://doi.org/10.1177/0956797614557867>
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal*, 22(2), 338-342. <https://doi.org/10.1096/fj.07-9492LSF>
- Feeley, T. H. (2008). A bibliometric analysis of communication journals from 2002 to 2005. *Human Communication Research*, 34(3), 505-520. <https://doi.org/10.1111/j.1468-2958.2008.00330.x>

- Gan, C., & Wang, W. (2015). Research characteristics and status on social media in China: A bibliometric and co-word analysis. *Scientometrics*, *105*(2), 1167-1182. <https://doi.org/10.1007/s11192-015-1723-2>
- Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*, 129-152. <https://doi.org/10.1146/annurev-soc-071913-043145>
- Hasan, S., & Ukkusuri, S. V. (2014). Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies*, *44*, 363-381. <https://doi.org/10.1016/j.trc.2014.04.003>
- Ishikawa, H. (2015). *Social big data mining*. Florida, USA: CRC Press.
- Kalantari, A., Kamsin, A., Kamaruddin, H. S., Ebrahim, N. A., Gani, A., Ebrahimi, A., & Shamshirband, S. (2017). A bibliometric approach to tracking big data research trends. *Journal of Big Data*, *4*(1), 30. <https://doi.org/10.1186/s40537-017-0088-1>
- Kaur, N., & Sood, S. K. (2017). Dynamic resource allocation for big data streams based on data characteristics (5Vs). *International Journal of Network Management*, *27*(4), 1-16. <https://doi.org/10.1002/nem.1978>
- Kollanyi, B., Howard, P., & Woolley, S. C. (2016). Bots and automation over Twitter during the first U.S. presidential debate. *Comprom Data Memo 2016.1*. Retrieved from <https://regmedia.co.uk/2016/10/19/data-memo-first-presidential-debate.pdf>
- Koseoglu, M. A., Rahimi, R., Okumus, F., & Liu, J. (2016). Bibliometric studies in tourism. *Annals of Tourism Research*, *61*, 180-198. <https://doi.org/10.1016/j.annals.2016.10.006>
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788-8790. <https://doi.org/10.1073/pnas.1320040111>
- Kreiss, D., & Jasinski, C. (2016). The tech industry meets presidential politics: Explaining the Democratic Party's technological advantage in electoral campaigning, 2004–2012. *Political Communication*, *33*(4), 544-562. <https://doi.org/10.1080/10584609.2015.1121941>
- Larson, K., & Watson, R. (2011). The value of social media: toward measuring social media strategies. In *Proceedings of ICIS 2011*, Shanghai, China.
- Lee, K., Jung, H., & Song, M. (2016). Subject–method topic network analysis in communication studies. *Scientometrics*, *109*(3), 1761-1787. <https://doi.org/10.1007/s11192-016-2135-7>
- Leeflang, P. S., Verhoef, P. C., Dahlström, P., & Freundt, T. (2014). Challenges and solutions for marketing in a digital era. *European Management Journal*, *32*(1), 1-12. <https://doi.org/10.1016/j.emj.2013.12.001>
- Leung, X. Y., Sun, J., & Bai, B. (2017). Bibliometrics of social media research: A co-citation and co-word analysis. *International Journal of Hospitality Management*, *66*, 35-45. <https://doi.org/10.1016/j.ijhm.2017.06.012>
- Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, *57*(1), 34–52. <https://doi.org/10.1080/08838151.2012.761702>

