




Big Data and Business Intelligence on Twitter and Instagram for digital inclusion

Big Data y Business Intelligence en Twitter e Instagram para la inclusión digital

-  Carlos Barroso-Moreno. Predoctoral Research Fellow, Department of Educational Studies, Complutense University of Madrid (Spain) (carbarro@ucm.es) (<https://orcid.org/0000-0002-1609-2267>)
-  Dr. Laura Rayon-Rumayor. Professor, Department of Educational Studies, Complutense University of Madrid (Spain) (larayon@ucm.es) (<https://orcid.org/0000-0003-0339-8221>)
-  Dr. Antonio Bautista García-Vera. Full Professor, Department of Educational Studies, Complutense University of Madrid (Spain) (bautista@edu.ucm.es) (<https://orcid.org/0000-0002-5194-7419>)

ABSTRACT

Social media can contribute to an inclusive society, but they are also asymmetrical and polarised communication spaces. This requires competent teachers to build critical digital citizenship. The aim of this article is twofold: to present web scraping and text analytics as tools that define teachers' digital competences, and to investigate which posts on Twitter and Instagram are most viral in relation to education, disability and inclusion. A total of 48,991 publications in Spanish and English were analysed, corresponding to the period from 13 October 2021 to 1 May 2022. The 100 most viral posts were selected, and correlations were identified between the sentiment, gender and influence associated with the content, its temporal and geographic space. The results show that economic and political influence groups are the most viral, relegating non-profit organisations or individuals with altruistic outreach to second place; only on international days is this trend reversed. Bots do not interfere to impose messages; it is artificial intelligence algorithms that overshadow vindictive and humanistic content. The most influential people are predominantly male, associated with institutional accounts in the political sphere. It is concluded that Big Data and Business Intelligence tools help teachers to analyse relevant educational and social issues, and to acquire a collective ethic in the face of new educational challenges.

RESUMEN

Las redes sociales pueden contribuir a una sociedad inclusiva, pero también son espacios de comunicación asimétricos y polarizados. Ello requiere de un profesorado competente para la construcción de una ciudadanía digital crítica. Este artículo tiene un doble objetivo, presentar las técnicas «Web scraping» y «text analytics» como herramientas que definen competencias digitales docentes, e indagar sobre qué publicaciones, en Twitter e Instagram, son más virales en relación con educación, discapacidad e inclusión. Se analizaron 48.991 publicaciones en español e inglés, correspondientes al periodo del 13 de octubre de 2021 al 1 de mayo de 2022. Se seleccionaron las 100 más virales, e identificaron las correlaciones entre el sentimiento, género e influencia asociado al contenido, su espacio temporal y geográfico. Los resultados evidencian que los grupos de influencia económica y política son los más virales, relegando a un segundo plano a las organizaciones sin ánimo de lucro o particulares con difusión altruista; solo en los días internacionales se invierte esta tendencia. Los «bots» no interfieren para imponer mensajes, son los algoritmos de inteligencia artificial los que opacan contenido reivindicativo y humanístico. Las personas más influyentes tienen una prevalencia de género masculino asociadas a cuentas institucionales de ámbito político. Se concluye que las herramientas de «Big Data» y «Business Intelligence» ayudan al profesorado a analizar temas educativos y sociales relevantes, y a adquirir una ética colectiva frente a los nuevos retos educativos.

KEYWORDS | PALABRAS CLAVE

Social network analysis, Big Data, education, disability, digital inclusion, influence groups.
Análisis de redes sociales, Big Data, educación, discapacidad, inclusión digital, grupos de influencia.

1. Introduction and state of the art

Current reports (We are social, 2021; Ditrendia, 2020) point to the exponential increase of users connected to social networks worldwide, being not only individuals, but also professional and institutional groups and media that impact on the construction of reality and how meanings are shared (Dellwing, 2021; Del-Fresno-García, 2014, 2019; Ladogina et al., 2020).

In terms of studies on the functions, roles and relationships between members (Awidi et al., 2019; Brunner et al., 2019; De-Groot et al., 2022; Grace et al., 2019; Tuzel & Hobbs, 2017; White & Forrester-Jones, 2020), we are particularly concerned with the approach that emphasises social networks as asymmetric spaces of communication. Specifically, Barberá (2015), Barberá et al. (2015) and Brady et al. (2019) point out that users share messages that represent beliefs, opinions and values that they endorse and follow profiles they trust ideologically. Another relevant factor that conditions communication is the moral-emotional language that political leaders use and its moral contagion effect, as well as the asymmetry in communication between content creators and their potential followers, depending on the content of the messages, an asymmetry amplified by bots (Robles et al., 2022).

There is no doubt about Twitter's potential for certain users and groups, such as political elites and corporations, to reach large audiences, potential voters and consumers, through direct and indirect links of influence that define interaction on this network, a predictive factor of influence and social impact outside the network (Brady et al., 2019). In this context, hate speech is of particular relevance, as it is generating significant polarisation among people based on their ideology, with effects on the "selective perception bias" that favours the positive evaluation of the message of issuers with whom there is ideological affinity and the rejection of speeches with an opposing ideology. This is an important predictive factor for behaviour towards certain offline groups, which are represented through exclusionary and anti-democratic messages on social media (Ortega-Sánchez et al., 2021).

