Research into the behaviour of minors and young people on social networks using Social Big Data techniques

Investigación del comportamiento de menores y jóvenes en las redes sociales mediante técnicas de Social Big Data

ISSN: 1696-019X / e-ISSN: 2386-3978

Rebeca Suárez-Álvarez. Assistant Professor at University Rey Juan Carlos (Department of Communication and Sociology). She also holds a PhD in Social Communication from CEU San Pablo University, as well as a Bachelor's Degree in Journalism from Complutense University of Madrid (*UCM*). In addition, Professor Suárez has completed two master's degrees: one in Radio from CEU San Pablo University and another in Communication of Public Institutions and Policies from *UCM*. Her main lines of research are communication and vulnerable audiences (specifically minors), digital media, media literacy and digital competence. She has participated in diverse research projects, and is currently working as a researcher on the project entitled, "New scenarios of digital vulnerability: media literacy for an inclusive society" (*PROVULDIG-2-CM*) (ref. H2019/HUM5775), funded by the Autonomous Region of Madrid (CAM) and the European Social Fund. She is also a member of the research group focused on Communication, Society and Culture (*GICOMSOC*).
Universidad Rey Juan Carlos, Madrid, Spain
rebeca.suarez@urjc.es
ORCID: 0000-0002-0102-4472

Antonio García-Jiménez. Journalism Chair at University Rey Juan Carlos (Department of Communication and Sociology). Professor García-Jiménez holds a PhD in Information Sciences, and is the former Dean of the Faculty of Communication Sciences at University Rey Juan Carlos (2008-2014). In addition, he has held the position of director of the Master's Degree in Communication and Sociocultural Problems (2015-2018), and has served as a Lecturer for the Bachelor's Degree in Journalism programme, as well as several Master's Degree Programmes in Communication (Social Media, Data Journalism). Professor García-Jiménez also holds the post of Principal Investigator of the research group focused on communication, society, and culture (*GICOMSOC*). He has led or participated in 19 research projects related to digital use and risk for adolescents, and is the author of a large number of influential publications, both national and international.
Universidad Rey Juan Carlos, Madrid, Spain
antonio.garcia@urjc.es
ORCID: 0000-0002-8423-9486

Received: 15/09/2020 - Accepted: 07/04/2021 - Early access: 05/05/2021 - Published: 14/06/2021

Recibido: 15/09/2020 - Aceptado: 07/04/2021 - En edición: 05/05/2021 - Publicado: 14/06/2021

Abstract:

The aim of this study is to identify the academic research related to the behaviour and communication consumption of minors and young people on the internet through the use of a Social Big Data (SBD) methodology. A systematic review has identified 58 scholarly works published between 2010 and 2020 (May). This article has

Resumen:

El objetivo es identificar la producción científica sobre el comportamiento y consumo comunicativo de los menores y los jóvenes en internet utilizando una metodología Social Big Data (SBD). Mediante una revisión sistemática se han identificado 58 documentos académicos publicados entre 2010-2020 (mayo). Se compendian las

How to cite this article:
Suárez-Álvarez, R. y García-Jiménez, A. (2021). Research into the behaviour of minors and young people on social networks using Social Big Data techniques. *Doxa Comunicación*, *32*, pp. 95-113.

https://doi.org/10.31921/doxacom.n32a5

summarised the most widely studied aspect of the issue, as well as the countries that have conducted the largest amount of research, the academic profile of the journals, research techniques based on SBD, and the most relevant findings. The main conclusions of the study are that scientists currently use SBD to understand the uses and effects of adolescent online activity by analysing large amounts of data in real time, and also to create algorithms that make it possible to identify trends in adolescent risk. It has been confirmed that there is a scarce amount of academic research related to this topic in social science journals, and there is also a shortage of an association and co-authorship between communicologists and sociologists, on the one hand, and scientists with a technical background on the other hand, which is necessary in order to achieve a greater understanding of the situation and increase the number of publications related to this issue in the area of social science.

Keywords:

Social Big Data; adolescents; young people; datafication: social networks.

dimensiones más investigadas, los países con mayor producción científica, el perfil académico de las revistas, las técnicas de investigación basadas en SBD y los hallazgos más relevantes. Las principales conclusiones son que los científicos están utilizando el SBD para, mediante gran cantidad de datos analizados en tiempo real, conocer los usos y efectos de las acciones de los adolescentes en la red así como para crear algoritmos que posibiliten la identificación de tendencias de riesgo adolescente. Se confirma que la producción científica es escasa en revistas de ciencias sociales y la necesaria asociación y coautoría de comunicólogos y sociólogos con científicos de perfil técnico para lograr mayor compresión de la realidad e incrementar las publicaciones en ciencias sociales.

Palabras clave:

Social Big Data; adolescentes; jóvenes; datificación; redes sociales.

ISSN: 1696-019X / e-ISSN: 2386-3978

1. Introduction

This article presents a systematic review of published research on children's online behaviour using Big Data (BD) techniques. Depending on the research objective, BD techniques can be classified as descriptive, predictive, or prescriptive. Descriptive techniques aim to understand reality by means of techniques such as variation range, frequency tables, A/B tests, factorial and cluster analysis, and so on. Predictive techniques anticipate events (time series, regression techniques, neural networks, machine learning and deep learning, as well as algorithms used for boosting, such as XGBoost). Finally, prescriptive techniques attempt to find a recommendation by identifying cause-and-effect rules or optimisation algorithms (conditional probability methods, regression techniques, association rules, stochastic simulation and Monte Carlo methods, genetic algorithms, and spatial optimisation techniques).

In order to apply these techniques and be capable of converting the millions of data into information, scientists need specifically designed technology, some of which include the following: Cassandra, Extract Transform and Load, Hadoop, HBase, MapReduce, Linguistic Inquiry and Word Count (LIWC), Hamlet, WordStat, QDAMiner, or Python for automatic language analysis, among others (Pérez, 2015; Joyanes, 2016; Arcila-Calderón, Barbosa-Caro & Cabezuelo-Lorenzo, 2016; Hernández-Leal, Duque-Méndez & Moreno-Cadavid, 2017).

The aim of this paper is to delve deeper into this type of research and determine its potential for incorporation into the study of the communicative behaviour and consumption of young people on the internet, in order to complement the knowledge we have of this phenomenon.