- Manovich, L. (2011). Trending: The promises and the challenges of big social data. *Debates in the Digital Humanities*, 2, 460-475. <https://doi.org/10.5749/minnesota/9780816677948.003.0047>
- Marine-Roig, E. (2017). Measuring destination image through travel reviews in search engines. *Sustainability*, 9(8), 1425. <https://doi.org/10.3390/su9081425>
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60-68. Retrieved from <https://wiki.uib.no/info310/images/4/4c/McAfeeBrynjolfsson2012-BigData-TheManagementRevolution-HBR.pdf>
- McBurney, M. K., & Novak, P. L. (2002, September). What is bibliometrics and why should you care?. In *Proceedings. IEEE International Professional Communication Conference* (pp. 108-114). IEEE. Retrieved from <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1049094>
- Mishra, D., Gunasekaran, A., Papadopoulos, T., & Childe, S. J. (2018). Big Data and supply chain management: a review and bibliometric analysis. *Annals of Operations Research*, 270(1-2), 313-336. <https://doi.org/10.1007/s10479-016-2236-y>
- O'Leary, D. E. (2013). Artificial intelligence and big data. *IEEE Intelligent Systems*, 28(2), 96-99. <https://doi.org/10.1109/MIS.2013.39>
- Olshannikova, E., Olsson, T., Huhtamäki, J., & Kärrkäinen, H. (2017). Conceptualizing big social data. *Journal of Big Data*, 4(1), 3. <https://doi.org/10.1186/s40537-017-0063-x>
- Oracle. (2012). *Oracle: Big Data for the Enterprise*. Retrieved from <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>
- Özköse, H., Ari, E. S., & Gencer, C. (2015). Yesterday, today and tomorrow of big data. *Procedia-Social and Behavioral Sciences*, 195, 1042-1050. <https://doi.org/10.1016/j.sbspro.2015.06.147>
- Paris, C., & Wan, S. (2011, May). Listening to the community: social media monitoring tasks for improving government services. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2095-2100). ACM. Retrieved from <https://dl.acm.org/doi/pdf/10.1145/1979742.1979878>
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934. <https://doi.org/10.1037/pspp0000020>
- Parks, M. R. (2014). Big data in communication research: Its contents and discontents. *Journal of Communication*, 64(2), 355-360. <https://doi.org/10.1111/jcom.12090>
- Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of documentation*, 25(4), 348-349.
- Rauchfleisch, A. (2017). The public sphere as an essentially contested concept: A co-citation analysis of the last 20 years of public sphere research. *Communication and the Public*, 2(1), 3-18. <https://doi.org/10.1177/2057047317691054>
- Russell Neuman, W., Guggenheim, L., Mo Jang, S., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, 64(2), 193-214. <https://doi.org/10.1111/jcom.12088>

- Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter, 19*, 40. Retrieved from <https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf>
- Rust, R. T., & Huang, M. H. (2014). The service revolution and the transformation of marketing science. *Marketing Science, 33*(2), 206-221. <https://doi.org/10.1287/mksc.2013.0836>
- Sainaghi, R., Phillips, P., Baggio, R., & Mauri, A. (2018). Cross-citation and authorship analysis of hotel performance studies. *International Journal of Hospitality Management, 73*, 75-84. <https://doi.org/10.1016/j.ijhm.2018.02.004>
- Schroeder, R. (2016). Big data and communication research. *Oxford Research Encyclopedia of Communication*. <https://doi.org/10.1093/acrefore/9780190228613.013.276>
- Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Toward computational social science: Exploiting big data in the digital age. Philadelphia PA: *The ANNALS of the American Academy of Political and Social Science*.
- Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and urban planning, 142*, 198-211. <https://doi.org/10.1016/j.landurbplan.2015.02.020>
- Shelton, T., Poorthuis, A., Graham, M., & Zook, M. (2014). Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data'. *Geoforum, 52*, 167-179. <https://doi.org/10.1016/j.geoforum.2014.01.006>
- Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social network analysis and mining, 3*(4), 1277-1291. <https://doi.org/10.1007/s13278-012-0079-3>
- Stieglitz, S., Brockmann, T., & Dang-Xuan, L. (2012, July). Usage Of Social Media For Political Communication. In *PACIS* (p. 22). Retrieved from <https://aisel.aisnet.org/pacis2012/22>
- Trilling, D. (2017). Big data, analysis of. In J. Matthes, C. S. Davis, R. F. Potter (Eds.) *International Encyclopedia of Communication Research Methods*. New York, NY: Wiley Online Library
- Tsou, M. H. (2011, January). Mapping cyberspace: Tracking the spread of ideas on the internet. In *Proceeding of the 25th International Cartographic Conference*. Retrieved from https://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/D3-Internet,%20web%20services%20and%20web%20mapping/CO-354.pdf
- van Atteveldt, W., & Peng, T. Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures, 12*(2-3), 81-92. <https://doi.org/10.1080/19312458.2018.1458084>
- Vogel, R., & Güttel, W. H. (2013). The dynamic capability view in strategic management: A bibliometric review. *International Journal of Management Reviews, 15*(4), 426-446. <https://doi.org/10.1111/ijmr.12000>
- We are Social and Hootsuite, Global Digital 2019 Report, Retrieved from <https://dijilopedi.com/2019-turkiye-internet-kullanim-ve-sosyal-medya-istatistikleri/>
- White, M. (2012). Digital workplaces: Vision and reality. *Business Information Review, 29*(4), 205-214. <https://doi.org/10.1177/0266382112470412>

- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51-65. <https://doi.org/10.1016/j.tourman.2016.10.001>
- Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120-130. <https://doi.org/10.1016/j.ijhm.2014.10.013>
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036-1040. <https://doi.org/10.1073/pnas.1418680112>
- Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429-472. <https://doi.org/10.1177/1094428114562629>

TABLES AND FIGURES

Table 1: Descriptive Statistics of Articles Published

Description	Results
Sources (Journals)	285
Article	511
Review	27
Total Documents	538
Period	2014 - 2018
Average citations per documents	16.39
Authors of single-authored documents	96
Authors of multi-authored documents	1395
Single-authored documents	100
Author's Keywords (DE)	1703
Authors	1491
Author Appearances	1724
Documents per Author	0.361
Authors per Document	2.77
Co-Authors per Documents	3.2
Collaboration Index	3.18

Table 2: Journal-Article Frequencies

Total Articles	Journal Frequency
1	196
2	33
3	24
4	14
5	6
6	3
7	3
8	1
11	2
12	1
16	1
21	1
538	285

Table 3: Most Productive Journals

Journal	Total Articles	Total Citations	First Publication Year
<i>Digital Journalism</i>	21	364	2014
<i>International Journal of Communication</i>	16	232	2014
<i>Sustainability</i>	12	53	2016
<i>Information Communication & Society</i>	11	240	2014
<i>New Media & Society</i>	11	112	2016
<i>Computers Environment and Urban Systems</i>	8	113	2015
<i>Computers In Human Behavior</i>	7	110	2015
<i>Isprs International Journal of Geo-Information</i>	7	67	2015
<i>Journal of Communication</i>	7	431	2014
<i>Epj Data Science</i>	6	55	2015
<i>Plos One</i>	6	71	2016
<i>Technological Forecasting and Social Change</i>	6	43	2015
<i>Communication Methods and Measures</i>	5	24	2016
<i>Information & Management</i>	5	152	2015
<i>Journal of Advertising</i>	5	53	2017
<i>Journal of Medical Internet Research</i>	5	102	2014
<i>Social Science Computer Review</i>	5	57	2017
<i>Transportation Research Part C-Emerging Technologies</i>	5	111	2014
<i>Annals of The American Association of Geographers</i>	4	30	2016
<i>Decision Support Systems</i>	4	137	2014

Table 4: Most Productive 20 Authors

Authors	Articles Authors-Frac	Articles Fractionalized
Chen, Y.	7	Vargo, C.J. 2.117
Wang, Z.	7	Zhang, Y. 2.071
Ye, X.	7	Bail, C.A. 2.000
Vargo, C.J.	6	Carlson, M. 2.000
Eichstaedt, J.C.	5	Couldry, N. 2.000
Guo, L.	5	Jang, S.M. 2.000
Jang, S.M.	5	Papacharissi, Z. 2.000
Li, Q.	5	Uldam, J. 2.000
Poorthuis, A.	5	Wang, Z. 1.933
Zhang, Y.	5	Poorthuis, A. 1.917
Kern, M.L.	4	Ye, X. 1.833
Kosinski, M.	4	Li, Q. 1.783
Li, J.	4	Guo, L. 1.733
Liu, Y.	4	Chae, B.K. 1.667
Schwartz, H.A.	4	Choi, T.M. 1.500
Seligman, M.E.P.	4	Lewis, S.C. 1.500
Ungar, L.H.	4	Marine-Roig, E. 1.500
Wang, J.	4	Neuman, Wr. 1.500
Wang, X.	4	Chen, Y. 1.467
Wang, Y.	4	Ch'ng, E. 1.450

Table 5: Most Productive 20 Authors' Impact

Author	H-index	G-index	M-index	Total Citations	Number of Papers	First Publication Year
Poorthuis, A.	5	5	0.833	203	5	2014
Vargo, C.J.	5	6	1.250	83	6	2016
Kosinski, M.	4	4	0.667	340	4	2014
Eichstaedt, J.C.	4	5	0.667	273	5	2014
Schwartz, H.A.	4	4	0.667	271	4	2014
Ungar, L.H.	4	4	0.667	271	4	2014
Kern, M.L.	4	4	0.667	258	4	2014
Seligman, M.E.P.	4	4	0.667	258	4	2014
Jang, S.M.	4	5	0.667	223	5	2014
Guo, L.	4	5	0.800	86	5	2015
Ye, X.	4	7	1.000	64	7	2016
Wang, Y.	4	4	1.000	47	4	2016
Chen, Y.	4	5	1.333	28	6	2017
Wang, X.	3	4	0.600	116	4	2015
Wang, Z.	3	7	0.500	66	7	2014
Li, Q.	3	3	1.000	16	5	2017
Wang, J.	2	4	0.500	18	4	2016
Zhang, Y.	2	3	0.667	15	5	2017
Liu, Y.	2	3	0.667	14	4	2017
Li, J.	2	3	0.500	10	4	2016