To address this concern, we understand that teachers have a key role to play in building a digital citizenship with the capacity to participate in an informed and responsible way, and thus contribute to a democratic and inclusive networked society (Bautista, 2021; Carlsson, 2019; Ortega-Sánchez et al., 2021). As evidenced by Tuzel and Hoobs (2017: 64), the use of social media for intercultural citizenship requires teachers to have a "solid appreciation of the asymmetries and inequalities inherent in information flows", and an understanding of how these digital platforms function as spaces for dissemination, amplification of ideas and mobilisation of actions for groups and individuals with unequal rights, such as those with disabilities, in order to disseminate and critique ideas, as well as publicise their achievements for a more inclusive society (Hemsley et al., 2018).

So, how should teachers work on these digital competences that prepare them for inclusive education? What tools and procedures will help them in the complex and beautiful task of knowing the meanings, beliefs and attitudes that circulate on social networks with thousands and thousands of participants? We understand that an appropriate way to investigate and generate the knowledge that teachers must have in order to promote relationships that lead to creating feelings of inclusion and belonging of students to the reference group, is through the tracking of social networks with techniques and tools such as Web Scrapping and text analytics that we present in this article.

In view of the above, concerns and proposals derived from the review of the state of the art, this article has a dual purpose. One, to exemplify in a research context the use of these tools that make up one of the digital teaching competences to encourage debate, and to inspire and illuminate evidence of the value of the techniques mentioned in the analysis of social media and interactive data visualisation. The other is to answer three questions/hypotheses on the processes of asymmetric relationships that help to better understand the processes of remodelling and relegation of these subjects on social media, useful knowledge for defining the content of the digital teaching competence in particular:

- H1: The influence of economic and political power groups together with an interest of digital platforms reshape the issues associated with education, disability and inclusion according to their own interests.
- H2: Artificial intelligence algorithms relegate minority ideas or altruistic broadcasts to second or third place.

- H3: Posts with negative sentiment are associated with political groups and leaders, compared to positive posts linked to associations and individual content creators advocating for the human rights of persons with disabilities, with no influence by gender.

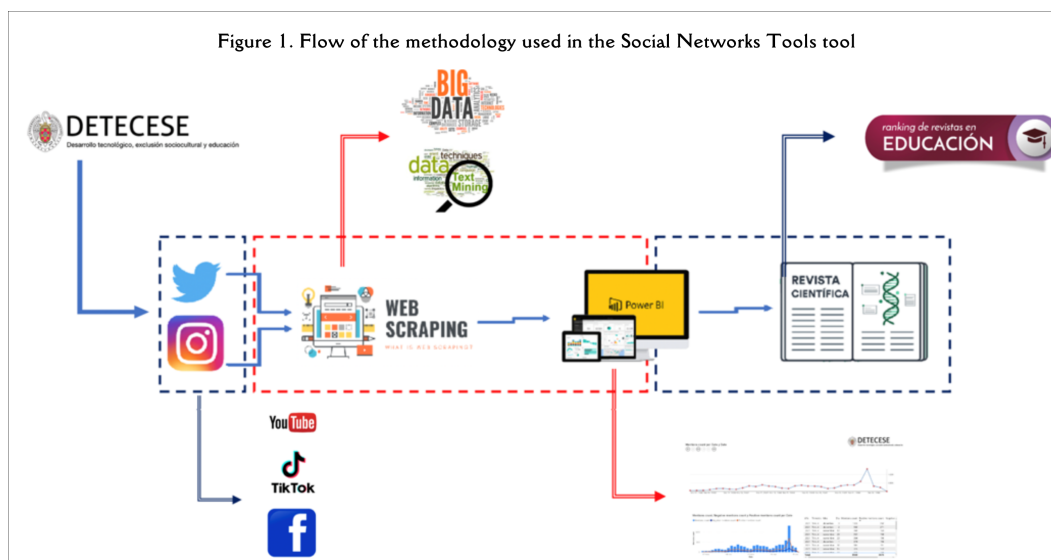
An innovative contribution of the article is the access to the data and results obtained, which are made available to researchers, teachers and other professionals so that they can dynamically and interactively read them.

2. Material and methods

In this article we analyse which posts on Twitter and Instagram are most viral in relation to education and disability and inclusion: in this sense, it will allow us to better identify correlations between content, sentiment, gender, influence, and their temporal and geographic space. The database is composed of posts downloaded from both social networks on education content related to inclusion or disability over 200 days. Social Big Data Analysis techniques are applied, such as Web Scraping techniques to extract the information and analyse it with Big Data and Text Mining algorithms. Subsequently, the results are represented using the Power BI business tool, allowing readers to interact dynamically.

2.1. Process flow in the methodology

The set of research processes, from the origin of the data to the graphic representation, is conceptually referred to as the "Social Networks Tools", which is represented in Figure 1. The processes are divided into two large blocks; those delimited in blue are modifiable by the researcher, such as the origin of the data or the keywords used. The block delimited with red dashed lines refers to Web Scraping techniques, Big Data algorithms, Data Mining and part of the Business Intelligence tool, belonging to the research group with intellectual property, in which the interested parties cannot make modifications or have full access.



The first process corresponds to the selection of social networks with the possibility of data extraction, such as Twitter, Instagram, YouTube, TikTok or Facebook. At this point, the necessary keywords are selected in order to consider the storage of the information. The second process applies Web Scraping techniques for data extraction in social media with Python, which literally means scraping from the web (Mitchell, 2018). Subsequently, Big Data techniques are applied with the information obtained to process the large amount of data collected, and data mining is used to analyse text from the publications. Then, variables specific to the research on the communication profile such as gender, professional activity, geolocation and temporal space are included. The next process is to load the data into the Microsoft Power BI data analysis service in order to provide interactive visualisations for researchers and faculty in order to generate their own reports and dashboards (Becker & Gould., 2019). Finally, the user interprets

the generated graphs for the understanding of the subject matter addressed, in search of a scientific and/or educational use, with a real impact on society.

