

Introduction to computational social science

Jacob Habinek



LINKÖPING
UNIVERSITY

Institute for Analytical Sociology

What is computational social science?

“An emerging field that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.”

From Lazer *et al.* (2009)

“Computational social science” <https://doi.org/10.1126/science.1167742>

An interdisciplinary field that advances theories of human behavior by applying computational techniques to large datasets from social media sites, the Internet, or other digitized archives such as administrative records.”

From Edelmann *et al.* (2020: 15)

“CSS and sociology” <https://doi.org/10.1146/annurev-soc-121919-054621>

"Anything that's cool."

From Salganik (2018)

What is computational social science?

Coherent... more or less
Contentious
Organized
Episodic

“A general theory of scientific/intellectual movements” <https://doi.org/10.1177/000312240507000202>

RESEARCH

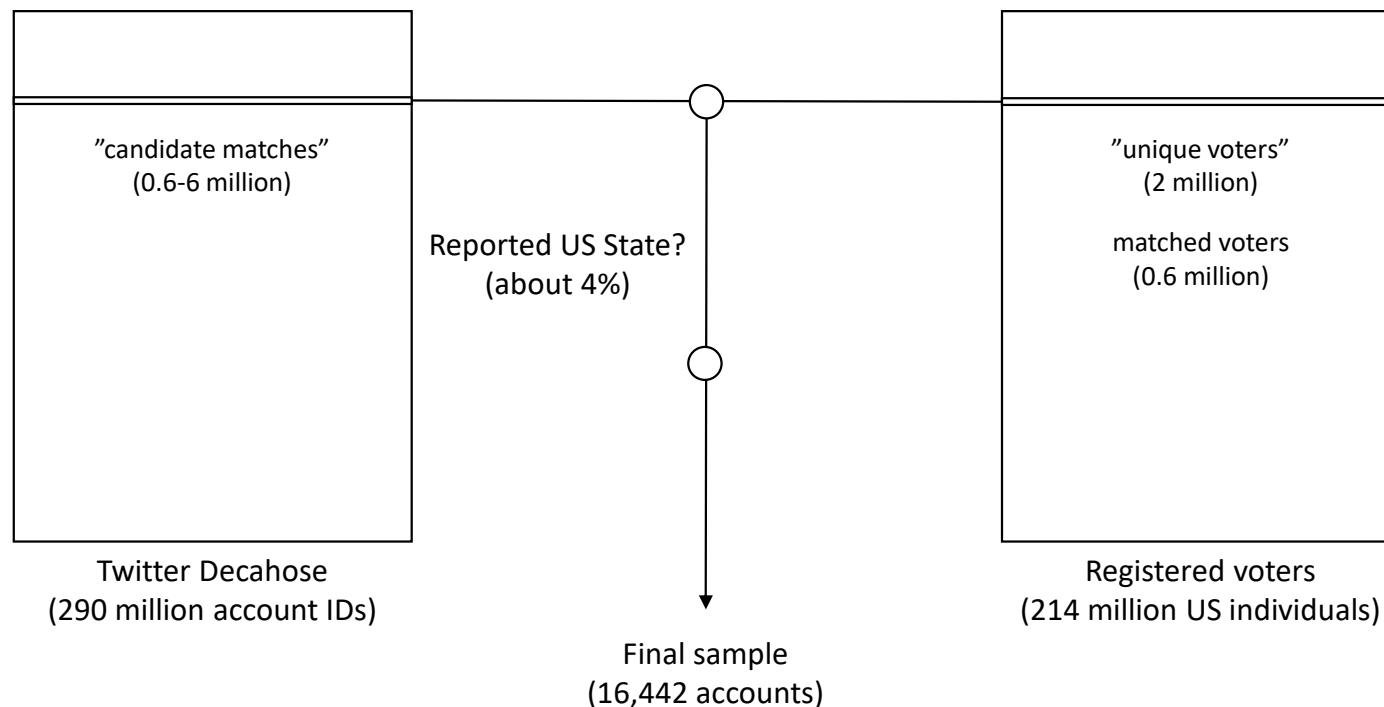
POLITICAL SCIENCE

Fake news on Twitter during the 2016 U.S. presidential election

Nir Grinberg^{1,2*}, Kenneth Joseph^{3*}, Lisa Friedland^{1*},
Briony Swire-Thompson^{1,2}, David Lazer^{1,2†}

The spread of fake news on social media became a public concern in the United States after the 2016 presidential election. We examined exposure to and sharing of fake news by registered voters on Twitter and found that engagement with fake news sources was extremely concentrated. Only 1% of individuals accounted for 80% of fake news source exposures, and 0.1% accounted for nearly 80% of fake news sources shared. Individuals most likely to engage with fake news sources were conservative leaning, older, and highly engaged with political news. A cluster of fake news sources shared overlapping audiences on the extreme right, but for people across the political spectrum, most political news exposure still came from mainstream media outlets.

<https://doi.org/10.1126/science.aau2706>



<https://doi.org/10.1126/science.aau2706>

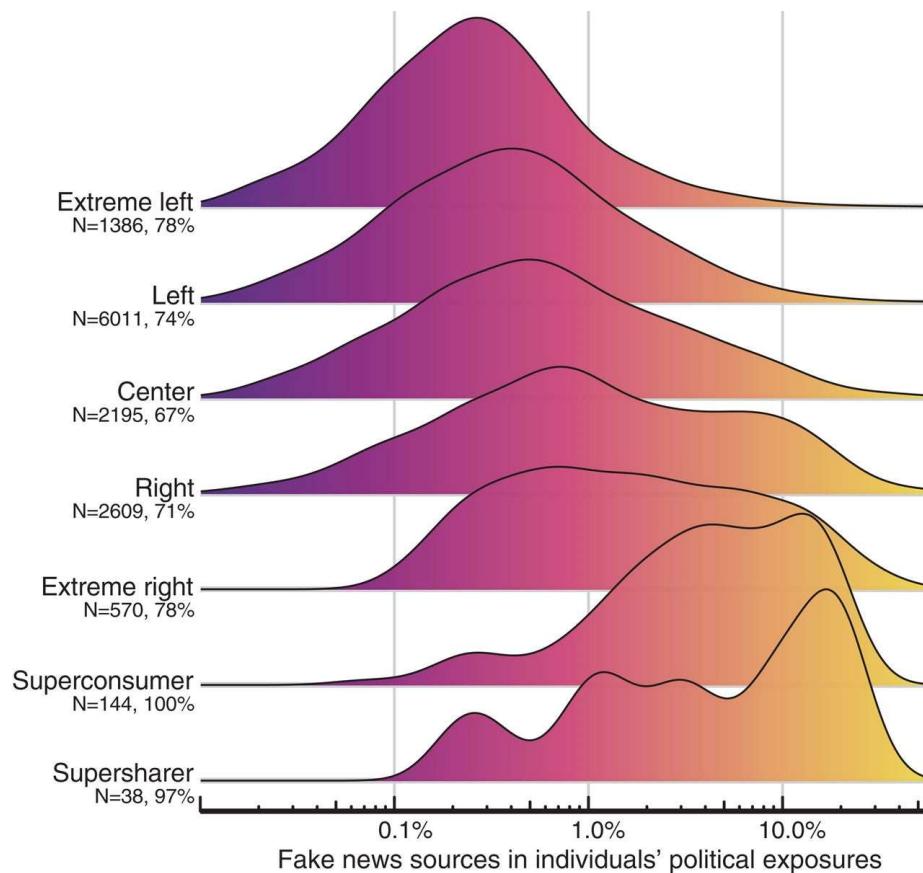


Fig. 3 Probability density estimates for the percentage of content from fake news sources in people's news feeds (for people with any fake news exposures).

The number of individuals in each subgroup (N) and the percent with nonzero exposures to fake news sources are shown.

