Comunicación y Sociedad
Departamento de Estudios de la Comunicación Social
Universidad de Guadalajara

Dead end? The sampling problem in the study of digital social networks[1]

¿Callejón sin salida?: El problema muestral en el estudio de las redes sociales digitales

Beco sem saída? O problema de amostragem no estudo das redes sociais digitais

DOI: https://doi.org/10.32870/cys.v2024.8580

DANIEL PATTIER[2]
https://orcid.org/0000-0003-3426-922X

The objectives of this work are to delve into the sampling problem in the study of digital social networks and to show the possibilities and solutions that the literature has provided to deal with this problem. Through a theoretical-reflexive methodology, the subject is investigated highlighting and analyzing the procedural options used by the literature. In addition, some methodological improvements are proposed that can contribute to the development of more reliable research on the subject.
KEYWORDS: Digital communication, social networks, research methodology, sample size.

Los objetivos de este trabajo son profundizar en la problemática muestral en el estudio de las redes sociales digitales y mostrar las posibilidades y soluciones que aporta la literatura para hacer frente a este problema. A través de una metodología teórico-reflexiva se indaga sobre el tema, destacando y analizando las opciones procedimentales utilizadas en la literatura. Se plantean, además, algunas mejoras metodológicas que pueden contribuir al desarrollo de investigaciones más fiables sobre la temática.
PALABRAS CLAVE: Comunicación digital, redes sociales, metodología de investigación, tamaño de la muestra.

Os objetivos deste trabalho são aprofundar os problemas exemplares no estudo das redes sociais digitais e mostrar as possibilidades e soluções que a literatura oferece para resolver este problema. Através de uma metodologia teórico-reflexiva, o tema é investigado, destacando e analisando as opções processuais utilizadas na literatura. Além disso, são propostas algumas melhorias metodológicas que podem contribuir para o desenvolvimento de pesquisas mais confiáveis sobre o tema.
PALAVRAS-CHAVE: Comunicação digital, redes sociais, metodologia de pesquisa, tamanho de amostra.

How to cite:
Pattier, D. (2024). Dead end? The sampling problem in the study of digital social networks. *Comunicación y Sociedad*, e8580. https://doi.org/10.32870/cys.v2024.8580

[1] Work funded by the Ministry of Science and Innovation of Spain through the R&D&i project "The imperative of educational innovation: analysis of its reception and articulation in the Spanish educational system (IMP-NOVA)" with reference PID2022-138878NA-I00.

[2] Universidad Complutense de Madrid, Spain.
dpattier@ucm.es
Submitted: 02/16/23. Accepted: 09/12/23. Published: 02/14/24.

Introduction

Social networks, as digital platforms for global communication, connect a large number of users around the world. They promote the creation of environments where users create their own profiles in order to communicate with other people, express themselves, view or comment on content created by other users, or share their own (Phua et al., 2016). They are dynamic environments, as the users join, use and leave communities depending on their interests over time (Mahmoudi et al., 2020).

Social networks have transformed the way we communicate, learn, talk or interact with others. The accessibility offered by technological devices has led to the frequent, habitual use of all types of webpages and social media. According to the *Digital 2022 Global Overview Report* (Kemp, 2022), which analyses worldwide digital use, each passing year there is an increase in the time that we spend using technological devices connected to the Internet. In 2022 this reached an average of almost seven hours a day. In this way, if we take into account that a person usually sleeps between seven and eight hours a day, we can say that we spend over 40% of our waking hours online.

This report provides an in-depth analysis of the use of digital social networks worldwide and reveals that 95.2% of people between 16 and 64 years old have used them in the last month. Furthermore, it is significant that the average user visits 7.5 different social networks a month, which shows a wide-ranging use of this type of digital platform. Lastly, it is significant that four of the ten most-visited webpages belong to social networks: YouTube (second place), Facebook (third place), Instagram (sixth place) and Twitter (ninth place).

This growing use of digital social networks has contributed to the increase of literature focusing on research into pages, profiles, accounts and channels belonging to a great number of these networks, to obtain data relating to numerous fields of knowledge. Thus, we can find research on this theme in the fields of psychiatry (Brooks et al., 2022), social care (Fernández-Peña et al., 2022), physical activity (Prochnow & Patterson, 2022), politics and economy (Faciroli et al., 2022), education (Marcelo & Marcelo, 2021), sociology (Tobón Berrio et al.,

2022), communication (Fernández et al., 2021) or medicine (Kumar et al., 2021), among others.

The study of digital social networks

The literature has approached the theme of social networks from numerous viewpoints, from simply examining the quantitative data to conducting thorough discourse network analysis (DNA) of semantic networks or even social-ecological networks (Tindall et al., 2022).

However, in general, the research approaches the theme of social networks in two ways. One way focuses on the relationships reported by a focal person, and the other considers a whole network using a specific criterion to define the population (Vatrapu et al., 2016). Moreover, the dynamic nature of this type of environment requires the network to be fragmented into separate time windows. As indicated by Mahmoudi et al. (2018), many studies have only used a single time window as an inseparable piece of information, which could lead to misinterpretation of the data.