Extract, Transform and Load (ETL) processes allow data from multiple sources to be mixed. In our particular case, data from Twitter and Instagram is extracted, transformed by Python script with a series of requirements and loaded into a data warehouse to channel the data in a homogeneous way and analyse it using different algorithms.

2.2. Data extraction

Twitter data is downloaded from the Twitter API using the "tweepy" library and Instagram data is downloaded using private algorithms developed specifically for the research. In both cases, real-time data is collected using the Python programming language. Given the need to capture all traffic in real time, unlike most previous studies, we use posts that have the demanded keywords; we do not focus on what a particular group of profiles are saying to avoid bias. This has another implication: the data capture starts at the moment of the activation of these words; we do not collect data retroactively.

The database generated consists of publications in the period from 13 October 2021 to 1 May 2022, with a total of 48,991 publications in the Spanish and English-speaking world. The requirement is to contain a number of keywords regardless of the user disseminating them. The common keyword used is education, and at least one disability or inclusion word, as well as words derived from them. In addition, words without accents are included due to the virality of some news items with such spelling. Therefore, any publication that has the word 'education' together with one or more words for disability or inclusion on the social network Twitter or Instagram is considered. The choice of the word 'education' together with 'disability' or 'inclusion' is motivated by the fact that certain publications that deal with disability do not refer to this term, but to the idea of inclusion; in addition, inclusion does reflect different social sensitivities around the topic addressed. The publications are updated with all the necessary information one week after publication; this situation does not significantly affect the data because the publications have a significant impact in the first days of dissemination, which is empirically proven. In fact, the reader can corroborate in later sections how the number of likes, followers, followings or comments does not increase significantly; therefore, this situation does not influence the subsequent analysis.

2.3. Data transformation

Once the publications were stored in raw form, data mining was performed with Natural Language Processing algorithms from the NLTK and Scikit-Learn Python libraries, a branch of artificial intelligence to determine the interaction between computers and humans (Cheng & Tsai., 2019). At this point, two transformations are differentiated: manual and automatic. The automatic transformations are applied to the 48,991 research publications; the manual transformations are applied by the research team to the 100 publications of greatest interest with particular classifications for the subject matter addressed. All the variables available in the database are detailed by group, type, category and example (Table 1).

2.3.1. Automatic

The application of text mining begins with the process of tokenisation of the content, which allows words to be separated by the spaces that make up the sentence. Next, the words known as "stopwords" are eliminated, consisting of prepositions, determiners or particular words, among others. This separation and cleaning of the text allows for the analysis of repetition frequency, word clouds, etc. Regarding sentiment analysis, Liu's (2010) dictionary is applied to detect the positive, negative or neutral content of words, providing a final value to the sentence as a whole. Among the limitations of this type of dictionary is the invisibility or misclassification of ironies or puns. In reference to the detection of thematic topics, tweets are grouped by the hashtags for Instagram content, depending on the most frequent words, so that they can be grouped under one of the hashtags. Based on bot detection, the Botometer API for Twitter profiles is used to extract more than 1,200 features such as activity patterns, language, sentiment, social structure or friends, assigning a 1 if it is a bot or a 0 if the account is real. High criteria are set to consider it a bot and not to include real users in that category. Therefore, a cross validation for the Area Under the ROC Curve (AUC) of 0.99 is established.

2.3.2. Manual

In order to answer the hypotheses, put forward, it is necessary to generate specific variables, which cannot be automated. The research team distributes the labelling with homogeneous criteria to ensure a correct classification. This requires a thorough analysis of user profiles, based on tracking publicly disseminated information on other networks or personal blogs to determine variables such as: gender, field, profession, multimodality or estimated locations.

2.4. Data upload

To analyse the data, once the data mark, known as education, is available, with all the variables cleaned, the data is loaded into the Power BI programme. The objective is to analyse the data and indicators to test hypotheses through different visual data analysis. Some of the graphs deal with simple and complex statistics, including maps (geographical or heat maps) on the locations of the most viral senders (Arcila-Calderón et al., 2022), statistical correlation to analyse whether there is a strong or weak relationship, multidimensional graphs, time series or word clouds, among others.

Group of variables	Variables	Categories	Type	Examples
Identifier	Identifier	-	Numeric	02186348225, 31186048213...
	Label		Categorical	T associated with Top 10 V associated with viral 10 to 50 P associated with publications 50 to 100
	Profile name	-	Text	Mzavalagc
	Real name	-	Text	Margarita Zavala
Transactional data	Origin	-	URL	www.instagram.com/p/CZKkN4jp13G/
	Domain		Categorical	Twitter or Instagram
Temporary	Day	200	Date	2022-01-24
	Time	1440	Date	21:32
Content	Title	-	Text	Real inclusion
	Content	-	Text	Inclusion is achieved through...
	Hashtags	-	Text	#education #education #inclusion
Virality	Retweets (twitter)	-	Numeric	1957, 34...
	Comment (Instagram)	-	Numeric	33971, 4664...
	Reproductions (Video)	-	Numeric	139103, 83400...
	Following	-	Numeric	826, 453...
	Follower	-	Numeric	1596670, 320...
	Publications (Instagram)	-	Numeric	6068, 234...
Behaviour	Sentiment		Categorical	Positive, neutral and negative
	Profession	>10	Categorical	Politician, digital creator, teacher...
Scope	Scope		Categorical	1. Institutional 2. Group 3. Individual
	Subarea		Categorical	11 Local, 12 Supranational (political), 13 Media 21 Academies, clinics, or companies 22 Non-profit associations 23 Science communicators 31 Professional 32 Professional with disabilities 33 Particular
Multimodality	Modal		Categorical	Single-modal and multimodal
	Submultimodal		Categorical	Text and video; text and image; text audio and audio and image
Demographic	Gender		Categorical	Male, female and neutral
	Estimated location of the person	195	Category	United States, Spain