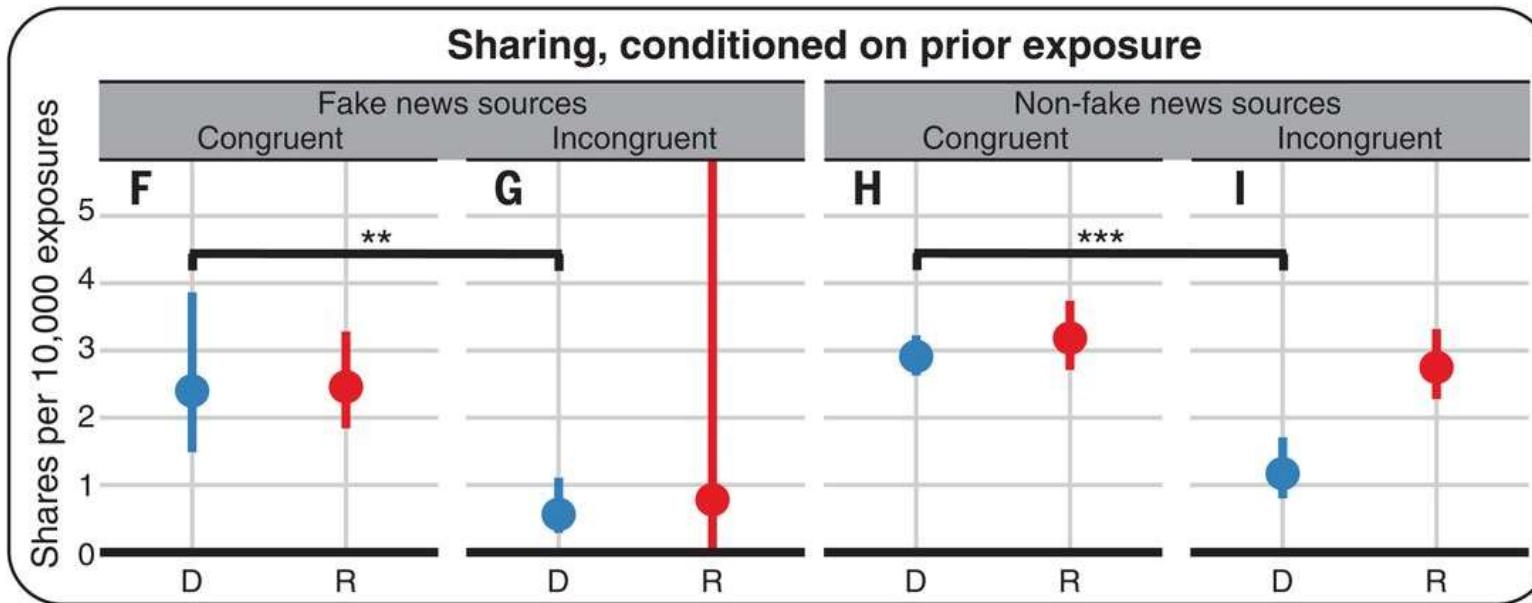


Fig. 4 Key individual characteristics associated with exposure to and sharing of fake news sources.

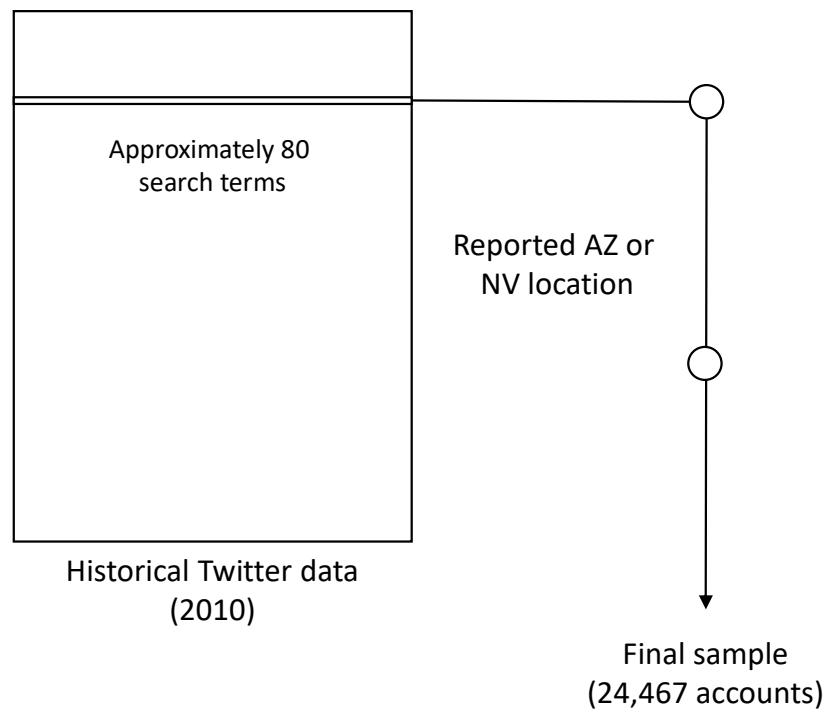
The proportion of an individual's political exposures coming from fake news sources as a function of (F to I) Likelihood of a liberal (D) or conservative (R) individual sharing a political URL to which they have been exposed, depending on the political congruency and veracity of the source: (F) congruent and fake, (G) incongruent and fake, (H) congruent and nonfake, and (I) incongruent and nonfake. Brackets indicate significantly different pairs: ** $P < 0.01$, *** $P < 0.001$.

Do Anti-Immigrant Laws Shape Public Sentiment? A Study of Arizona's SB 1070 Using Twitter Data¹

René D. Flores
University of Washington

Scholars have debated whether laws can influence public opinion, but evidence of these “feedback” effects is scant. This article examines the effect of Arizona’s 2010 high-profile anti-immigrant law, SB 1070, on both public attitudes and behaviors toward immigrants. Using sentiment analysis and a difference-in-difference approach to analyze more than 250,000 tweets, the author finds that SB 1070 had a negative impact on the average sentiment of tweets regarding immigrants, Mexicans, and Hispanics, but not on those about Asians or blacks. However, these changes in public discourse were not caused by shifting attitudes toward immigrants but by the mobilization of anti-immigrant users and by motivating new users to begin tweeting. While some scholars propose that punitive laws can shape people’s attitudes toward targeted groups, this study shows that policies are more likely to influence behaviors. Rather than placating the electorate, anti-immigrant laws may stir the pot further, mobilizing individuals already critical of immigrants.

<https://doi.org/10.1086/692983>



<https://doi.org/10.1086/692983>

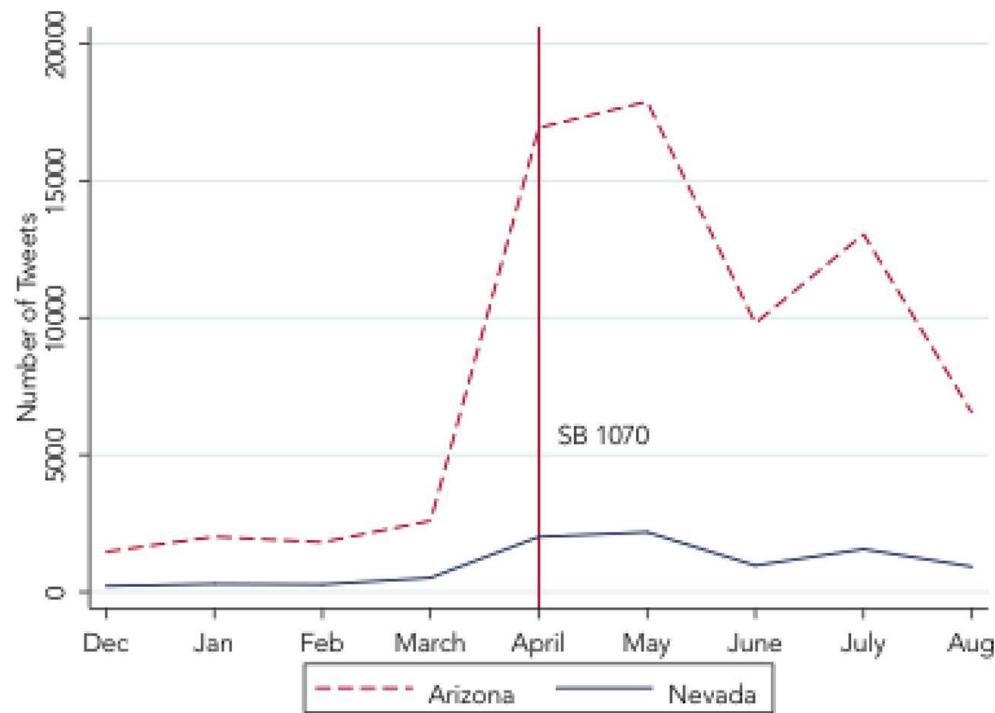


Fig. 2. Number of Twitter messages related to immigrants per month in Arizona and Nevada (December 2010–August 2011). The vertical line on April 2010 indicates when the Arizona governor approved SB 1070.