It is important to mention, additionally, the current debate involving research that employs two types of processes in their use of data. On the one hand, the use of big data, whereby specific analysis programs manage huge quantities of data quantitatively. Problems may arise in these studies due to factors of heterogeneity, noise accumulation, spurious correlations and endogeneity (Hernández-Leal et al., 2017). On the other hand, soft data are based on a qualitative perspective involving opinions and interpretations, which can produce a subjective deviation of the research results or conclusions.

Furthermore, along the same lines as Tindall et al. (2022), it is significant that there are specific structural forms in the architecture of social networks where several processes that simultaneously determine each other are at work, such as reciprocity, homophily, transitivity or preferential attachment. As their research indicates, if we consider generative models for data from cross-sectional networks, we can emphasize the current usage of exponential random graph models (ERGMs).

Likewise, when collecting data from social networks or webpages, the literature reflects two reliable ways of doing so: web scraping and

sending requests to application programming interfaces (APIs). This may be a complex process, depending on the research being conducted (Tindall et al., 2022). However, the lack of information about sampling frames (Rocheva et al., 2022) or the enormous amount of data generated by the users in real time makes sampling unavoidable and leads to the development of methodological alternatives, such as social mining (Piña-García et al., 2016). Hence the importance of having a clear aim when sampling and knowing how to define the sample efficiently, in view of the impossibility of using the whole population in most research processes (Pihlaja, 2022).

This usage of a certain part of the sample may lead to sampling errors with results that do not represent the ones that could be obtained through the analysis of the entire population. The literature indicates certain kinds of sampling errors. Firstly, the population-specific error, which arises when the total research population is not defined correctly. Secondly, the selection error, which may occur when the sample is self-selected or when only people who are interested in the subject participate. Thirdly, the sampling frame error, produced when the population subset is selected incorrectly. Fourthly, the non-response error, when the research does not obtain useful responses from the participants.

As the literature indicates, research on social networks needs to meet the challenge of conducting in-depth examination of a field as changeable as digital media (Jones et al., 2015). In view of this, it is important to mention the contribution made by Vásquez (2022), who offers a multifaceted viewpoint, by advocating the efficiency of different research methodologies to approach the reality of social networks and digital discourse, both from a quantitative and qualitative perspective.

Finally, there are several notable studies that have been conducted on the subject of privacy and research ethics in data mining, selection and usage in research into social networks (Tindall et al., 2022). There are sensitive factors relating to privacy, data processing, compliance with the regulatory framework and confidentiality (Zhang et al., 2010). In this respect, it is worth noting the contribution made by Wang et al. (2022), who present a method to ensure privacy in data processing using two tasks: detection and preservation. The task of

detection identifies images and captions in order to locate the specific objects. The task of preservation uses methods of hiding or masking to manage the images and captions detected, whilst at the same time guaranteeing the usefulness of the resources.

Lastly, we should not forget the possibility of biased and malicious user behavior, involving an added problem for current research (Sirisala et al., 2022).

Research objectives and questions

In this study, we will establish a methodological discussion regarding a problem that arises in studies on digital communication, specifically regarding research into digital social networks. Therefore, this article has the following two objectives: 1) to examine the problematic aspect of sampling in the study of digital social networks, and 2) to demonstrate the possibilities and solutions found in the literature to address these problems.

The research questions are as follow: Which methodological problems arise when designing a sample for studies of social networks? How do studies address these problems in order to obtain data and results?

To answer these research questions, we will refer to the methodology used. At a later stage, we will present the sampling problem in the study of digital social networks and we will investigate the different solutions proposed by the literature to date. Lastly, there will be a discussion section, ending with the conclusions.

Methodology

This work describes a theoretical, reflexive methodology that aims to examine the essential relationships involved in the problems of various procedural phenomena in the field of communication research (Del Sol Fabregat et al., 2017). Due to the similarity of the difficulties that arise in the study of digital social networks, our basis is a general perspective that offers an overall vision and frames an understanding of the question.

The approaches described in this text have emerged after an extensive study of the above-mentioned state-of-the-art review,

which will be further reinforced with specific examples of studies that use each of the methodological options described, in their attempt to solve the problems examined in this work.

The Sampling Problem

Studies on digital social networks in the different subject areas face a methodological problem that is inherent in the very nature of these technological pages and platforms: sample establishment.

In the context of research, the sample is an essential feature, as it is the foundation that provides the data that will be interpreted to prove or disprove the hypothesis or hypotheses or to answer the different research questions. The use of samples with poor methodological quality may mean results are obtained that are not generalizable or that are even incorrect (Mucha-Hospinal et al., 2021). Consequently, it is important to establish rigorous methodological and quality processes that provide valuable data for the overall discussion about the subject of study (Vásquez, 2022).

