

Annual Review of Sociology

Computational Social Science and Sociology

Achim Edelmann,^{1,2} Tom Wolff,³ Danielle Montagne,³
and Christopher A. Bail³

¹Institute of Sociology, University of Bern, 3012 Bern, Switzerland;
email: achim.edelmann@soz.unibe.ch

²Department of Sociology, London School of Economics and Political Science,
London WC2A 2AE, United Kingdom

³Department of Sociology, Duke University, Durham, North Carolina 27708, USA;
email: christopher.bail@duke.edu

Annu. Rev. Sociol. 2020. 46:61–81

First published as a Review in Advance on
April 28, 2020

The *Annual Review of Sociology* is online at
soc.annualreviews.org

<https://doi.org/10.1146/annurev-soc-121919-054621>

Copyright © 2020 by Annual Reviews.
All rights reserved

Keywords

computational social science, machine learning, network analysis, text analysis, demography, social psychology, economic sociology, political sociology, cultural sociology, sociology of knowledge

Abstract

The integration of social science with computer science and engineering fields has produced a new area of study: computational social science. This field applies computational methods to novel sources of digital data such as social media, administrative records, and historical archives to develop theories of human behavior. We review the evolution of this field within sociology via bibliometric analysis and in-depth analysis of the following subfields where this new work is appearing most rapidly: (a) social network analysis and group formation; (b) collective behavior and political sociology; (c) the sociology of knowledge; (d) cultural sociology, social psychology, and emotions; (e) the production of culture; (f) economic sociology and organizations; and (g) demography and population studies. Our review reveals that sociologists are not only at the center of cutting-edge research that addresses longstanding questions about human behavior but also developing new lines of inquiry about digital spaces as well. We conclude by discussing challenging new obstacles in the field, calling for increased attention to sociological theory, and identifying new areas where computational social science might be further integrated into mainstream sociology.

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

INTRODUCTION

The rise of the Internet and the mass digitization of administrative records and historical archives have unleashed an unprecedented amount of digital data in recent years. Unlike conventional datasets collected by social scientists, these new digital sources often provide rich detail about the evolution of social relationships across large populations as they unfold (Bail 2014, Golder & Macy 2011, Lazer et al. 2009, Salganik 2018). Meanwhile, a variety of new techniques are now available to analyze these large, complex datasets. These include various forms of automated text analysis, online field experiments, mass collaboration, and many others inspired by machine learning (Evans & Aceves 2016, Molina & Garip 2019, Nelson 2017, Salganik 2018). The rapid increase in digital data—alongside new methods to analyze them—has given birth to a new interdisciplinary field called computational social science.

The term computational social science emerged in the final quarter of the twentieth century within social science disciplines as well as science, technology, engineering, and mathematics (STEM) disciplines. Within social science, the term originally described agent-based modeling—or the use of computer programs to simulate human behavior within artificial populations (Bruch & Atwell 2015, Macy & Willer 2002). This work led to fundamental theoretical advances in the study of social psychology, network analysis, and many other subjects (e.g., Baldassarri & Bearman 2007, Centola & Macy 2007, Watts 1999). Within STEM fields, by contrast, any study that employs large datasets that describe human behavior is often described as computational social science (e.g., Helbing et al. 2000, Pentland 2015). Although many of these studies applied elegant theories from physics and mathematics to analyze collective dynamics such as crowd behavior, they were largely disconnected from social science theory (see McFarland et al. 2015)—despite early efforts to synchronize them (e.g., Carley 1991, Macy & Willer 2002).

Acknowledging its diverse origins across multiple disciplines (Lazer et al. 2009), we provide the following definition of the field for this review: Computational social science is 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. Our definition forefronts sociological theory because we believe the future of the field within sociology depends not only on novel data sources and methods, but also on its capacity to produce new theories of human behavior or elaborate on existing explanations of the social world. Although we applaud recent calls for solution-oriented social science that aims to predict human behavior for practical purposes (Macy 2016, Watts 2017), our review examines only studies that also aim to explain human behavior to advance social science theory.¹ Readers should take note, however, that ours is not a consensus view among computational social scientists—particularly those outside of sociology. Such a consensus might not be possible given the remarkably rapid growth of the field across so many disciplines.

Another defining feature of our review is our focus on the evolution of computational social science within the discipline of sociology. Because of space limitations, we are unable to provide a comprehensive overview of computational social science across parallel social science disciplines such as political science or economics that may be of interest to sociologists. Although earlier reviews have examined the growth of new data sources or methods within the field of computational social science (Bail 2014, Evans & Aceves 2016, Golder & Macy 2014, Molina & Garip 2019), our goal is to map how such tools are currently being applied in empirical settings across sociology. Using a combination of bibliometric methods and in-depth analyses of individual studies, our review surveys work that is addressing longstanding questions about human behavior that predate

¹For more on the distinction between prediction-oriented and explanation-oriented computational social science, see McFarland et al. (2015).

the field, as well as new questions about human behavior that have emerged alongside the rapid integration of digital data into nearly every quarter of our lives.

Our principal conclusion is that computational social science is spreading rapidly into many different subfields within sociology. Our review, however, focuses on seven substantive areas where at least five studies have been published that meet our definition of the field. These include: (a) social networks and group formation; (b) collective behavior and political sociology; (c) the sociology of knowledge; (d) cultural sociology, social psychology, and emotions; (e) the production of culture; (f) economic sociology and organizations; and (g) demography and population studies. Work in these areas appears in flagship journals, inspires panels at major conferences, and helps communicate the public value of sociology as well. Nevertheless, our review also identifies numerous new challenges that range from ethics to the increasing opacity of data production in digital spaces. We conclude by identifying promising lines of inquiry for future studies, more effectively integrating sociological theory into the agenda of computational social science, and calling for increased engagement with other social science fields to ensure the continued spread of computational social science into mainstream sociology.

MAPPING THE FIELD

One of the many domains where data digitization has exploded in recent years is scholarly archives. This, combined with recent innovations in automated text analysis, inspired us to use the tools of computational social science to map the field itself. Our analysis is multi-faceted—combining data from a large bibliometric database, data from popular conferences in the field, and multiple stages of human coding—to identify all articles by sociologists that meet the definition of computational social science we proposed above.

We began by querying the Web of Science, the largest database of scholarly publications currently available.² Unfortunately, this resource omits most books and does not cover all journals—particularly some conference proceedings, where a considerable amount of computational social science research appears. Still, it provides perhaps the best possible point of departure currently available. Our sampling strategy began by searching for all mentions of the terms “computational social science” or “big data” (a neologism that was used in the early days of the field) in the title, abstract, or keywords of articles designated with a social science classification by the database.³ Because computational social science scholarship often appears in interdisciplinary journals that might not receive a social science designation, we also searched for our keywords in five additional publications that feature prominent research in the field: *Science*, *Nature*, *Proceedings of the National Academy of Sciences*, *Science Advances*, and *Nature Human Behavior*. To further improve the breadth of our sample, we collected the names of all scholars who presented publications or posters at the largest conference in the field: the International Conference in Social Science (2018). Next, we collected all articles written by these individuals within the Web of Science database and added them to our sample. Finally—to capture influential articles that emerged before our keywords became popular—we inspected the 300 most-cited articles within our database and identified 79 additional articles that met our definition of the field.

²We use the Science Citation Index Expanded; Social Sciences Citation Index and Arts & Humanities Citation Index; Conference Proceedings Citation Index–Science and Conference Proceedings Citation Index–Social Science & Humanities; and Emerging Sources Citation Index.

³This includes author-assigned keywords and so-called keywords that Web of Science derives from phrases that occur frequently in the titles of an article’s references. We searched for orthographic variants and plural forms, including big data, big-data, and computational social sciences.