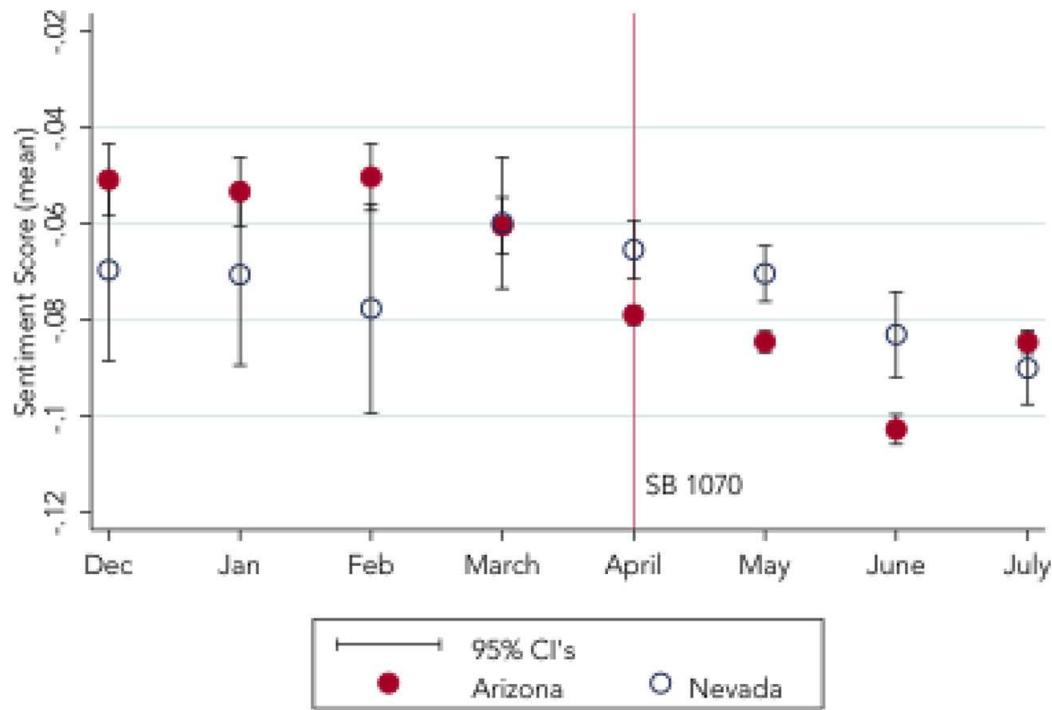


Fig. 3. Average sentiment score of tweets about immigrants. The vertical lines represent 95% confidence intervals. The vertical line on April 2010 indicates when the Arizona governor approved SB 1070.

What is computational social science?

- Computational *and* social science
- Involves complex ethical questions
- Combines readymades and custommades

Computer science

Study anything

Methods driven

Large found data

Prediction

Social science

Study social things

Question driven

Small designed data

Explanation

http://videolectures.net/icml2015_wallach_social_science/



readymade



custommade

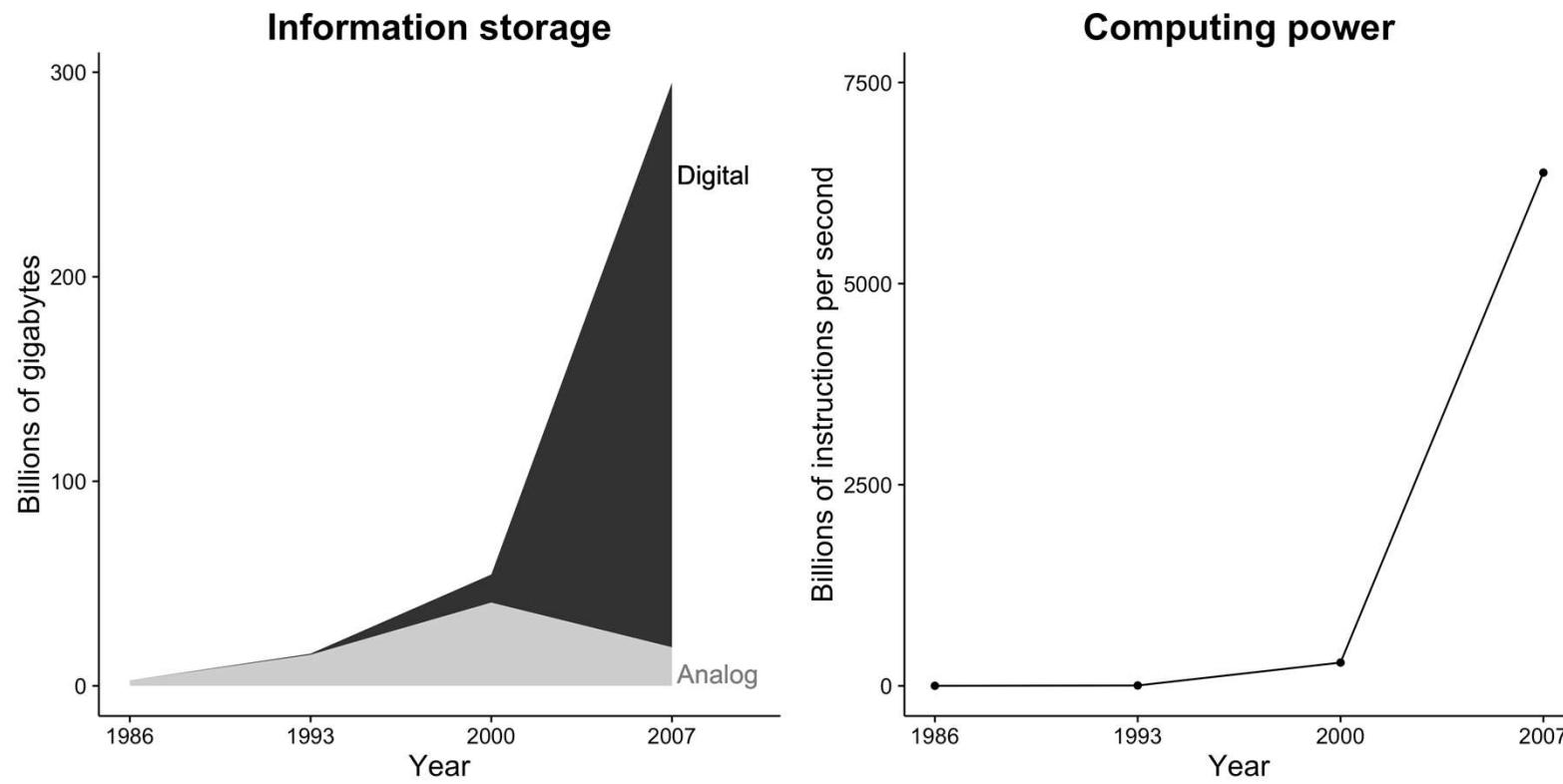
https://commons.wikimedia.org/wiki/File:Duchamp_Fountaine.jpg/

