Computational social science with confidence

Carolina E.S. Mattsson[1]*

*Correspondence:
mattsson.c@northeastern.edu
[1]Leiden Institute of Advanced
Computer Science, Leiden
University, Leiden, The Netherlands

Abstract

There is an ongoing shift in computational social science towards validating our methodologies and improving the reliability of our findings. This is tremendously exciting in that we are moving beyond exploration, towards a fuller integration with theory in social science. We stand poised to advance also new, better theory. But, as we look towards this future we must also work to update our conventions around training, hiring, and funding to suit our maturing field.

Keywords: Computational social science; Collective emotion; Social media

1 Introduction

There is an ongoing shift in computational social science towards validating our methodologies and improving the reliability of our findings. This commentary invokes the 2021 International Conference for Computational Social Science (IC2S2) keynote address by Dr. David Garcia as a springboard for discussing these developments and considering their implications. Collective emotion is an excellent example of an area where computational social science has built on early successes and is moving beyond exploration; I put forward companion examples in economics and epidemiology where similar efforts are at an earlier or later stage. Drawing on points raised in the keynote, I argue that more mature methodologies allow for fuller integration with theory in social science. We now stand poised to advance new, better theory. At the same time, this promising future for computational social science will challenge our conventions around training, hiring, and funding in specific ways. Dr. Garcia's keynote hints at the pressures facing early-career researchers and I discuss these explicitly. I go on to suggest that we might draw inspiration from mature scientific fields with established divisions of labor and institutionalized investment in research infrastructures. In this view, social "macroscopes" are scientific instruments and those with the expertise to build, maintain, and calibrate these valuable tools should be in permanent roles. Proper support for the broad use of maturing methodologies would further computational social science, with confidence.

2 Commentary

In his 2021 IC2S2 keynote address, Dr. David Garcia shares with us his recipe for serving up trustworthy findings in computational social science. The secret is to develop

dependable methods that can stand up to even the highest form of scientific scrutiny: pre-registered predictions. Preregistration is a particular study design where an analysis methodology and specific hypotheses are published prior to data collection. In adopting this approach for methodological validation, Garcia and co-authors pre-registered a research plan to asses the question: Is there a positive correlation between large-scale aggregates of emotional expression on Twitter and self-reported mood in UK representative surveys? The research plan was made publicly available via the pre-registration service aspredicted.org, outlining the data collection approach, filtering steps, and analyses to be conducted plus the criteria for assessing success. Notably, the study design was published online before the start of the period over which the results would be assessed (https://aspredicted.org/blind.php?x=r89nv2).

The answer to the proposed question is, thankfully, yes [1, 2]. Social media can be used as a so-called macroscope to capture collective emotional experiences at the scale of entire countries. But, this is far from automatic. Several key methodological steps are used to transform raw Twitter data into timeseries that reflect the expression of particular emotions, including filtering for original content from the relevant geographic area and measures to exclude spam and mass media. Aspects of this has perhaps become standard, but that is the entire point. Solid methods ought to be standardized! Dr. Garcia and co-authors also make sure to include some of the latest tools for evaluating emotional content in Tweets; deep learning has not been around long enough to be standardized. Moreover, one crucial element of the methodology—gender re-weighting—is certainly not obvious. The authors note that much of the emotion expressed on Twitter is likely to be sensed in others rather than felt by the user themselves. That this helps alleviate other demographic biases in the Twitter user base, but not gender, is intriguing. Definitely important to note in the development of appropriately calibrated macroscopes for collective emotions.

Advances in calibration and validation are happening also in other corners of computational social science. For example, substantial progress has been made in using card payment data to first reconstruct and then disaggregate official economic statistics about consumer spending [3]. Also here, this is far from automatic. Considerable creativity is needed to overcome instability in the user base of the data provider and to filter relevant cardpayments, including a bespoke mapping between the merchant categorizations used by the data provider and those used by the US Census Bureau. This emerging effort within macroeconomics would benefit from more exposure to methodological research in collective emotion on how to reliably incorporate digital trace data. But, digital trace data doesn't always come out ahead, either. Over in the field of epidemiology, ongoing efforts at validation have come to favor wastewater surveillance over social media data for monitoring contagion [4].

Across computational social science, streamlined validation methods become especially valuable as the world shifts beneath our feet. It is too early to tell whether recent changes at Twitter, and the growth of competing platforms, have moved our macroscopes meaningfully out of focus. Undoubtedly, the first studies to this effect are already in preparation. While changes in platform governance are especially dramatic, there are also more pernicious threats to the validity of our tools over time [5]. Ever-shifting patterns in user composition, in media use, in platform operations, and perhaps in many other factors add up to an overarching issue of drift [6]. Fitted models will tend to experience a steady erosion of predictive power as users turn over, language changes, and priorities shift. Our macro-

scopes need ongoing calibration and this implies a need for routinized and transparent validation.