3. Analysis and results

The publication of the data can be found in the Power Bi tool, which should preferably be opened with the Microsoft Edge browser from a computer (<https://bit.ly/3z9wDU6>), or with the QR Code, presented in Figure 2.

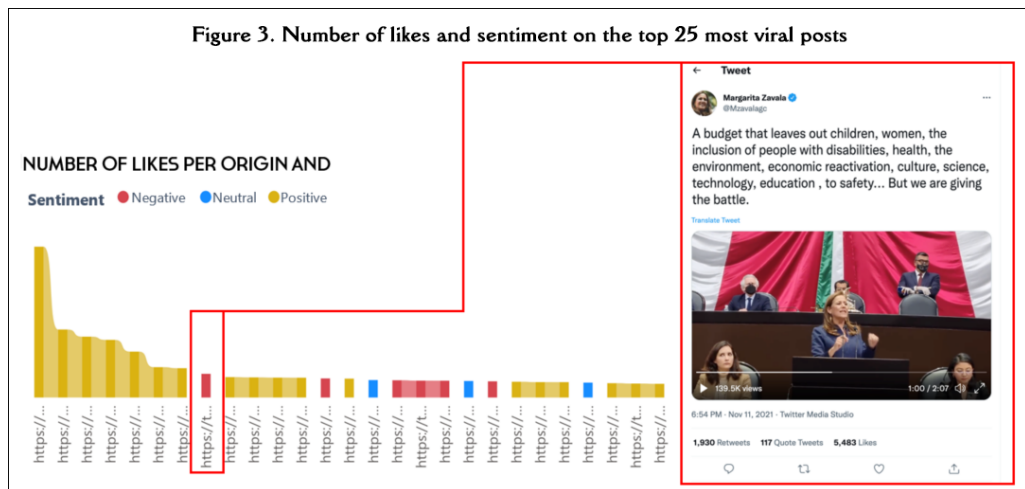
We extract an initial snapshot of the publications (N=48,991), which are distributed on Twitter with 59.38% (N=29,095) and Instagram with 40.61% (N= 19,896). Twitter is the most popular, but Instagram is the most popular in terms of the number of likes and comments.

The results of the analysis of the most viral publications, sentiment and the type of associated profile, show five main ideas presented in Figure 3. (i) The publications with polarised sentiment are the most viral, specifically the positive ones, and of the 10 most viral, 9 are positive. (ii) The 25 most viral publications are mainly campaigns by relevant power groups, orchestrated by multinationals, politicians,

influencers or digital creators with economic interests. Proof of this is the most viral communications [T1] (<https://bit.ly/3z5bEBO>), [T4] and [V11] related to McDonald's, located in accounts that do not claim any social or educational aspect in relation to disability and inclusion, published on the same date, with the same image and text content.

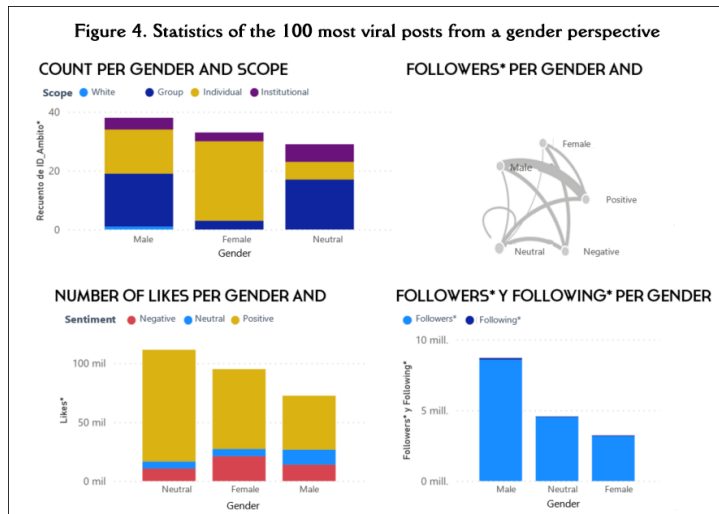


To orchestrate such an impactful campaign requires accounts with thousands of followers; even if it has a low diffusion rate per follower the total impact is very high. (iii) We have to relegate ourselves to position 6 [T6] and 28 [V28] by number of likes to find a social media space. However, [T6] is defined as a space for learning about diversity and LGBTQ+; however, some exclusive content requires a financial outlay (<https://bit.ly/3N0Vv4G>). One has to go down to position 34 [T34] to find the Only You Are Missing Foundation, a non-profit Non-Governmental Organisation (NGO) for autism spectrum awareness (<https://bit.ly/3t2Gpno>). (iv) Cases of denunciation of situations of exclusion are almost non-existent and do not have a presence on social networks, mainly due to denunciation from networks with a small number of followers, and a neutral and institutional feel to the denunciations. (v) Mostly, the most disseminated messages are positive or negative, but neutral messages are not so widespread. In light of these results, we can affirm that the hypothesis (H1) is supported: certain groups and political leaders reshape the issues analysed for their own interests, relegating groups and individual actors with altruistic interests to the background.