https://commons.wikimedia.org/wiki/File:%27David%27_by_Michelangelo_JBU0001.JPG/

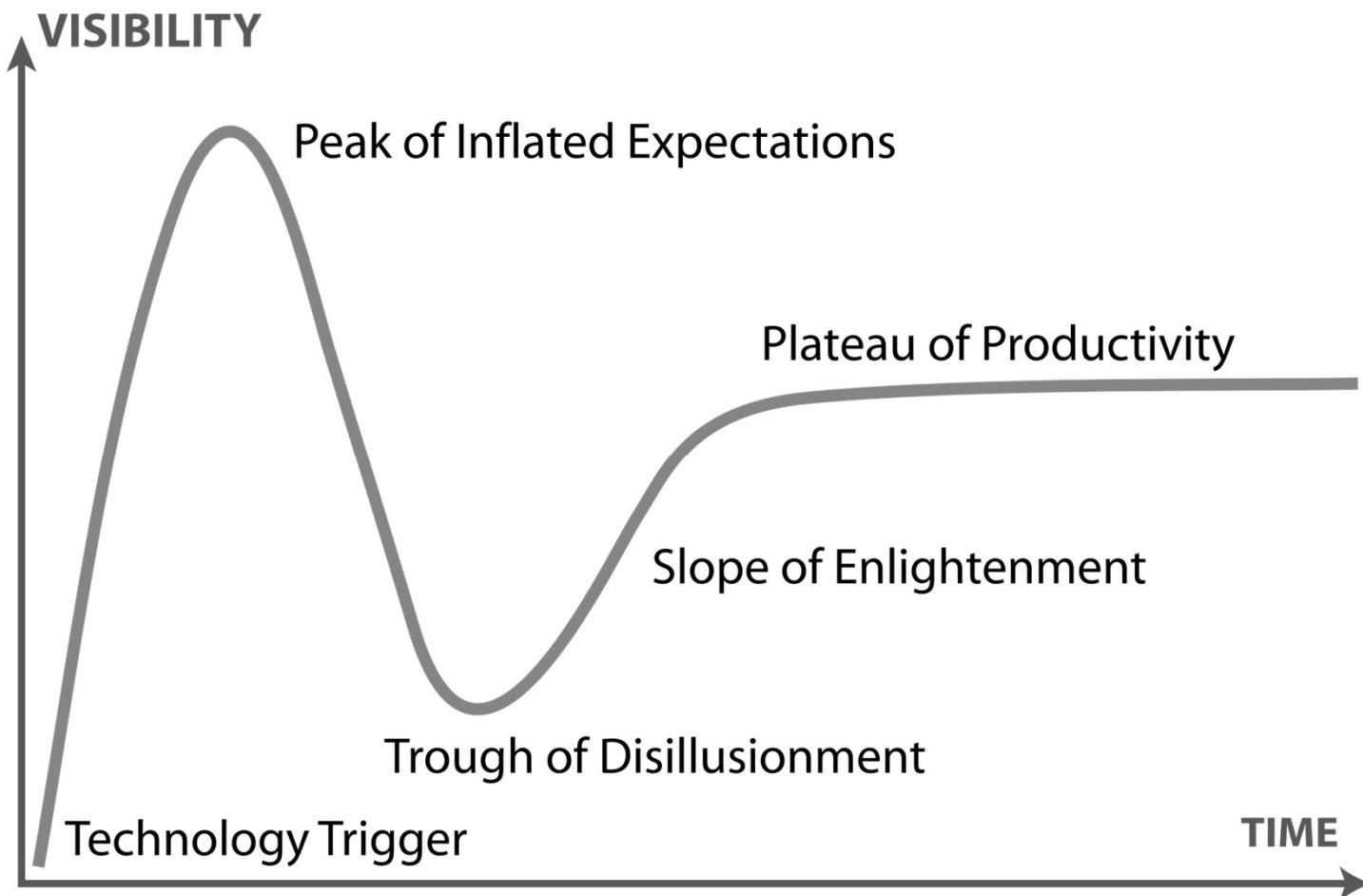
What is computational social science?

- Computational *and* social science
- Involves complex ethical questions
- Combines readymades and custommades

Is computational social science a fad?



Information storage capacity and computing power are increasing dramatically. Further, information storage is now almost exclusively digital. These changes create incredible opportunities for social researchers. Adapted from Hilbert and López (2011), figures 2 and 5.



https://commons.wikimedia.org/wiki/File:Gartner_Hype_Cycle.svg

Article

Papers and patents are becoming less disruptive over time

<https://doi.org/10.1038/s41586-022-05543-x>

Michael Park¹, Erin Leahy² & Russell J. Funk^{1,2,3}

Received: 14 February 2022

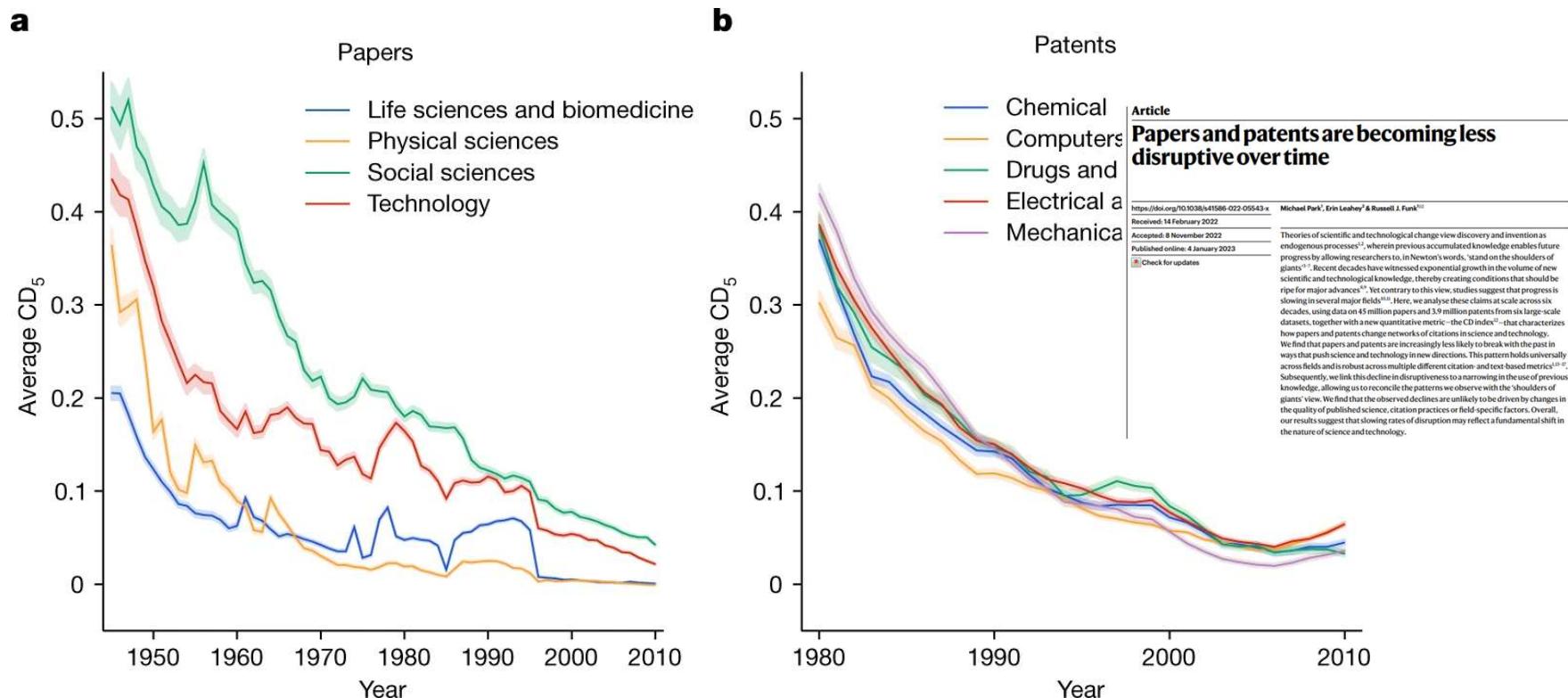
Accepted: 8 November 2022

Published online: 4 January 2023

 Check for updates

Theories of scientific and technological change view discovery and invention as endogenous processes^{1,2}, wherein previous accumulated knowledge enables future progress by allowing researchers to, in Newton's words, 'stand on the shoulders of giants'^{3–7}. Recent decades have witnessed exponential growth in the volume of new scientific and technological knowledge, thereby creating conditions that should be ripe for major advances^{8,9}. Yet contrary to this view, studies suggest that progress is slowing in several major fields^{10,11}. Here, we analyse these claims at scale across six decades, using data on 45 million papers and 3.9 million patents from six large-scale datasets, together with a new quantitative metric—the CD index¹²—that characterizes how papers and patents change networks of citations in science and technology. We find that papers and patents are increasingly less likely to break with the past in ways that push science and technology in new directions. This pattern holds universally across fields and is robust across multiple different citation- and text-based metrics^{13–17}. Subsequently, we link this decline in disruptiveness to a narrowing in the use of previous knowledge, allowing us to reconcile the patterns we observe with the 'shoulders of giants' view. We find that the observed declines are unlikely to be driven by changes in the quality of published science, citation practices or field-specific factors. Overall, our results suggest that slowing rates of disruption may reflect a fundamental shift in the nature of science and technology.