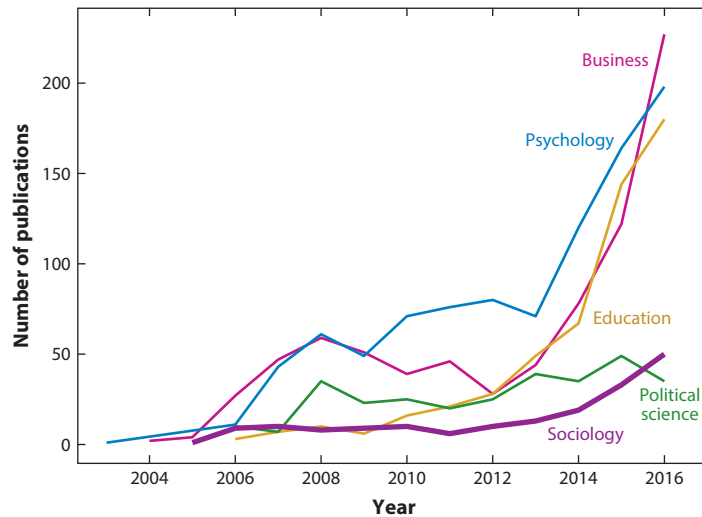


Figure 1

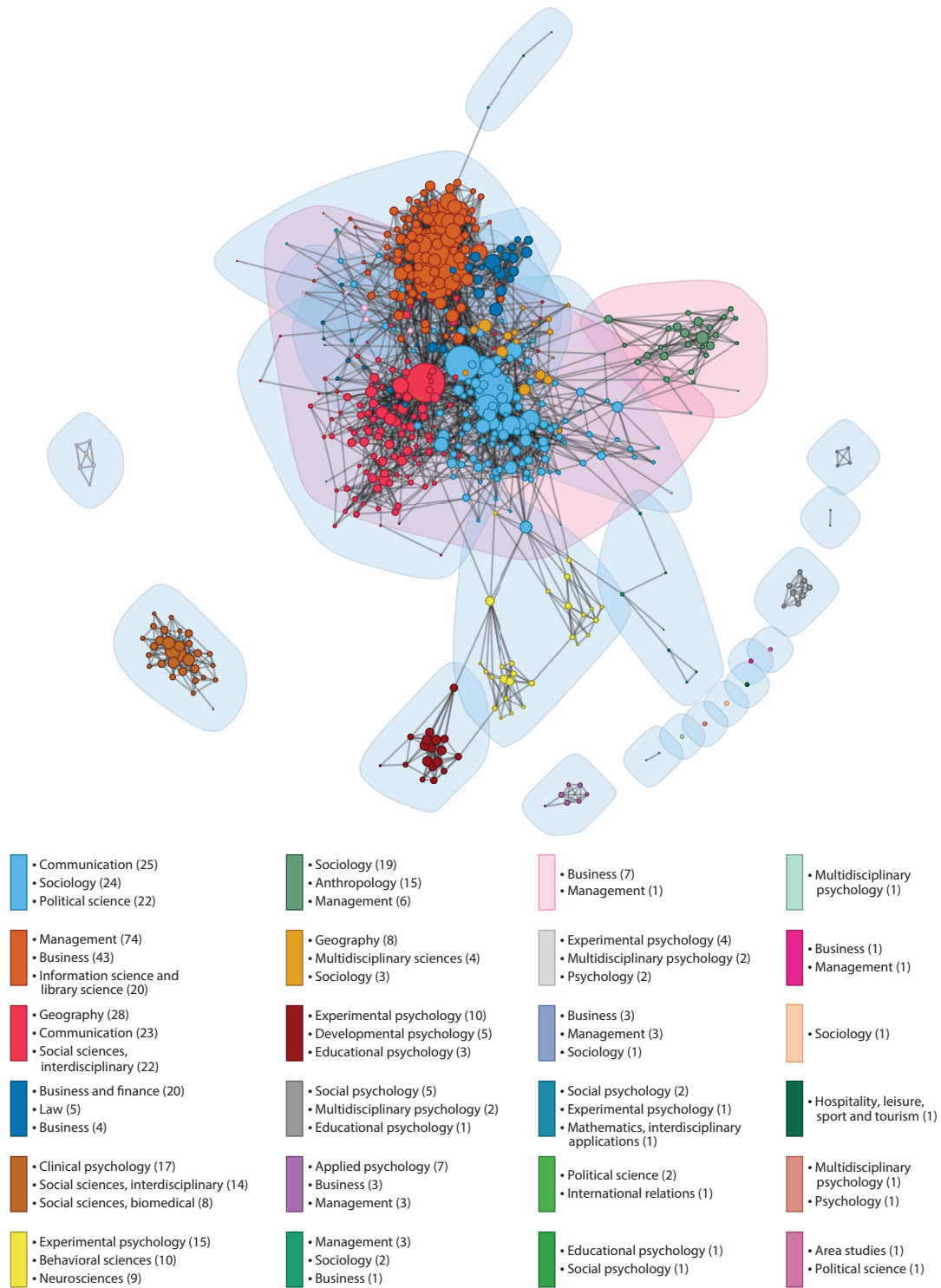
Number of computational social science publications by year—2003–2016—across five scholarly disciplines.

Before we present our results for the discipline of sociology, we provide an overview of the evolution of computational social science across a broader set of fields. **Figure 1** is a time series graph that describes the number of publications within five scholarly disciplines where scholarship mentioned the terms computational social science or big data between 2000 and 2016. This figure should be taken as a rough approximation of the field, given that individual articles were not reviewed by human coders to confirm that they are on the subject of computational social science. Still, several things are noteworthy about this figure. First, there has been a remarkable explosion of work in computational social science across many different fields since 2012. Second, the disciplines in which this research agenda is most active are clearly business, psychology, and education. Although computational social science is growing more slowly in sociology according to this broad view, there has been an exponential increase in work by sociologists since 2010 as well.

Next, we produced a citation network of all articles by discipline (see **Figure 2**). Each node describes a paper in our corpus, and edges between papers indicate that they cite each other.⁴ We identified 24 areas of scholarship using the Louvain community detection algorithm. Four large communities can be identified at the core of this network. The first large community connects communications, sociology, and political science. The second large community is primarily comprised of work in geography and communications. The third sizeable community ties together business and library science. A fourth, notable community connects business, finance, and law. Although sociology is prominently featured in the center of this visualization, it also appears within another community connected to anthropology and business management. We highlight with pink shading in **Figure 2** communities in which sociology is a prominent member.

Although **Figures 1** and **2** provide a useful panorama of computational social science, all keyword-based sampling procedures suffer from both false positives and false negatives. Because our focus in the remainder of this review is on sociology, we took additional steps to ensure the accuracy of our sample within this discipline. First, we obtained a crowd-sourced list of scholars

⁴We construct these edges using a sequence of matches based on regular expressions that involve increasingly more lenient keys to counter the messiness of the Web of Science reference format.



(Caption appears on following page)

Figure 2 (*Figure appears on preceding page*)

Computational social science citation network. Nodes are colored according to community membership. Numbers in labels indicate number of papers from each discipline in the community. Communities highlighted in pink are those in which sociology appears prominently.

who work in the field of computational social science produced by participants of the Summer Institutes in Computational Social Science (SICSS)—a major training event in the field funded by the Russell Sage Foundation and the Alfred P. Sloan Foundation. We then identified all sociologists on this list, collected their curricula vitae, and added to our database any articles that met our definition of the field. Second, we hand-coded a refined list of articles in our database classified as sociology to remove false positives, which yielded a total of 248 articles. The remainder of our article discusses the seven subfields of sociology where we were able to identify at least five articles within this database that employ both digital data and computational methods to develop theories of human behavior.

SOCIAL NETWORKS AND GROUP FORMATION

One of the first areas where computational social science emerged in sociology is the study of social networks and group formation. This is not surprising given that conventional research techniques such as surveys struggle to capture the evolution of social relationships in situ. Although many early studies employed agent-based models for these reasons, data from Internet sites, social media platforms, and telecommunications inspired some of the first population-level research in the field. Watts and colleagues, for example, employed email data to demonstrate and elaborate on core social science theories in digital spaces such as the principle of six degrees of separation (Dodds et al. 2003, Watts 2004), network dynamics and equilibrium (Kossinets & Watts 2006), and opinion leadership (Watts & Dodds 2007). Similarly, Macy and colleagues used telecommunications data to demonstrate a strong relationship between network diversity and economic development and various other theories of group formation (e.g., Eagle et al. 2010).

More recent studies employ digital data sources to examine the diffusion of complex contagions in social networks. Centola (2010), for example, created an online community where he could control the topology of social networks. This work shows that diffusions were much more likely to reach individuals in tightly clustered networks than those that were organized at random. Other research showed that network topology can also shape the adoption of social conventions (Centola & Baronchelli 2015). Social media and telecommunications data also enable analyses of networks with unprecedented scale. Bail et al. (2017) employ Facebook data to identify synergy between the diffusion of emotional and rational communicative styles in a large network of people discussing public health issues. In another study, Bail et al. (2019) use Google Search data to track the diffusion of cultural products across global networks. Together, these studies show that microlevel interactions between individuals can generate macrolevel diffusion patterns anticipated by early social science theorists such as Gabriel de Tarde. Park et al. (2018) study extremely long-range ties using data produced on Twitter and via international phone calls. Their findings challenge the long-held belief that social relationships that connect clusters of individuals within networks are usually weak (i.e., based on acquaintances rather than close friends). In contrast, they show that very long-range network connections—although rare—are often as strong as those that connect a close circle of friends.