The results of analysing the 100 most viral posts from a gender perspective are shown in Figure 4, and are distributed as follows: 38% male, 33% female and 29% neutral. Significant results are as follows: (i) Male accounts have more followers and their comments are harsher (negative) than those of the female gender, the latter being associated with positive comments. In figures, the male gender has 8,582,921 followers and 118,981 following; the female gender has 4,523,345 followers and 43,307 following. (ii) Female accounts have fewer followers, but more likes. (iii) Institutional accounts disseminate more neutral or positive messages. (iv) The male gender has the most institutional accounts on social media. Therefore, hypothesis (H3) is only partially confirmed in that negative sentiments are linked to influential groups

and political leaders, but these accounts are male, which is evidence of a gendered influence. Figure 4 shows the correlation between the variables gender, domain, feelings and multimodality. The results are as follows: (i) Through the nodes we observe a relationship (thick thickness) between the institutional groups and the male gender. The female gender has less presence in the institutional sphere, it appears linked to individual profiles (thin thickness). (ii) Nicolás Maduro has 4 million followers, a paradigmatic case that contributes to the association between the male gender and institutional variables, as happens with other political leaders. The publication associated with this case [P80] compiles the publication (<https://bit.ly/3N0QOro>), and stands out for its low dissemination ratio with 824 likes in relation to its high number of followers. (iii) Most are multimodal messages (text and image) and monomodal publications are scarce.



With regard to the detection of bots, with a probability of over 98% that they correspond to automated accounts, 23 accounts were detected. This represents less than 0.07% of the total number of profiles corresponding to 89 posts that were deleted. These publications have themes focused on political issues with a load of negative sentiment in favour of or against certain ideologies and political parties during the election campaign period. Another worrying aspect, as is the case with all social media, is fake news. This is the case of the publication [P60], associated with the Instagram account "adhd_understood" with a significant number of followers, whose owner is Ms. Donna Giachino, who defines herself as a doctor specialising in ADHD (<https://bit.ly/3z5yIFV>). However, the College of Physicians in Vancouver in 2019 had to rule that she was not registered and therefore could not practice as a speech and language professional (<https://bit.ly/3wTC7Rt>).

Figure 5 exemplifies the results presented so far, by means of two publications in the context of the last electoral campaign in Chile, published in a close temporal space. They show how the male gender compared to the female gender has a quantitative difference of 560,321 followers and 80,002 following, in line with the gender results (i) of a greater number of followers and following in the male gender. Another relevant aspect of the publications analysed is the number of likes: a publication by José Antonio Kast [V32] has 3,019 likes and 1,265 retweets compared to the 3,854 likes and 3,681 retweets of Claudia Aldana [V17]. In other words, with a lower number of female followers, the publication achieves a greater social impact on Twitter. At this point, the secondary idea of gender (ii) is exemplified as it is women who have more loyal followers in the interaction with the publications. If we look at the number of views of the video [V17], it stands at 85,693 views, a figure higher than the number of followers. These results show that it is the algorithms of the digital platforms themselves that suggest publications to other users, even if they are not followers of the profile, motivated by the polarisation of sentiment associated with electoral campaigns, which leads to segmented loyalty. This situation evidences that algorithms generate echo chambers. In other words, users are presented with suggestions based on their thematic interests and

ideological affinity on a recurring basis, and this procedure is repeated by listening only to information that the digital platforms consider to be of interest to the user. This manipulation of information dissemination leads us to the cognitive biases of the users analysed in the Power BI tool.



Digital platforms play a relevant role in the educational and social sphere because they are a loudspeaker for international days whose purpose is to raise awareness of social issues. As can be seen in Figure 6, the count of publications per day according to sentiment shows the following results. A series of days with the highest number of publications can be seen, which are reflected in the graph in the peaks listed. If we look at the largest, it corresponds to 3 December 2021, the International Day of Persons with Disabilities, with 1,329 publications of which 702 and 199 are of positive and negative sentiments respectively. If we analyse the publications from that day, the most viral publication is positive (<https://bit.ly/3GwET20>) [T7], associated with a profile outside the traditional political and economic interest groups mentioned above. The creator of this publication is a person with Down's Syndrome, María Jose Paiz Arias, known on social media as Majo. Her publication has 7,034 likes and 55,463 reproductions, whose title conceptualises the vindictive idea of the international days analysed: "inclusion is achieved with fewer labels and more action". However, political leaders such as the Senator of the Republic of Mexico, President of the Commission on the Rights of Children and Adolescents with more than 100,000 followers, only received 13 likes and 6 retweets (<https://bit.ly/3wYaCVJ>).

In other words, the viral power of influential groups over minority groups is reversed, unlike on other days of the year. Other notable days that show the reversal of the trend are 4th of January World Braille Day, 13th of January World Day to Combat Depression, 26th of January World Environment Education Day and 21st of March World Down Syndrome Day. In reference to World Education Day, 24th of January 2022, the situation described above occurs: political leaders publish, but are relegated to the background regardless of their thousands or millions of followers, as is the case with the official account of the Government of Spain with 778,858 followers, which barely receives 74 likes (<https://bit.ly/3IUDyIV>). Another significant day, of relevance for this research because it does not directly contain the keywords determined for the study, is the 2nd of April, World Autism Awareness Day, where the most viral publication has a clearly altruistic component from an individual profile, a parent of a child with Autism Spectrum Disorder (ASD): <https://bit.ly/3tqQY3R> [P95]. In reference to the frequency of publications, weekdays stand out as the time periods of greatest intensity, but on Saturdays and Sundays for the topic analysed, publications are drastically reduced; the causes are not known at the moment. However, it is worth mentioning that, if the international day is a public holiday, it remains positioned as a loudspeaker for demands, due to the high number of publications.