<https://doi.org/10.1038/s41586-022-05543-x>



Decline in CD_5 over time, separately for papers (**a**, $n = 24,659,076$) and patents (**b**, $n = 3,912,353$). For papers, lines correspond to WoS research areas; from 1945 to 2010 the magnitude of decline ranges from 91.9% (social sciences) to 100% (physical sciences). For patents, lines correspond to National Bureau of Economic Research (NBER) technology categories; from 1980 to 2010 the magnitude of decline ranges from 93.5% (computers and communications) to 96.4% (drugs and medical). Shaded bands correspond to 95% confidence intervals.

<https://doi.org/10.1038/s41586-022-05543-x>

What is computational social science?

- Computational *and* social science
- Involves complex ethical questions
- Combines readymades and custommades
- Not a fad

Why SICSS?

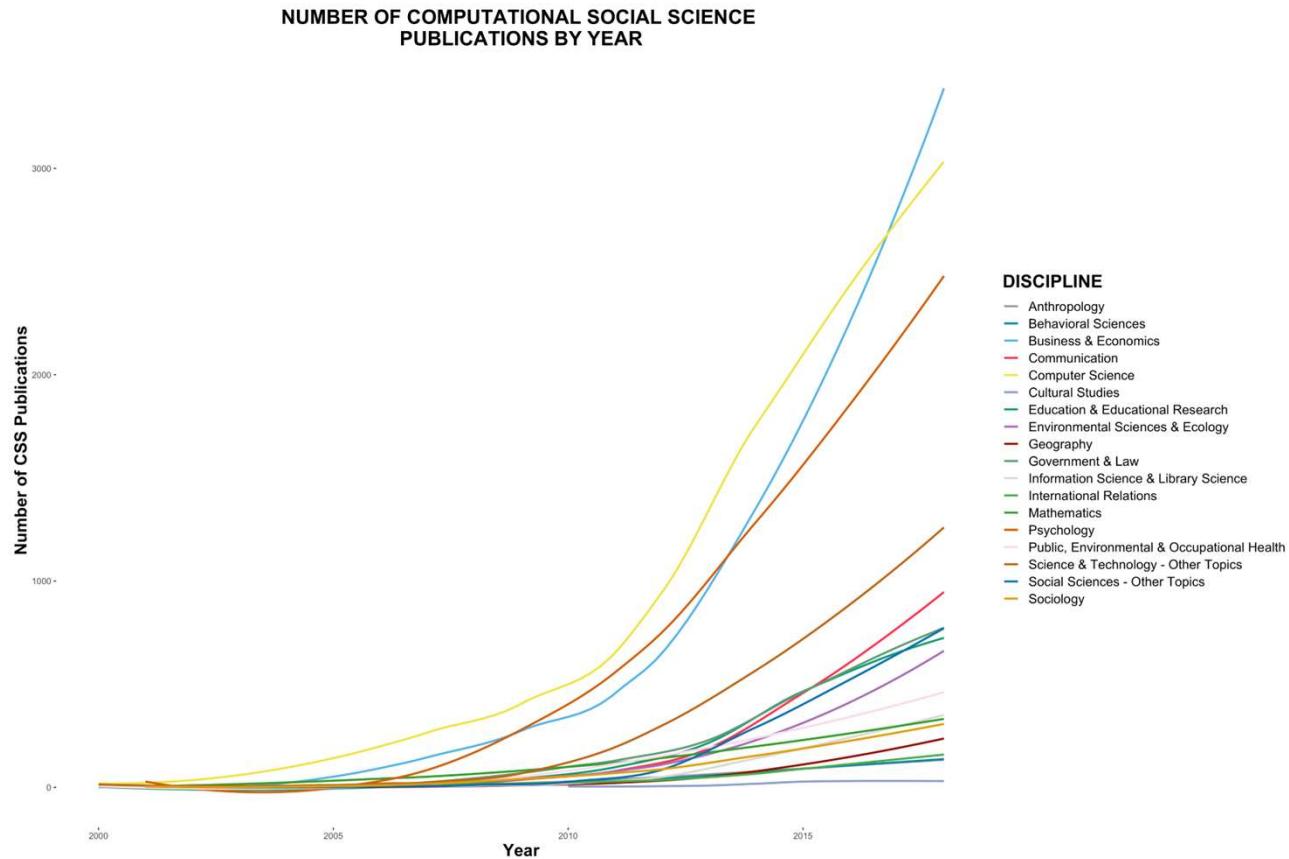
Jacob Habinek



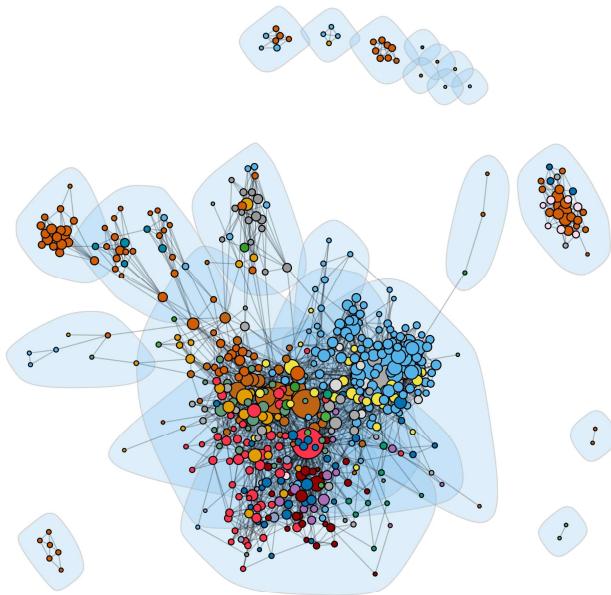
LINKÖPING
UNIVERSITY

Institute for Analytical Sociology

CSS is growing.