Another line of research examines network dynamics via online games. Shirado & Christakis (2017) recruited participants to play a game that required them to coordinate colors within a matrix. By introducing bots that simulated human agents playing the game poorly, they were able to

improve coordination among human respondents in the study. In another experiment where respondents occupied hypothetical neighborhoods and were asked to share Wi-Fi with each other, Shirado et al. (2019) examined how network brokerage shapes inequality. They find that well-connected individuals can suffer when too many others come to depend on them for services. Online games have also been influential in studying the wisdom of crowds and political beliefs. Guilbeault et al. (2018), for example, show that diverse political groups produce more accurate estimates of political facts if they are anonymous to each other, but less accurate estimates if their political identities are revealed to each other. Becker et al. (2019) show that the wisdom of crowds also improves estimation in politically homogeneous settings. Finally, games have been used to study collective dynamics more broadly. Centola et al. (2018) created a game where groups of respondents were asked to determine the appropriate names for avatars. They showed that network properties enable small factions of study participants to overturn pre-existing consensus among prevailing majorities.

COLLECTIVE BEHAVIOR AND POLITICAL SOCIOLOGY

Computational social science has inspired considerable work on collective behavior and politics. Data from social media and other communication platforms have advanced the study of collective behavior in particular. The Egyptian Revolution, the Indignados movement, and Occupy Wall Street, for example, show that digital tools now play a central role in protests and that data generated by these tools can inform studies of collective political efforts (Tufekci & Wilson 2012). Twitter has become a focus of research in this area because it provides large datasets that can be used to study how information spreads across networks. Case studies by González-Bailón and colleagues indicate peripheral users in movement networks can generate large cascades of information, but that leadership and hierarchy create larger information flows (Barberá et al. 2015, González-Bailón et al. 2013, González-Bailón & Wang 2016). However, some scholars have asked whether such findings can be generalized to other settings. Analyzing data from a Facebook application, Lewis et al. (2014) find that online supporters are less committed to movements than activists on the ground. Still, others find strong correspondence between online and offline activity among social movements (Abul-Fottouh & Fetner 2018, Hanna 2013). Even if online data do not capture all offline processes, they often provide a useful supplement to conventional data sources—and for an increasing number of movements may be the only source of information available (Zhang & Pan 2019).

Computational research has helped develop theories related to mobilization and behavior change in social movements as well. Vasi et al. (2015) use data from social media and Google to show how both online discussions about hydraulic fracturing and mobilizations against the practice increased in communities where screenings of a documentary took place. This illustrates how cultural artifacts might be used in social movements to motivate activists and shape public opinion. Online experiments have also proven valuable for testing existing theories about mobilization. Centola and colleagues' network experiments, described above, highlight the importance of critical masses for spreading new social conventions (Centola et al. 2018). van de Rijt's and others' studies leverage Change.org, an online petitioning platform, to conduct field experiments in which samples of petitions are assigned different quantities of signatures as treatments (Vaillant et al. 2015; van de Rijt et al. 2014, 2016). These experiments demonstrate the importance of cumulative advantage effects in mobilization (van de Rijt et al. 2016) and the potential for initially unpopular campaigns to encounter unexpected revivals (Vaillant et al. 2015).

Elsewhere, the proliferation of text-based data has inspired new work on political discourse. Recent work by Bail (2015) and Bonikowski & Gidron (2016) demonstrates how fringe discourse

can enter the mainstream by analyzing the social position of actors within discursive fields and the emotional valence of their language. Scholars also use text data to examine how certain discursive styles prove more effective than others for advocacy groups (Bail 2015, 2016a; Bail et al. 2017), as we discuss in additional detail below. Other studies examine the interaction between political discourse and policy change. Flores (2017), for example, uses text data from Twitter to show how anti-immigration laws hardened public opinion against immigration in Arizona. Text-based research has also examined how elites shape political discourse—both across elite fields (e.g., journalism, politics, and celebrities) and across countries (AlMaghlouth et al. 2015, Wells et al. 2016).

Finally, many studies examine political polarization and persuasion using online data, automated text analysis, experiments, and agent-based modeling. These studies investigate how political tribes form and highlight the role of “echo chambers” that create selective exposure to information. Network-based research suggests that polarization arises from homophily among conservatives and extremists (Boutyline & Willer 2017) as well as more general homophily and peer influence processes (DellaPosta et al. 2015). Research using other methods notes that patterns in donations to candidates have grown more polarized in recent decades (Heerwig 2017) and that donations from elites drive belief polarization (Farrell 2016a). Computational sociologists have also performed experiments to learn how partisans can change their beliefs. As noted above, Becker et al. (2019) find that politically homogeneous groups make better decisions in online games. At the same time, other research reveals that exposure to opposing views can create backfire effects. Bail et al. (2018b) paid Twitter users to follow bots that exposed them to opposing political views and found this treatment increased partisanship. Still, other studies indicate that minimizing signals of partisan identity (Guilbeault et al. 2018), using specific moral language (Feinberg & Willer 2015), and matching the linguistic styles (Romero et al. 2015) may help reduce polarization.

SOCIOLOGY OF KNOWLEDGE

Computational social science has become a central part of the sociology of knowledge. One strand of research employs citation data to study consensus formation within science. Building on Shwed & Bearman's (2010) influential study, adams & Light (2015) analyze citation networks to determine how consensus about the outcomes of children with same-sex parents arises among scholars. Focusing on temporal patterns, they identify the point when the scientific consensus emerged that children of same-sex parents show no marked differences in their sexual orientation compared to those from other parental configurations. Bruggeman et al. (2012) demonstrate the importance of differentiating between citations that signal agreement and citations that signal disagreement. Using simulations, they find that small proportions of citations that are contentious have significant effects on whether citation networks signal consensus.

Now that bibliographic data have become easier to collect and analyze at scale, scholars can map and model processes that shape scientific fields as a whole. The Metaknowledge program of University of Chicago sociologist James Evans exemplifies such research (e.g., Evans & Foster 2011). This includes studies that trace the increasing dominance of teamwork in science. For example, Wuchty et al. (2007) show that, in the social sciences, the propensity to work in teams has more than doubled over the past five decades. Others show that small teams tend to produce work that introduces novel and disruptive ideas in science and technology, whereas large teams tend to develop existing ideas further (Wu et al. 2019). Advances in text analysis have enabled scholars to shed new light on similarities and differences between disciplines as well (e.g., Evans et al. 2016, McMahan & Evans 2018, Vilhena et al. 2014). For example, McMahan & Evans (2018) develop a measure to capture the ambiguity of language within scientific articles. They find that expressions are used most consistently in the biological and chemical sciences and least consistently in

the humanities, law, and environmental sciences. Ambiguous language also produces more integrated citation streams, which stimulates more involved academic debate. Vilhena et al. (2014) draw on the concept of cultural holes to map differences in disciplinary jargon within and across fields. Although these language-based gaps do not map neatly onto structural holes within citation networks, they nevertheless inhibit efficient communication between scientists. Shi et al. (2015) go a step further and model the generative process underlying an entire field. They show that biomedical research can be conceptualized as a dynamic network that evolves according to the ways scientists link theory and methods across time.

Another strand of literature examines how academic work gains impact and prestige. Uzzi et al. (2013), for example, show that high-impact publications build on prior work yet simultaneously feature unusual and novel combinations. Within sociology, Leahey & Moody (2014) show that articles that span subfields garner more citations and are more likely to appear in prestigious journals. Still others have created large databases on academic prizes (e.g., Li et al. 2019). Using such data, Ma & Uzzi (2018) study the worldwide network of academic prize-winners over 100 years. They show that a relatively small number of prizes and scholars push the boundaries of science, whereas the increasing number of prizes remains concentrated in a small circle of scientific elites.

