Data practices in Computational Social Science. A scoping review.

Keywords: scoping review, computational social sciences, data practices, research methods, research ethics

Extended Abstract

Aided by the technological progress of the past decades, the age of "computational everything" is upon us. Computational sub-fields are emerging in disciplines across the board, from biology to humanities, and everything in between. Computational social science is no exception, with the term recording an exponential boom in literature usage starting around year 2000, according to Google Ngrams. But what is computational social science? And more importantly, what are the methodological, data, and ethical practices associated with the research literature that uses the composite term "computational social science" as a keyword. This mapping exercise is not only informative as a snapshot for the current state of the field, but more importatly, it enables us to identify and critically assess the data practices that underpin the coming together of priorly distinct research traditions.

We place the emphasis of our inquiry onto data practices in research built on data pertaining to human subjects, narrowing in on both methodological approaches and ethical considerations. We expect the relationshp between research built on data pertaining to human subjects and the ethical considerations under which such research is developed to be mediated by a number of factors, such as teritorial jurisdictions (eg. GDPR, or Bundesdatenschutzgesetz, just to name a few), or ethical codes of conduct that regulate research practices at universities. Our proclivity towards analyzing specifically research practices on data pertaining to human subjects comes from the need to grasp critically the extent to which social science research traditions can, and/or should, embrace computational modes of processing increasingly large amounts of data generated by human activity online or through personal devices and technologies.

Searching Scopus, we identified 849 publications that have used the term "computational social science" in the period 1999-2023 as a keyword. The majority of the publications are indexed under the subject computer science (35.1%), followed by social sciences (21.4%). Mathematics and engineering together make a total of 15% of the publications, and we expect this to be where the methodological innovations are springing. Other social science disciplines, as well as business, make up the other 30% of the publications. From this corpus of literature spanning across disciplines, a relatively small percentage engages directly with data pertaining to human subjects. Conversely, the studies that do, are largely social media studies.

This paper is trying to illuminate what stands in between the obvious and the possible, with the obvious being the application of computational approaches to social media data, and the possible being applications of such approaches to potentially more consequential, yet sensitive, human matters. We see computational social science methods, and the ethical considerations that play into those, on the one hand as a potential building block at the forefront of coming developments in working with data at scale to learn about people and societies, and on the other hand as a critical alternative to the commodification of personal data in the service of commercial interests. With data thirsty technologies (eg. AI) becoming (at least) discursively ubiquitous, we assume a normative stance in attempting to outline ethical data practices that serve, rather than hamper, human interests, both presently and in the long term.

References (preliminary)

Awad, E., Levine, S., Anderson, M., Anderson, S.L., Conitzer, V., Crockett, M.J., Everett, J.A., Evgeniou, T., Gopnik, A., Jamison, J.C., Kim, T.W., Liao, S.M., Meyer, M.N., Mikhail, J., Opoku-Agyemang, K., Borg, J.S., Schroeder, J., Sinnott-Armstrong, W., Slavkovik, M., & Tenenbaum, J.B. (2022). Computational ethics. *Trends in Cognitive Sciences, 26*, 388-405.

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology: Theory and Practice, 8(1), 19-32. doi:10.1080/1364557032000119616

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertész, J., Loreto, V., Moat, S., Nadal, J., Sanchez, A., Nowak, A., Flache, A., San Miguel, M.C., & Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics, 214*, 325-346.

Edelmann A, Wolff T, Montagne D, Bail CA. (2020). Computational Social Science and Sociology. Annu Rev Sociol. Jul;46(1):61-81. doi: 10.1146/annurev-soc-121919-054621

Lazer, D., Pentland, A., Watts, D.J., Aral, S., Athey, S., Contractor, N.S., Freelon, D., González-Bailón, S., King, G., Margetts, H.Z., Nelson, A., Salganik, M.J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science, 369*, 1060 - 1062.

Leslie, D. (2022). Don't "research fast and break things": On the ethics of Computational Social Science. *ArXiv, abs/2206.06370*.

Lundberg, I., Narayanan, A., Levy, K.E., & Salganik, M.J. (2018). Privacy, Ethics, and Data Access: A Case Study of the Fragile Families Challenge. Socius, 5.

Radford, J., & Joseph, K. (2020). Theory In, Theory Out: The Uses of Social Theory in Machine Learning for Social Science. *Frontiers in Big Data, 3*.