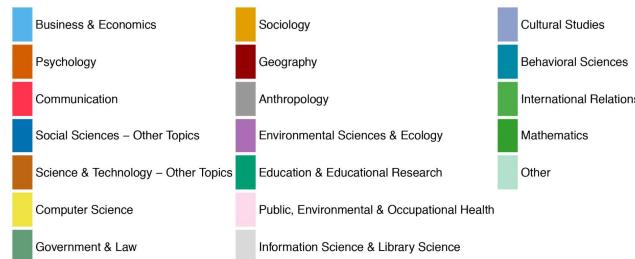


“CSS and sociology” <https://doi.org/10.1146/annurev-soc-121919-054621>



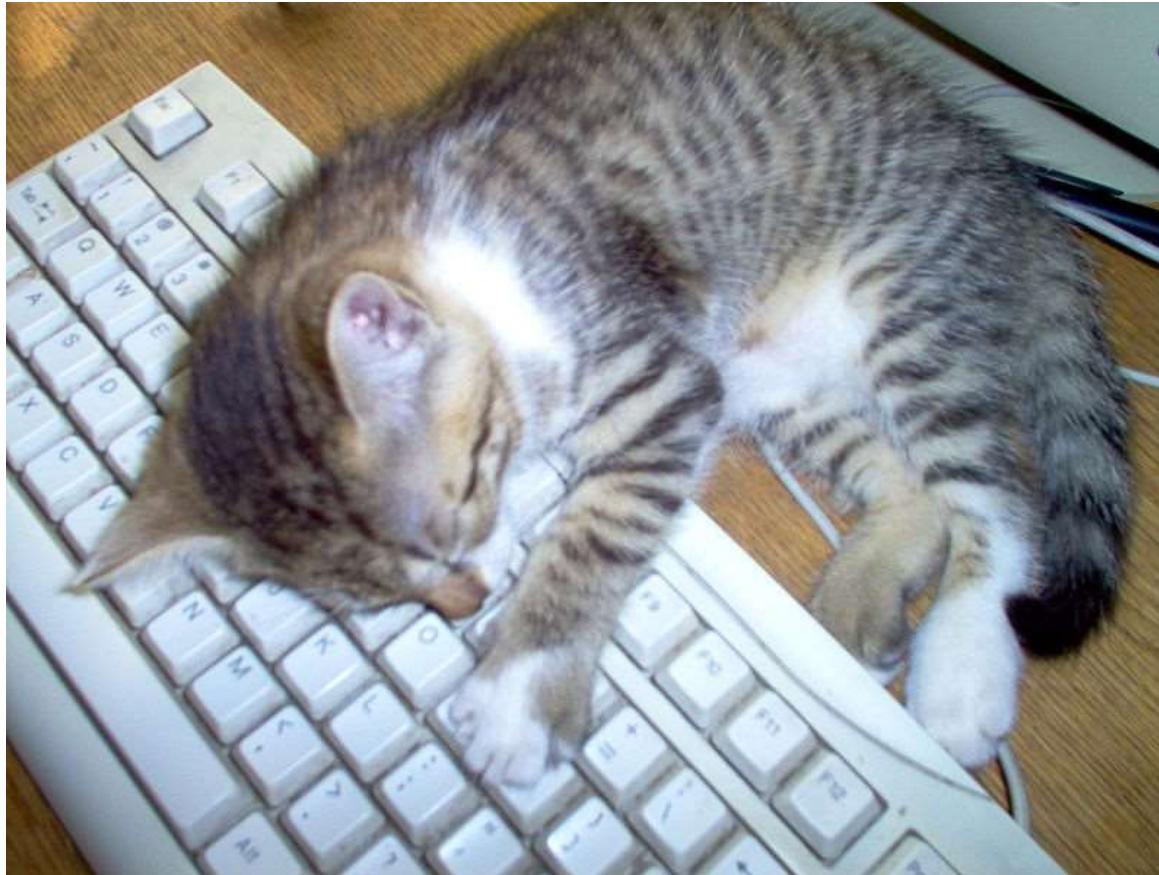
COMPUTATIONAL SOCIAL SCIENCE AS A NETWORK

Nodes colored by first-listed discipline



“CSS and sociology” <https://doi.org/10.1146/annurev-soc-121919-054621>

CSS training is rare.



https://commons.wikimedia.org/wiki/File:Wikipedians_cat.jpg

See also: https://commons.wikimedia.org/wiki/Category:Cats_with_computers

SICSS core principles

Provide state-of-the-art training



<https://commons.wikimedia.org/wiki/File:Cat-and-computer.JPG>

See also: https://commons.wikimedia.org/wiki/Category:Cats_with_computers

Challenge disciplinary divides

Atypical Combinations and Scientific Impact

Brian Uzzi,^{1,2} Satyam Mukherjee,^{1,2} Michael Stringer,^{2,3} Ben Jones^{1,4*}

Novelty is an essential feature of creative ideas, yet the building blocks of new ideas are often embodied in existing knowledge. From this perspective, balancing atypical knowledge with conventional knowledge may be critical to the link between innovativeness and impact. Our analysis of 17.9 million papers spanning all scientific fields suggests that science follows a nearly universal pattern: The highest-impact science is primarily grounded in exceptionally conventional combinations of prior work yet simultaneously features an intrusion of unusual combinations. Papers of this type were twice as likely to be highly cited works. Novel combinations of prior work are rare, yet teams are 37.7% more likely than solo authors to insert novel combinations into familiar knowledge domains.

<https://doi.org/10.1126/science.1240474>

Challenge disciplinary divides

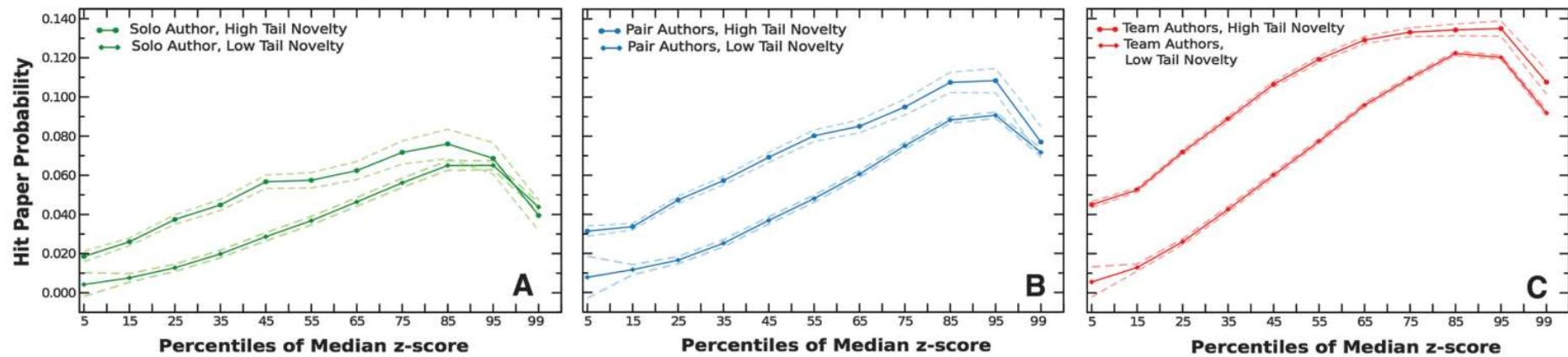


Fig. 4. Novel and conventional combinations in the production of science. (A to C) The interplay between tail novelty, median conventionality, and hit paper probabilities shows remarkable empirical regularities. First, high tail novelty papers have higher impact than low tail novelty papers at (i) any level of conventionality and (ii) regardless of authorship structure. Second, increasing median conventionality is associated with higher impact up to the

85th to 95th percentile of median conventionality, after which the relationship reverses. Third, larger teams obtain higher impact given the right mix of tail novelty and median conventionality. Nonetheless, at low levels of median convention and tail novelty, even teams have low impact, further emphasizing the fundamental relationship between novelty, conventionality, and impact in science.

Reach a broad audience



Keep it open source

The screenshot shows the GitHub repository page for `compsocialscience/summer-institute`. The repository is public, has 337 branches, and 0 tags. It contains 15,641 commits. The repository is associated with `sicss.io` and follows an MIT license. It has 287 stars and 45 watchers, with 250 forks. The repository is used to report on summer institutes in Computational Social Science.

Code | **Issues** 3 | **Pull requests** 7 | **Actions** | **Projects** | **Security** | **Insights**

Code | **About**

Summer Institutes in Computational Social Science

sicss.io

Readme

MIT license

287 stars

45 watching

250 forks

Report repository

Releases

No releases published

Packages

No packages published

Contributors 105

Commit	Message	Date
hsarmiento Merge pull request #3725 from compsocialscience/s...	upgrade: Ruby 2.7.1, bundler 2.1.4	2 hours ago
.circleci	update to every 2 hours	3 years ago
.github	png compression	8 months ago
2017	png compression	8 months ago
2018	png compression	8 months ago
2019	png compression	8 months ago
2020	png compression	8 months ago
2021	remove old application link	5 months ago
2022	Add 3 more posts to In the News	5 months ago
2023	Update schedule.md	last week
_data	Update teaching_assistants.yml	4 hours ago
_includes	Added first week of schedule and changed the date representa...	2 weeks ago
_layouts	Update location_detail.html	4 months ago
_videos	added remaining 2022 videos	5 months ago

<https://github.com/compsocialscience/summer-institute>

Teach the teachers



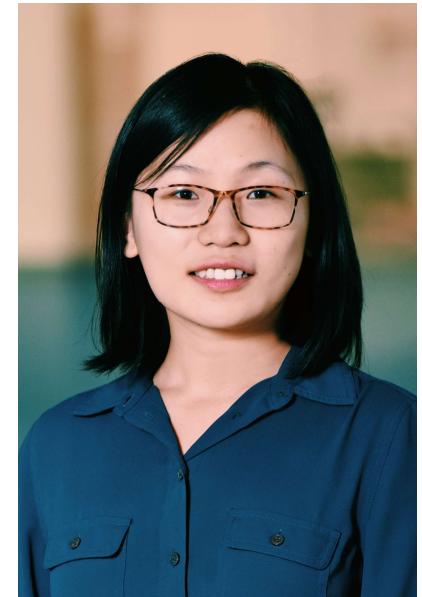
SICSS-London 2021



SICSS-Oxford 2021



SICSS-Paris 2022



SICSS-Chicago 2019

Create a diverse community



Research School in Computational Social Science

[Applications](#) | [News and Events](#) | [Contact us](#)

© THOR BALKHED

The Swedish Interdisciplinary Research School in Computational Social Science provides courses, seminars, and networking opportunities for doctoral students interested in the emerging field of computational social science.