Table 6: Paper Frequencies of Countries

Country	Articles	Country	Articles
USA	196	Ireland	4
China	62	Turkey	4
United Kingdom	61	New Zealand	3
Australia	23	Portugal	3
Germany	23	Romania	3
Italy	18	Singapore	3
Canada	17	Austria	2
Netherlands	16	Poland	2
Korea	13	South Africa	2
Spain	12	Switzerland	2
Finland	8	Chile	1
France	6	Lebanon	1
Israel	6	Malaysia	1
Denmark	5	Mexico	1
Greece	5	Nigeria	1
Norway	5	Pakistan	1
Sweden	5	Russia	1
Taiwan	5	Saudi Arabia	1
Belgium	4	Serbia	1
Brazil	4	Slovenia	1
India	4	---	-

Table 7: Most Frequent 50 Author Keywords

Words	Occurrences	Words	Occurrences
big data	257	social network analysis	7
social media	157	user generated content	7
twitter	66	computer mediated communication	6
sentiment analysis	24	methodology	6
machine learning	23	natural language processing	6
social networks	19	online news	6
data mining	17	social media analysis	6
content analysis	14	analytics	5
facebook	14	artificial intelligence	5
big data analytics	12	china	5
internet	11	computational journalism	5
text mining	11	customer relationship management	5
computational social science	10	data visualization	5
algorithms	9	decision making	5
data journalism	9	deep learning	5
privacy	9	disaster management	5
surveillance	9	journalism	5
data analytics	8	media	5
online reviews	8	power	5
social media analytics	8	social network	5
technology	8	topic modeling	5
crowdsourcing	7	urban planning	5
ethics	7	big data analysis	4
research methods	7	communication	4
social big data	7	data	4

Table 8: Top 20 Most Cited Papers in Research List

Paper	Total Citations	Total Citations per Year
Kramer, Guillory and Hancock, (2014)	713	118.83
Bello-Orgaz, Jung, and Camacho (2016).	215	53.75
Youyou, Kosinski and Stillwell (2015).	191	38.2
Xiang et al. (2015)	185	37
Colleoni, Rozza and Arvidsson (2014)	183	30.5
Russell Neuman et al. (2014)	140	23.333
Eichstaedt et al. (2015)	137	27.4
Leeflang et al. (2014)	115	19.167
Chang, Kauffman and Kwon (2014)	107	17.833
Shelton, Poorthuis and Zook (2015)	105	21
Coddington (2015)	101	20.2
Xiang et al. (2017)	100	33.333
Chae (2015)	99	19.8
Park et al. (2015)	95	19
Hasan and Ukkusuri (2014)	82	13.667
Marine-Roig (2017)	80	16
Rust and Huang (2014)	78	13
Shelton et al. (2014)	74	12.333
Golder and Macy (2014)	73	12.167
Conway, Kenski and Wang (2015)	72	14.4