The problems relating to sampling in studies on social networks are based on two factors that directly affect this methodological aspect. Firstly, social networks are constantly changing, as they depend on their users, who upload content or interact on a daily basis, thereby generating millions of pieces of new content on the network (Jones et al., 2015; Mahmoudi et al., 2020). As we can see, this creates a problematic area as far as sample definition is concerned. Which profiles to examine? For how long? Which kinds of content should be considered and which should not? These are some of the questions that any study on social networks needs to consider in order to establish a rigorous methodological framework (Vásquez, 2022).

Secondly, there is an issue in terms of sample establishment, as there are no official lists of themes or subject areas that include all the profiles, accounts or channels on a social network. We should understand that digital social networks are not created or designed in a structured, methodological way, but rather based on ranking algorithms that seek to obtain the greatest audience interaction and engagement. This explains why the suitability or quality of contents does not depend on their positioning on the platform (Burgos et al., 2020; Osman et al., 2022).

In this way, when researchers aim to study, for example, influencers in the area of science dissemination on Twitch (Buitrago & Torres-Ortiz, 2022), this platform does not create an official list of all the profiles based on the fact that they focus on informal education and, in particular, science dissemination. This issue arises on each and every one of the social networks available to date and in all subject areas.

Furthermore, it is important to emphasize the difficulty in categorizing, on the one hand, certain channels or profiles that focus on several areas or very diverse topics, and, on the other, those that fall between two categories and, depending on the content creator's own interpretation or that of the researchers, could be classified under one category or another. For example, a YouTube channel for uploading educational songs could be categorized under "Training" or "Music".

As it is not possible to gain access to an official list from the platform itself that indicates the channels or profiles relating to subject of study, this creates a problem in terms of sample definition and, therefore, is at the heart of the methodology of the different studies on digital social networks (Pihlaja, 2022).

Solutions provided by the literature

As we have indicated above, when establishing the study sample regarding social networks, there is a problematic methodological aspect that is based on two factors: the constant changes and updates applied to contents and the lack of officially categorized public lists on the different digital platforms.

In this section, we will describe the different solutions provided by the literature over the years that have attempted to solve this methodological problem by taking different decisions from a creative perspective and aiming to improve the quality of research (Flores-Márquez & González Reyes, 2021).

Case study

Firstly, several studies have examined the subject in depth by analyzing certain profiles or channels on social networks using the case study methodology. The definition of the sample, therefore, does not involve

the above-mentioned problematic aspect, as it focuses on one or several specific cases.

Throughout the literature, studies that have used this method have done so out of convenience and due to the ease of accessing the sample. Thus, by simply analyzing the public data provided by the different accounts, a case study can be conducted on any particular subject. For example, Martorell and Serra Folch (2018) analyzed the presence of NGOs on Instagram by conducting a case study on the organization Proactiva Open Arms. Their conclusions reveal the evolution of the NGO's activity on the social network, with an increase in the frequency, regularity and use of images, which generated greater audience engagement, thereby boosting profile growth on Instagram.

Furthermore, obtaining permission from a social network, channel or profile not only provides access to the data that can be publicly observed and collated on the social network, but also makes some data available that the different platforms offer privately to content creators or users of social network profiles. To this effect, some researchers have analyzed the public and private data of their own profiles or channels, as they have unlimited access to all the data provided by the platform. For example, Saurabh and Gautam (2019) analyzed their own YouTube channel to demonstrate that the ranking of videos and the number of visualizations follow the Zipf distribution for educational videos. Moreover, they observed a strong correlation between the geographical location of the audience and the location of the industry addressed by the channel.

Nevertheless, as these case studies only involve one or several profiles or accounts, there is a limited amount of data and it is difficult to generalize the results and conclusions of this research. The concordance of the results with other case studies of other social network channels or profiles could indicate the possibility of generalization but, as a very limited sample is involved, this could produce mixed results, which would need to be clarified in the overall discussion of this subject.

Use of samples from other studies

Secondly, one solution to avoid this problem of sample definition in research on social networks is to use samples established by other

studies. An article that has passed the filter of being published by a journal can be assumed to have applied a certain rigor to establishing the sample for a social network study. In this way, the establishment of the sample is easy because it has been obtained directly from another study.

This is the option used, for example, by Buitrago and Torres-Ortiz (2022) in their research on science influencers on Twitch, where they established a series of criteria to filter a list created by Zaragoza and Roca Marín (2020) regarding informative channels on YouTube in Spain. This option led them to conclude that the social network Twitch is an efficient means of drawing young people's attention to scientific content and demonstrating the commitment of the content creators on this platform to a discourse based on scientific rigor.

Some studies even use lists, in part, from less reliable sources such as webpages that do not disclose which methodology they employ, but rather suggest a series of profiles or channels on different study topics. This is the case, for example, for Feijoo Fernández and Fernández Gómez (2021) who use the list provided by a webpage to create part of the sample for their study on child influencers on YouTube and Instagram during the Covid-19 pandemic. The conclusion of this research addresses these content creators' frequent use of pandemic-related themes, through the deployment of scripted stories based on healthy routines, tips and habits.