Vast troves of digital data are opening up entirely new possibilities for answering fundamental questions in the social sciences. Social media, smartphones, online libraries, and other digital devices produce continually updated records of human behaviour at unprecedented scales. The emerging interdisciplinary field of computational social science uses methods such as predictive modelling, text and image analysis, and network analysis to analyse massive datasets in new and insightful ways.

The next generation of researchers, however, face substantial barriers in taking full advantage of the tools of computational social science. Most fundamentally, the required knowledge and training often falls outside of traditional doctoral programs. The primary aim of the Swedish Interdisciplinary Research School in Computational Social Science is to provide doctoral students with the up-to-date knowledge and skills necessary to make the best use of digital data on human behavior.

Apply now

We are now accepting applications for the 2023 student cohort. Deadline 3 April 2023.

<https://liu.se/en/research/research-school-in-computational-social-science>

How SICSS-IAS works

Schedule

Morning	Lectures			Group Project Time	Project Presentations	MID	
	Workshops		Project Match-making				
	Week 1 (12 to 16 June 2023)		Week 2 (19 to 23 June 2023)			SUM	
Evening							MER

SICSS Core Curriculum

Topic	Speaker
Introduction and Ethics	Partner Site Leaders
Digital Trace Data	Partner Site Leaders
Computational Text Analysis	Partner Site Leaders
Surveys in the Digital Age	Partner Site Leaders
Mass Collaboration and Experiments	Partner Site Leaders

SICSS-IAS Curriculum

Topic	Speaker
Introduction and Ethics	Jacob Habinek
Digital Trace Data	Jacob Habinek
Computational Text Analysis	Hendrik Erz
Social Network Analysis	Carl Nordlund
Spatial Data Analysis	Maël Lecoursonnais

Our Speakers

Speaker	Speaker
Carly Knight (New York University)	How Should an Employee Be?
Hui Sun (Stockholm School of Economics)	Cognitive Search and Social Network
Andrea Voyer (Stockholm University)	TBA
Adel Daoud (Linköping University)	Statistical Modelling and the Three Cultures
Martin Arvidsson (Linköping University)	TBA

Group Project (and Course Credit!)

Presentation on Thursday, June 22

For credit (3 ECTS): 4 page description or project proposal

Expectations

Openness

Patience

Community

Generosity

Anonymous feedback forms

- For SICSS headquarters: <https://forms.gle/uwB17RpvFU5RkJgHA>
- For the local partners: TBA.

Ethics

Jacob Habinek



LINKÖPING
UNIVERSITY

Institute for Analytical Sociology

Computational social scientists should
care about ethics

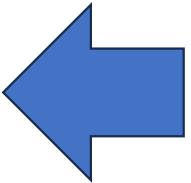
Computational social scientists should care about ethics

- Ad-hoc approaches
- Rules-based approaches
- Principles-based approaches

Facebook emotion study breached ethical guidelines, researchers say

Lack of 'informed consent' means that Facebook experiment on nearly 700,000 news feeds broke rules on tests on human subjects, say scientists

Poll: Facebook's secret mood experiment: have you lost trust in the social network?



PNAS

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer^{a,1}, Jamie E. Guillory^{b,2}, and Jeffrey T. Hancock^{b,c}

^aCore Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of ^bCommunication and ^cInformation Science, Cornell University, Ithaca, NY 14853

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

Emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. Emotional contagion is well established in laboratory experiments, with people transferring positive and negative emotions to others. Data from a large real-world social network, collected over a 20-y period suggests that longer-lasting moods (e.g., depression, happiness) can be transferred through networks [Fowler JH, Christakis NA (2008) *BMJ* 337:a2338], although the results are controversial. In an experiment with people who use Facebook, we test whether emotional contagion occurs outside of in-person interaction between individuals by reducing the amount of emotional content in the News Feed. When positive expressions were reduced, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred. These results indicate that emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks. This work also suggests that, in contrast to prevailing assumptions, in-person interaction and non-verbal cues are not strictly necessary for emotional contagion, and that the observation of others' positive experiences constitutes a positive experience for people.

computer-mediated communication | social media | big data

demonstrated that (i) emotional contagion occurs via text-based computer-mediated communication (7); (ii) contagion of psychological and physiological qualities has been suggested based on correlational data for social networks generally (7, 8); and (iii) people's emotional expressions on Facebook predict friends' emotional expressions, even days later (7) (although some shared experiences may in fact last several days). To date, however, there is no experimental evidence that emotions or moods are contagious in the absence of direct interaction between experimenter and target.

On Facebook, people frequently express emotions, which are later seen by their friends via Facebook's "News Feed" product (8). Because people's friends frequently produce much more content than one person can view, the News Feed filters posts, stories, and activities undertaken by friends. News Feed is the primary manner by which people see content that friends share. Which content is shown or omitted in the News Feed is determined via a ranking algorithm that Facebook continually develops and tests in the interest of showing viewers the content they will find most relevant and engaging. One such test is reported in this study: A test of whether posts with emotional content are more engaging.

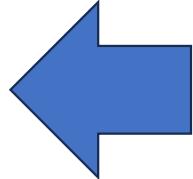
The experiment manipulated the extent to which people ($N = 689,003$) were exposed to emotional expressions in their News Feed. This tested whether exposure to emotions led people to

<https://doi.org/10.1073/pnas.1320040111>

<https://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say>



In 2006, Harvard also conducted a Facebook study that went too far



Social Networks 30 (2008) 330–342

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Tastes, ties, and time: A new social network dataset using Facebook.com

Kevin Lewis^{a,*}, Jason Kaufman^a, Marco Gonzalez^a, Andreas Wimmer^b, Nicholas Christakis^a

^a Department of Sociology, Harvard University, United States

^b Department of Sociology, University of California, Los Angeles, United States

ARTICLE INFO

Keywords:
Internet
Network data
Facebook
Culture
Race/ethnicity
Higher education
Tastes

ABSTRACT

Scholars have long recognized the potential of Internet-based communication technologies for improving network research—potential that, to date, remains largely underexploited. In the first half of this paper, we introduce a new public dataset based on manipulations and embellishments of a popular social network site, Facebook.com. We emphasize five distinctive features of this dataset and highlight its advantages and limitations vis-à-vis other kinds of network data. In the second half of this paper, we present descriptive findings from our first wave of data. Subgroups defined by gender, race/ethnicity, and socioeconomic status are characterized by distinct network behaviors, and students sharing social relationships as well as demographic traits tend to share a significant number of cultural preferences. These findings exemplify the scientific and pedagogical potential of this new network resource and provide a starting point for future analyses.

© 2008 Elsevier B.V. All rights reserved.