Computational approaches also reveal how scientific discoveries are driven by the choices of individual scientists and their organizational contexts. Rzhetsky et al. (2015), for example, identify the strategies that scientists in biomedicine follow in choosing which relationships between molecules to study. They show that these strategies reflect personal career choices and demonstrate that increased risk-taking and publishing of research failures could improve scientific discovery in this field as a whole (see also Foster et al. 2015). Others focus on organizational context. For example, Rawlings et al. (2015) and Rawlings & McFarland (2011) identify how the organizational structure of one prominent university shapes peer influence in academic funding proposals, noting the increasing dominance of a few faculty members in the flow of knowledge as measured in terms of their citation histories.

Another strand of work in this area employs computational techniques to study the public face of science and how science interfaces with industry. For example, Shwed (2015) shows how attempts by the tobacco industry to hamper scientific inquiry paradoxically enabled scientists to discover the perils of smoking. Farrell (2016a,b) examines how private-sector influence shaped the debate about anthropogenic climate change. His analysis reveals how corporate funding influences both the content and prevalence of polarizing themes in this debate over time. Still others combine network analysis with text analysis to better understand how scientists position themselves in public debates. For example, Edelmann et al. (2017) study the controversy about the use of potentially pandemic pathogens in biomedical research. They find that peer effects and research specializations shape the positions scientists publicly support in this debate. Scholars have also explored the public consumption of science by turning toward online purchasing data. Studying millions of online book purchases, for example, Shi et al. (2017) reveal partisan interests in science. Consumers of liberal-leaning political books prefer basic science, whereas customers of conservative books tend to prefer applied, commercial science. This implies that science might both bridge and reinforce political differences within the general public.

Finally, computational work reveals gender asymmetries and inequalities within science and the representation of knowledge online more generally. For example, drawing on data from JSTOR (one of the largest digital repositories of academic journals), West et al. (2013) trace gender inequalities in scholarly authorship across the natural sciences, social sciences, and humanities since 1545. Even when women and men appear to have similar publication counts, this study shows that men continue to dominate single-authored papers or prestigious first or last author positions. In another study, King et al. (2017) show that men are more likely to cite themselves than women—a

tendency that increased over the past two decades. Finally, Wagner et al. (2016) focus on the representation of women on Wikipedia. They find that ceiling effects curb the entry of women and identify linguistic differences in how achievements of women and men are described—as well as differences in meta-data that imply asymmetries in the wider reception of articles written by men and women online.

CULTURAL SOCIOLOGY, SOCIAL PSYCHOLOGY, AND EMOTIONS

Computational social science also appears in the study of cultural sociology, social psychology, and emotions. Within cultural sociology, several studies examine broad processes of cultural change. Bail (2015) shows how automated text analysis can be used to examine the transformation of discursive fields during what Ann Swidler calls “unsettled times” in order to develop a systematic theory of resonance—or why certain cultural messages have a natural advantage over others. In follow-up work, Bail (2016b) identifies a “carrying capacity” within discursive fields: Messaging strategies attract more attention if they cover a diverse range of subjects, but messages that combine too many themes are less impactful. Bail (2016a) also develops a theory of “cultural bridging” which indicates organizations that adopt brokerage positions within discursive fields are more likely to attract large audiences. Other studies document similar cultural processes in scientific fields, markets, and corporations (Goldberg et al. 2016a,b, Vilhena et al. 2014), as we discuss elsewhere in this article. Finally, Kozlowski et al. (2019) use word embeddings to examine large-scale shifts in the meaning of terms associated with class and gender in the United States and United Kingdom over time.

The growth of text-based data has also opened new lines of inquiry about how cultural messages are expressed. Much of this research involves data collected from Facebook or Twitter. For example, Golder & Macy (2011) track the frequency of language associated with different types of emotions and discovered evidence of diurnal rhythms. Other work reveals emotional contagion in public discussions about public health issues on Facebook, or the likelihood that exposure to emotional language makes social media users more likely to become emotional themselves—and thus more likely to interact with emotional content as well (Bail 2016c). Bail et al. (2017) identify what they call “cognitive-emotional currents” in such discussions, or alternating surges of rational and emotional styles of language. This work thus feeds into broader discussions about the role of “hot” and “cold” processing within the human brain and explains how social context can shape how expressive styles spread across social networks over time.

In addition to text analysis, scholars are pioneering the use of virtual reality to further examine core questions in social psychology. For example, van Loon et al. (2018) paid university students to perform collaborative tasks, but randomized half of them into a treatment condition where they used virtual reality to take the perspective of the other research participants. This intervention increased prosocial behavior. Another line of research contributes to theories of phenomenology and small-group processes using virtual reality (Schroeder 2010). Although virtual reality-based research is exploding in adjacent social science fields such as psychology, sociology has been slow to integrate this new technology. This is surprising not only because virtual reality holds great promise for studying how people respond to social settings within tightly controlled experiments, but also because sociologists conducted some of the earliest research on virtual communities (e.g., Gamson & Peppers 1966).

PRODUCTION OF CULTURE

Computational approaches are often used to study how people evaluate cultural products such as music, art, or films. In an influential study, Salganik et al. (2006) created an online “music lab”

to study how peer influence shapes musical preferences of new artists. In their experiment, participants rated music from obscure bands. In the control condition, participants received no information about its popularity, whereas those in the treatment condition were shown how many times songs were downloaded. Treated respondents were more likely to listen to songs with high download rates and to rate such songs more positively. Although most songs in the treatment condition followed a self-fulfilling prophecy—where false popularity became real over time—a follow-up study revealed the highest-quality songs recovered popularity over time (Salganik & Watts 2008, van de Rijt 2019). Other studies explore the dynamics of critical acclaim. For example, previous work finds the age of music or film, its genre, and sponsorship by a major production company increase the likelihood of industry accolades for both music and film (Light & Odden 2017, Rossman & Schilke 2014).

Computational work also measures how combinations of themes in cultural products shape their reception. Using data from Spotify and Billboard, Askin & Mauskopf (2017) find songs perform best when they sound similar to songs that were on the chart the year before but include minor diversions from such themes as well. Other research suggests tastes for atypicality vary across demographics, such as class, race, and geographical location. Some people prefer products—such as food and film—that fit cleanly into a single category, whereas others prefer novel combinations of themes from multiple categories. Those that prefer items from single categories at a time represent the traditional omnivore hypothesis. They have a taste for authenticity in multiple categories—no categorical mixing—which signals high status (Goldberg 2016a). Yet studies also indicate tastes are stratified by regional differences, as interest in certain ingredients and recipes is influenced by location (Wagner et al. 2014). By considering the characteristics of products and consumers in tandem, researchers are developing a more comprehensive theory of how taste and consumption patterns align.

Another line of studies examines how social networks between individuals shape the creation of cultural objects—often with a focus on novelty and innovation. For instance, de Vaan et al. (2015) use a large database of video-game production credits to demonstrate that cross-cutting teams of developers create products that are both more creative and popular. Analyzing networks of Hollywood actors, Rossman et al. (2010) show how status is transferred between people who work together. In this case, an actor is more likely to receive an Oscar nomination—an important consecration event in the film industry—when they appear in a film alongside a high-status actor.

Finally, recent studies examine how gender, class, and political identities shape the production of cultural products. Shor et al. (2015), for example, find men are featured more often in the media than women because of journalists' focus on high-status topics that typically focus on men. A large-scale historical analysis of Google Books data shows that mentions of social class, class struggle, and other terms that describe social class increase alongside the economic misery index—a measure of inflation and unemployment (Chen & Yan 2016). Finally, Hoffman (2019) uses a combination of text analysis and network analysis to study how reading patterns shape political ideology and vice versa using records from a large public library.

ECONOMIC SOCIOLOGY AND ORGANIZATIONS

The intersection of computational and network methods is also flourishing within economic sociology. Email and instant messaging archives create large, dynamic networks of individuals in organizations—a significant improvement on data generated via self-reports. Early exchanges among individuals and organizations—for instance, patient transfers between hospitals—shape the evolution of social networks and the allocation of resources (Horvát et al. 2015, Kitts et al. 2017). Elsewhere, instant messaging data reveal that social balance theory predicts streaks of high

performance among day traders (Askarisichani et al. 2019). Messaging data also allow researchers to measure the level of coordination among employees in economic decision making and show that synchronous communication increases the profitability of trades (Saavedra et al. 2011b). Finally, network data have improved the study of social capital and career outcomes. For example, studies indicate certain network compositions benefit women differently than men (Lutter 2015). For example, women are more likely to succeed if they are part of an inner circle within a network that is predominantly female (Yang et al. 2019). Network data are useful for studying not only exchange and career outcomes, but also organizational culture. Goldberg et al. (2016b) analyze email messages within a large corporation to measure whether employees' communications fit within prevailing norms and behavior. Employees who integrate themselves into organizational culture receive a variety of rewards and are less likely to be fired (Srivastava et al. 2018).