Bearing this in mind, it should be mentioned that this option to solve the problematic aspect of sample definition in the study of social networks means that the research methodology depends on the methodological basis of other research. Therefore, if the methodology used as a basis for establishing the study sample is not suitably rigorous, all the studies whose samples are taken directly from this sample will not achieve the rigor required by academic and scientific spheres.

Definition of a small sample

Thirdly, there is the possibility of using a relatively small sample to analyze the profiles or channels of social networks and to obtain valuable data for the overall discussion of this subject. This sample can be established randomly, for convenience, or by using some kind of

systematic process that leads to the definition of a sample that we can consider to be relatively small.

This is the case, for example, in the study by Beltrán-Pellicer et al. (2018) that, in order to define the study sample of educational videos about mathematics on YouTube, presented the use of the keywords "proportional distribution" in the YouTube general search engine, and subsequently filtered the list by videos with over a thousand views and by the specific subject addressed in the content. In doing so, their study used a total of 31 videos, the analysis of which led them to conclude that these videos had a low degree of epistemic suitability and that this was not related to the level of popularity and ranking on the platform.

Eslen-Ziya (2022) presents another example of the use of this option, by analyzing a total of 1 295 tweets obtained at random over the course of a week, on the subject of climate change on the social network Twitter. This research reveals the presence of profiles who are skeptical about their country's government taking a more active stance on climate change.

As can be seen in these two examples that may serve as a reference, this methodological option for solving the problem of sample definition in the study of social networks produces studies that take advantage of easy public access to certain data and content on digital platforms to establish a small sample out of convenience or affinity. However, it is important to emphasize that smaller samples make it difficult to generalize results and conclusions. Therefore, overall data collection is limited by the low numbers in the sample.

Definition of a large sample

Fourthly, another solution to the problematic aspect analyzed in this article is the definition of a large sample of profiles or channels on the digital social networks to be studied. As with the definition of small samples, the methodology used may lean towards either convenience or a more systematic process, but the end result is the establishment of a sample that can be considered to be large.

The impossibility of knowing all of the profiles or channels on social networks relating to a topic or subject area means that the larger

the sample is, the closer it is to the total number and the results may be more generalizable and used in an overall discussion (Pihlaja, 2022; Rocheva et al., 2022). The problem with this possibility is that it is more difficult to obtain a sample, as significantly more time is required to conduct the search for channels or profiles relating to the subject under study.

For example, the research by Pattier (2021) acknowledges that it took over two years to complete the search for the sample of educational channels on YouTube according to a series of parameters and guidelines. As we can see, this is a significant amount of time that means that this methodological aspect involves greater difficulty. In this way, the research managed to compile a list of the 204 best educational channels on YouTube Spain, a much larger sample than in most of the studies on the subject. This study proved the existence of a gender gap in the community of edutubers (educational YouTubers) with a significantly greater presence of men than women among creators of audiovisual content on the platform.

As this suggests, the results and conclusions obtained by a study using a sample of 204 educational channels on YouTube are more generalizable than others on the same subject involving a smaller sample (De Souza & Costa, 2021; Dos Santos Galvão, 2019; López et al., 2020; Moreira Queiroga & Spyer Dulci, 2019; Zaragoza & Roca Marín, 2020).

Use of social network web readers or compilers

Fifthly, in recent years a previously undeveloped possibility has emerged, based on the use of pages or platforms that are unrelated to the social network being analyzed, to establish samples from digital social networks using the categorization and systematic search provided by the pages themselves.

This is the case of social network web readers or compilers, such as Social Blade, Noxinfluencer or Feedspot, which enhance this possibility of searching for samples of profiles or channels on digital social networks. By indicating a series of categories on the website or using different keywords, a list is obtained that can be used as the basis for a study sample. This is a replicable methodology, although it should be kept in mind that social networks are constantly changing

(Jones et al., 2015) and therefore search results using the same analysis categories may vary if they are collected at different times (Mahmoudi et al., 2020).

This solution is used, for example, by Ojeda-Serna and García-Ruiz (2022), who used different keywords in the Social Blade tool to identify the museums with the greatest number of subscribers on the YouTube platform and Noxinfluencer to search for youtubers that address the subject. Their work discusses the need for higher education centers and museums to consider the strategies used by youtubers in order to improve their effectiveness as scientific communicators. Likewise, this approach was employed by Vizcaíno-Verdú and Contreras Pulido (2019), who managed to create a list of YouTube channels according to different fields of knowledge, by using the Feedspot tool that considers the impact and measurement of trending algorithmic systems.