Figure 1: Total Yearly Production of Big Data and Communication Articles

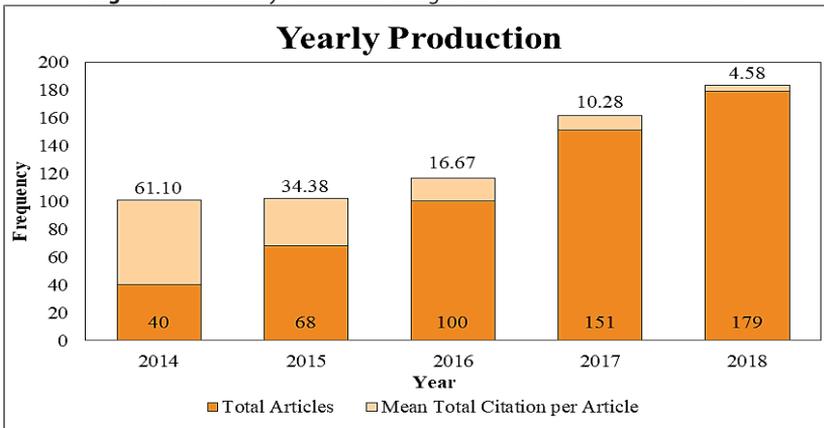


Figure 2: Top 20 Authors' Production over Time

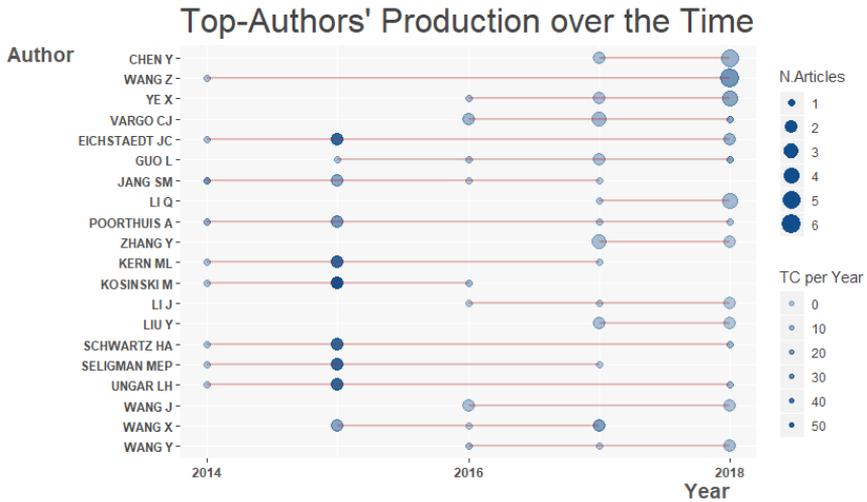


Figure 3: Corresponding Author's Country

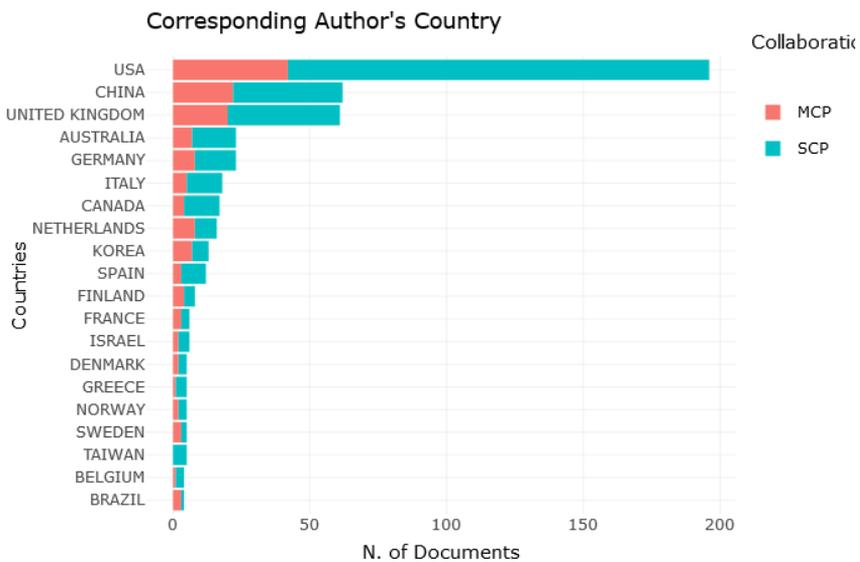


Figure 4: Country Collaboration Network