2. State of the art

2.1. An approach to Big Data

As pointed out by Pérez (2016) and Schwab (2017), society is undergoing the "fourth industrial revolution", or "digital revolution", in which technological innovation and digital devices are changing the social paradigm. This revolution is characterised by the exponential increase in the amount of digital data being generated by individuals in real time automatically, routinely, and through various devices (Batty, 2013), all of which is known as Big Data (BD).

Paredes-Moreno (2015) points out that the sphere of activity of BD is related to information that cannot be adequately analysed by other means. Moreover, Arcila-Calderón, Barbosa-Caro and Cabezuelo-Lorenzo (2016) describe it in terms of immense flows and quantities of information, both structured and unstructured, to which computational tools and methods are applied for the purpose of extracting knowledge.

2.2. Big Data within the scientific paradigm of social science

The analysis of big data has emerged as a new communication paradigm that goes beyond being considered a purely technological factor, but instead it has acquired social, political and economic dimensions as well (Malvicino & Yogue, 2014). This has led to the perspective that digital technology is not only a by-product of social relations, but it also has a social and communicative role (Tapia, 2018) in the codification of data that should be considered as well.

From this point of view, the "fourth scientific paradigm" proposed by Hey, Tansley and Tolle (2009) has emerged, in which ICT and scientists have converged, leading to the discovery of a new research paradigm based on massive data-driven science (Bell, Hey & Szalay, 2009), through the gathering of information and techniques such as data mining that were unthinkable merely a decade ago.

This approach has sparked a debate on the boundaries and interrelationships between the method, the object of study, and the datum (Tapia, 2018). It will soon be necessary to rethink the key questions about knowledge construction and social research processes in order to categorise reality (Boyd & Crawford, 2012). Manovich (2011) proposes the way in which BD is changing research methodology in the social sciences and humanities as the world becomes more digital. Thus, new techniques will be needed to investigate, analyse and understand a large volume of data, the impact of which can be measured in qualitative and quantitative terms (Mayer-Schönberger & Cukier, 2013), and according to Cerezo (2015), this is leading to a type of science that is relevant for both the present and the future as well. Qin (2014) identifies the present day as the "Big Data era", which is now gaining relevance in scientific research, with implications for social theory itself. Along the same lines, Mayer-Schönberger and Cukier (2013) assert that BD is not only providing massive data. They believe it is also prompting three modifications in the social science research paradigm: greater availability and accessibility of data; greater acceptance of the levels of imprecision and disorder in the data; and the increased possibility of focusing more on correlations rather than seeking causality.

ISSN: 1696-019X / e-ISSN: 2386-3978

ISSN: 1696-019X / e-ISSN: 2386-3978

2.3. Social Big Data

The massive increase in data reveals the need for research methodologies to evolve in order to turn data into knowledge that is transferred to society. Social Big Data (SBD) (Manovich, 2011) has emerged from the possibility of quantifying social reality online. It focuses on the study of public content, expressions of taste, states or trends of opinion, topics discussed, profile descriptions, and interaction between people, among other aspects (Mayer-Schönberger & Cukier, 2013). According to Russell (2013), SBD offers the option of capturing the details of online communication through mobile devices and provides multiple possibilities for the analysis of data stored on websites.

It is the result of a convergence of three major areas, which are social media, data analysis, and big data (Bello-Orgaz, Jung & Camacho, 2016), and their fusion is generating a new discipline in which knowledge is generated as a result of processing and analysing the information and data derived from social networks (Gualda & Rebollo, 2020). Through the use of SBD, it is possible to analyse and discover the knowledge held by applications and web services that are not limited to text messages, but also include sound, image and video as well (Kumar, Sangwan & Nayyar, 2020).

2.4. Criticism of Big Data use in social science

This techno-optimist view of applying BD to research is viewed with scepticism by social scientists as well. López (2018) points out that the promise of knowledge enhancement and a paradigm shift is supported by the techno neo-positivist position promoted by Silicon Valley and driven by today's growing data processing capability. For their part, Boyd and Crawford (2012) and Taylor-Sakyi (2016), who do not question the benefits of big data management, argue that the assumptions and biases of BD need to be critically questioned. At the heart of the issue, we are confronting a myth built on the assumption that big data is some kind of superior intelligence model that guarantees more accurate and objective results (Boyd & Crawford, 2012; Martínez-Martínez & Lara-Navarra, 2014).

In the same vein, Pybus, Coté and Blanke (2015) question the faith in the datafication of society theorised by Mayer-Schönberger and Cukier (2013), as the former group consider that it lacks "plausibility resolution". They question whether data analysis through an algorithm necessarily produces valid information. On the other hand, van Dijck (2014) points out the ontological and epistemological problems encountered in working with data, due to the fact that its collection, processing, and analysis is not a neutral process. Moreover, O'Neil (2013) warns that data analysis must take into account three factors: who controls the data, how it is structured, and for what purpose.

Likewise, it must be understood that large-scale data collection requires care with regard to the representativeness of the data, as it can lead to erroneous decisions (UN Global Pulse, 2018; Uman, 2018), which is not foreign to the computational social science perspective (Törnberg & Törnberg, 2018). In addition, there are potential risks related to invasion of privacy and intimacy, curtailment of free expression, and impairment of the right to free access of information (Arellano, 2014; Shuijing, 2015).

2.5. Research regarding children, adolescents and young people on the internet

In this paper, SBD is used as the starting point for this research in order to investigate children's use of the Internet and social networks. Although use and consumption by this sector of the population is taken into consideration, the interest of this work goes beyond an approach focused merely on communicology and the social realm. The literature on children, adolescents and young people on the Internet is extensive. With regard to research in recent years, the EU Kids Online project (http://www.lse.ac.uk/media-and-ommunications/research/research-projects/eu-kids-online) and its director Sonia Livingstone (Livingstone, Mascheroni & Staksrud, 2018) are the leading exponents of studies focused on adolescents, along with many other national and international initiatives. The techniques commonly used have been surveys and focus groups. In addition to the risks and habits of use and consumption, issues such as parental control, education in the digital age, and related policies have been addressed, among others (Jiménez, Garmendia & Casado, 2018).

