



Eder Ávila Barrientos*

The Analysis of Big Data Generated by Digital Communities

Introduction

Digital communities constantly generate data published on different platforms and have contributed to creating massive amounts of varied data of different typology. Analyzing them allows us to obtain estimates and predictions that help better understand a particular event or problem. Nevertheless, a large amount of massive data are not trustworthy and do not have the quality needed to be able to use and consume them, which affects their effective management.

Digital communities make it possible for people with common interests to connect with each other, regardless of their geographic location; this broadens their opportunities to interact, learn, and cooperate with individuals

in ways that would not have been conceivable in another time. In addition, they foster the exchange of data and information since they bring together people with experiences and knowledge in different areas, prompting both questions and answers. They also often function as platforms for sharing resources, tutorials, and advice. Joaquín Gairín Sallán writes in this regard that people who belong to them learn to participate in culturally important activities thanks to mutual collaboration and the construction of collective knowledge,¹ and they obtain, for example:

- *Support and empowerment.* In these communities, they can find back-up and understanding in difficult situations. If someone is having a similar experience or has questions about a specific topic, he/she can

* Eder is a researcher at the UNAM Library Science and Information Research Institute (IIBI); you can contact him at eder@iibi.unam.mx.

go to the group to get advice, share experiences, and feel they have received feedback. These spaces can also be empowering, since they allow people to find their voice, express themselves, and advocate for causes important to them.

- *Opportunities for collaboration.* They can be spaces where people come together to work on projects, develop joint ideas, or solve problems. Cooperation in these groups can lead to new professional or creative opportunities or undertakings.
- *Access to resources and opportunities.* They often offer access to valuable resources, such as tools, tutorials, online courses, information about activities, and job opportunities. These resources can be beneficial both personally and professionally since they help broaden skills, knowledge, and networks of contacts. People can find their place in these online spaces in an increasingly interconnected digital world.

What Are Digital Communities Like?

Some examples are discussion forums about specific topics like technology, music, film, or videogames, as well as social networks centered on activities like cooking, photography, or travel, and collaborative platforms for artists and/or software developers. Some have opted to enter areas such as academic and scientific research, encouraging border research and the creation of networks.

Digital communities have grown in popularity because of their advantages, such as a wide gamut of perspectives, the possibility of connecting with people from all around the world, and the capacity to find data, information, and specialized resources. These groups play an important role in the creation of networks, collaborative learning, and the development of joint projects in the digital sphere, which has prompted greater generation and expansion of data worldwide.

The Data Tsunami

This term refers to massive data accumulation due to factors such as an increase in the use of electronic devices, the expansion of the Internet, the growth of social media, the advance of sensor technology, and the digitalization

of various processes of human activity. In his book *Dataism*, Steve Lohr writes that the capitalism of data shows an economic system in which data are considered a valuable economic resource and are used as a form of capital.² In this context, companies and organizations gather, analyze, and utilize large amounts of data to achieve competitive advantages, improve their products and services, and make more informed commercial decisions. This means that through the data they generate and publish, digital communities are contributing the raw materials so these organizations can find out people's tastes and preferences when they interact in the digital world.

In accordance with this premise, the "data tsunami" consists of the exponential generation of pieces of information in different formats (text, images, videos, audios, commercial transactions, medical logs, and sensor data) from different sources, such as social media, mobile devices, IoT (Internet of Things) sensors, online transactions, governmental logs, and scientific research, among others.

This phenomenon poses important challenges since the amount of data created surpasses traditional human technology and skills' ability to process and analyze them. As the amount of data available increases, the need arises for tools and techniques to extract relevant information, discern patterns, predict, and make informed decisions.

In this sense, the "data tsunami" has also sparked the development of fields such as big data analysis, artificial intelligence, and automatic learning. These seek to take advantage of these data to obtain knowledge and improve processes in different sectors and social contexts such as health, commerce, industry, scientific research, transportation, and many others.

Added to the challenge of generating relevant information, others exist associated with privacy security, and storage. Effective management of large volumes of data requires solid infrastructures, efficient algorithms, and clear policies to guarantee the protection of people's privacy and the security of sensitive data.

Analysis of Big Data Generated by Digital Communities

This analysis involves the gathering, examination, and extraction of valuable information from huge volumes of data generated by online communities such as Facebook,

The appropriate handling of these data would help combat phenomena like fake news and the propagation of disinformation in the digital world, since, when they are organized and analyzed, anomalies and the sources used to generate them can be identified, making it possible to gauge the emitter's intent.

Twitter (now X), Instagram, online forums, and blogs, among others. This involves the use of advanced processing techniques such as data mining, automatic learning, and data visualization to understand the patterns, tendencies, and people's behavior inside these communities. Some of the key aspects of this process are as follows:

- *Data gathering.* This can be done using techniques such as web scraping, the application programming interface (API) of platforms, or access to data bases.³
- *Pre-processing.* The data gathered are often unstructured and can contain noise; therefore, they require pre-processing to clean and organize them before analysis.
- *Analysis of feelings.* This can determine the emotional polarity associated with the data; that is, if comments, publications, or messages are positive, negative, or neutral. This can be useful for understanding the community's opinion about a specific topic.
- *Analysis of social networks.* An analysis of the relations among digital community members may reveal the network's structure and the nature of these interactions. This can help identify opinion leaders, thematic communities, and the dissemination of information within the network, as well as the origin of the data when they are published and go viral through a given trend.
- *Analysis of trends.* When observing the data generated by the community over time, analysts can identify patterns and trends, which can help understand users' changing preferences and predict changes.
- *Personalization and recommendation.* The analysis of the data makes it possible to personalize the user's experience, offering relevant recommendations according to the community's interests and previous behavior.

The analysis of big data generated by digital communities has applications in different areas, such as marketing, market research, entrepreneurial decision-making, fraud detection, managing online reputations, and identifying emerging problems. However, it is important to include ethical and privacy considerations when working with these data, since many of them can be sensitive and can be used counterproductively.

Final Considerations

Digital communities have the potential for generating different kinds of data because of their broad interaction with social platforms and media. Many are sought after by external agents like companies and digital services that want to find out preferences and opinions regarding a specific topic or trend.

Data analysis, then, is fundamental in the study of digital communities because it provides valuable information and makes it possible to generate knowledge based on the data uploaded by users on these platforms.

The appropriate handling of these data would help combat phenomena like fake news and the propagation of disinformation in the digital world, since, when they are organized and analyzed, anomalies and the sources used to generate them can be identified, making it possible to gauge the emitter's intent.

Therefore, it is important to become literate in data issues in an increasingly digitalized world. Those most aware and knowledgeable in this area will be prepared to make informed decisions, resolve complex problems, communicate data-based ideas, and actively participate in an increasingly interconnected society. ■■■

Notes

¹ Joaquín Gairín Sallán, "Las comunidades virtuales de aprendizaje," *Educar* vol. 37 (2006), pp. 41-64, <https://www.redalyc.org/articulo.oa?id=342130826004>.

² Steve Lohr, *Data-ism* (New York: Harper Collins, 2015), p. 207.

³ An application programming interface (api) is a piece of code for two or more computer programs or components to communicate with each other; it is an intermediary between two systems that allows an application to communicate with another and request data or specific actions. Information from Wikipedia, <https://en.wikipedia.org/wiki/API>. [Editor's Note.]