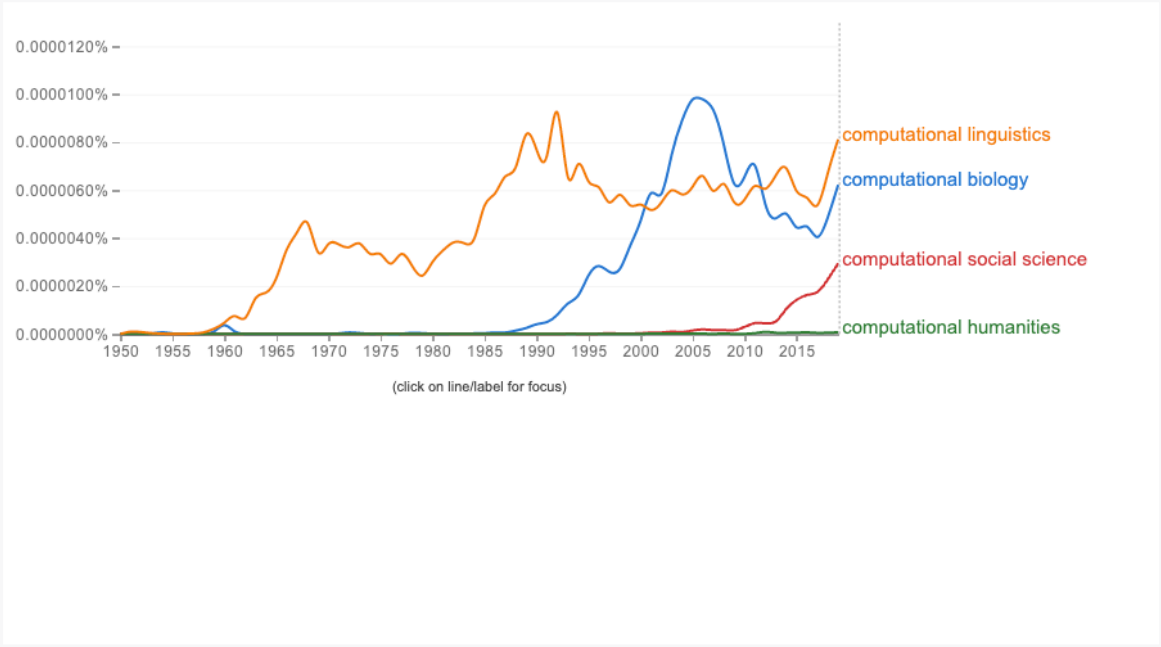

Figure 1: "Computational Everything"

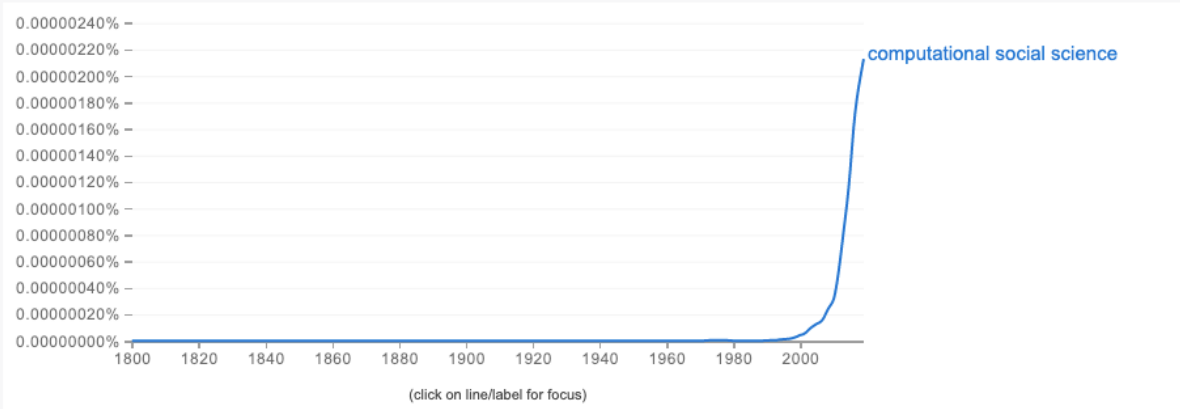

Figure 2: "Computational Social Science"