In turn, this is an object of study that involves other domains such as psychology (Alsehaima & Alanazi, 2018), health (Falzone, et al., 2017), and education (Giménez, Luengo & Bartrina, 2017), with a clear proliferation of literature in recent years. At the same time, this field not only allows multidisciplinary approaches, but requires them as well, which has provided the opportunity for fields of study to be blended, an example of which is Educommunication (Aparici & García, 2016; Nupairoj, 2016).

In this article, we have examined the characteristics of existing publications related to the use and behaviour of minors and young people on the Internet and social networks, which have used a methodology that can be linked to SBD. We have also analysed the contributions of this academic literature, and we will attempt to discover the potential methodological contribution of SBD to the study of the social and communicational behaviour of this sector of the population on the internet.

3. Methodology

This research has used bibliometric analysis (Díaz-Campo, 2015), which studies the flow of knowledge within a specific field over a period of time (Leung, Sun & Bai, 2017), through quantitative statistical analysis of the existing academic literature and the citations obtained (van Eck & Waltman, 2014). The idea is to identify and analyse the extant literature in three phases: literature search, extraction of the most relevant data, and analysis of the content addressed by the literature under study (Ramírez-Montoya & García-Peñalvo, 2018). To this end, this research aims to answer the following questions:

Q1) How many studies have examined the object of analysis? In which publications have they appeared? When have they appeared?

Q2) Which concepts and aspects of the research have been the most studied?

Q3) What is the knowledge area of the publications that approach the object of study?

Q4) Which SBD-based techniques have been applied, and what are the main samples analysed?

Q5) What are the main contributions of the research?

ISSN: 1696-019X / e-ISSN: 2386-3978

ISSN: 1696-019X / e-ISSN: 2386-3978

The documentary search was carried out in the WoS, Scopus, and IEEE Xplore databases, and by using the Google Scholar search engine for all scientific articles published between 2010 and 2020 (May), both in English and Spanish, in scientific journals (articles), reports, books (chapters), and various theses as sources of research.

The documentation strategy included the following search terms and their combinations with Boolean operators and/or the following terms: "young*", "children", "minor*", "youth", "child*", "childhood", "adolesc*" and "social big data". Documents without scientific conclusions have been discarded and, through hypertext, a snowball sampling research methodology was applied in order to locate more documents with the selected search references in the same databases (Wohlin, 2014). The register variables that have been used are the following:

- Complete reference of the document: title, year, author/s, journal/conference, keywords, abstract, and number of citations.
- Authors: name, country, and institution.
- Investigation: research hypothesis, objectives, variables studied.
- Results: findings, limitations, and recommendations of the authors.

Likewise, the SCImago Journal Rank database was used to access the descriptors of the areas and categories addressed by each publication, as well as to determine the quarter in which each one is positioned. The authors also used the Publish or Peish programme (http://www.harzing.com/pop.htm) and the Excel programme of the Microsoft Office package to classify and identify the information as well as to record each element of the sample.

After obtaining the results, we proceeded to their necessary filtering in which a preliminary review was carried out to eliminate false positives such as duplicates, as well as to find the extent to which the searches did not address the object of study, were focused on theoretical writings or application techniques, did not offer full texts, did not address issues related to adolescents, or bibliographic reviews and data that did not coincide with the purposes of the research. Subsequently, all the articles were read in order to ensure their object of study and corroborate the methodology and samples selected in order to analyse their content and answer the research questions posed.

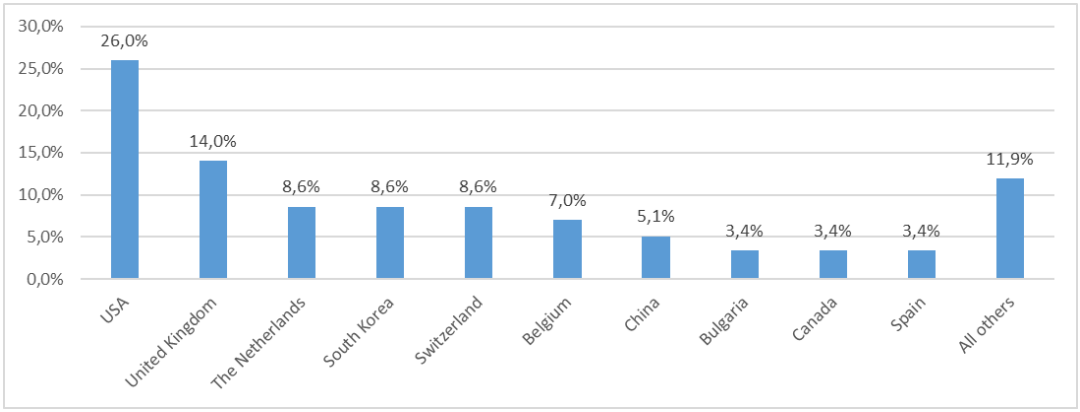
4. Results

4.1. Number of publications, sources, and dates

The final result of the bibliographic screening process offers 58 (N) results that make up the sample. A total of 63.8% (n=37) are journal articles, 12% (n=7) are book chapters, 10.3% are conferences (n=6), and theses and reports comprise 7% (n=4) each. Three journals stand out for their citations: *Yonsei Medical Journal* has 57 citations (2014), *Telecommunications Policy* has 46 (2015), and the *Journal of Adolescent Health* has 29 citations (2016). The United States is the country with the largest amount of scientific research (26%; n=15), followed by the United Kingdom (14%; n=8), and the Netherlands, South Korea and Switzerland all stand at 8.6% (n=5). Brussels comes in at 7% (n=4), China at 5.1% (n=3), and Bulgaria, Canada

and Spain at 3.4% (n=2) each. The remaining countries –Australia, Burma, Finland, India, New Zealand, Singapore and Vietnam– published 1.7% (n=1) (see Figure 1).

