COMPUTATION METHODS AND APPLICATIONS IN COMMUNICATION RESEARCH: MAPPING THE CONCEPTUAL, SOCIAL AND INTELLECTUAL STRUCTURE

PABLO SÁNCHEZ-NÚÑEZ
Universidad de Málaga

GUSTAVO VACCARO
Universidad de Málaga

ANDREA PELÁEZ-REPISO
Universidad de Málaga

FRANCISCO E. CABRERA
Universidad de Málaga

1. INTRODUCTION

Communication research supported by computational techniques is a relatively novel scientific field that aims to deepen the understanding of communication phenomena by taking advantage of computational procedures to process vast amounts of data, find patterns in heterogeneous datasets of raw text and analyse complex biometric information. Furthermore, in many cases, studies within this field involve advanced computational techniques that require specialized training and equipment (Hu et al., 2014; Hussain & Howard, 2013; Jackson & Foucault Welles, 2015; Waldherr et al., 2021).

Computational methods include (but are not limited to) methods such as text analysis, semantic/social network analysis, online experiments, machine learning, visual analysis, and agent-based modelling and simulations. Computational methods can be applied to "big data" and social or behavioural (online) data, but can also be used to provide a more

insightful understanding of "small data" or for theoretical explorations (Freelon, 2010; Granovetter, 1978; Habermas, 2006; Hargittai, 2015).

Computational methods have the potential to greatly enhance the scientific study of communication because they allow us to move towards collaborative large-N studies of actual behaviour in its social context. This requires us to develop new skills and infrastructure and meet the challenges of open, valid, reliable, and ethical "big data" research (Andersen, 2018; Bennett et al., 2018; Broido & Clauset, 2019; Centola, 2013; Choi, 2020; Conte et al., 2012; Couldry & Hepp, 2013; Waldherr et al., 2021).

This strong focus on multidisciplinary causes the related scientific and academic literature to be spread across various sources of different nature, each one specialized in different fields such as social sciences, communication sciences, computational sciences, engineering, and mathematics. Therefore, the analysis of the relevant literature in this field is not an easy task, and a clear vision of the spectrum of communication research involving computational methods is needed. In this regard, we explore the possibility of using informetric analysis to address this issue (Burrell, 2007; Cobo et al., 2011; Daraio & Glänzel, 2020).

Informetric research studies the quantitative aspects of information, from the production to the dissemination and usage. It considers a broad spectrum of information sources, not only those compiled in bibliographic records, but also covers all aspects of formal or informal, oral, or written communication, that is, regardless of the form in which it is recorded and how it is generated (Bornmann et al., 2012; Daraio & Glänzel, 2020; Moed, 2017). Currently, no work or project includes a broad analysis of this type that focuses on computation methods and applications in Communication Research.

The knowledge provided by an informetric analysis provides support for further research and development of scientific fields, the improvement in the formulation of scientific evaluation, research and funding policies, the establishment of R&D&I agreements and the creation of synergies between authors, countries, and institutions. In addition, it

provides insight into what, who, how, when, and why a field of research is enhanced and developed (Mingers & Leydesdorff, 2015; Sooryamoorthy, 2020).

2. OBJECTIVES

The book chapter is divided into an analysis of the knowledge structures or k-structures (conceptual, social, and intellectual) of communication. In this sense, science mapping analysis, the use of tools such as Bibliometrix will be analysed and used as a methodology, and the Web of Science database will be used as a source of information (period of study from 2011-2020) to find out who are the most relevant authors in non-profit communication, which institutions produce the most notable research in non-profit communication, which countries stand out in this field of knowledge, which documents are the most highly cited papers, and the evolution and key concepts of research in non-profit communication will be analysed. Finally, the chapter will deal with social relations and an analysis of international collaboration between authors, institutions and countries working in Computation methods and applications in Communication Research.

3. MATERIALS AND METHODS

3.1. Bibliographic database

The source of the bibliometric data was the Web of Science (WoS) database. WoS, owned by Clarivate Analytics, is a collection of databases of bibliographic references and citations from periodicals that gathers information from 1900 to the present. The choice of WoS as the data source was based on two main characteristics of the database: it offers highly accurate and reliable research information, and it provides numerous analysis tools for processing the data.