Nevertheless, it is important to note that the categorization offered by this type of tool is lacking as far as research methodology is concerned. The categories are too simple and it is not possible to generate new categories or divisions that would enable deeper analysis of the subjects under study. Furthermore, these pages that collect information on social network profiles and accounts use the data provided by the content creators themselves and no human participates in filtering, but rather algorithmic systems read the data. For this reason, profiles that do not specify their content category or that can be placed in different categories may not be included in the list offered by the page. Therefore, and although in methodological terms this solution comes close to being a replicable and reliable process for sample establishment, it still does not cover the potential totality of the population to be analyzed (Pihlaja, 2022).

As can be observed, according to some of the examples outlined in this work, there are channels on YouTube created before 2019 that appear in the study by Pattier (2021), which used the approach of establishing a large sample by means of systematic searches of all kinds, but do not appear in the study by Vizcaíno-Verdú and Contreras Pulido (2019), who deployed the option of the Feedspot tool, thereby demonstrating the inadequacy of the last mentioned possible solution to the sample establishment problem in the study of digital social networks.

In this way, as our analysis reveals, five possibilities have evolved throughout the literature to solve this problem, as shown in Table 1.

TABLE 1
ADVANTAGES AND DISADVANTAGES OF THE SOLUTIONS USED BY THE LITERATURE TO THE PROBLEM OF SAMPLE DELIMITATION IN THE STUDY OF DIGITAL SOCIAL NETWORKS

Solution	Advantages	Disadvantages
Case study	Ease of access. Choice by convenience. Possibility of obtaining private data.	Difficulty of generalizing results.
Use of samples from other studies	Ease of definition of the sample.	Dependence on the methodological rigor of other research.
Definition of a small sample	Ease of searching the sample. Choice by convenience.	Difficulty of generalizing results.
Definition of a large sample	Viability of generalization of results.	Difficulty of searching for the sample. Spending more time searching.
Use of social network web readers or compilers	Ease of searching the sample.	Poor categorization. Continuous updating of readers or compilers. Possible generalization errors.

Source: Own elaboration.

DISCUSSION

The widespread use of social networks among the general population (Kemp, 2022) has produced numerous studies on a considerable number of subjects and fields of knowledge (Brooks et al., 2022; Faciroli et al., 2022; Fernández et al., 2021; Fernández-Peña et al., 2022; Kumar

et al., 2021; Marcelo & Marcelo, 2021; Prochnow & Patterson, 2022; Tobón Berrio et al., 2022).

The establishment of a sample, as part of the methodological process of such studies, is fundamental to obtain a reliable basis for results that promote the knowledge that we have about the different study topics (Vásquez, 2022). Therefore, it is important to analyze and reflect on the rigor and possibilities provided by the different solutions that have arisen over time with respect to this problematic aspect of research into social networks.

It is not a question of analyzing which solution is better or which is worse. As can be seen in Table 1, all the approaches present advantages and disadvantages, given that, at present, we do not have the necessary tools to obtain the complete sample of profiles or channels relating to a specific subject.

Moreover, depending on the research questions and the researchers' aims, it may be that the case study or the establishment of small samples achieve the results needed to answer these questions and to advance knowledge of the subject further (Beltrán-Pellicer et al., 2018; Eslen-Ziya, 2022; Martorell & Serra Folch, 2018; Saurabh & Gautam, 2019). Hence the importance of reflection and limiting the research objectives (Pihlaja, 2022).

The problem may arise when generalizing results, where studies with larger samples, whether established through searches on the social network itself or through the use of social network web readers or compilers, could offer more rigorous results in this respect (Pattier, 2021; Vizcaíno-Verdú & Pulido, 2019). In recent years, a tendency has emerged to use artificial intelligence and data mining in an attempt to detect patterns in large amounts of data sets, which could benefit the interpretation of results on a large scale, but these data are the result of using a sample for which the total population is unknown (Hassan & Behadili, 2022; Jones et al., 2015; Pihlaja, 2022).

Certainly, the possibility of establishing a systematic, replicable search process for samples in studies on digital social networks is an aspect that should be evaluated and, therefore, the use of web readers or compilers could come close to solving the problem, although, as we mentioned previously, it would be necessary to make a series of changes that we will outline below.

Firstly, there should be greater categorization in searches. To solve this difficulty, digital social networks should offer content creators the possibility of categorizing the type of resource they upload in a more specific way. For example, YouTube could offer the option for content creators of not only indicating that their channels are aimed at education but also of indicating the field of knowledge. We could thereby obtain more reliable data, for example, from YouTube channels aimed at science dissemination (Buitrago & Torres-Ortiz, 2022; Zaragoza & Roca Marín, 2020).

Secondly, and as we indicated earlier, there is the possibility of profiles or channels on social networks not sharing their categorization as information for use by web readers or compilers, or being in different categories to those indicated by the user. To this effect, and in line with the approach outlined by Osman et al. (2022) in terms of ranking on digital social networks, one suggestion could be the participation of experts who would judge the different profiles and channels by introducing these evaluation data into the classification algorithm. In this way, the lists offered by the pages that provide tools such as social network web readers or compilers would pass through a methodological filter that would be a considerable improvement on the current results.