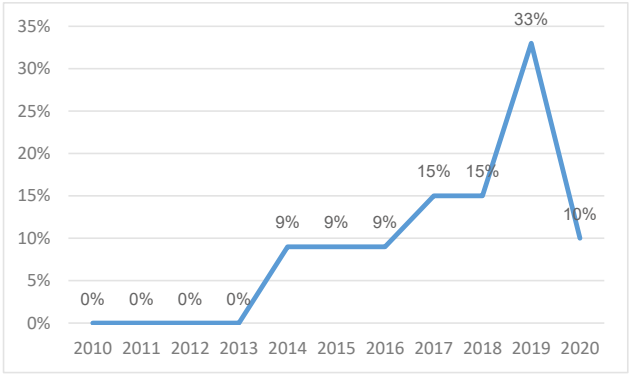
Figure 1. Countries with the largest amount of scientific production related to SBD and minors



Source: prepared by the authors

In relation to publications by year (see Figure 2), a temporary trend toward greater scientific interest in studying issues related to child-centred SBD can be observed, rising from 18% (n=10) of the articles between 2010 and 2015 to 82% (n=48) between 2016 and 2020.

Figure 2. Evolution of scientific production after refining the number of articles (2010-2020)



Source: prepared by the authors

ISSN: 1696-019X / e-ISSN: 2386-3978

ISSN: 1696-019X / e-ISSN: 2386-3978

Among the authors with the most scientific publications are researchers Tae Min Song (Korea Institute for Health and Social Affairs) and Juyoung Song (University of West Georgia, USA) who co-authored 18.6% (n=11) of the articles with other scientists from South Korea and the United States, with the most citations (n=57) having been achieved from their research entitled "*Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends*" (2014), and "*Data mining of web-based documents on social networking sites that included suicide-related words among Korean adolescents*", which was the third most cited study (n=29). Of the scientific articles published in journals of science, 67.5% (n=25) appear in the *Scimago Journal & Country Rank* database. Of these, more than half (59.5%, n=22) are Q1 and Q2, and 8.1% (n=3) are Q3. The rest are not indexed in this database.

Concepts and dimensions studied

The universe of concepts included in the title and keywords of the documents registered demonstrates the correct choice of terms for the documentary search and highlights the relevant and significant aspects of this topic (see Figure 3).

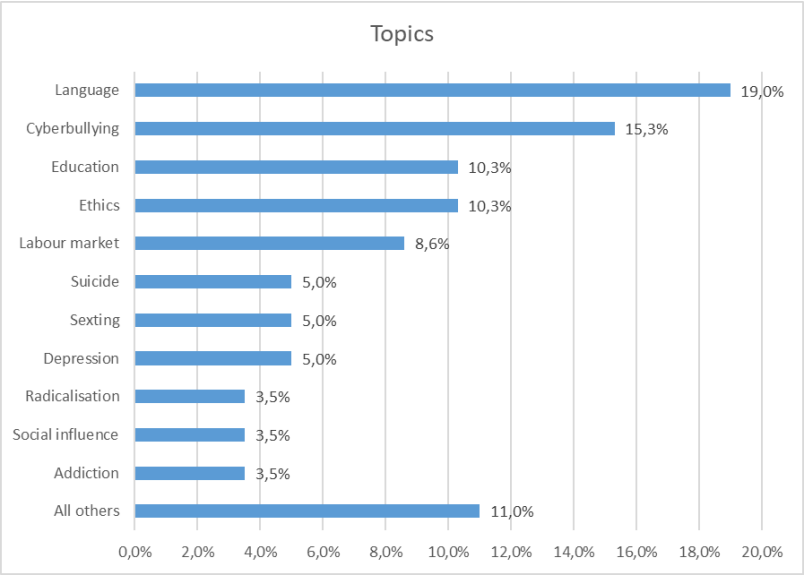
Figure 3. Recurring main concepts in the research

Source: prepared by the authors using Word it out.

With regard to the dimensions identified in the scientific literature related to topics concerning minors and their interests (United Nations Children's Fund [UNICEF], 2019), which have been investigated using SBD, those that stand out are studies focusing on the analysis of language and opinions on social networks (19%; n=11), online bullying (15.3%; n=9), risky practices that involve the invasion of privacy of adolescents, and educational policies (10.3%; n=6).

The remaining entries (see Figure 4) address topics as disparate as the situation and forecast of the labour market at 8.6% (n=5), depression, sexting, and teenage suicide, each at 5% (n=3), as well as the social influence of social networks and youth radicalisation at 3.5% (n=2). These are followed by juvenile delinquency, sport and leisure, teenage pregnancy, environment, mobile phones, fashion, and privacy, each with 1.7% (n=1) (identified as "All others" in the graph)

Figure 4. Dimensions addressed in the research on SBD and minors



Source: prepared by the authors

4.3. Knowledge domain of the studies

Another of the aspects examined in this bibliographic research was the knowledge domain to which the publications belong. To this end, the publications have been segmented into four categories in accordance with the profile of the journals in the sample:

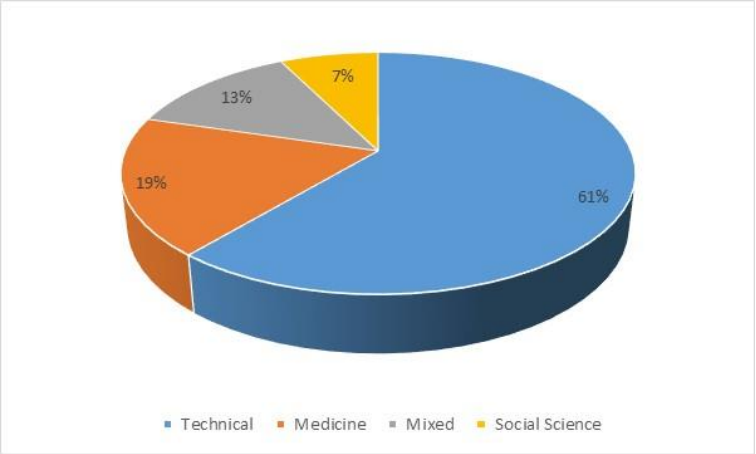
- Technical: deals with issues such as computing, engineering and informatics.
- Social sciences: these are composed of 13 main branches: Anthropology, History, Legal Science, Economics, Sociology, Psychology, Linguistics, Semiology, Archaeology, Demography, Human Ecology, Pedagogy, and International Relations (*Centro de Estudios Cervantinos*, 2020).
- Mixed: this combines the categories of technical and social science journals.
- Medicine.

This classification does not include the four European Union reports that do not have a topic bias. Of the remaining 93% (n=54), 61% (n=33) are technical journals, 19% (n=10) are medical journals, 13% (n=7) are mixed, and 7% (n=4) are social science journals (see Figure 5).