3.2. Delimitation of the research area, data querying, data acquisition and data filtering

The search was conducted in September 2021. A specific search query was formulated to delimit the research area under study and collect those academic publications related to computational social science methods in communication research indexed in WoS from 2011 to 2020. Table 1 shows the query design, indexes, timespan, and dataset download date.

TABLE 1. *Search equation.*

Indexes	Timespan	Query	Documents	Download
Web of Science Core Collection: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.	2011-2020	((((((((((((((((((((TS=(Modelling and Simulation in Social Sciences)) OR TS=(Computational Economics)) OR TS=(Quantitative methods and their application in Social Sciences)) OR TS=(Spatial statistics)) OR TS=(Software tools and algorithms for statistical data processing)) OR TS=(Statistical methodology for censuses and surveys)) OR TS=(Agent-based modelling)) OR TS=(Analysis of social networks)) OR TS=(Computers and Education)) OR TS=(Rough Systems)) OR TS=(Database and Data Mining in Social Sciences)) OR TS=(Distributed and Parallel Systems & Algorithms)) OR TS=(Grid and Scalable Computing)) OR TS=(Mobile Computing)) OR TS=(Cloud Computing)) OR TS=(Soft Computing with applications in Social Sciences)) OR TS=(Fuzzy Systems)) OR TS=(Neural Networks)) OR TS=(Artificial Intelligence)) OR TS=(Web Analytics and Internet Computing)) AND WC=(Communication)	3288	10.09.2021

Source: Own elaboration.

3.3. Science Mapping Analysis Tool: Bibliometrix R Package

To perform the analysis, a scientometric tool known as Bibliometrix was used first. Bibliometrix version 3.0.5 is an open-source tool, programmed in R (Aria & Cuccurullo, 2017). This online tool is designed for informetrics research and includes several bibliometric methods of analysis: co-citation, coupling and scientific collaboration. To perform the descriptive analysis and study the scientific production, we imported the dataset of computational methods of social sciences in communication research into the Bibliometrix package. We used Bibliometrix intending to study the main knowledge structures of scientific research (social, intellectual, and conceptual structure). The authors

used the informetric visualization tool Biblioshiny R package (https://bibliometrix.org/Biblioshiny.html), a web interface to Bibliometrix R Package.

4. RESULTS

4.1. MAIN INFORMATION ABOUT THE DATASET

Of the 3288 documents analysed in the research field, from 2011 to 2021, 2650 (80.5%) were articles, followed by 332 (10.1%) proceeding papers and 138 (5.2%) book chapters. Furthermore, the average number of citations per document was 4.23, accounting for 1.878 average citations per year. There were 6176 different authors; in this regard, it is interesting to notice that 1026 documents, almost a third of the database, were single-authored. The total number of references accumulated by all the documents included in this research was 109327. Table 2 summarizes the main descriptive information about the dataset of analysed documents.

TABLE 2. *Main information about the dataset.*

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2011:2021
Sources (Journals, Books, etc)	360
Documents	3288
Average years from publication	4.23
Average citations per document	11.05
Average citations per year per doc	1.878
References	109327
DOCUMENT TYPES	
article	2650
article; book chapter	138
article; early access	52
article; proceedings paper	18
book	4

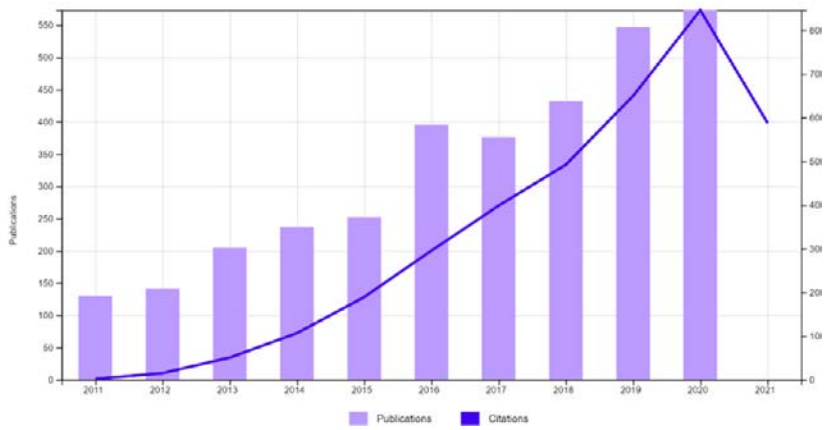
book review	12
book review; early access	1
correction	1
editorial material	31
editorial material; book chapter	9
editorial material; early access	1
proceedings paper	332
review	38
review; early access	1
DOCUMENT CONTENTS	
Keywords Plus (ID)	2782
Author's Keywords (DE)	7995
AUTHORS	
Authors	6176
Author Appearances	7442
Authors of single-authored documents	951
Authors of multi-authored documents	5225
AUTHORS COLLABORATION	
Single-authored documents	1026
Documents per Author	0.532
Authors per Document	1.88
Co-Authors per Documents	2.26
Collaboration Index	2.31

Source: Own elaboration.