In addition to network data, digitized records of textual communications provide a richer understanding of the role of culture and emotions in the marketplace. Using text analysis, studies reveal how market volatility influences whether traders are discussing current or future conditions, and how these shifts shape trading behavior (Saavedra et al. 2011a). Other studies indicate individuals who exhibit moderate levels of emotions in their communication with other employees make the most profitable stock trades (Liu et al. 2016). Finally, Schnable (2016) combines qualitative methods with automated text analysis to study how religion provides grassroots NGOs with cultural frames that allow them to claim legitimacy, build social networks, and secure funding and resources.

DEMOGRAPHY AND POPULATION STUDIES

Computational social science has emerged within demography only relatively recently, but it is rapidly gaining popularity (Cesare et al. 2018). Unsurprisingly, computational approaches are most often used to produce high-quality population estimates. Mobile phone data, for example, are employed to produce more dynamic population estimates—particularly in areas where national statistics are not reliable (Cesare et al. 2018, Eagle et al. 2010, Palmer et al. 2013). Others use Google Street View and deep learning to estimate demographic characteristics of neighborhoods (Gebru et al. 2017), crowd-sourcing techniques to measure demographics of social media populations (McCormick et al. 2017), websites to map large-scale genealogy trees (Kaplanis et al. 2018), and online image data to predict age (Helleringer et al. 2019). Computational approaches are also being developed to study the “holy trinity” of demography: fertility, mortality, and migration. To study fertility and mortality, studies employ Google Search data and Facebook data (Hobbs et al. 2016, Kashyap & Villavicencio 2016, Ojala et al. 2017). Other studies use data from Twitter and LinkedIn to develop more accurate estimates of migration within and between countries (Palmer et al. 2013, State et al. 2014, Zagheni et al. 2017).

Demographers also use computational approaches to probe the microdynamics of population processes such as dating and marriage. Several recent studies use data from a large Internet dating platform to examine how relationships form across racial and ethnic divides. Lewis (2013) finds strong evidence of racial homogamy within Internet dating, but also shows that if people are invited on a date by someone they do not know, they are more likely to go on the date if the person is from another racial or ethnic group. Lin & Lundquist (2013), however, find that women are more likely to respond to dating invitations from members of dominant racial and ethnic groups, regardless of social distance. Outside of the US context, Potâră & Mills (2015) identify very strong racial and ethnic homogamy across European countries—and particularly those with ethnically homogeneous populations. More recent work by Bruch and colleagues (Bruch et al. 2016, Bruch & Newman 2018) examines how population-level factors—such as the number of

single men or women in the city—shape individual decision-making processes above and beyond physical attraction.

A final strand of computational social science research by demographers uses digital data sources to study socially undesirable health behaviors and enumerate populations that are difficult to access. For example, Kashyap & Villavicencio (2016) employ Google Search data to study the prevalence of selective abortion in India. Bail et al. (2018a) use Google Search data to examine how demographic factors contribute to violent radicalization. Moreno et al. (2012) use Facebook data linked to surveys to measure the prevalence of binge drinking in US colleges. Chakrabarti & Frye (2017) use text analysis of diary data to study AIDS prevention. Araujo et al. (2017) track the prevalence of lifestyle diseases in 47 countries using Facebook's advertising tools. Other studies leverage these same data to revisit core questions in demography about migration, fertility, and gender segregation (Fatehkia et al. 2018, Rampazzo et al. 2018, Stewart et al. 2019). Stewart et al. (2019), for example, study the cultural assimilation of undocumented immigrants in the southern United States by tracking the size of audiences interested in non-US soccer teams.

Other scholars have developed entirely new models for population research inspired by conventions within the field of data science. Within these fields, it is common to create a competition wherein multiple teams compete to build models for a cash prize. A popular example is the so-called Netflix Prize, for which teams of data scientists competed to build a model that provides the most high-quality recommendations of what Netflix users might like to watch. Princeton sociologist Matthew Salganik coordinated the Fragile Families Challenge using a similar model that staggered the release of a new wave of the Fragile Families and Child Wellbeing Study—a multi-wave study of inequality and the family. He invited teams of researchers to develop models based on a subsample of the new wave of the dataset—commonly referred to as a training dataset—to predict outcomes of interest on the full dataset that was released later (Salganik et al. 2020, Lundberg et al. 2018). Although this effort did not produce major advances over extant models, it provided a critical litmus test for machine learning within population science—and social science more broadly.

CONCLUSION

As this review illustrates, the field of computational social science is expanding in many exciting new directions—indeed, it is expanding so rapidly that any review of the literature will become outmoded within short order. Still, this preliminary review indicates that computational social science has become a central part of research across many different subfields. Perhaps more importantly, the field has moved far beyond the descriptive social media research that characterized much early work in the field. Indeed, sociologists have developed a range of hybrid methodologies that combine computational approaches with more conventional techniques to take advantage of the promise of digital data sources while addressing their limitations as well. Yet sociologists are also rapidly pursuing the cutting edge of new technologies as well. Although machine learning has yet to have its watershed moment within sociology, artificial intelligence, bots, and virtual reality already offer sociologists a powerful new suite of techniques to study how social relationships form and evolve across contexts.

Using Computational Social Science to Build Theories of Human Behavior

Importantly, our review also indicates that sociologists are contributing not only new methods for computational social science, but also major theoretical advances as well—which we view as key to the continued expansion of the field into mainstream sociology. There are multiple ways in which theoretical advances can be made via computational social science. First, many studies are using new types of data and methods to revisit old sociological questions that were once thought

impossible to study. Examples include new macrolevel theories of social networks and cultural change as well as microlevel theories of human decision making. Yet such work only scratches the surface of the full potential of computational social science to advance theory development across multiple levels of analysis. In our view, the most influential work within computational social science in the coming years will be the type that is able to link macro levels of theories about topics such as cultural change to microlevel processes of decision making. Such efforts will most likely not be possible with social media data, administrative records, or other new sources of digital data alone. Rather, they will require creative hybrid methods that illuminate the space between different levels of analysis.

A second way that computational social science can advance the theoretical progress of sociology is to develop new theories of the social *terra incognita* created by the rise of digital technology itself. These include novel theories of digital protest, consumption of cultural products, and the spread of knowledge online—to name but a few. Many of the studies above have examined these new spaces—particularly within the realm of politics, networks, and communication. But the number of new social spaces—and new forms of social relationships they enable—are currently outpacing the evolution of sociological theory. Perhaps the most obvious example is the manner in which machine learning and artificial intelligence are currently recasting the way we communicate with each other—which may ultimately create new forms of social segregation that shape the evolution of our networks themselves. Sociologists have also left mostly untheorized the ways that interaction between machines themselves can shape patterns of human behavior and social organization as well. If sociologists do not rise to these new challenges, other fields will. Indeed, the majority of theorizing in the broader field of computational social science outside of sociology either is inattentive to sociological theory or focuses on a handful of influential ideas. We believe that sociologists need to insert themselves into such conversations more proactively by demonstrating how core tenets of sociological theory such as the social construction of reality or self-fulfilling prophecies can advance the study of the many new social spaces created in recent decades.

Finally, sociologists can use the tools of computational social science to develop new ways of creating theory itself. Although it is becoming increasingly clear that machine learning will not easily allow us to reverse-engineer collective behavior, culture, or human decision making, we can systematically use such tools to identify new dimensions of social behavior or test the robustness of our existing accounts. This will probably proceed most naturally by comparing the gap between the predictions produced by such models and the underlying reality—either on a variable-by-variable basis or by systematically testing how permutations of different variables might enable us to identify unforeseen complexity of human behavior or otherwise identify our blind spots. Finally, scholars are also creating theories by interfacing with other fields. Comparing theories of human behavior, computing, and artificial intelligence, for example, opens exciting new lines of inquiry about the nature of social life (e.g., Foster 2018). Taken to the extreme, we can also use this line of theoretical development to compare precisely how differences between artificial and human intelligence help us develop deeper understanding of social processes—particularly those that are not easily articulated or may be difficult for a machine to observe.