ISSN: 1696-019X / e-ISSN: 2386-3978

ISSN: 1696-019X / e-ISSN: 2386-3978

Figure 5. Knowledge domain of the publications



Source: prepared by the authors

4.4. Applied SBD techniques and samples

The techniques applied vary according to the object of study, and the authors do not claim that there are differences between SBD and BD techniques. Researchers conduct their studies by applying statistical techniques such as frequency analysis to detect the importance of one or several words in the network, as well as diverse correlational statistical tests. Moreover, the application of various algorithms to detect trends have also been found, in addition to the use of bivariate statistical analysis, cluster analysis, temporal analysis and structural equations, among others. Similarly, studies have also been found that employ the vocabulary used in social networks to apply sentiment analysis and machine learning, which allows for the analysis and coding of large-scale texts. In other cases, data mining, text mining, and opinion mining are used to identify and analyse the comments, posts and messages that young people leave on social networks.

Thanks to the use of SBD, the samples that researchers work with are very large. Among them, Kim, et al. (2019) stand out. In order to understand the current and future experiences of adolescent bullying, these researchers analysed 436,508 web documents from 279 online channels between 2013 and 2017, in which they identified 177 terms related to specific forms of bullying. For their part, Song et al. (2016) reviewed more than 2.35 billion posts over a period of two years from 163 websites, and they consolidated 99,693 records of minors concerning youth suicide. Han, et al. (2019) studied 2,400 terms in three categories concerning cyberbullying, and Song, et al. (2014) compiled web documents from 227 online channels from which they extracted 435,563 cases related to cyberbullying and identified 118 terms as causes of cyberbullying. Finally, Matsumoto, Yoshida and Kita (2019) have created their own corpus for researching youthful jargon-based tweets when communicating on social networks. They have succeeded in creating lists and vectors of emotional expression of up to 30,000 items.

4.5. Main proposals and contributions

In this section, we have focused on the studies that have been the most influential among the corpus analysed with regard to the sector of the population relevant to this research. Of all the records found, 16 have examined specific samples of adolescents or have referred to them explicitly when preparing and developing their corpus related to the topics and dimensions listed below.

4.5.1. Detection of cyberbullying

By using automatic language identification systems in the context of social networks, Kim, et al. (2019) have pointed out that the problems of online bullying are not limited to schools, but have spread to other environments such as teenage "fandom" communities, which bring together young fans of a hobby or pastime. In this regard, they recommend government systems to control the emerging modes of diverse types of bullying. Han, et al. (2019) have also analysed bullying as a social phenomenon by examining the terms used by minors on social networks. This coincides with the study by Kim et al. (2019), which advocates the creation of data security and privacy protection policies for individuals who are investigated by using SBD.

Song, at al. (2014) studied the risk factors for each type of online bullying with the aim of automatically generating a predictive model to explain the variables that affect this type of behaviour by young people. They also suggest that parental education programmes should be put in place to help children in these situations. Tarwani, Chorasia & Shukla (2017), who have extracted the opinions of internet users on Facebook, MySpace and Twitter using machine learning techniques, offer the possibility of instant detection of bullying messages on social networks and propose a solution to the problem of cyberbullying based on the identification of bullying terms.

For their part, Lara and Giancinto (2017) have worked with statistical models for the detection of offensive expressions, as well as with a system that collects these individual communications. The key lies in creating an application that recognises comments and statements with negative connotations. Its main limitation is the interpretation of the contextual analysis of conversations, which can lead to false positives, especially when language can be sarcastic or vulgar, yet inoffensive, in a given situation.

4.5.2. Sexting as a virtual practice among adolescents

Sexting is an emerging phenomenon that is attracting significant attention for its potential to cause harm to minors (Song & Song, 2015). By applying SBD to the communications of young people under 18 years of age on Twitter, Song, Song and Lee (2018) have concluded that adolescents engage in sexting through the use of smartphone text messaging in the hope of gaining increased attention from friends. They also suggest that expert advice and the promotion of children's integrity are factors that help to minimise the risk involved in sexting. With regard to the use of SBD, they say that while this methodology is useful and effective, it has challenges such as real-time monitoring and individual data analysis. They consider it essential to conduct further research using this technology with a focus on the risk factors related to this practice.

ISSN: 1696-019X / e-ISSN: 2386-3978

ISSN: 1696-019X / e-ISSN: 2386-3978

4.5.3. Suicide risk among young people

This aspect of research using SBD has become particularly relevant in South Korea, as the suicide rate among youth in that country is one of the highest in the world (Song, et al., 2016), and conducting searches of suicide-related terms on Google have been growing steadily since 2007 in that country.

Song et al., (2014) point out that real-time investigation of big data is useful for detecting suicidal tendencies. Specifically, these authors have created a statistical model based on analysing the relationship between the number of Google searches for the keywords *stress*, *drinking*, *exercise*, and *suicide*, and the detection of adolescent suicide risk. They have determined a correlation between a higher number of searches for the words *stress* and *drinking alcohol* and the actual number of suicides, which is significant. They are aware that the use and application of SBD for this type of research is in the early stages, and their advice is that to improve this situation and be able to establish national suicide prevention plans, it will be necessary to create training programmes for professionals in the management of big data, improved infrastructures, and the implementation of cloud computing services.

Song, et al. (2016) have confirmed that adolescents vent their stress, depressive emotions, and suicidal thoughts in cyberspace, and they communicate these thoughts and feelings with other internet users. Among the factors associated with increased activity in searching for words related to teenage suicide, some of the terms that stand out include the following: academic pressure, poor body image, being a victim of bullying, preoccupation with illness, low employment rates, increased rental price index, and intimidation. To reduce the problem, the authors say it is necessary to develop a system for monitoring and responding to real-time searches for suicide-related words on social media via bots and instant messaging.

4.5.4. Adolescent language in social networks

Among the articles that specifically study adolescent conversations that take place in social networks are those of Matsumoto, et al. (2019), and Matsumoto, Yoshida and Kita (2019), who focus on the vocabulary and short expressions that young people use in social networks that do not appear in dictionaries. Their work involves designing models based on groups of documents that include these youthful expressions on Twitter and propose a classification method based on the coding of their characteristics. With this data-driven, slang-clustering model, they have been able to identify 40 terms and short expressions related to the way young people express themselves on social networks. Specifically, Matsumoto, et al. (2019) work with emotional recognition and have created a method they call the "bag of concepts", the purpose of which is to build an emotional corpus through the collection of weblog texts that include these new words. They point out that most of the words in youthful slang that express the feelings of young people are too ambiguous to be recorded in dictionaries.