4.2. Distribution of publications by year and record count /Average citations per year and record count

Figure 1 shows the document production and citations per year in the analysed period (2011-2021). 2020 was the year with both the highest document production and several citations, followed by 2019 and 2018. In this regard, the number of documents produced each year as well as the citations has increased steadily for the past 10 years.

FIGURE 1. *Distribution of publications/citations by year and record count (2011-2020).*



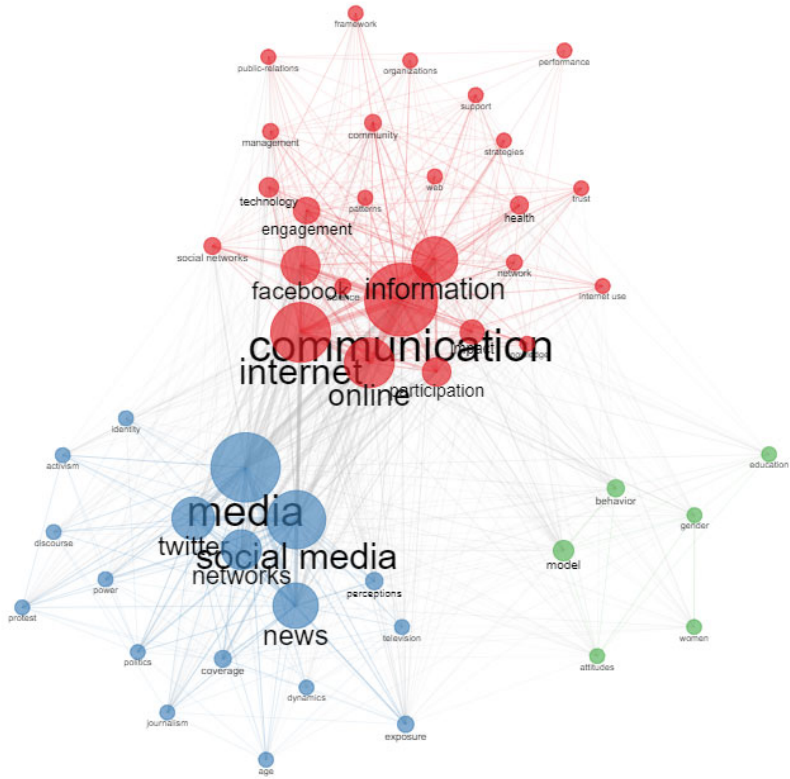
Source: Own elaboration.

4.3. Knowledge Structures: Conceptual Structure, Intellectual Structure, and Social Structure

4.3.1. Conceptual Structure (Co-occurrence Network)

The co-occurrence network is visually represented in Figure 1. The analysis identified 3 clusters that shared significant conceptual structure, ordered by size. The first cluster of the network is related to studies that involve online communication and information; and is characterized by the terms: communication, information, internet, online, Facebook and participation. The second cluster is related to social media and social network studies and is characterized by the terms: media, social media, news, networks, and Twitter. The third cluster is related to behaviour models and gender studies, characterized by the terms: model, behaviour, gender, and women.