Therefore, it is evident that: (i) international days give visibility to social demands, invisible during the rest of the year in the eyes of social media users, in order to raise awareness, guide and vindicate the social cause. (ii) The profiles involved in social demands are the most viral, relegating influence groups to second

place, although these groups commemorate these days with publications. Therefore, the hypothesis (H1) posed of the dominant prevalence of influential groups has an exception on international days.

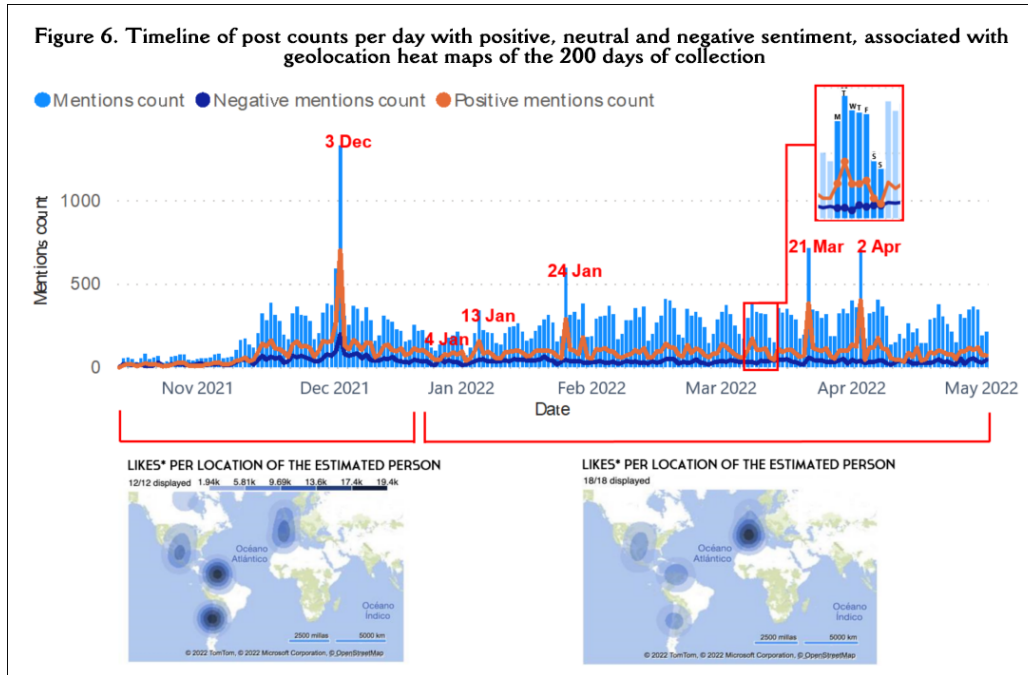


Figure 6 shows a chronogram, associated with heat maps by location of the most viral profiles, which shows the following results: the chronological division to generate these maps is motivated to argue hypothesis H1, where the interests of the influential groups are superimposed on the altruistic groups. If we look at the heat maps generated by the geolocations of the publications, in the dates corresponding to 2021, Argentina and Chile stand out. The reason for this is because the 12th of November was the first round of elections in Chile and on 14th of November the elections in Argentina. In 2022, the focus is on Europe, where different electoral events take place in Portugal, France and Germany. Therefore, (i) events of political interest also overshadow social demands and altruistic content.

If we look at the 100 most mentioned Instagram hashtags (#) in the whole time frame, represented in Figure 6, the first positions are obviously established by the key words of the research. The following hashtags, in the eighth position, are the word autism, #autism (N=2,117) and #autism (N=1,962). According to the World Health Organisation (WHO), the average prevalence of autism is estimated at 1/160 people, therefore, the presence of the topic is logical. Other hashtags reflect #adhd (21) N=913, #dyslexia (23) N=876 #tea (29) N=801 #downsyndorme (35) N=733 or #asperger (86) N=457. In addition, hashtags related to culture, values and respect such as #diversity (10) N=1927, #culture (19) N=1024, #teacher (31) N=773, #specialeducations N=762 stand out. Other hashtags reflecting the tuning of the most social posts in the top 100 are #diversity #equality #equality #values #equity #respect #inclusiveeducation #accessibility. (i) Aligned with the H1 of influencers among the most repeated hashtags is #influencer, related to digital creators. (ii) In line with the philosophy that education professionals play a fundamental role in these issues, reflected in the following hashtags: #teachers #teachers #education #primary #primary #neuroeducation #students #school #teacher #libraries.

In relation to the professions of the most viral profiles, (i) the one that receives the highest number of likes is linked to digital creators [T1], which is logical, as their aim is to achieve impact on social media. Unlike an association that is hidden in these social media whose purpose is to raise awareness and collaborate with the cause that moves them [T34], but given the neutral messages and content they publish, they are relegated to the background due to a "glass algorithm" in social networks, i.e. they are visible to any user who explicitly searches for them, but not by suggestions from digital platforms.

