Yannis Skarpelos, Kostas Karpouzis**, Martha Michailidou****

INTRODUCTION

ABSTRACT

Computational Social Science (CSS) has emerged as a fast-developing scientific field. Its growth has been driven by the increasing availability of computational power, coordinated digitization, annotation and interoperability efforts, and large-scale data on social activities. computing, data sciences and the social sciences created a loop, opening new fields of inquiry and forming new methodologies, making way for innovative approaches, and potentially leading to fresh discoveries. The papers in this special issue cover a wide range of phenomena using computational methods, such as social media, simulation, creative and cultural industries, as well as infrastructures for CSS. They apply CSS methods including social network analysis, virtual encounter simulations, sentiment and emotion analysis, and corpus linguistics, to case studies ranging from Covid-19 pandemic to electoral behaviour, political charisma, cultural affinities and value orientation to border regions, esports networks, to museums, to national and international infrastructures, thus providing a spectrum of possibilities (and limitations) for social scientists.

Keywords: *computational social science, virtual encounter simulations, social network analysis, emotion analysis, political charisma*

* Professor, Panteion University of Social and Political Sciences, e-mail: gskarp@panteion.gr

** Assistant Professor, Panteion University of Social and Political Sciences, e-mail: kkarpou@panteion.gr

*** Associate Professor, Panteion University of Social and Political Sciences, e-mail: marthami@panteion.gr

Γιάννης Σκαρπέλος*, *Κώστας Καρπούζης***, *Μάρθα Μιχαηλίδου****

ΕΙΣΑΓΩΓΗ

ΠΕΡΙΛΗΨΗ

Οι Υπολογιστικές Κοινωνικές Επιστήμες είναι ένα ταχύτατα αναπτυσσό-μενο επιστημονικό πεδίο, χάρη στην αυξανόμενη διαθεσιμότητα υπολογι-στικής ισχύος, τα συντονισμένα εγχειρήματα ψηφιοποίησης, επισημείωσης και διαλειτουργικότητας, και τα μεγάλης κλίμακας δεδομένα που αφορούν κοινωνικές συμπεριφορές. Η επιστήμη των υπολογιστών, η επιστήμη δεδο-μένων και οι κοινωνικές επιστήμες διαμόρφωσαν έναν ανατροφοδοτούμενο κύκλο διανοίγοντας νέα πεδία έρευνας και διαμορφώνοντας νέες μεθοδο-λογίες, καινοτόμες προσεγγίσεις και, ενδεχομένως, νέες ανακαλύψεις. Τα άρθρα σε αυτό το ειδικό τεύχος καλύπτουν ένα ευρύ φάσμα φαινομένων όπως τα μέσα κοινωνικής δικτύωσης, οι προσομοιώσεις, οι δημιουργικές και πολιτιστικές βιομηχανίες, καθώς και οι υποδομές για τις υπολογιστικές κοινωνικές επιστήμες. Εφαρμόζουν μεθόδους όπως η ανάλυση κοινωνικών δικτύων, προσομοιώσεις εικονικών συναντήσεων, ανάλυση συναισθήματος και αισθημάτων, και η υπολογιστική γλωσσολογία, σε μελέτες περιπτώσε-ων που εκτείνονται από την πανδημία Covid-19 ως την εκλογική συμπε-ριφορά, το πολιτικό χάρισμα, τις πολιτισμικές συγγένειες και τον αξιακό προσανατολισμό σε περιοχές κοντά στα σύνορα, τα δίκτυα ψηφιακών σπορ (esports), τα μουσεία, τις εθνικές και διεθνείς υποδομές, παρουσιάζοντας έτσι ένα φάσμα δυνατοτήτων (και περιορισμών) που προσφέρονται στους κοινωνικούς επιστήμονες.

Λέξεις κλειδιά: υπολογιστικές κοινωνικές επιστήμες, προσομοιώσεις εικονικών συναντήσεων, ανάλυση κοινωνικών δικτύων, ανάλυση συναισθήματος, πολιτικό χάρισμα

---

* Καθηγητής, Πάντειο Πανεπιστήμιο Κοινωνικών και Πολιτικών Επιστημών, e-mail: gskarp@panteion.gr

** Επίκουρος Καθηγητής, Πάντειο Πανεπιστήμιο Κοινωνικών και Πολιτικών Επιστημών, e-mail: kkarpou@panteion.gr

* Αναπληρώτρια Καθηγήτρια, Πάντειο Πανεπιστήμιο Κοινωνικών και Πολιτικών Επιστημών, e-mail: marthami@panteion.gr

As early as 2009, Lazer et al. mentioned the scarce presence of the term as well as of the scientific field in leading journals in economics, sociology and political science, and suggested that Computational Social Sciences (CSS) "is occurring – in Internet companies such as Google and Yahoo, and government agencies such as the U.S. National Security Agency" (2009, p. 721). They warned that either as "the exclusive domain of private companies and government agencies," or with the emergence of "a privileged set of academic researchers presiding over private data from which they produce papers that cannot be critiqued or replicated," this field would not serve the long-term public interest nor the production of new verifiable knowledge. Stuetzer, Welker and Egger comment that "large multinational telecommunications companies and online service providers install their departments of data, thus making science proprietary instead of publicly debatable and verifiable" (2018, p. 12). Isn't it the case in the worries and lawsuits aimed at OpenAI and Microsoft for copyright infringement (The Guardian 2024), as well as for using private emails to train their large language models?