FIGURE 2. *Co-occurrence Network. Field: Keyword Plus, Network Parameters: Hide Network Parameters and Graphical Parameters: Hide Graphical Parameters.*

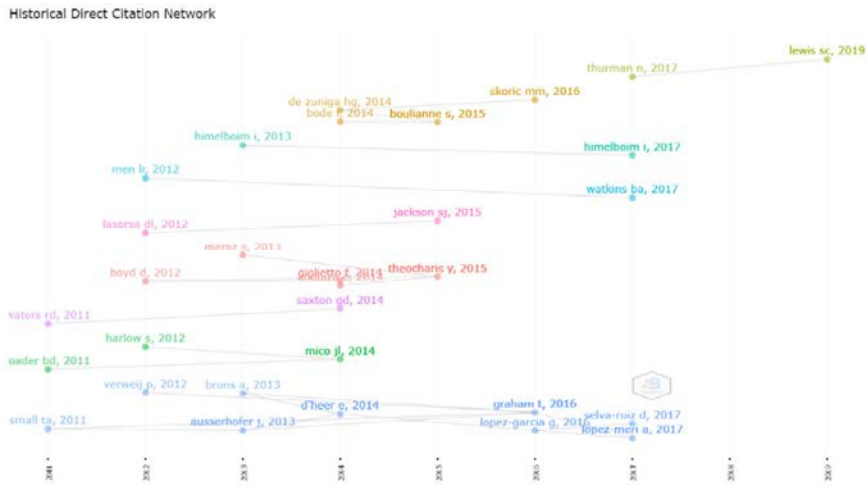


Source: Own elaboration.

4.3.2. Intellectual structure (Historiography: Historical Direct Citation Network)

Figure 3 shows a graphical representation of the Historical Direct Citation Network. The results of the analysis suggest that the citation network can be traced back to Small TA, Oader BD and Vaters RD (2011). The work of Small TA (2011) exposes the largest direct citation network that can be traced up to the documents published by Selva-Ruiz D and Lopez-Mena A (2017).

FIGURE 3. *Historiography. Number of Nodes:50, Graphical Parameters: Short-id (1st author, Year), Label size: 4 and Node size: 2.*

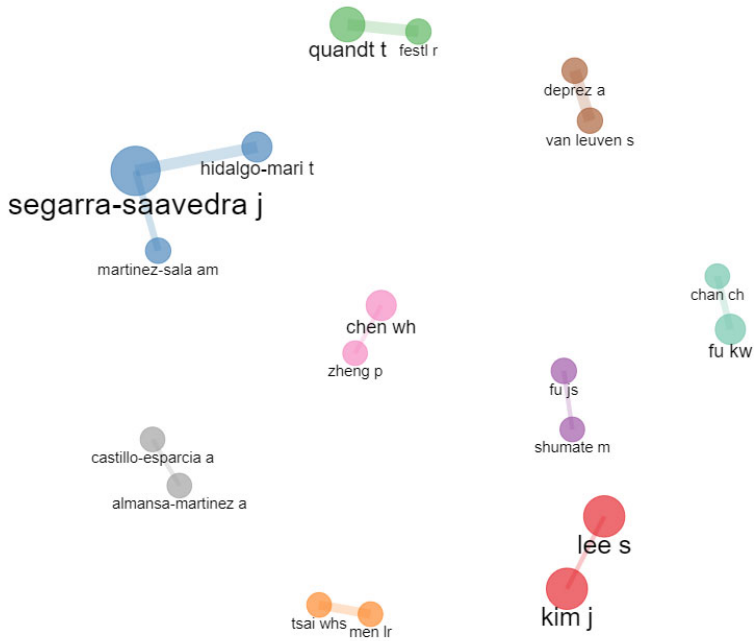


Source: Own elaboration.

4.3.3. Social Structure (Collaboration Network: authors, countries and institutions)

Figure 4 shows a structure of 9 clusters of the author's collaborations. A small cluster with the greatest impact in terms of scientific publications is the one led by Segarra-Saavedra, J collaborating with Martínez-Sala, A.M. and Hidalgo-Marí, T. In the rest of the collaborations there are only two members, so we could not consider them as solid clusters.

FIGURE 4. *Collaboration Network: authors. Network Parameters: Hide Network Parameters and Graphical Parameters: Hide Graphical Parameters.*