4. Discussion and conclusions

Influencing groups plays a relevant role as actors who are phagocytising the influence of other actors with more altruistic and humanist interests. This is an important issue if we take into account that the issues analysed affect groups and people who are often unequal, and who need to make their demands, experiences and achievements visible. Social networks as spaces in which perception and reality are intermediated, with effects on the ways of understanding it and what is or is not important for the collective public debate, plays in favour of groups and actors with political interests (Del-Fresno-García, 2019). In this sense, the role of social media as spaces for affinity, participation and collaboration, as evidenced by the work of Grace et al. (2019) and De-Groot et al. (2022), has not been confirmed in this study and for the subject under analysis. The role of social media as opportunities for users to interact and share information in influential relational structures of support and monitoring has not been validated in this research either.

The only way to reverse the viral power of influencers is twofold: to change in the algorithms of digital platforms, and make users on social media aware. For the first possibility, it would require state and supra-state public policies that promote some law to establish some control over influencers, with the paradox, as demonstrated in this study, that they are the most influential actors in the networks. For the second possibility, a digitally competent teaching staff is required. The results obtained point to the need to raise awareness among digital citizens with a critical vision in order to understand how algorithms impose the messages of certain actors, as opposed to other groups interested in making situations related to disability visible, and to claim their educational and social value from inclusive references. Some of the 100 most viral Instagram hashtags (#) confirm these interests. Teachers must make new generations understand the asymmetry in communication, as represented by Barberá et al. (2015) and Brady et al. (2019). This will enable them to recognise that certain publications, and the profiles associated with them, define invaluable human and social values.

In relation to this, it is important that teachers help the new generations to become progressively aware of the risks of falling into bubble filters or echo chambers. This would avoid a dominance as a factor of destabilisation of democracies worldwide, in the face of disinformation promoted by influential groups according to their interests, which defines another important challenge for teachers (Ortega-Sánchez et al., 2021). In reference to the study Robles et al. (2022) on bots and negative feelings associated with them as fuel for political polarisation, it has no presence in the social and educational subject matter under analysis. This may be due to public awareness, which would not assume and accept explicit confrontation, as is the case with topics such as immigration, the economy, security, politics or health. These topics are prone to bots that polarise publications, and thus, citizens. The relegation of the power of influence of influential groups on international days would support this conclusion. Other results that would validate these conclusions are that, although the male gender has more followers than the female gender, it is women who get more likes. This could be linked to the positive feelings that prevail in their publications for the content analysed. These results would indicate another important task for teachers: teaching new generations to post on social media with constructive attitudes, away from negative emotions (Arcila-Calderón, 2022).

The innovative aspect of the use of the Business Intelligence tool justifies its consideration as content of the digital competence of teachers for education in the coming years. The scientific-technical impact implies a radical change in the way of conceptualising and using social media as spaces for communication and construction of discourses and actions in the field of inclusive education and disability. Among other reasons, this occurs because our analysis no longer views social media as a tool for the uncritical and unreflective consumption of discourses and superficial contacts, but as valuable media that support relationships and content of educational and social value. Another important fact that supports the innovative nature of the tools and techniques applied lies in the case studies we have initiated, because they will allow us to systematise the value and function of multimodal representation to communicate and promote awareness, debates, and narratives that generate counter-hegemonic actions. Another reason lies in the methodological impact inherent to the study, defined by the analytical techniques we will carry out. In the scientific debate on communication and human thought, this will contribute to define, the value of interdisciplinary work in two fields, that of Educational Technology and that of Telecommunications.

Finally, it should be noted that the analysis process carried out with Big Data and Business Intelligence tools is in itself a digital competence in inclusive relationship spaces in educational institutions, which will help teachers to understand the relational structure of social networks and to know how to arrive at meanings about the content of the exchanges made by the participants in these networks. This implies that such teaching competence includes knowing how to identify the most influential profiles and how to investigate the meaning given to topics and behaviours, which are not always visible (Del-Fresno-García, 2014), linking them to messages on social media and the type of profiles associated with them. This knowledge will lead teachers to acquire a collective ethic, for which it will sometimes be necessary to unlearn ideas and beliefs built in the history of each of the participants in their life contexts. This is the only way to build a shared morality based on reason, and not only on emotions, in the contemporary, mixed and diverse relationships that characterise educational centres that seek inclusive teaching situations. This requirement in teaching practice entails the need to assume the inquiring-innovative dimension of social networks, of a critical nature, as one of the objectives of teachers' digital competence (Bautista, 2021).

Authors' Contribution

Idea, C.B.; Literature review (state of the art), C.B., L.R; Methodology, C.B; Data analysis, C.B. L.R; Results, C.B, L.R, A.B; Discussion and conclusions, C.B, L.R, A.B.; Drafting (original draft), C.B, L.R, A.B.; Final revisions, C.B, L.R, A.B; Project design and sponsorships, C.B, L.R, A.B.

Funding Agency

This work has been supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with Universidad Complutense de Madrid in the line Excellence Programme for university teaching staff, in the context of the V PRICIT (Regional Programme of Research and Technological Innovation).