Lazer et al. (2009) propose micro-sociological applications, such as examining group interactions with online data, as well as macro-sociological applications like social networks beyond the confines of a company or an organization's employees, as well as their dynamics and evolution over time. While such applications are now quite common in social sciences, this early paper remains significant in stressing the importance of access and privacy, along with institutional obstacles related to Ethics Committees unable to understand the intricacies of complex data, as well as the challenge for developing a paradigm for training new scholars in the field.

A decade later, Salganic would attempt to provide such a paradigm, while using a rather loose definition of CSS: "The intersection of social science and data science is sometimes called computational social science" (Salganic 2019, p. xviii). He proposed that two different styles can be employed in CSS: ready-made, i.e. "repurposing of big data sources that were created by companies and governments," and custom-made, i.e., "a researcher started with a specific question and then used the tools of the digital age to create the data needed to answer that question" (Salganic 2018, p. 7).

In between the two publications, Computational Social Science has emerged as a fast-developing scientific field. Its growth has been driven by the increasing availability of computational power, coordinated

digitization, annotation and interoperability efforts, and large-scale data on social activities. Social media have been instrumental in its development, while the importance and potential impact of CSS grew during the COVID-19 pandemic (cf. Skarpelos et al. 2024). Artificial Intelligence (AI) and Large Language Models (LLMs) open new possibilities for CSS, allowing for new ways of data collection, Natural Language Processing, the analysis of big visual corpora etc., while enabling a deeper look into the biases it incorporates (cf. Karpouzis 2024). It can be argued that computing, data sciences and the social sciences created a loop, opening new fields of inquiry and forming new methodologies, making way for innovative approaches, and potentially leading to fresh discoveries. Such approaches and discoveries usually lie in the middle level of theorizing, and several steps remain to be taken to achieve the full potential of that area of enquiry.

Those developments led to an increasing concern about ethics in CSS research. Ethics Committees at Universities are still following a former logic in understanding crucial issues such as privacy, or human subjects' protection while following a positivistic pattern where predictability is both desirable and achievable. In CSS research though, data complexity is more often than not leading to unpredictability, to moving grounds about privacy, where GDPR is a framework for a constantly fluid and oscillating set of national, international and corporate rules and decisions impacting research. A recent example was the decision of X (former Twitter) to close the academic accounts, leaving funded research and PhD candidates without the data necessary to complete their projects. In a recent paper, Lazer et al. note that "Access to data from private companies is thus rarely available to academics, and when it is, it is typically granted through a patchwork system in which some data are available through public application programming interfaces (APIs), other data only by working with (and often physically in) the company in question, and still other data through personal connections and one-off arrangements, often governed by nondisclosure agreements and subject to potential conflicts of interest" (Lazer et al. 2020, p. 1060).

The papers in this special issue attempt to showcase diverse applications of computational social sciences. The first four papers focus on social media as a primary source of data for social scientists. Nikos Smyrnaios, Panos Tsimpoukis, and Charis Papaevangelou's paper explores the discourses developed in Twitter around the controversial promotion of Hydroxychloroquine as a treatment for COVID-19 by French professor

Didier Raoult. Using a network and lexicometric analysis of 1.2 million tweets, the study identifies peaks in Twitter reactions linked to significant media events. It shows how influential political figures and media outlets shaped the discourse, revealing that pro-HCQ messages often combined anti-science conspiracy theories with critiques of liberal political economy. Notably, the paper finds that supporters of Donald Trump and Jair Bolsonaro connected with pro-Raoult French-speaking groups, illustrating the global and politically charged nature of the debate.

In the next paper, Dimitris Elafropoulos investigates the ways Covid-19 pandemic was presented in Greek political and scientific discourse. Analyzing a corpus from March 2020 to May 2022 highlights how different actors – government representatives, the major opposition party, and the Greek Public Health Organization – communicated the crisis. The study employs a triangulated framework of corpus linguistics and critical discourse analysis to reveal patterns and trends in language use. The findings indicate distinct narratives: the government framed the pandemic as a 'rescue mission,' often used metaphors of war and paternalistic narratives, positioning itself as a saviour and emphasizing individual responsibility; the opposition criticized the management of the pandemic and highlighted the adversities faced by citizens and working-class priorities; and the Greek Public Health Organization maintained a more scientific and detached tone but occasionally adopted a familiar style to humanize the pandemic's impact.