Thus, the problem regarding sample establishment in studies on social networks has been addressed by research throughout the literature, using different techniques, and we should acknowledge, as of today, that there is still no solution. For this reason, the approach described in this work is significant as it would greatly improve the use of web readers or compilers on technological social networks.

At present, it seems that we are at a dead end, since there is no real possibility of obtaining complete, rigorous lists on different subjects of fields of knowledge from digital social networks. However, and as we have proved in this work, research continues to seek creative ways to mitigate this methodological problem and solve, at least partially, this difficulty (Flores-Márquez & González Reyes, 2021). The focus is now on the future, where this possibility may arise through the improvement of one of the options used in research into social networks over the years or the appearance on the scene of a new possibility due

to advances in technological innovation, where the use of artificial intelligence is an element to be taken into consideration.

Conclusions

We can confirm that there is a methodological problem in terms of sample establishment in studies on digital social networks that use profiles, accounts or channels. As of today, it is not possible to obtain suitably rigorous lists of topics or subject areas and, therefore, research that has conducted this type of studies on social networks has needed to juggle several methodological options, as categorized in this article.

Firstly, there is the case study, used fairly frequently in the literature due to the ease of accessing one or several profiles and channels and the possibility of choosing the case study for the sake of convenience. Although it provides valuable data, it is important to note the difficulty in generalizing the results, which is inherent to this methodological approach.

Secondly, there is the use of a sample from other studies, which involves ease of sample establishment, given that it is already available in the publications consulted. Nevertheless, this methodological option means that the rigor of the methodology depends on the process of obtaining the sample from these publications.

Thirdly, there is the definition of a small sample. The impossibility of deploying search mechanisms with social network web compilation tools before these existed, along with the non-existence of reliable, official lists, has fostered this methodological option of obtaining a small sample with ease and for the sake of convenience. However, precisely for this very reason, there is a limitation to generalizing the results.

Fourthly, there is the definition of a large sample that provides viability to the generalization of results. The larger the sample is, the less likely the occurrence of the typical deviations that occur when using small samples. Even so, it is advisable to consider the difficulty involved in finding the sample and the need to spend a longer time on the investigation process.

Fifthly, there is the use of social network web readers or compilers, which provide a replicable, systematic option for obtaining samples

in a simple way. Despite this, it should be noted that, these days, the categorization of digital social networks and, therefore, of these tools, is inadequate. Contents, moreover, are constantly being updated, which makes it necessary to specify the time the search was conducted. And, lastly, and taking these two points into account, we cannot forget that the lists provided by these pages are neither as complete nor as accurate as academic research requires. For this reason, we suggest two possible improvements to this methodological option: greater categorization of profiles and accounts on digital social networks, and the participation of experts to filter these lists in order to provide more rigorous data.

The limitation of this research lies in the vast number of studies on social networks that exist within the literature. It may be possible to complete our categorization of the possibilities and approaches that have been created in an attempt to solve the problem of sample establishment in such studies by considering some new aspect that has not been considered in this article. Regarding the future of this work, we suggest that interviews be conducted with researchers who have used several of the possibilities categorized in this article during the course of their academic career. Nobody is better placed than they are to examine, in even more depth, if possible, the advantages and disadvantages of each of the methodological options.

Bibliographic references

Beltrán-Pellicer, P., Giacomone, M. B. & Navarro, M. B. (2018). Online educational videos according to specific didactics: The case of Mathematics. *Culture and Education*, *30*(4), 633-662. https://doi.org/10.1080/11356405.2018.1524651

Brooks, H., Devereux-Fitzgerald, A., Richmond, L., Bee, P., Lovell, K., Caton, N., Cherry, M. G., Edwards, B. M., Downs, J., Bush, L., Vassilev, I. & Young, B. (2022). Assessing the effectiveness of social network interventions for adults with a diagnosis of mental health problems: a systematic review and narrative synthesis of impact. *Social Psychiatry and Psychiatric Epidemiology*, *57*(5), 907-925. https://doi.org/10.1007/s00127-022-02242-w

Buitrago, A. & Torres-Ortiz, L. (2022). Influencers de ciencia en Twitch. Divulgación científica a través de vídeo-streaming en tiempos de

COVID-19. *Teknokultura*, *19*(2), 165-176. https://doi.org/10.5209/tekn.77941

Burgos, M., Beltrán-Pellicer, P. & Godino, J. D. (2020). La cuestión de la idoneidad de los vídeos educativos de matemáticas. *Revista Española de Pedagogía*, *78*(275), 27-50. https://doi.org/10.22550/rep78-1-2020-07

Del Sol Fabregat, L. A., Tejeda Castañeda, E. & Mirabal Díaz, J. M. (2017). Los métodos teóricos: una necesidad de conocimiento en la investigación científico-pedagógica. *Edumecentro*, *9*(4), 250-253. https://revedumecentro.sld.cu/index.php/edumc/article/view/934

De Souza, S. T. & Costa, P. H. (2021). Educommunicative culture and new narrative creations. The use of video and storytelling language in an educative context. *Indagatio Didactica*, *13*(3), 215-232. https://doi.org/10.34624/id.v13i3.25539