This need for ongoing calibration reflects a broader shift in the field. Slowly but surely we have amassed a considerable collection of neat findings about collective attention, collective emotion, and collective behavior. About the social world at scale. Much of the initial promise of computational social science [7] has come to pass and we have begun to establish new facts about the world that we had not known before. By setting ourselves higher methodological standards—properly calibrating our instruments—we strengthen the empirical basis for this knowledge. Nicely validated methods for measuring collective emotion and consumer spending, in particular, have both demonstrated their promise in this regard by allowing spatially and temporally fine-grained documentation of the societal response to the COVID-19 pandemic [8, 9]. There are major opportunities on the horizon, and now is the time to take a longer view.

New facts lead to new questions, and Dr. Garcia is especially bullish about new possibilities to test previously untestable social theories. For example, the effect of collective emotional expression on social solidarity becomes an empirical question when these things can be reliably monitored via social media. Garcia & Rimé (2017) examine the longer-term effect of participating in collective emotional expression during and after terrorist attacks in France in November 2015. Considering tweets from a large sample of affected users located in France, they find that those who participated in the collective negative emotional response to the attacks shortly after they happened also used more prosocial terms for many months thereafter [10]. This has clear implications as a mechanism whereby social capital translates into societal resilience [11]. While social synchronization after terrorist attacks is well-documented using digital traces [12–14] its salutary effect on the participating collective had been tricky to study. Now, a century after Durkheim published about his principle of 'collective effervescence' it has become possible to conduct straightforward scientific observation of this specific social phenomenon at the relevant scale.

As Dr. Garcia touches on in his keynote, we also now have the raw material for developing more complete theories about societies at scale. For instance, the closer we look the more profound are the heterogineities we find. A truly minute share of Twitter users are responsible for introducing most of the news coming from questionable sources [15] and, potentially, the most-shared negative content [16]. Heterogeneity is also becoming highly relevant in other corners of social science. There is growing evidence that even basic monetary indicators, such as inflation, obscure dramatic differences among people and between places [3, 17, 18]. Heterogineity is clearly under-theorized across a wide range of social phenomena, and computational social science is especially well-suited to change this. Notably, the scale and resolution of the data that computational social scientists use can often provide plentiful representation of small subpopulations. Foucault Welles (2014) nicely articulates this point, noting that small subsets of very large datasets still leave substantial study populations. Freed from the constraints around respondent sampling, minorities and statistical outliers can be better studied in and of themselves [19]. And while digital records offer a very partial portrait of society as a whole, they sometimes allow insight into pockets that are otherwise truly difficult to observe [20]. We stand poised to advance new, better theories that offer a more complete understanding of heterogeneous societies.

While this shift towards integration with theory is undoubtedly promising, it also stresses our conception of what progress in computational social science looks like. Dr. Garcia jokes that he could only stand to open the results of his preregistered study because he already had tenure. If the predicted results failed to materialize, for whatever reason, his career would survive it. The joke is funny because there lies an element of truth in it, and those of us in more junior roles can't lose our sense of humor about this element of truth or else we'd despair. Collective practices in training, in hiring, and in funding computational social science have established themselves during a decade of remarkable exploration. Our collective success at uncovering exciting new things has left us with conventions that favor flexible research designs; all the better to draw maximum advantage from novel datasets, trendy methods, and timely funding calls.

But maturing fields require research of a markedly different flavor, with more established methodologies and more careful research designs. As Dr. Garcia notes, it takes time to do it right. This means we need to give our junior researchers the time to do it right. Since more mature fields rest on a larger body of knowledge, students may need more time for training simply because there is more to learn. I've previously written about how universal challenges are magnified when a field draws on several distinct disciplines, and this is certainly the case in computational social science [21]. So perhaps we can turn to exceptionally mature fields for a bit of inspiration?

The word "macroscope" evokes an exceedingly apt comparison to laboratory sciences, such as microbiology or condensed matter physics. These mature fields treat their microscopes as valuable scientific instruments, which has both obvious and less obvious implications. Most obviously, the production or purchase of scientific instruments can itself be expensive, sometimes enormously so, and this is understood by funding agencies. Less obviously, the ongoing maintenance of valuable scientific instruments is institutionally prioritized. Least obviously, valuable scientific instruments enable (or, require) substantial division of labor. Nobody expects a doctoral student to build their own electron microscope. Instead, there are mechanisms in place to ensure that students have access to the instruments they need and receive appropriate training in their operation. Calibration is routine, following standardized procedures, and much emphasis is placed on validating individual findings. In a mature laboratory, there can be a constellation of lab managers, technical staff, and technicians in place to support the scientific work that takes place using the valuable scientific instruments... and that's not to mention the funding and specialized personnel behind radio telescopes or inertial confinement fusion reactors.