Sophia Messini in her paper examines the relationship between emotions expressed on Twitter and voting behavior during the critical Greek elections of 2015 and 2019. Moving beyond sentiment analysis, the study combines opinion mining with social network analysis to assess the impact of specific emotions on electoral outcomes, finding that certain emotions, such as anger and disgust, were dominant during these periods and could indicate a desire for political change. The intensity of emotional expression was identified as a potential predictor of shifts in voting behaviour, highlighting the complex interplay between online discourse and electoral dynamics.

Finally, Kaimaki et al. turn to measuring the political charisma of Greek journalists on X (Twitter). The authors propose an algorithm based on the Analytic Hierarchy Process (AHP) method to quantify journalists' charisma, considering various metrics such as impressions, likes, quotes, replies, retweets, and interactions with politicians. They highlight the distinction between influence and charisma, emphasizing that while

influence can be measured through engagement metrics, charisma involves qualitative aspects like symbolic capital and charismatic authority.

Simulations have been also part of computational applications in social science. Georg P. Mueller's paper examines the political opinions of the Arc Lémanique region in Switzerland, comparing them to the rest of Switzerland and France, through a novel methodology called Virtual Encounter Simulations. This method uses survey data to create random dyads of individuals, allowing for a comparison of inter-individual differences in political attitudes. By analyzing the 2016 International Social Survey Programme data, the study aims to understand the ideological proximity between these regions. Contrary to popular belief, the findings reveal that Arc Lémanique is ideologically closer to German-speaking Switzerland than previously thought.

The next two papers turn to cultural and creative industries as their field of application. Stavroula Dargonaki examines the Greek professional gaming scene using Social Network Analysis to visualize and explore its structure and dynamics. Professional gaming, or esports, is an expanding cultural, social, and economic activity. The analysis identifies key players within the Greek scene, including professional gamers, gaming companies, and related entities, revealing the rapid commercialization and highly gendered nature of this space. The study underscores the central role of content producers and the marginalization of non-masculine gender identities within the professional gaming community.

On the other hand, Ioanna Zouli focuses on the methodological and metaphorical use of front-end and back-end processes in the context of art production and research in cultural institutions, specifically through the lens of the Tate Museum's BMW Tate Live: Performance Room project. Zouli explores how Tate adapted to digital platforms, revealing the complexities and challenges faced in maintaining institutional authority and control over art experiences in an online environment. By examining both the front-end (public-facing elements like live performances and audience interaction) and back-end (internal processes and decision-making), the study provides insights into the museum's attempts to reconcile traditional curatorial practices with the demands of digital engagement.

The last two papers turn to issues of infrastructure supporting computational social science, data management, responsible and ethical research practices, and the configuration of repositories for social data. Kondyli and Klironomos discuss the emergence of New Data Types, particularly big data, in social sciences over the past fifteen years, and

the implications for research methodologies and data management. The paper addresses the challenges and advantages of integrating these data into existing Research Infrastructures and Data Repositories, emphasizing the need for new skills and capabilities at the intersection of computational and social sciences. It highlights the potential for collaborations between data organizations and researchers to foster a culture of data sharing and reuse, which is critical for advancing empirical social research.

Finally, Linardis et al. outline the efforts of SoDaNet to enhance its data repositories by adopting and customizing the Dataverse software to improve data Findability, Accessibility, Interoperability, and Reusability (FAIR). The report details the specific customizations made to Dataverse, such as the addition of new metadata fields, the development of tools for online statistical analysis, and enhancements to user interface and multilingual support. These improvements aim to meet the needs of both national and international research communities, ensuring that the data hosted by SoDaNet repositories are accessible and usable according to FAIR principles.

REFERENCES

Karpouzis, K. (2024). Plato's shadows in the digital cave: Controlling cultural bias in Generative AI. *Electronics*, *13*(8), p. 1457.

Lazer, D. M., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., et al. (2020). Computational social science: Obstacles and opportunities. *Science*, *369*(6507), pp. 1060-1062.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., et al. (2009). Computational social science. *Science*, *323*(5915), pp. 721-723.

Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.

Skarpelos, Y., Messini, S., Roinioti, E., Karpouzis, K., Kaperonis, S., & Marazoti, M. G. (2024). Emotions during the Pandemic's First Wave: The Case of Greek Tweets. *Digital*, *4*(1), pp. 126-151.

The Guardian (2024). Eight US newspapers sued OpenAI and Microsoft for copyright infringement. Tue 30 Apr 2024 20.29 CEST, https://www.theguardian.com/technology/2024/apr/30/us-newspaper-openai-lawsuit