4.5.5. Other research dimensions

Other dimensions of research have also been addressed using SBD. According to Jung, Park and Song (2016, 2017), social networks contain an enormous amount of information about children's feelings, thoughts, interests, and behaviour patterns that can be used to understand their tendency toward depression. In their 2017 research, these authors point out

that the ontology they have developed is especially innovative compared to similar studies for three main reasons: Firstly, they have included previously unidentified factors of adolescent depression related not only to individual characteristics, but also to environmental factors such as family, school and community, as well as low academic performance, delinquency and truancy; secondly, they have taken into account risky situations, symptoms, diagnoses, preventive measures, and treatments of adolescent depression; and thirdly, they have also considered slang, buzzwords and neologisms used by adolescents, which makes SBD analysis especially suitable for analysing information generated by adolescents on social networks.

With regard to practices that violate the privacy of teenagers when online, Montgomery (2015) highlights that one of the keys to Facebook's business model is the fact that it has managed to datify social relationships through the data analysis system known as "The Social Graph". This has allowed it to study and control all the movements that take place in its network, which is a crucial element in its diverse commercial operations. This omnipresent commercial surveillance is unprecedented in its reach and penetration into the lives of young people. At the same time, this author believes it is necessary to increase public policies in order to protect minors from Facebook's intrusion into their lives and data, and to implement privacy laws to safeguard against the excesses of BD.

Finally, DeJonckheere, et al., (2019) have integrated several research techniques with regard to young pregnant women: the use of a survey, corresponding text message analysis, and social data mining with natural language processing. Their aim is to identify the beliefs and social norms of these young women in relation to weight gain during pregnancy. The SBD part of the study targets accounts on Facebook, Twitter and Instagram of the sample of selected young women at two precise moments: during their visit to the medical centre in the first trimester, and again in the third. By analysing messages and images posted on social media sites, the researchers studied the beliefs and public statements made by these women with regard to pregnancy, weight gain, body image, and nutrition during the months that the women were pregnant.

5. Conclusions and discussion

This paper has approached publications regarding the online behaviour of children and young people in which techniques linked to SBD have been used. Analysis related to the object of study is constantly growing in terms of the number of academic articles found, publication dates, and volume of citations. By carrying out analyses using big social data, researchers are working to discover the activities of minors online, and to create algorithms to identify risky trends affecting adolescents.

The dimensions that have received the most attention are the recognition of emotions and expressions on social networks, cyberbullying, the monitoring and implementation of educational policies, and the study of unethical practices that interfere with the virtual lives of adolescents, such as virtual manipulation, suicide, and depression among young people. Other issues identified include juvenile delinquency, the perception of teenage pregnancy, sexting, dangerous contacts on social networks, sexual abuse of minors on the Internet, the commercialisation of private data, and the arbitrary way in which social networks invade the privacy of minors.

ISSN: 1696-019X / e-ISSN: 2386-3978

ISSN: 1696-019X / e-ISSN: 2386-3978

Specialised authors in the field who are working to exploit data from social networks on issues that directly affect minors or young people have also been documented. By using SBD, scientists can investigate large amounts of real-time data with samples of up to 2.35 billion posts (Song, et al., 2016). They can also create algorithms and automated models to detect children's behaviour and feelings and identify trends.

What stands out is that only 6.8% (n=4) of the publications have appeared in social science journals, and 13.5% (n=8) were found in mixed journals. The rest were published in technical research journals related to Information Technology, Computer Science, and Engineering, and in medical journals as well. In addition to highlighting the value of perspectives that are more technophile and systematic, which have been provided by the sample under study of a subject as complex as the online behaviour of human beings when faced with multiple screens, an issue still pending is the merger and comprehension of other disciplines such as Communication, Sociology and Education.

In this paper, we have chosen to undertake a search focused on SBD. However, there is some discrepancy in the naming of studies and techniques related to this field. Most authors title their research and include the term SBD or BD in their keywords when they investigate social networks, and they use research techniques such as data mining, machine learning, semantic analysis, and sentiment analysis. However, they do so without actually stating that they are applying a specific SBD or BD methodology.

We expect to see an increase in the number of studies that address the virtual lives of minors through the use of these research techniques, as there is still a gap between their potential and their actual use. Of course, SBD is only one approach (Malvicino & Yoguel, 2014), and data alone is not a magic solution for analysing the social reality. One should not fall into quantitative reductionism (Sánchez-Bayón, 2017) and abandon cultural aspects, among others, when investigating social movements. Researchers must approach big data with caution to ensure social progress (Kosinski & Behrend, 2017), and there is no doubt that they will continue to be necessary in order to develop theories and provide the legitimacy that science needs as a social entity that sets standards (Sætra, 2018).

This research has two limitations. On the one hand, it does not incorporate all possible databases that can be accessed to locate scholarly literature in this field. On the other hand, a tag-based search strategy may at times fail to include some aspects that could be of interest.

6. Funding

This study has been funded by the programme entitled, "New scenarios of digital vulnerability: media literacy for an inclusive society" (PROVULDIG-2-CM) (ref. H2019/HUM5775), CAM and the European Social Fund, and by the National Project of the R&D&I programme, "Social networks, adolescents and young people: media convergence and digital culture" (CSO2016-74980-C2-2-R).

7. Bibliographic references

Alsehaima, A. O., & Alanazi, A. A. (2018). Psychological and social risks to children of using the internet: Literature Review. *Journal of Child Adolescent Behavior, 6*(380), 2. http://dx.doi.org/10.4172/2375-4494.1000380

Aparici, R., & García Matilla, A. (2016). ¿Qué ha ocurrido con la educación en comunicación en los últimos 35 años?: pensar el futuro. *Espacios en Blanco. Revista de Educación (Serie Indagaciones)*. https://bit.ly/2Twn6ko

Arcila-Calderón, C., Barbosa-Caro, E., & Cabezuelo-Lorenzo, F. (2016). Técnicas BD: análisis de textos a gran escala para la investigación científica y periodística. *El profesional de la información, 25*(4), 623-631. https://doi.org/10.3145/epi.2016.jul.12

Arellano Toledo, W. (2014). Gobierno abierto y privacidad: la problemática del Big data y el cómputo en la nube. *Virtualis, 5*(10), 34-59. https://bit.ly/2FADr3y

Batty, M. (2013). Big Data, smart cities and city planning. *Dialogues in Human Geography, 3*(3), 274-279. http://dx.doi.org/10.1177/2043820613513390

Bell, G., Hey, T. & Szalay, A. (2009). Beyond the data deluge. *Science,* 323, 1297–1298. http://dx.doi.org/10.1126/science.1170411.