Computational social science would do well to embrace the division of labor, and begin growing our field in more responsible ways. From the perspective of a single research lab, standing up a data-gathering infrastructure is no longer a project fit for a doctoral student, no matter how dedicated or enthusiastic. Data-gathering infrastructure has become an investment into a long-term scientific capability and doing it right requires considerable expertise. While this makes it undoubtedly tempting to hire a postdoc, that is also a lose-lose proposition. Hiring someone into a temporary position to set up a valuable scientific instrument means they will not have long to train other lab members on how to use it properly, much less how to maintain it over time. Nor will the postdoc be around to see the long-term benefits of their work. There are a growing number of ways for infrastructural efforts to be reflected on a researcher's Google Scholar profile, with new journals offering peer-review for data documentation and open source software. However, those

who take on this crucial work also need tangible career opportunities [22]. Permanent roles for those building and maintaining the valuable instruments of computational social science, as we see in more mature sciences, are a must.

Beyond individual research labs, there is a potentially important role here for university libraries and outside institutions. University libraries have relevant experience in the procurement and curation of digital resources, and in offering appropriate training; there may be overlapping needs with research programs in other disciplines (e.g., digital humanities, population health, or even marketing). There are also institutions outside universities with a history of supporting social science research, specifically. One such organization with initiatives in this direction is the Leibniz Institute for the Social Sciences (GESIS), where Dr. Garcia is a part of the Digital Behavioral Data Coordination Group. GESIS offers training in computational social science and is building out infrastructure for collecting, hosting, and processing digital behavioral data for the benefit of social science researchers. There are also national initiatives that may be well poised to support research infrastructures in computational social science, such as social "macroscopes". One that comes to mind is ODISSEI (Open Data Infrastructure for Social Science and Economic Innovations) in the Netherlands, which has facilities in place to run representative surveys and to perform compute-intensive analyses on highly sensitive data. Surely there are more, and the path forward for such institutions should be clear. The data needed to study collective attention, collective emotion, and collective behavior is out there, and the methods are maturing. What's needed is institutionalized support.

3 Conclusions

The study of collective attention, collective emotion, and collective behavior has become an established area of research within computational social science, which is itself a maturing field. The standardization of methods and best practices is a welcome development; the future is bright. Increased confidence in our approaches will help computational social scientists look increasingly outward. We stand poised to contribute back to theories in the social sciences and forwards to formulating more complete theories. At the same time, mature fields in most disciplines have developed institutions that support a productive division of labor. It is no longer appropriate to expect that students (or worse, postdocs) should build up a data-gathering infrastructure as a prelude to the science itself. "Macroscopes" that can measure collective attention, emotion, or behavior at scale—accurately and consistently—are valuable to science in and of themselves. Computational social science, as a field, ought to treat these as valuable scientific instruments. This will require institutional investment and, crucially, permanent roles for those who have become experts in building, maintaining, and calibrating our data-gathering infrastructures.

Abbreviations
IC2S2, International Conference for Computational Social Science; GESIS, Leibniz Institute for the Social Sciences; ODISSEI, Open Data Infrastructure for Social Science and Economic Innovations.

Data availability
Not applicable.