References

- Arcila-Calderón, C., Sánchez-Holgado, P., Quintana-Moreno, C., Amores, J., & Blanco-Herrero, D. (2022). Hate speech and social acceptance of migrants in Europe: Analysis of tweets with geolocation. [Discurso de odio y aceptación social hacia migrantes en Europa: Análisis de tuits con geolocalización]. *Comunicar*, 71, 21-35. <https://doi.org/10.3916/C71-2022-02>
- Avidi, I.T., Paynter, M., & Vujosevic, T. (2019). Facebook group in the learning design of a higher education course: An analysis of factors influencing positive learning experience for students. *Computers & Education*, 129, 106-121. <https://doi.org/10.1016/j.compedu.2018.10.018>
- Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23(1), 76-91. <https://doi.org/10.1093/pan/mpu011>
- Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531-1542. <https://doi.org/10.1177/0956797615594620>
- Bautista, A. (2021). Functional resignification and technological innovation as a digital teaching competence. *IEEE. Revista Iberoamericana de Tecnologías del Aprendizaje*, 16(1), 93-99. <https://doi.org/10.1109/RITA.2021.3052679>
- Becker, L.T., & Gould, E.M. (2019). Microsoft power BI: Extending excel to manipulate, analyze, and visualize diverse data. *Serials Review*, 45(3), 184-188. <https://doi.org/10.1080/00987913.2019.1644891>
- Brady, W.J., Wills, J.A., Burkart, D., Jost, J.T., & Van-Bavel, J.J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, 148(10), 1802-1813. <https://doi.org/10.1037/xge0000532>
- Brunner, M., Palmer, S., Togher, L., & Hemsley, B. (2019). I kind of figured it out': The views and experiences of people with traumatic brain injury in using social media-self-determination for participation and inclusion online. *International Journal of Language & Communication Disorders*, 54(2), 221-233. <https://doi.org/10.1111/1460-6984.12405>
- Carlsson (Ed.) (2019). *Understanding media and information literacy (MIL) in the digital age. A question of democracy*. University of Gothenburg. <https://bit.ly/3xWVsCz2>
- Cheng, L.C., & Tsai, S.L. (2019). Deep learning for automated sentiment analysis of social media. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 1001-1004). Association for Computing Machinery. <https://doi.org/10.1145/3341161.3344821>
- De-Groot, R., Kaal, L.H., & Wouter, P.S. (2022). The online lives of adolescents with mild or borderline intellectual disabilities in the Netherlands: Care staff knowledge and perceptions. *Journal of Intellectual & Developmental Disability*. <https://doi.org/10.3109/13668250.2021.2004635>
- Del-Fresno-García, M. (2014). Haciendo visible lo invisible: Visualización de la estructura de las relaciones en red en Twitter por medio del análisis de redes sociales. *Profesional de la Información*, 23, 246-252. <https://doi.org/10.3145/epi.2014.may.04>
- Del-Fresno-García, M. (2019). Desórdenes informativos: sobreexposiciones e infrainformados en la era de la posverdad. *Profesional de la Información*, 28(3), 1-11. <https://doi.org/10.3145/epi.2019.may.02>
- Dellwing, M. (2021). *The continuities of Twitter strategies and algorithmic terror. Society for the study of symbolic interaction*. Sage. <https://doi.org/10.1002/symb.585>

- Ditrendia (Ed.) (2009). *Móvil en España y en el Mundo*. Ditrendia Digital Marketing Trends. <https://bit.ly/3MNim3j>
- Grace, E., Raghavendra, P., Mcmillan, J., & Gunson, J. (2019). Exploring participation experiences of youth who use AAC in social media settings: Impact of an e-mentoring intervention. *Augmentative and Alternative Communication*, 35(2), 132-141. <https://doi.org/10.1080/07434618.2018.1557250>
- Hemsley, B., Dann, D., Palmer, S., Allan, M., & Balandin, S. (2018). Using Twitter to access the human right of communication for people who use Augmentative and Alternative Communication. *International Journal of Speech-Language Pathology*, 20(1), 50-58. <https://doi.org/10.1080/17549507.2017.1413137>
- Ladogina, A., Samoylenko, I., Golovina, V., Razina, N., & Petushkova, E. (2020). Communication effectiveness in social networks of leading universities. *Diálogo*, 43, 35-50. <https://doi.org/10.18316/dialogo.v0i43.6497>
- Liu, B. (2010). Sentiment analysis and subjectivity. In N. Indurkha, & F. J. Damerou (Eds.), *Handbook of Natural Language Processing*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781420085938>
- Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. O'Reilly Media: O'Reilly Media, Inc. <https://bit.ly/3zw5nOn>
- Ortega-Sánchez, D., Blanch, J.P., Quintana, J.I., Cal, E., Fuente-Anuncibay, C., & Hate, R. (2021). Speech, Emotions, and Gender Identities: A Study of Social Narratives on Twitter with Trainee Teachers. *International Journal of Environmental Research and Public Health*, 18(8), 4055-4055. <https://doi.org/10.3390/ijerph18084055>
- Robles, J., Guevara, J., Casas-Mas, B., & Gómez, D. (2022). When negativity is the fuel. Bots and Political Polarization in the COVID-19 debate. [Cuando la negatividad es el combustible. Bots y polarización política en el debate sobre el COVID-19]. *Comunicar*, 71, 63-75. <https://doi.org/10.3916/C71-2022-05>
- Tuzel, S., & Hobbs, R. (2017). The use of social media and popular culture to advance cross-cultural understanding. [El uso de las redes sociales y la cultura popular para una mejor comprensión intercultural]. *Comunicar*, 51, 63-72. <https://doi.org/10.3916/C51-2017-06>
- We are social (Ed.) (2021). *Digital 2021. The latest insights into the 'state of digital'*. We Are Social Ltd. <https://bit.ly/3POIhJV>
- White, P., & Forrester-Jones, R. (2020). Valuing e-inclusion: Social media and the social networks of adolescents with intellectual disability. *Journal of Intellectual Disabilities*, 24(3), 381-397. <https://doi.org/10.1177/1744629518821240>