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

Boyd, D. & Crawford, K. (2012). Critical Questions for Big Data. Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society, 15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Centro de Estudios Cervantinos (2020). Cuáles son las ramas de las ciencias sociales. https://www.centroestudioscervantinos.es/ramas-de-las-ciencias-sociales/

Cerezo Gilarranz, J. (2015). Big Data, la nueva ciencia del siglo XXI. En: *Pérez-Hernández y Torra (eds). Big Data. Revista de Ciencias y Humanidades de la Fundación Ramón Areces, 14,* 7-11. https://bit.ly/3hsHTyA

DeJonckheere, M., Nichols, L. P., Vydiswaran, V. V., Zhao, X., Collins-Thompson, K., Resnicow, K., & Chang, T. (2019). Using text messaging, social media, and interviews to understand what pregnant youth think about weight gain during pregnancy. *JMIR formative research, 3*(2): e11397. https://doi.org/10.2196/11397

Díaz-Campo, J. (2015). Análisis bibliométrico de las tesis doctorales sobre Ética de los Medios de Comunicación presentadas en España (1979-2013). *Doxa Comunicación,* 65-88. https://doi.org/10.31921/doxacom.n20a3

Falzone, A. E., Brindis, C. D., Chren, M. M., Junn, A., Pagoto, S., Wehner, M., & Linos, E. (2017). Teens, tweets, and tanning beds: rethinking the use of social media for skin cancer prevention. *American journal of preventive medicine, 53*(3), S86-S94. https://doi.org/10.1016/j.amepre.2017.04.027

ISSN: 1696-019X / e-ISSN: 2386-3978

Fondo de las Naciones Unidas para la Infancia (UNICEF, 2019). *Barómetro de Opinión de Infancia y Adolescencia 2019. https://www.unicef.es/publicacion/que-opinan-los-ninos-y-las-ninas*

Giménez, A. M., Luengo, J. A., & Bartrina, M. J. (2017). ¿Qué hacen los menores en internet? Usos de las TIC, estrategias de supervisión parental y exposición a riesgos. *Electronic Journal of Research in Education Psychology, 15*(3), 533-552. http://dx.doi.org/10.14204/ejrep.43.16123

Gualda, E., & Rebollo, C. (2020). Big Data y Twitter para el estudio de procesos migratorios: Métodos, técnicas de investigación y software. *Empiria. Revista de metodología de ciencias sociales, 46*, 147-177. https://doi.org/10.5944/empiria.46.2020.26970

Han, Y., Kim, H., Song, J., & Song, T. M. (2019). Ontology development of school bullying for social big data collection and analysis. *The Journal of the Korea Contents Association, 19*(6), 10-23. https://doi.org/10.3390/ijerph16142596

Hernández-Leal, E. J., Duque-Méndez, N. D., & Moreno-Cadavid, J. (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoLógicas, 20*(39), 17-24. http://www.scielo.org.co/pdf/teclo/v20n39/v20n39a02.pdf

Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Washington: Microsoft Research.

Jiménez, E., Garmendia, M., & Casado, M. A. (2018). *Entre selfies y whatsapps. Oportunidades y riesgos para la infancia y la adolescencia conectada*. Barcelona: Gedisa.

Joyanes Aguilar, L. (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor.

Jung, H., Park, HA., & Song, TM. (2016). Development and Evaluation of an Adolescents´ Depression Ontology for an Analysing Social Data. En: *SERMEUS, et al., (Eds.) Nursing Informatics 2016: EHealth for All: Every Level Collaboration–From Project to Realization,* 225-442. Amsterdam: IOS Press BW.. https://doi.org/10.3233/978-1-61499-658-3-442

Jung, H., Park, HA., & Song, TM. (2017). Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. *Journal of medical internet research, 19*(7), e259. https://doi.org/10.2196/jmir.7452

Kim, H., Han, Y., Song, J., & Song, T. M. (2019). Application of social big data to identify trends of school bullying forms in South Korea. *International journal of environmental research and public health, 16*(14), 2596. https://doi.org/10.3390/ijerph16142596

Kosinski, M., & Behrend, T. (2017). Editorial overview: Big data in the behavioral sciences. *Current Opinion in Behavioral Sciences, 18*, iv–vi. https://doi.org/10.1016/j.cobeha.2017.11.007

Kumar A., Sangwan S.R., Nayyar A. (2020) Multimedia Social Big Data: Mining. In: *Tanwar S., Tyagi S., Kumar N. (eds) Multimedia Big Data Computing for IoT Applications*. Intelligent Systems Reference Library, 163. Springer, Singapore. https://doi.org/10.1007/978-981-13-8759-3_11

ISSN: 1696-019X / e-ISSN: 2386-3978

Lara Palma, A. M., & Giancinto, R. (2017). Ethical Machines against Unethical Computer-mediated Social Interactions. *Journal of Information and Management, 37*(2), 22-38. https://doi.org/10.20627/jsim.37.2_22

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

Livingstone, S., Mascheroni, G., & Staksrud, E. (2018). European research on children's internet use: Assessing the past and anticipating the future. *New Media & Society, 20*(3), 1103-1122 https://doi.org/10.1177/1461444816685930

López Cantos, F. J. (2018). *Cultura visual y conocimiento científico: Comunicación transmedia de la ciencia en la era Big Data*. Barcelona: Editorial UOC.

Malvicino, F., & Yoguel, G. (2014). Big Data. Avances recientes a nivel internacional y perspectivas para el desarrollo local. *Documento de Trabajo. Centro Interdisciplinario de Estudios en Ciencia Tecnología e Innovación (CIECTI-MinCyT)*. Buenos Aires. https://bit.ly/3cFeizD

Manovich, L. (2011). Trending: The promises and the challenges of big social data. En *Gold (eds) Debates in the digital humanities*, 460-475. London: University Minnesota Press.