Outstanding Challenges and Directions for Future Research

Despite the tremendous potential of computational social science to advance sociological theory, it also creates numerous pressing challenges. Many of the most valuable data sources in the field—such as data from social media companies—are either inaccessible to researchers or only accessible by those with high status. Although a variety of initiatives are underway to address this problem (King & Persily 2019), the ongoing controversy about privacy and data within the public

domain will make such collaborations difficult for years to come (and perhaps appropriately so). This is because the field of computational social science also creates a range of complex ethical issues (Salganik 2018). Among others, these include balancing the principles of open science and protection of human subjects, linking information across datasets without violating confidentiality, and asking a broader set of questions about how deeply researchers should probe into the online lives of their subjects—particularly in cases where people may not know that the data they generate online could be analyzed by others.

Another set of challenges concerns access to training in computational social science. Coding in open-source software, embedding field experiments in online platforms, and dealing with unconventional data structures are not part of regular training within most sociology departments. Although numerous online resources offer training in such techniques, they nevertheless require extensive domain knowledge and are not geared toward social scientists who aim to repurpose digital data for the study of human behavior. The SICSS are designed to fill this gap, by providing training to graduate students, postdoctoral fellows, and junior faculty in a range of techniques including web scraping, automated text analysis, digital experimentation, ethics, mass collaboration, and a range of other subjects. The SICSS also seed interdisciplinary research by connecting young scholars in a range of social science and STEM fields in more than 11 locations around the world and encouraging them to start research projects together. Although these events provide a temporary solution to the lack of training in computational social science, we hope that more departments will consider including such material within their core curricula.

Perhaps because of the issues of data access, ethics, and training just described, computational social science has not yet been broadly integrated into many of the largest subfields in sociology. These include race and ethnicity, criminology, education, inequality and stratification, religion, sex and gender, law, medical sociology, historical sociology, religion, and many other subfields. It is possible that computational social science has not yet permeated these subfields because the initial wave of data created by new communication technologies was more conducive to pressing questions in the fields we reviewed above. Yet there is tremendous potential to link such new data sources—alongside the widespread digitization of historical and administrative archives—to address longstanding questions outside of these fields as well. To give only a few examples, the digitization of medical records might enable a host of new questions within medical sociology, and new data on policing and crime create fertile ground for new theorizing within criminology as well (see Brayne 2020). In such cases, computational data and methods might be fruitfully combined with the panoply of existing methods—from conventional surveys to ethnography.

Just as we hope future research will span the subfields of sociology, we also call for even greater engagement with other disciplines where computational social science continues to blossom. As the bibliometric data we presented at the outset of this article showed, research is exploding in other social science disciplines as well as computer science and engineering. Sociologists are fortunate to occupy a central position—if not the most central position—within this interdisciplinary conversation. We believe that sociologists must continue to capitalize on this position—not only because computational social science offers so much to sociology, but also because many of the most pressing problems studied by computational social scientists are inherently sociological—from the role of social networks in the spread of misinformation or the emergence of echo chambers to the ways that algorithms create or reproduce social inequality.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Elizabeth Bruch, Nicolo Cavalli, Karen Cook, James Evans, Ridhi Kashyap, Matt Salganik, and Emilio Zagheni for helpful comments on previous drafts. We thank Joshua Becker and Alanna Lazarowich for providing data about the International Conference on Computational Social Science.

LITERATURE CITED

- Abul-Fottouh D, Fetner T. 2018. Solidarity or schism: ideological congruence and the Twitter networks of Egyptian activists. *Mobil.: Int. Q.* 23:23–44
- adams J, Light R. 2015. Scientific consensus, the law, and same sex parenting outcomes. *Soc. Sci. Res.* 53:300–10
- AlMaghlouth N, Arvanitis R, Cointet JP, Hanafi S. 2015. Who frames the debate on the Arab uprisings? Analysis of Arabic, English, and French academic scholarship. *Int. Sociol.* 30:418–41
- Araujo M, Mejova Y, Weber I, Benevenuto F. 2017. Using Facebook ads audiences for global lifestyle disease surveillance: promises and limitations. ArXiv:1705.04045 [cs.CY]
- Askarisichani O, Lane JN, Bullo F, Friedkin NE, Singh AK, Uzzi B. 2019. Structural balance emerges and explains performance in risky decision-making. *Nat. Commun.* 10:2648
- Askin N, Mauskopf M. 2017. What makes popular culture popular? Product features and optimal differentiation in music. *Am. Sociol. Rev.* 82:910–44
- Bail CA. 2014. The cultural environment: measuring culture with big data. *Theory Soc.* 43:465–82
- Bail CA. 2015. *Terrified: How Anti-Muslim Fringe Organizations Became Mainstream*. Princeton, NJ: Princeton Univ. Press
- Bail CA. 2016a. Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media. *PNAS* 113:11823–28
- Bail CA. 2016b. Cultural carrying capacity: organ donation advocacy, discursive framing, and social media engagement. *Soc. Sci. Med.* 165:280–88
- Bail CA. 2016c. Emotional feedback and the viral spread of social media messages about autism spectrum disorders. *Am. J. Public Health* 106:1173–80
- Bail CA, Brown TW, Mann M. 2017. Channeling hearts and minds: advocacy organizations, cognitive-emotional currents, and public conversation. *Am. Sociol. Rev.* 82:1188–213
- Bail CA, Brown TW, Wimmer A. 2019. Prestige, proximity, and prejudice: how Google search terms diffuse across the world. *Am. J. Sociol.* 124:1496–548
- Bail CA, Merhout F, Ding P. 2018a. Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in U.S. counties. *Sci. Adv.* 4:eaao5948
- Bail CA, Volfovsky A, Argyle LP, Brown TW, Bumpus JP, et al. 2018b. Exposure to opposing views on social media can increase political polarization. *PNAS* 115:9216–21
- Baldassarri D, Bearman P. 2007. Dynamics of political polarization. *Am. Sociol. Rev.* 72:784–811
- Barberá P, Wang N, Bonneau R, Jost JT, Nagler J, et al. 2015. The critical periphery in the growth of social protests. *PLOS ONE* 10:e0143611
- Becker J, Porter E, Centola D. 2019. The wisdom of partisan crowds. *PNAS* 116:10717–22
- Bonikowski B, Gidron N. 2016. The populist style in American politics: presidential campaign discourse, 1952–96. *Soc. Forces* 94:1593–621
- Boutyline A, Willer R. 2017. The social structure of political echo chambers: variation in ideological homophily in online networks. *Political Psychol.* 38:551–69
- Brayne S. 2020. *Policing Data*. Oxford: Oxford Univ. Press
- Bruch E, Atwell J. 2015. Agent-based models in empirical social research. *Sociol. Methods Res.* 44:186–221
- Bruch E, Feinberg F, Lee KY. 2016. Extracting multistage screening rules from online dating activity data. *PNAS* 113:10530–35
- Bruch E, Newman MEJ. 2018. Aspirational pursuit of mates in online dating markets. *Sci. Adv.* 4:eaap9815
- Bruggeman J, Traag VA, Uitermark J. 2012. Detecting communities through network data. *Am. Sociol. Rev.* 77:1050–63