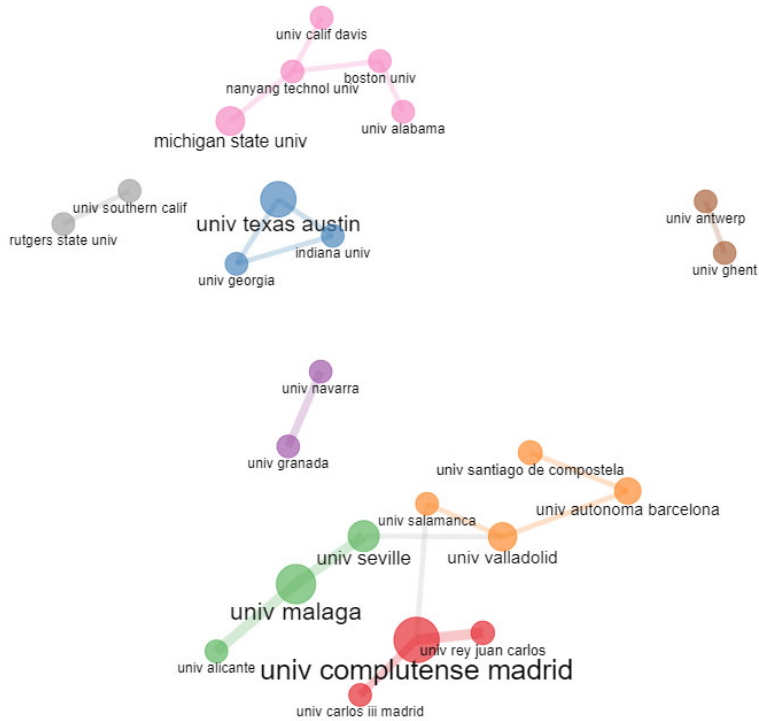


Source: Own elaboration.

Figure 5 shows a total set of 8 clusters of institutions collaboration. The most relevant cluster is the one in red, led by the Complutense University of Madrid, followed by the Rey Juan Carlos University and the Carlos III University of Madrid with close co-authorships. This cluster is joined by the connections of the University of Malaga with working groups related to the University of Alicante and the University of Seville. To this last Spanish collaboration cluster, there is an orange cluster where there are collaboration groups between the University of Salamanca, the University of Valladolid, the Autonomous University of Barcelona, and the University of Santiago de Compostela. There is a Spanish satellite collaboration cluster between the University of Granada and the University of Navarra. The rest of the collaborations on

the map show co-authorships between Belgian institutions (the University of Antwerp and the University of Ghent), and the rest are three collaborative clusters of U.S. institutions and Singapore.

FIGURE 5. *Collaboration Network: institutions. Network Parameters: Hide Network Parameters and Graphical Parameters: Hide Graphical Parameters.*

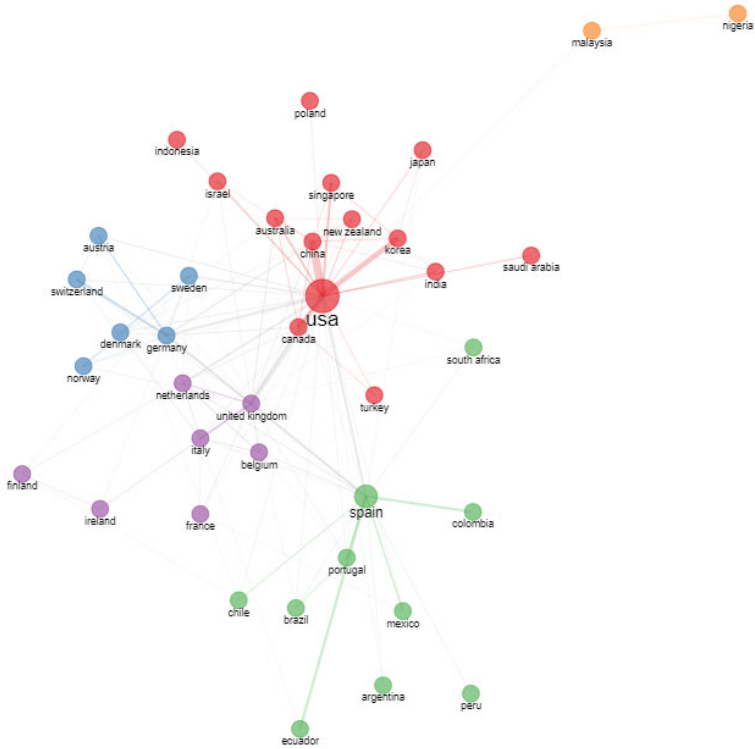


Source: Own elaboration.