Martínez-Martínez, S., & Lara-Navarra, P. (2014). El big data transforma la interpretación de los medios sociales. *El profesional de la información, 23*(6), 575-581. https://doi.org/10.3145/epi.2014.nov.03

Matsumoto, K., Ren, F., Matsuoka, M., Yoshida, M., & Kita, K. (2019). Slang feature extraction by analysing topic change on social media. *CAAI Transactions on Intelligence Technology, 4*(1), 64-71. https://doi.org/10.1049/trit.2018.1060

Matsumoto, K., Yoshida, M., & Kita, K. (2019). Emotion Recognition for Japanese Short Sentences Including Slangs Based on Bag of Concepts Feature Trained by Large Web Text. *Current Analysis on Instrumentation and Control, 2019*(2), 9-18. https://bit.ly/2XeP6dk

Mayer-Schönberger, V. & Cukier, K. (2013). Big Data. La revolución de los datos masivos. London: Turner.

Montgomery, K. C. (2015). Youth and surveillance in the Facebook era: Policy interventions and social implications. *Telecommunications Policy, 39*(9), 771-786. http://dx.doi.org/10.1016/j.telpol.2014.12.006

Nupairoj, N. (2016). El ecosistema de la alfabetización mediática: Un enfoque integral y sistemático para divulgar la educomunicación. *Comunicar, 24*(49), 29-37. https://doi.org/10.3916/C49-2016-03

O'Neil C. (2013). *The rise of big data, big brother.* https://bit.ly/3eR7Jvi

Paredes-Moreno, A. (2015). Big Data: Estado de la cuestión. *International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC), 2*(1), 38-59.

Pérez, M. (2015). *BIG DATA-Técnicas, herramientas y aplicaciones*. Alfaomega Grupo Editor.

Pérez, M. J. (2016). Davos y la cuarta revolución industrial. *Nueva Revista, 157*. https://bit.ly/3wIUnLp

ISSN: 1696-019X / e-ISSN: 2386-3978

Pybus, J., Coté, M., & Blanke, T. (2015). Hacking the social life of big data. *Big Data & Society*, *2*(2), 1-10. https://doi.org/10.1177/2053951715616649

Qin, S. J. (2014). Process data analytics in the era of bid data. *AIChE Journal, 60*(9), 3092-3100. https://doi.org/10.1002/aic.14523

Ramírez-Montoya, M., & García-Peñalvo, F. (2018). Co-creation and open innovation: Systematic literature review. *Comunicar. Media Education Research Journal*, 54, 9-18. https://doi.org/10.3916/C54-2018-01

Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. Sebastopol: O'Reilly Media, Inc.

Sætra, H. S. (2018). Science as a vocation in the era of big data: The philosophy of science behind big data and humanity's continued part in science. *Integrative Psychological and Behavioral Science*, *52*, 508-522. https://doi.org/10.1007/s12124-018-9447-5

Sánchez-Bayón, A. (2017). Apuntes para una teoría crítica humanista y su praxis económico-empresarial en la posglobalización. *Miscelánea Comillas. Revista de Ciencias Humanas y Sociales, 75*(147), 305-329. https://bit.ly/2RtY188

Schwab, K. (2017). *The fourth industrial revolution*. London: Currency.

Shuijing, H. (2015). Big Data Mining: Essential Technologies and Concerns. In *Proceedings of the 2015 Sixth International Conference on Digital Manufacturing and Automation*. 88-92. https://bit.ly/3dA77dd

Song, J., Song, T-M., & Lee, J. R. (2018). Stay alert: Forecasting the risks of sexting in Korea using social big data. *Computers in Human Behavior, 81*, 294-302. https://doi.org/10.1016/j.chb.2017.12.035

Song, J., Song, T-M., Seo, D. C., & Jin, J. H. (2016). Data mining of web-based documents on social networking sites that included suicide-related words among Korean adolescents. *Journal of Adolescent Health, 59*(6), 668-673. http://dx.doi.org/10.1016/j.jadohealth.2016.07.025

Song, T., & Song, J. (2015). Social Big Data Analysis and Utilization Methodologies-With Special Reference to Forecasting the dangers of sexting in Korea using social big data. https://bit.ly/2Zbpq2d

Song, T., Song, J., An, J. Y., & Woo, J. M. (2014). Social Risk Factor Prediction Utilizing Social Big. Kihasa. https://bit.ly/3c4sq6D

Tapia Nava, E. (2018). *El uso del Big Data en los estudios de opinión pública*. Instituto Belisario Domínguez Dirección General de Análisis Legislativo. https://bit.ly/2z7HUb4

Tarwani, N., Chorasia, U., & Shukla, P. K. (2017). Survey of Cyberbulling Detection on Social Media Big-Data. *International Journal of Advanced Research in Computer Science, 8*(5), 831-835.

Taylor-Sakyi, K. (2016). *Big data: Understanding big data*. arXiv preprint arXiv:1601.04602.

ISSN: 1696-019X / e-ISSN: 2386-3978

Törnberg, P., & Törnberg, A. (2018). The limits of computation: A philosophical critique of contemporary Big Data research. *Big Data & Society, 5*(2), 1-12. https://doi.org/10.1177/2053951718811843

Uman, I. (2018). Big Data y Memoria Digital: claves para su exploración e investigación desde las ciencias sociales. *Avatares de la Comunicación y la Cultura, 15*. https://bit.ly/3gRtOeR

UN Global Pulse, IAPP. (2018). Building Ethics into Privacy Frameworks for Big Data and AI. https://bit.ly/3sYaySp

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & society, 12*(2), 197-208. https://doi.org/10.24908/ss.v12i2.4776

van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In: *Ding Y., Rousseau R., Wolfram D. (eds) Measuring Scholarly Impact*, 285-320. Springer, Cham. https://doi.org/10.1007/978-3-319-10377-8_13

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: *Proceedings of the 18th international conference on evaluation and assessment in software engineering, 38*, 1-10. https://doi.org/10.1145/2601248.2601268

ISSN: 1696-019X / e-ISSN: 2386-3978