Dos Santos Galvão, N. M. (2019). The YouTube as a study tool for the sufficiency test in accounting. *Revista EDaPECI*, *19*(2), 76-84. https://periodicos.ufs.br/edapeci/article/view/10738

Eslen-Ziya, H. (2022). Humour and sarcasm: expressions of global warming on Twitter. *Humanities and Social Sciences Communications*, *9*(1), 1-8. https://doi.org/10.1057/s41599-022-01236-y

Faciroli, J., Freguglia, R. D. S., Sirqueira, T. F. M., Vieira, M. D. T. (2022). Social networks effects on outcomes of government programs: a systematic review. *Brazilian Journal of Political Economy*, *42*(1), 222-243. https://doi.org/10.1590/0101-31572022-3056

Feijoo Fernández, B. & Fernández-Gómez, E. (2021). Niños y niñas influyentes en YouTube e Instagram: contenidos y presencia de marcas durante el confinamiento. *Cuadernos.info*, (49), 300-328. https://doi.org/10.7764/cdi.49.27309

Fernández, V. F., Suárez, M. Á. S. & Machas, R. C. (2021). Communication in museums through social media during the pandemic: unveiling new opportunities for interaction. *Fonseca, Journal of Communication*, *23*, 129-149. https://doi.org/10.14201/fjc202123129149

Fernández-Peña, R., Orvalle-Perandones, M. A., Marqués-Sánchez, P., Ortego-Maté, C. & Serrano-Fuentes, N. (2022). The use of social network analysis in social support and care: a systematic sco-

ping review protocol. *Systematic Reviews*, *11*(1), 1-6. https://doi.org/10.1186/s13643-021-01876-2

Flores-Márquez, D. & González Reyes, R. (2021). En busca de coordenadas metodológicas para estudiar la cultura digital. In D. Flores-Márquez & R. González Reyes (Coords.), *La imaginación metodológica. Coordenadas, rutas y apuestas para el estudio de la cultura digital* (pp. 15-23). Tintable.

Hassan, F. & Behadili, S. F. (2022). Modeling Social Networks using Data Mining Approaches-Review. *Iraqi Journal of Science*, *63*(3), 1313-1338. https://doi.org/10.24996/ijs.2022.63.3.35

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), 15-38. https://doi.org/10.22430/22565337.685

Jones, R., Chik, A. & Hafner, C. (Eds.). (2015). *Discourse Analysis and Digital Practices*. Routledge.

Kemp, S. (2022). *Digital 2022 Global Overview Report*. https://bit.ly/3OVJRYP

Kumar, N., Oles, W., Howell, B., Janmohamed, K., Lee, S. T., Funaro, M. C., O'Connor, P. G. & Alexander, M. (2021). The role of social network support in treatment outcomes for medication for opioid use disorder: A systematic review. *Journal of Substance Abuse Treatment*, *127*, 1-9. https://doi.org/10.1016/j.jsat.2021.108367

López, L., Maza-Córdoba, J. & Tusa, F. (2020). Educar en el contexto digital: el reto de ser edutuber. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E25), 188-200. https://bit.ly/3qnyHBS

Mahmoudi, A., Yaakub, M. R. & Abu Bakar, A. (2018). A new method to discretize time to identify the milestones of online social networks. *Social Network Analysis and Mining*, *8*(1), 1-20. https://doi.org/10.1007/s13278-018-0511-4

Mahmoudi, A., Yaakub, M. R. & Abu Bakar, A. (2020). A new real-time link prediction method based on user community changes in online social networks. *The Computer Journal*, *63*(3), 448-459. https://doi.org/10.1093/comjnl/bxz050

Marcelo, C. & Marcelo, P. (2021). Educational influencers on Twitter. Analysis of hashtags and relationship structure. *Comunicar*, *68*, 73-83. https://doi.org/10.3916/C68-2021-06

Martorell, C. & Serra Folch, C. (2018). La presencia de las ONG en Instagram: estudio de caso de Proactiva Open Arms. *Trípodos*, *43*, 83-97. http://hdl.handle.net/20.500.12328/1363

Moreira Queiroga, T. & Spyer Dulci, T. M. (2019). "Teachers youtubers": Analysis of three channels of YouTube dedicated to the teaching of History. *Escritas do Tempo*, *1*(1), 4-29. https://doi.org/10.47694/issn.2674-7758.v1.i1.2019.0429

Mucha-Hospinal, L. F., Chamorro-Mejía, R., Oseda-Lazo, M. E. & Alania-Contreras, R. D. (2021). Evaluación de procedimientos empleados para determinar la población y muestra en trabajos de investigación de posgrado. *Desafíos*, *12*(1), 44-51. https://doi.org/10.37711/desafios.2021.12.1.253