Figure 6 shows the countries where the documents were originated and the collaborations between them. The countries with the most related documents were the United States of America and Spain. Furthermore, there are 5 clusters within the collaboration network with significant cohesion. The first cluster is represented by the United States of America, where its closest collaborators were Canada, China, Australia, and Korea. The second cluster is represented by Spain, where its closest

collaborators were Portugal, Colombia, and Mexico. The third cluster is represented by Germany, where its closest collaborators were Denmark and Sweden. The fourth cluster is represented by the United Kingdom, where its closest collaborators were Belgium, The Netherlands and Italy. Finally, the fifth cluster is composed of Malaysia and Nigeria.

FIGURE 6. *Collaboration Network: countries. Network Parameters: Hide Network Parameters and Graphical Parameters: Hide Graphical Parameters.*



Source: Own elaboration.

5. DISCUSSION AND CONCLUSION

This book chapter mapped the conceptual, social, and intellectual structure of computation methods and applications in communication research. Research in computational methodologies applied to communication is an emerging and growing field, as demonstrated by the analysis of scientific production and the distribution of publications over the last decade.

- The scientific production has practically quadrupled if we compare the first year of the study (2011) and the last (2020). This production growth has been due to the improvement and appearance of emerging technologies that have facilitated research in computing, as well as the evolution and expansion of what we know as the information and knowledge society and its impact on the communication industry. This fact, together with the democratization of access to digital technology and the increase in research in artificial intelligence, data processing and analysis and specialization in communication has led to this increase in publications, among others.

- The conceptual structure has been given by the analysis of co-occurrence of keywords, which has allowed the identification of 3 large groups of terms. These are headed by the two main clusters, red and blue, followed by green. On the one hand, the red cluster seems to be determined by communication research and its impact on social networks and the internet, as well as participation, public relations research and research on strategy and organizations. The blue cluster, on the other hand, focuses on media, news, perception and television, politics or activism. Finally, the green cluster, without a relevant presence and with a rather diluted appearance, we find that handles terms such as attitude, behaviour, women, gender, education, as well as the design of models and frameworks.

- The analysis of the intellectual structure, obtained through the historical chains of citation or historiography, has allowed us to identify the most notorious authors as well as to discover

who are the most relevant actors in this field of knowledge and who were the precursors of this discipline. On the one hand, we see that Vaters, Oader and Small are the three authors who have distinct lines that converge at different points over the years, and we can see the citation chains derived from what we could say their lines of research. We note, for example, that Oader's line comes to a halt in 2014, as does that of Vaters, however, the one initiated by Vaters continues until practically the present day. Producing some very interesting author citation clusters. Other emerging lines emerge such as the one initiated by Thurman in 2017 and finished by Lewis in 2019.

– The social structure analyzed through the collaboration networks of authors countries and institutions has allowed us to identify the most relevant actors and the relationships established between them. On the one hand, we observe that although there is a thematic network on communication research through computational methodologies, this network is still very fragmented, and we only see small clusters of two or three authors. Research in this area, although incipient, is not yet consolidated. We found many people working on the subject, but with a rather reduced collaboration, being therefore somewhat hermetic groups. We found that the common nexus of collaboration between authors is language (groups of Spanish, Chinese or Belgian researchers, among others). Something similar happens with the analysis of institutions, where we find well-differentiated networks of universities that share the same working language, as well as belonging to the same country. We note that the Spanish cluster is the largest, since it brings together almost 4 collaborative sub-networks, and even within these sub-networks there are collaborations of interest or territorial proximity. Universities close to each other are the ones that collaborate the most. Behind this large subgroup of networks, we find another cluster of Belgian universities, with only two institutions and a large U.S. cluster with satellite collaborations with Singapore.

– To the cluster of collaboration between countries, we note that there are specific collaborations, although they are not yet consolidated. As in the previous networks, we find that the main link of collaboration and the point of union is language since we find the green cluster that we can call Ibero-American, the violet cluster that would be a cluster of Western European collaboration, another of Baltic collaboration and another large red cluster of American collaboration with collaborations with Turkey, Canada, Australia, India, Saudi Arabia, Israel, and Indonesia. We could say that this last cluster is the most diverse and dynamic with collaborations with practically all continents.

The results suggest that the last decade exhibited an increasing trend in synergies between computation and communication sciences, demonstrated by the number of documents published per year. Moreover, the interest of the scientific community in this field has also increased with a similar trend, reaching its current peak in 2020. Therefore, it is safe to assume that this tendency will remain in the following years.