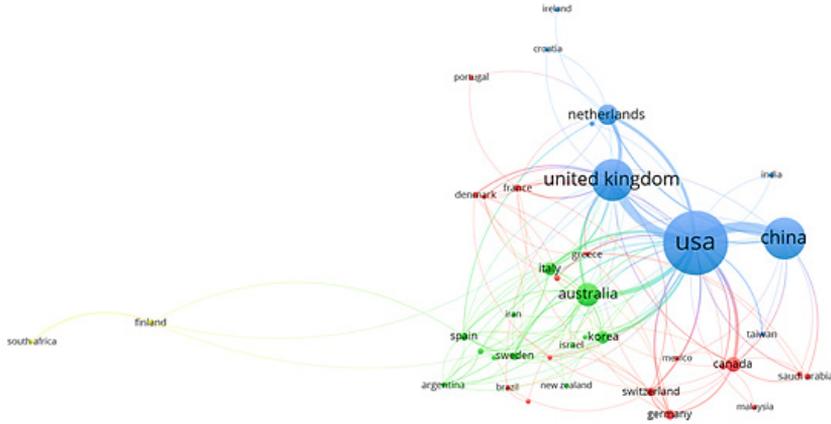


Figure 5: Trend Topic Analysis

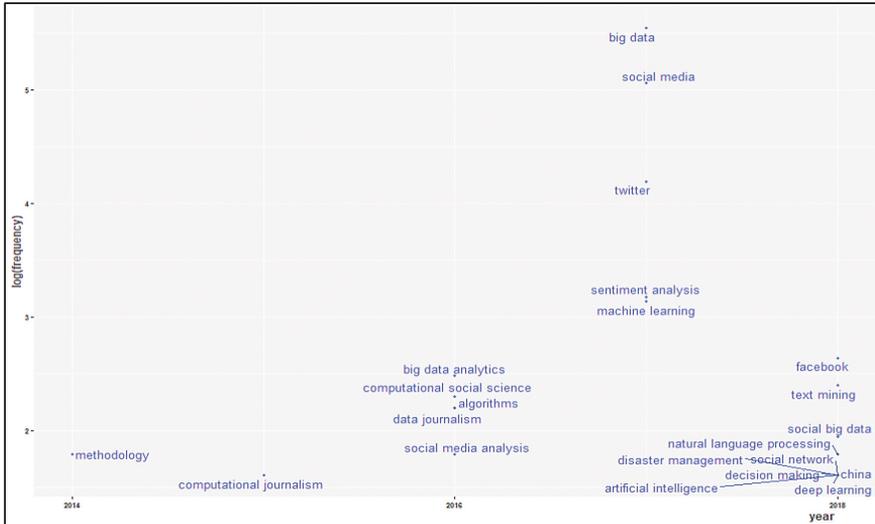


Figure 6: Co-word Analysis

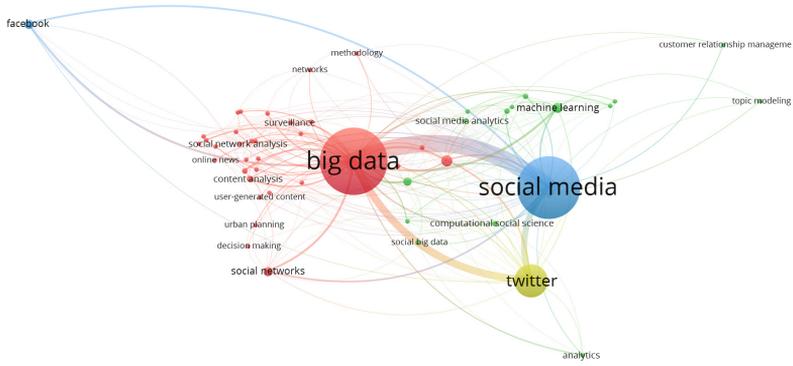


Figure 7: Co-citation Network

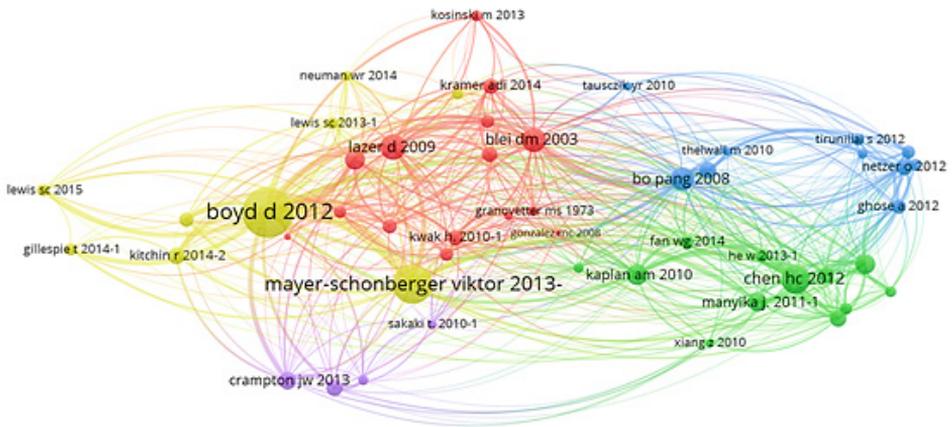


Figure 8: Journal Co-citation Network