Ojeda-Serna, V. & García-Ruiz, R. (2022). Divulgación científica en YouTube en Latinoamérica. Estudio de Casos de universidades, museos y YouTubers. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, *19*(2), 2204-2204. https://doi.org/10.25267/rev_eureka_ensen_divulg_cienc.2022.v19.i2.2204

Osman, W., Mohamed, F., Elhassan, M. & Shoufan, A. (2022). Is YouTube a reliable source of health-related information? A systematic review. *BMC Medical Education*, *22*(1), 1-12. https://doi.org/10.1186/s12909-022-03446-z

Pattier, D. (2021). The gender gap among EduTubers and the factors significantly influencing it. *Journal of New Approaches in Educational Research*, *10*(2), 313-329. https://doi.org/10.7821/naer.2021.7.732

Phua, J., Jin, S. V. & Kim, J. J. (2016). Gratifications of Using Facebook, Twitter, Instagram, or Snapchat to Follow Brands: The Moderating Effect of Social Comparison, Trust, Tie Strength, and Network Homophily on. *Telematics and Informatics*, *34*(1), 412-424. https://doi.org/10.1016/j.tele.2016.06.004

Pihlaja, S. (2022). Data sampling and digital discourse. In C. Vásquez (Ed.), *Research methods for digital discourse analysis* (pp. 55-71). Bloomsbury Publishing.

Piña-García, C. A., Gershenson, C. & Siqueiros-García, J. M. (2016). Towards a standard sampling methodology on online social networks: collecting global trends on Twitter. *Applied Network Science*, *1*(3), 1-19. https://doi.org/10.1007/s41109-016-0004-1

Prochnow, T. & Patterson, M. S. (2022). Assessing Social Network Influences on Adult Physical Activity Using Social Network Analysis: A Systematic Review. *American Journal of Health Promotion*, *36*(3), 537-558. https://doi.org/10.1177/08901171211060701

Rocheva, A., Varshaver, E. & Ivanova, N. (2022). Targeting on Social Networking Sites as Sampling Strategy for Online Migrant Surveys: The Challenge of Biases and Search for Possible Solutions. In S. Pötzschke & S. Rinken (Eds.), *Migration Research in a Digitized World*. IMISCOE Research Series, Springer, Cham. https://doi.org/10.1007/978-3-031-01319-5_3

Saurabh, S. & Gautam, S. (2019). Modelling and statistical analysis of YouTube's educational videos: A channel Owner's perspective. *Computers & Education*, *128*, 145-158. https://doi.org/10.1016/j.compedu.2018.09.003

Sirisala, N., Yarava, A., Reddy, Y. P. & Poola, V. (2022). A novel trust recommendation model in online social networks using soft computing methods. *Concurrency and Computation: Practice and Experience*, *34*(22), e7153. https://doi.org/10.1002/cpe.7153

Tindall, D., McLevey, J., Koop-Monteiro, Y. & Graham, A. (2022). Big data, computational social science, and other recent innovations in social network analysis. *Canadian Review of Sociology/Revue Canadienne de Sociologie*, *59*(2), 271-288. https://doi.org/10.1111/cars.12377

Tobón Berrio, L. E., Sabatier, C., Palacio Sañudo, J. E. & Navarro, O. (2021). The social representation of the rights of children: legality built by fathers and mothers. *Revista Colombiana de Sociología*, *44*(2), 243-265. https://doi.org/10.15446/rcs.v44n2.87666

Vásquez, C. (Ed.). (2022). *Research methods for digital discourse analysis*. Bloomsbury Publishing.

Vatrapu, R., Mukkamala, R. R., Hussain, A. & Flesch, B. (2016). Social set analysis: A set theoretical approach to big data analytics. *IEEE Access*, *4*, 2542-2571. https://doi.org/10.1109/access.2016.2559584

Vizcaíno-Verdú, A. & Contreras Pulido, P. (2019). From the classroom to the universal online school: Thematic dimensions on YouTube channels. *HAMUT'AY*, *6*(3), 12-25. http://dx.doi.org/10.21503/hamu.v6i3.1841

Wang, C., Tianqing, Z., Xiong, P., Ren, W. & Choo, K. K. R. (2022). A privacy preservation method for multiple-source unstructured data in online social networks. *Computers & Security*, *113*, 102574. https://doi.org/10.1016/j.cose.2021.102574

Zaragoza, J. C. & Roca Marín, D. (2020). El movimiento youtuber en la divulgación científica española. *Revista Prisma Social*, (31), 212-238. https://revistaprismasocial.es/article/view/3942

Zhang, C., Sun, J., Zhu, X. & Fang, Y. (2010). Privacy and security for online social networks: challenges and opportunities. *IEEE Network*, *24*(4), 13-18. https://doi.org/10.1109/mnet.2010.5510913

DANIEL PATTIER

PhD in education and assistant professor in the Department of Educational Studies of the Faculty of Education –Teacher Training Center of the Complutense University of Madrid. His main line of research focuses on the study of the use of social networks in the educational field. He is the author of several scientific papers published in indexed journals.