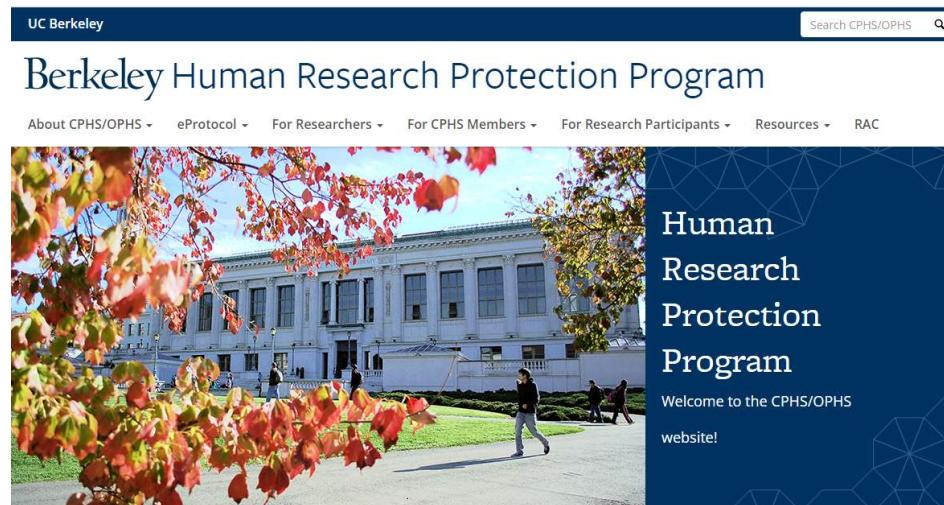
<https://doi.org/10.1016/j.socnet.2008.07.002>

<https://www.dailydot.com/debug/facebook-t3-study-tastes-ties-time/>

What's the problem?

- For ad-hoc approaches: increasing power
- For rules-based approaches:
 - Rules are inconsistent and slow to change
 - Little agreement about key concepts (e.g. privacy)
 - Blending of contexts.

Rules in the United States



The Committee for Protection of Human Subjects (CPHS) is comprised of two groups that serve as Institutional Review Boards (IRBs) for the University of California, Berkeley. The primary mission of the IRB is to ensure the protection of the rights and welfare of all human participants in research conducted by university faculty, staff, and students. The Office for Protection of Human Subjects (OPHS) is the administrative office that supports the CPHS.

<https://cphs.berkeley.edu/>

Rules in Sweden

[Skip to content](#)

 Etikprövnings myndigheten

Ethix Swedish Ethical Review Authority About the Authority How it works 

Protecting the individual in research



Ethical review
If you are going to do research on people, human tissue or sensitive personal data, you may need to apply for ethical review.

Welcome to Ethix!
If you are applying for ethical review, log in to the Ethix portal. This is where, as well as applying, you receive notifications, decisions and other messages related to your application.
The application must be made in Swedish. To enter Ethix, just click on the link below.

[Ethix](#)

The application process
[Initial application \(Grundansökan\)](#)
[Amendment \(Ändringsansökan\)](#)
[Fee for ethical review](#)
[What the act says](#)

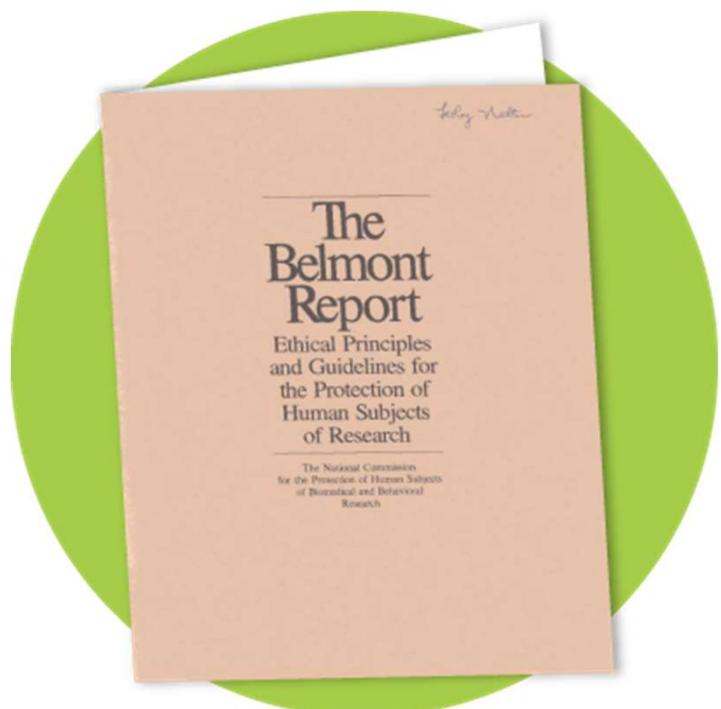
Support template informed consent form
Revised 2021-06-01.
[Stödmall forskningspersonsinformation](#)
[Stödmall samtyckesblankett](#)
[Stödmall samtyckesblankett framtida forskning på pröver](#)

<https://etikprovningsmyndigheten.se/en/>

Rules in Sweden

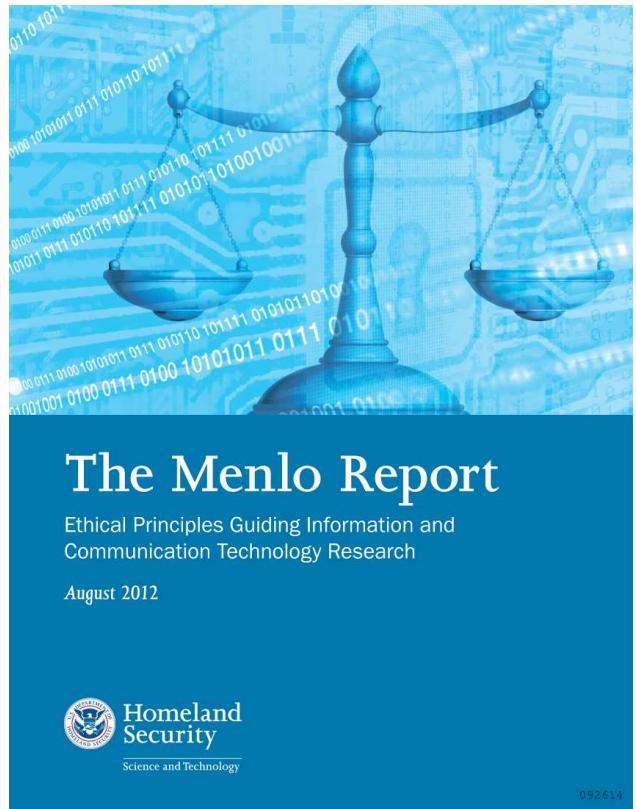
- The Ethical Review Act (2003: 460) covers all research conducted in Sweden involving humans, “sensitive personal data,” or criminal records.
 - Racial or ethnic origin
 - Political opinions and religious or philosophical beliefs
 - Trade union membership
 - Information about health, sex life and sexual orientation
 - Genetic and biometric data that can lead to the identification of a living person.

Meanwhile, in 1978



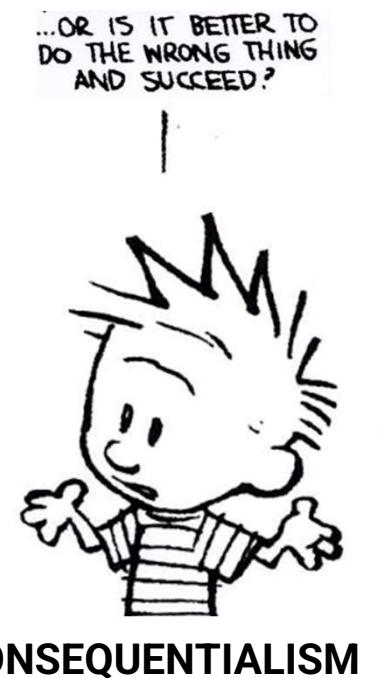
- Respect for persons: participants decide
- Beneficence: minimize risk, maximize benefits
- Justice: distribute the burdens and benefits of research

Fast forward to 2012



- Respect for persons: participants decide
- Beneficence: minimize risk, maximize benefits
- Justice: distribute the burdens and benefits of research
- Respect for law and public interest: compliance and transparency

How do you balance ethical principles?



Apologies to Bill Waterson ("Calvin & Hobbes," 12-9-1993).

Areas of difficulty

- Informed consent: some form of consent for most research
- Informational risk: all data are potentially identifiable and sensitive
- Privacy: consider flows of information
- Uncertainty: there is no risk free approach

Practical advice

- Rules are a floor, not a ceiling
- Put yourself in others' shoes
- Ethics are a continuum

After lunch

- An ethics case study
- More case studies available here:
 - <https://datasociety.net/library/data-ethics-case-studies/>
- Also consider the examples from lecture!



LINKÖPING
UNIVERSITY

Institute for Analytical Sociology