Furthermore, the insights provided by the co-occurrence network analysis proved that the internet, social media, and online communities are driving factors of this research field. It would be very interesting, for future lines of research, to investigate these relationships through other bibliometric indicators or with alternative metrics, delving deeper into the type of studies developed by the different actors that have defined and redefined this field of knowledge.

6. REFERENCES

Andersen, J. (2018). Archiving, ordering, and searching: search engines, algorithms, databases, and deep mediatization. *Media, Culture & Society*, .. (8), 1135–1150. https://doi.org/10.1177/0163443718754652

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

Bennett, W. L., Segerberg, A., & Yang, Y. (2018). The Strength of Peripheral Networks: Negotiating Attention and Meaning in Complex Media Ecologies. *Journal of Communication*, *68*(4), 659–684. https://doi.org/10.1093/joc/jqy032

Bornmann, L., de Moya Anegón, F., & Leydesdorff, L. (2012). The new Excellence Indicator in the World Report of the SCImago Institutions Rankings 2011. *Journal of Informetrics*, *6*(2), 333–335. https://doi.org/10.1016/j.joi.2011.11.006

Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, *10*(1), 1017. https://doi.org/10.1038/s41467-019-08746-5

Burrell, Q. L. (2007). On the h-index, the size of the Hirsch core and Jin's A-index. *Journal of Informetrics*, *1*(2), 170–177. https://doi.org/10.1016/j.joi.2007.01.003

Centola, D. M. (2013). Homophily, networks, and critical mass: Solving the start-up problem in large group collective action. *Rationality and Society*, *25*(1), 3–40. https://doi.org/10.1177/1043463112473734

Choi, S. (2020). When Digital Trace Data Meet Traditional Communication Theory: Theoretical/Methodological Directions. *Social Science Computer Review*, *38*(1), 91–107. https://doi.org/10.1177/0894439318788618

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

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J.-P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., & Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, *214*(1), 325–346. https://doi.org/10.1140/epjst/e2012-01697-8

Couldry, N., & Hepp, A. (2013). Conceptualizing Mediatization: Contexts, Traditions, Arguments. *Communication Theory*, *23*(3), 191–202. https://doi.org/10.1111/comt.12019

Daraio, C., & Glänzel, W. (Eds.). (2020). *Evaluative Informetrics: The Art of Metrics-Based Research Assessment*. Springer International Publishing. https://doi.org/10.1007/978-3-030-47665-6

Freelon, D. G. (2010). Analyzing online political discussion using three models of democratic communication. *New Media & Society*, *12*(7), 1172–1190. https://doi.org/10.1177/1461444809357927

Granovetter, M. (1978). Threshold Models of Collective Behavior. *American Journal of Sociology*, *83*(6), 1420–1443. https://doi.org/10.1086/226707

Habermas, J. (2006). Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research. *Communication Theory*, *16*(4), 411–426. https://doi.org/10.1111/j.1468-2885.2006.00280.x

Hargittai, E. (2015). Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 63–76. https://doi.org/10.1177/0002716215570866

Hu, H., Cui, W., Lin, J., & Qian, Y. (2014). ICTs, Social Connectivity, and Collective Action: A Cultural-Political Perspective. *Journal of Artificial Societies and Social Simulation*, *17*(2). https://doi.org/10.18564/jasss.2486

Hussain, M. M., & Howard, P. N. (2013). What Best Explains Successful Protest Cascades? ICTs and the Fuzzy Causes of the Arab Spring. *International Studies Review*, *15*(1), 48–66. https://doi.org/10.1111/misr.12020

Jackson, S. J., & Foucault Welles, B. (2015). Hijacking #myNYPD: Social Media Dissent and Networked Counterpublics. *Journal of Communication*, *65*(6), 932–952. https://doi.org/10.1111/jcom.12185

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, *246*(1), 1–19. https://doi.org/10.1016/j.ejor.2015.04.002

Moed, H. F. (2017). *Applied Evaluative Informetrics*. Springer International Publishing. https://doi.org/10.1007/978-3-319-60522-7

Sooryamoorthy, R. (2020). *Scientometrics for the Humanities and Social Sciences*. Routledge. https://doi.org/10.4324/9781003110415

Waldherr, A., Geise, S., Mahrt, M., Katzenbach, C., & Nuernbergk, C. (2021). Toward a Stronger Theoretical Grounding of Computational Communication Science. *Computational Communication Research*, *3*(2), 1–28. https://doi.org/10.5117/CCR2021.02.002.WALD