Declarations

References

1. Garcia D, Pellert M, Lasser J, Metzler H (2021) Social media emotion macroscopes reflect emotional experiences in society at large. Technical Report. https://doi.org/10.48550/arXiv.2107.13236. http://arxiv.org/abs/2107.13236. Accessed 2022-06-14
2. Pellert M, Metzler H, Matzenberger M, Garcia D (2022) Validating daily social media macroscopes of emotions. Sci Rep 12(1):11236. https://doi.org/10.1038/s41598-022-14579-y. Accessed 2022-12-31
3. Aladangady A, Aron-Dine S, Dunn W, Feiveson L, Lengermann P, Sahm C (2019) From transactions data to economic statistics: constructing real-time, high-frequency, geographic measures of consumer spending. Big data for 21st century economic statistics. Accessed 2019-09-18
4. Diamond MB, Keshaviah A, Bento AI, Conroy-Ben O, Driver EM, Ensor KB, Halden RU, Hopkins LP, Kuhn KG, Moe CL, Rouchka EC, Smith T, Stevenson BS, Susswein Z, Vogel JR, Wolfe MK, Stadler LB, Scarpino SV (2022) Wastewater surveillance of pathogens can inform public health responses. Nat Med 28(10):1992–1995. https://doi.org/10.1038/s41591-022-01940-x. Accessed 2022-12-31
5. Lazer D, Kennedy R, King G, Vespignani A (2014) The parable of Google Flu: traps in big data analysis. Science 343(6176):1203–1205. https://doi.org/10.1126/science.1248506. Accessed 2019-08-28
6. Salganik MJ (2017) Bit by bit: social research in the digital age, Illustrated edn. Princeton University Press, Princeton
7. Lazer D, Pentland A, Adamic L, Aral S, Barabasi A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M (2009) Computational social science. Science 323(5915):721–723. https://doi.org/10.1126/science.1167742. Accessed 2015-07-30
8. Carvalho VM, Garcia JR, Hansen S, Ortiz A, Rodrigo T, Rodríguez Mora JV, Ruiz P (2021) Tracking the COVID-19 crisis with high-resolution transaction data. R Soc Open Sci 8(8):210218. https://doi.org/10.1098/rsos.210218. Accessed 2022-04-19
9. Metzler H, Rimé B, Pellert M, Niederkrotenthaler T, Di Natale A, Garcia D (2022) Collective emotions during the COVID-19 outbreak. Emotion 23(3):844–858. https://doi.org/10.1037/emo0001111
10. Garcia D, Rimé B (2019) Collective emotions and social resilience in the digital traces after a terrorist attack. Psychol Sci 30(4):617–628. https://doi.org/10.1177/0956797619831964. Accessed 2022-12-30
11. Aldrich DP, Meyer MA (2015) Social capital and community resilience. Am Behav Sci 59(2):254–269. https://doi.org/10.1177/0002764214550299. Accessed 2016-02-29
12. Bagrow JP, Wang D, Barabási A-L (2011) Collective response of human populations to large-scale emergencies. PLoS ONE 6(3):17680. https://doi.org/10.1371/journal.pone.0017680. Accessed 2015-10-27
13. Sundsoy PR, Bjelland J, Canright G, Engo-Monsen K, Ling R (2012) The activation of core social networks in the wake of the 22 July Oslo bombing. IEEE, Los Alamitos, pp 586–590. https://doi.org/10.1109/ASONAM.2012.99. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6425705. Accessed 2016-04-29
14. Eriksson M (2016) Managing collective trauma on social media: the role of Twitter after the 2011 Norway attacks. Media Cult Soc 38(3):365–380. https://doi.org/10.1177/0163443715608259. Accessed 2022-12-30
15. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on Twitter during the 2016 U.S. presidential election. Science 363(6425):374–378. https://doi.org/10.1126/science.aau2706. Accessed 2022-12-31
16. Schöne J, Garcia D, Parkinson B, Goldenberg A (2022) Negative expressions are shared more on Twitter for public figures than for ordinary users. https://doi.org/10.31234/osf.io/wng5v. PsyArXiv. https://psyarxiv.com/wng5v/. Accessed 2022-12-31
17. Argente D, Lee M (2021) Cost of living inequality during the great recession. J Eur Econ Assoc 19(2):913–952. https://doi.org/10.1093/jeea/jvaa018. Accessed 2022-01-13
18. Mattsson CES, Luedtke A, Takes FW (2023) Inverse estimation of the transfer velocity of money. https://doi.org/10.48550/arXiv.2209.01512. http://arxiv.org/abs/2209.01512. Accessed 2023-08-14
19. Foucault Welles B (2014) On minorities and outliers: the case for making Big Data small. Big Data Soc 1(1):205395171454061. https://doi.org/10.1177/2053951714540613. Accessed 2019-08-27
20. de Vries I, Radford J (2022) Identifying online risk markers of hard-to-observe crimes through semi-inductive triangulation: the case of human trafficking in the United States. Br J Criminol 62(3):639–658. https://doi.org/10.1093/bjc/azab077. Accessed 2023-08-14
21. Mattsson C (2019) Theory and tools in the age of big data. https://ocean.sagepub.com/blog/theory-and-tools-in-the-age-of-big-data. Accessed 2022-12-31
22. Bennett A, Garside D, Praag CGV, Hostler TJ, Garcia IK, Plomp E, Schettino A, Teplitzky S, Ye H (2023) A manifesto for rewarding and recognizing team infrastructure roles. J Trial Error. https://doi.org/10.36850/mr8. Accessed 2023-08-17

Publisher's Note