Dossier Editorial

Special Issue Computational Methods and Big Data in Communication Research

Carlos Arcila Calderón, Universidad de Salamanca, Salamanca, Spain (carcila@usal.es)

Wouter Van Atteveldt, Universidad libre de Amsterdam, Amsterdam, Netherlands (wouter.van.atteveldt@vu.nl)

Damian Trilling, Universidad de Amsterdam, Amsterdam, Netherlands (d.c.trilling@uva.nl)

Computational methods are becoming an exciting and promising area in the field of communication research, not only in the Global North but also in regions such as Ibero-America. Around the world, the availability of social media data, digital trace data, and other types of data that are larger than typical sample sizes in the social sciences have sparked academics' interest in the application of novel methodologies. As public discourse is more and more shaped by what is going online, this shift towards new data sources and new methodologies is also urgently needed.

This special issue of *Cuadernos.info* reflects how the computational approach is being included in the research agenda in countries such as Mexico, Chile, Brazil, or Spain. The data from these countries, analyzed in this special issue, represent a vast resource of contents and messages, mainly in Spanish and Portuguese, two of the most spoken languages in the world. This movement from English-based and Global North analysis to new geographical-linguistics contexts in the world depicts a significant challenge for the computational analysis of communication. First, techniques such as natural language processing, sentiment analysis, topic modelling or text classification perform differently in different languages. Specifically, they can be sensitive to language-specific features and the researcher's decisions (for instance, regarding pre-processing). Nevertheless, as most of the literature and examples are still dedicated to contents in English, there is a lack of reference material (and resources in general) for the computational analysis of communication in languages such as Spanish and Portuguese. Second, cultural realities in different parts of the world, such as in the Latin American countries, can stimulate the formulation of new research questions or hypotheses than might be answered using a computational approach. This, then, may in turn inspire new research lines in other parts of the world.

As guest editors of this issue, we were glad to receive many proposals to be included in the dossier dedicated to *Computational Methods and Big Data in Communication Research*. From this initial set of manuscripts, five papers were finally selected after a rigorous double-blind peer review process coordinated by the main editor and editorial team of the journal. The result is a well-balanced and select group of research reports that show evidence of the boost that Latin American and Spanish communication scholars are giving to the topic of this issue. We must say that, even though the call for papers was broader (i.e., by including agent-based modeling or data journalism), most of the received manuscripts dealt mainly with different types of computational content analysis in social media, such as Twitter and Instagram. This fact reflects the strong tradition of communication researchers on content analysis and their growing interest on how social media impacts the political and cultural sphere in the region. Along with automatic collection of tweets or images in Instagram, we find the application of sentiment analysis or supervised text classification as techniques to address relevant social questions. Moreover, the use of social network analysis (SNA) in some of the papers demonstrate how this well-established computational technique is fundamental to model, visualize, and understand the complex dynamics of user-generated contents.

To some readers, the scope of this special issue might seem highly technical. However, most of the received manuscripts addressed some relevant theoretical questions and had strong practical social implications: from polarization or hate speech to the reaction of the killing of an indigenous leader or the conversations during the COVID19 pandemic. This reveals the huge potential that the computational approach has to address both the traditional and the current concerns in communication research. Indeed, it is not a coincidence that the main associations of researchers in the region have already created permanent groups to discuss the application of these methods. This is the case of the section Digital communication, social networks and processes, in the Latin American Association of Communication Researchers (ALAIC), or the emerging section Computational Methods and Big Data Analysis in Communication, of the Spanish Association of Communication Researchers (AE-IC), both initiatives in line with the Computational Methods division of the International Communication Association (ICA), to which the three guess editors of this special issue belong to.

The first manuscript of this dossier, by the Mexican scholars Manuel Cebral Loureda and Gabriela Elisa Sued Palmeiro, is dedicated to a large-scale analysis of the public conversation in Spanish messages in Twitter during the beginning of the COVID19 pandemic. This paper is an excellent sample of the use of computational techniques for text analysis (such as topic modeling and sentiment analysis) to understand an ongoing complex social reality such as the pandemic.

The second paper –authored by Chilean scholars Marcelo Santos, Oscar Jaramillo Castro, and Daniel Aguirre Azócar–, combines SNA, social media analytics, and content analysis to cover the reaction in Twitter of the murder of a member of the Mapuche indigenous community. This research brings into scene the relevant theoretical discussion of the radicalization processes within the so-called echo chambers and filter bubbles. The third paper, written in Portuguese (all the others are in Spanish) by the Brazilian colleagues Rita de Cássia Romeiro Paulino and Mariane Pires Ventura, keeps its interest in Twitter and uses SNA, textual and emoji analysis, to look at the discussions maintained during an online protest (#Somos70porcento) in a moment of health crisis. This paper demonstrates the importance of social media as a new means for public discussion and debate. The fourth manuscript, by the Spanish scholars Miguel Varela and Miguel Vicente, focused on Instagram, by explaining a method to scrape images at scale. This report keeps the attention on a health communication process (an online campaign against cancer), but has a more general relevance as well, as its insights into the analysis of images of a very popular social medium can be useful for a larger audience. Finally, the paper by Spanish scholars Javier Amores, David Blanco, Patricia Sánchez, and Maximiliano Frías shows how to use supervised text classification (with shallow and deep algorithms) to detect hate speech motivated by political reasons in Spanish. This research shows how the combination of good-quality manual labeling and state-of-the-art algorithms (i.e., embeddings and deep neural networks) can produce excellent predictions when applied to contents in Spanish.

We believe that the emergence of a computational communication science is bringing more opportunities to researchers to address old and new research questions, and this in turn will contribute to a shift in the quality of media and communication research. We hope you enjoy and find these papers useful.

CARLOS ARCILA CALDERÓN, WOUTER VAN ATTEVELDT Y DAMIAN TRILLING

Guest Editors and authors of the forthcoming book

*Computational Analysis of Communication. A practical introduction
to the analysis of texts, networks, and images with code examples in Python and R.*