- Carley K. 1991. A theory of group stability. *Am. Sociol. Rev.* 56:331–54
- Centola D. 2010. The spread of behavior in an online social network experiment. *Science* 329:1194–97
- Centola D, Baronchelli A. 2015. The spontaneous emergence of conventions: an experimental study of cultural evolution. *PNAS* 112:1989–94
- Centola D, Becker J, Brackbill D, Baronchelli A. 2018. Experimental evidence for tipping points in social convention. *Science* 360:1116–19
- Centola D, Macy M. 2007. Complex contagions and the weakness of long ties. *Am. J. Sociol.* 113:702–34
- Cesare N, Lee H, McCormick T, Spiro E, Zagheni E. 2018. Promises and pitfalls of using digital traces for demographic research. *Demography* 55:1979–99
- Chakrabarti P, Frye M. 2017. A mixed-methods framework for analyzing text data: integrating computational techniques with qualitative methods in demography. *Demogr. Res.* 37:1351–82
- Chen Y, Yan F. 2016. Economic performance and public concerns about social class in twentieth-century books. *Soc. Sci. Res.* 59:37–51
- de Vaan M, Stark D, Vedres B. 2015. Game changer: the topology of creativity. *Am. J. Sociol.* 120:1144–94
- DellaPosta D, Shi Y, Macy M. 2015. Why do liberals drink lattes? *Am. J. Sociol.* 120:1473–511
- Dodds PS, Muhamad R, Watts DJ. 2003. An experimental study of search in global social networks. *Science* 301:827–29
- Eagle N, Macy M, Claxton R. 2010. Network diversity and economic development. *Science* 328:1029–31
- Edelmann A, Moody J, Light R. 2017. Disparate foundations of scientists' policy positions on contentious biomedical research. *PNAS* 114:6262–67
- Evans ED, Gomez CJ, McFarland DA. 2016. Measuring paradigmaticness of disciplines using text. *Sociol. Sci.* 3:757–78
- Evans JA, Aceves P. 2016. Machine translation: mining text for social theory. *Annu. Rev. Sociol.* 42:21–50
- Evans JA, Foster JG. 2011. Metaknowledge. *Science* 331:721–25
- Farrell J. 2016a. Corporate funding and ideological polarization about climate change. *PNAS* 113:92–97
- Farrell J. 2016b. Network structure and influence of the climate change counter-movement. *Nat. Clim. Change* 6:370–74
- Fatehkia M, Kashyap R, Weber I. 2018. Using Facebook ad data to track the global digital gender gap. *World Dev.* 107:189–209
- Feinberg M, Willer R. 2015. From gulf to bridge: When do moral arguments facilitate political influence? *Pers. Soc. Psychol. Bull.* 41:1665–81
- Flores RD. 2017. Do anti-immigrant laws shape public sentiment? A study of Arizona's SB 1070 using Twitter data. *Am. J. Sociol.* 123:333–84
- Foster JG. 2018. Culture and computation: steps to a probably approximately correct theory of culture. *Poetics* 68:144–54
- Foster JG, Rzhetsky A, Evans JA. 2015. Tradition and innovation in scientists' research strategies. *Am. Sociol. Rev.* 80:875–908
- Gamson W, Peppers L. 1966. *SIMSOC: Simulated Society, Coordinator's Manual*. New York: Simon & Schuster
- Geburu R, Krause J, Wang Y, Chen D, Deng J, et al. 2017. Using deep learning and Google street view to estimate the demographic makeup of neighborhoods across the United States. *PNAS* 114:13108–13
- Goldberg A, Hannan MT, Kovács B. 2016a. What does it mean to span cultural boundaries? Variety and atypicality in cultural consumption. *Am. Sociol. Rev.* 81:215–41
- Goldberg A, Srivastava SB, Manian VG, Monroe W, Potts C. 2016b. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. *Am. Sociol. Rev.* 81:1190–222
- Golder S, Macy M. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333(6051):1878–81
- Golder S, Macy M. 2014. Digital footprints: opportunities and challenges for social research. *Annu. Rev. Sociol.* 40:129–52
- González-Bailón S, Borge-Holthoefer J, Moreno Y. 2013. Broadcasters and hidden influentials in online protest diffusion. *Am. Behav. Sci.* 57:943–65
- González-Bailón S, Wang N. 2016. Networked discontent: the anatomy of protest campaigns in social media. *Soc. Netw.* 44:95–104

- Guilbeault D, Becker J, Centola D. 2018. Social learning and partisan bias in the interpretation of climate trends. *PNAS* 115:9714–19
- Hanna A. 2013. Computer-aided content analysis of digitally enabled movements. *Mobil.: Int. Q.* 18:367–88
- Heerwig JA. 2017. Money in the middle: contribution strategies among affluent donors to federal elections, 1980–2008. *Am. J. Sociol.* 123:1004–63
- Helbing D, Farkas, Vicsek T. 2000. Simulating dynamical features of escape panic. *Nature* 407:487–90
- Helleringer S, You C, Fleury L, Douillot L, Diouf I, et al. 2019. Improving age measurement in low- and middle-income countries through computer vision: a test in Senegal. *Demogr. Res.* 40:219–60
- Hobbs WR, Burke M, Christakis NA, Fowler JH. 2016. Online social integration is associated with reduced mortality risk. *PNAS* 113:12980–84
- Hoffman MA. 2019. The materiality of ideology: cultural consumption and political thought after the American Revolution. *Am. J. Sociol.* 125:1–62
- Horvát EÁ, Uparna J, Uzzi B. 2015. Network versus market relations: the effect of friends in crowdfunding. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015—ASONAM '15*, ed. J Pei, F Silvestri, J Tang, pp. 226–33. New York: Assoc. Comput. Mach.
- Kaplanis J, Gordon A, Shor T, Weissbrod O, Geiger D, et al. 2018. Quantitative analysis of population-scale family trees with millions of relatives. *Science* 360:171–75
- Kashyap R, Villavicencio F. 2016. The dynamics of son preference, technology diffusion, and fertility decline underlying distorted sex ratios at birth: a simulation approach. *Demography* 53:1261–81
- King G, Persily N. 2019. A new model for industry-academic partnerships. *PS: Political Sci. Politics*. <http://j.mp/2q1IQpH>
- King MM, Bergstrom CT, Correll SJ, Jacquet J, West JD. 2017. Men set their own cites high: gender and self-citation across fields and over time. *Socius* 3. <https://doi.org/10.1177/2378023117738903>
- Kitts JA, Lomi A, Mascia D, Pallotti F, Quintane E. 2017. Investigating the temporal dynamics of interorganizational exchange: patient transfers among Italian hospitals. *Am. J. Sociol.* 123:850–910
- Kossinets G, Watts DJ. 2006. Empirical analysis of an evolving social network. *Science* 311:88–90
- Kozlowski AC, Taddy M, Evans JA. 2019. The geometry of culture: analyzing meaning through word embeddings. *Am. Sociol. Rev.* 84:905–49
- Lazer D, Pentland A, Adamic L, Aral S, Barabasi AL, et al. 2009. Computational social science. *Science* 323:721–23
- Leahey E, Moody J. 2014. Sociological innovation through subfield integration. *Soc. Curr.* 1:228–56
- Lewis K. 2013. The limits of racial prejudice. *PNAS* 110:18814–19
- Lewis K, Gray K, Meierhenrich J. 2014. The structure of online activism. *Sociol. Sci.* 1:1–9
- Li J, Yin Y, Fortunato S, Wang D. 2019. A dataset of publication records for Nobel laureates. *Sci. Data* 6:33
- Light R, Odden C. 2017. Managing the boundaries of taste: culture, valuation, and computational social science. *Soc. Forces* 96:877–908
- Lin K-H, Lundquist J. 2013. Mate selection in cyberspace: the intersection of race, gender, and education. *Am. J. Sociol.* 119:183–215
- Liu B, Govindan R, Uzzi B. 2016. Do emotions expressed online correlate with actual changes in decision-making?: The case of stock day traders. *PLOS ONE* 11:e0144945
- Lundberg I, Narayanan A, Levy K, Salganik MJ. 2018. Privacy, ethics, and data access: a case study of the Fragile Families Challenge. ArXiv:1809.00103 [cs.CY]
- Lutter M. 2015. Do women suffer from network closure? The moderating effect of social capital on gender inequality in a project-based labor market, 1929 to 2010. *Am. Sociol. Rev.* 80:329–58
- Ma Y, Uzzi B. 2018. Scientific prize network predicts who pushes the boundaries of science. *PNAS* 115(50):12608–15
- Macy MW. 2016. An emerging trend: Is big data the end of theory? In *Emerging Trends in the Social and Behavioral Sciences*, RA Scott, S Kosslyn, pp. 1–14. Hoboken, NJ: Wiley
- Macy MW, Willer R. 2002. From factors to actors: computational sociology and agent-based modeling. *Annu. Rev. Sociol.* 28:143–66

- McCormick RH, Lee H, Cesare N, Shojaie A, Spiro ES. 2017. Using Twitter for demographic and social science research: tools for data collection and processing. *Sociol. Methods Res.* 46:390–421
- McFarland D, Lewis K, Goldberg A. 2015. Sociology in the era of big data: the ascent of forensic social science. *Am. Sociol.* 47:12–35
- McMahan P, Evans JA. 2018. Ambiguity and engagement. *Am. J. Sociol.* 124:860–912
- Molina M, Garip F. 2019. Machine learning for sociology. *Annu. Rev. Sociol.* 45:27–45
- Moreno MA, Christakis DA, Egan KG, Brockman LN, Becker R. 2012. Associations between displayed alcohol references on Facebook and problem drinking among college students. *Arch. Pediatr. Adolesc. Med.* 166:157–63
- Nelson LK. 2017. Computational grounded theory: a methodological framework. *Sociol. Methods Res.* 49:3–42
- Ojala J, Zagheni E, Billari FC, Weber I. 2017. Fertility and its meaning: evidence from search behavior. ArXiv:1703.03935 [cs.CY]
- Palmer JRB, Espenshade TJ, Bartumeus F, Chung CY, Ozgencil NE, Li K. 2013. New approaches to human mobility: using mobile phones for demographic research. *Demography* 50:1105–28
- Park PS, Blumenstock JE, Macy MW. 2018. The strength of long-range ties in population-scale social networks. *Science* 362(6421):1410–13
- Pentland A. 2015. *Social Physics: How Social Networks Can Make Us Smarter*. New York, NY: Penguin Books
- Potârca G, Mills M. 2015. Racial preferences in online dating across European countries. *Eur. Sociol. Rev.* 31:326–41
- Rampazzo F, Zagheni E, Weber I, Testa MR, Billari F. 2018. *Mater certa est, pater numquam*: What can Facebook advertising data tell us about male fertility rates? ArXiv:1804.04632 [cs.CY]
- Rawlings CM, McFarland DA. 2011. Influence flows in the academy: using affiliation networks to assess peer effects among researchers. *Soc. Sci. Res.* 40:1001–17
- Rawlings CM, McFarland DA, Dahlander L, Wang D. 2015. Streams of thought: knowledge flows and intellectual cohesion in a multidisciplinary era. *Soc. Forces* 93:1687–722
- Romero DM, Swaab RI, Uzzi B, Galinsky AD. 2015. Mimicry is presidential: linguistic style matching in presidential debates and improved polling numbers. *Pers. Soc. Psychol. Bull.* 41:1311–19
- Rossman G, Esparza N, Bonacich P. 2010. I'd like to thank the Academy, team spillovers, and network centrality. *Am. Sociol. Rev.* 75:31–51
- Rossman G, Schilke O. 2014. Close, but no cigar: the bimodal rewards to prize-seeking. *Am. Sociol. Rev.* 79:86–108
- Rzhetsky A, Foster JG, Foster IT, Evans JA. 2015. Choosing experiments to accelerate collective discovery. *PNAS* 112:14569–74
- Saavedra S, Duch J, Uzzi B. 2011a. Tracking traders' understanding of the market using e-communication data. *PLOS ONE* 6:e26705
- Saavedra S, Hagerty K, Uzzi B. 2011b. Synchronicity, instant messaging, and performance among financial traders. *PNAS* 108:5296–301
- Salganik M. 2018. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton Univ. Press
- Salganik MJ, Dodds PS, Watts DJ. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311:854–56
- Salganik MJ, Lundberg I, Kindel AT, Ahearn CE, Al-Ghoneim K, et al. 2020. Measuring the predictability of life outcomes with a scientific mass collaboration. *PNAS*. <https://doi.org/10.1073/pnas.1915006117>
- Salganik MJ, Watts DJ. 2008. Leading the herd astray: an experimental study of self-fulfilling prophecies in an artificial cultural market. *Soc. Psychol. Q.* 71:338–55
- Schnable A. 2016. What religion affords grassroots NGOs: frames, networks, modes of action. *J. Sci. Study Relig.* 55:216–32
- Schroeder R. 2010. *Being There Together: Social Interaction in Shared Virtual Environments*. Oxford, UK: Oxford Univ. Press
- Shi F, Foster JG, Evans JA. 2015. Weaving the fabric of science: dynamic network models of science's unfolding structure. *Soc. Netw.* 43:73–85
- Shi F, Shi Y, Dokshin FA, Evans JA, Macy MW. 2017. Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nat. Hum. Behav.* 1:0079

- Shirado H, Christakis NA. 2017. Locally noisy autonomous agents improve global human coordination in network experiments. *Nature* 545:370–74
- Shirado H, Iosifidis G, Tassioulas L, Christakis NA. 2019. Resource sharing in technologically defined social networks. *Nat. Commun.* 10:1079
- Shor E, van de Rijt A, Miltsov A, Kulkarni V, Skiena S. 2015. A paper ceiling: explaining the persistent underrepresentation of women in printed news. *Am. Sociol. Rev.* 80:960–84
- Shwed U. 2015. Robust science: passive smoking and scientific collaboration with the tobacco industry in the 1970s. *Sociol. Sci.* 2:158–85
- Shwed U, Bearman PS. 2010. The temporal structure of scientific consensus formation. *Am. Sociol. Rev.* 75:817–40
- Srivastava SB, Goldberg A, Manian VG, Potts C. 2018. Enculturation trajectories: language, cultural adaptation, and individual outcomes in organizations. *Manag. Sci.* 64:1348–64
- State B, Rodriguez M, Helbing D, Zagheni E. 2014. Migration of professionals to the U.S. In *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11–13, 2014. Proceedings, Lecture Notes in Computer Science*, ed. LM Aiello, D McFarland, pp. 531–43. Cham, Switz.: Springer
- Stewart I, Flores R, Riffe T, Weber I, Zagheni E. 2019. Rock, rap, or reggaeton?: Assessing Mexican immigrants' cultural assimilation using Facebook data. ArXiv:1902.09453 [cs.CY]
- Tufekci Z, Wilson C. 2012. Social media and the decision to participate in political protest: observations from Tahrir Square. *J. Commun.* 62:363–79
- Uzzi B, Mukherjee S, Stringer M, Jones B. 2013. Atypical combinations and scientific impact. *Science* 342:468–72
- Vaillant GG, Tyagi J, Akin IA, Poma FP, Schwartz M, van de Rijt A. 2015. A field-experimental study of emergent mobilization in online collective action. *Mobil.: Int. Q.* 20:281–303
- van de Rijt A. 2019. Self-correcting dynamics in social influence processes. *Am. J. Sociol.* 124:1468–95
- van de Rijt A, Akin IA, Willer R, Feinberg M. 2016. Success-breeds-success in collective political behavior: evidence from a field experiment. *Sociol. Sci.* 3:940–50
- van de Rijt A, Kang SM, Restivo M, Patil A. 2014. Field experiments of success-breeds-success dynamics. *PNAS* 111:6934–39
- van Loon A, Bailenson J, Zaki J, Bostick J, Willer R. 2018. Virtual reality perspective-taking increases cognitive empathy for specific others. *PLOS ONE* 13:e0202442
- Vasi IB, Walker ET, Johnson JS, Tan HF. 2015. “No Fracking Way!” Documentary film, discursive opportunity, and local opposition against hydraulic fracturing in the United States, 2010 to 2013. *Am. Sociol. Rev.* 80:934–59
- Vilhena DA, Foster JG, Rosvall M, West JD, Evans J, Bergstrom CT. 2014. Finding cultural holes: how structure and culture diverge in networks of scholarly communication. *Sociol. Sci.* 1:221–39
- Wagner C, Graells-Garrido E, Garcia D, Menczer F. 2016. Women through the glass ceiling: gender asymmetries in Wikipedia. *EPJ Data Sci.* 5:5
- Wagner C, Singer P, Strohmaier M. 2014. The nature and evolution of online food preferences. *EPJ Data Sci.* 3:38
- Watts DJ. 1999. Networks, dynamics, and the small-world phenomenon. *Am. J. Sociol.* 105:493–527
- Watts DJ. 2004. *Six Degrees: The Science of a Connected Age*. New York: W. W. Norton & Co.
- Watts DJ. 2017. Should social science be more solution-oriented? *Nat. Hum. Behav.* 1:0015
- Watts DJ, Dodds P. 2007. Influentials, networks, and public opinion formation. *J. Consumer Res.* 34:441–58
- Wells C, van Thomme J, Maurer P, Hanna A, Pevehouse J, et al. 2016. Coproduction or cooptation? Real-time spin and social media response during the 2012 French and US presidential debates. *French Politics* 14:206–33
- West JD, Jacquet J, King MM, Correll SJ, Bergstrom CT. 2013. The role of gender in scholarly authorship. *PLOS ONE* 8:e66212
- Wu L, Wang D, Evans JA. 2019. Large teams develop and small teams disrupt science and technology. *Nature* 566:378–82
- Wuchty S, Jones BF, Uzzi B. 2007. The increasing dominance of teams in production of knowledge. *Science* 316:1036–39

- Yang Y, Chawla NV, Uzzi B. 2019. A network's gender composition and communication pattern predict women's leadership success. *PNAS* 116:2033–38
- Zagheni E, Weber I, Gummadi K. 2017. Leveraging Facebook's advertising platform to monitor stocks of migrants. *Popul. Dev. Rev.* 43:721–34
- Zhang H, Pan J. 2019. CASM: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociol. Methodol.* 49:1–57