

Reports of the Workshops Held at the 2017 International AAAI Conference on Web and Social Media

*Jisun An, Giovanni Luca Ciampaglia, Nir Grinberg, Kenneth Joseph,
Alexios Mantzarlis, Gregory Maus, Filippo Menczer,
Nicholas Proferes, Brooke Foucault Welles*

■ *The Workshop Program of the Association for the Advancement of Artificial Intelligence's International Conference on Web and Social Media (AAAI-17) was held in Montréal, Québec, Canada, on Monday, May 15, 2017. There were eight workshops in the program: Digital Misinformation, Events Analytics Using Social Media Data, News and Public Opinion, Observational Studies through Social Media, Perceptual Biases and Social Media, Social Media and Demographic Research, Studying User Perceptions and Experiences with Algorithms, and The ICWSM Science Slam. Workshops were held on the first day of the conference. Workshop participants met and discussed issues with a selected focus — providing an informal setting for active exchange among researchers, developers, and users on topics of current interest. Organizers from two of the workshops chose to include papers in the AAAI Technical Reports series (Observational Studies through Social Media and News and Public Opinion). Their papers were included as a nonarchival part of the ICWSM proceedings. Organizers from four of the workshops (Digital Misinformation, News and Public Opinion, Perceptual Biases and Social Media, and Studying User Perceptions and Experiences with Algorithms) submitted reports, which are reproduced in this report. Brief summaries of the other four workshops have been reproduced from their website descriptions.*

Digital Misinformation

The deluge of online and offline misinformation is overloading the exchange of ideas upon which democracies depend. Fake news, conspiracy theories, and deceptive social bots proliferate, facilitating the manipulation of public opinion. Countering misinformation while protecting freedom of speech will require collaboration across industry, journalism, and academe. To foster such efforts, the Workshop on Digital Misinformation was held to discuss research challenges toward a reliable web.

The organizers opened the workshop by showcasing a growing suite of publicly available tools for studying digital misinformation, developed by the Center for Complex Networks and Systems Research and the IU Network Science Institute. They include Hoaxy, a system for tracking competing claims and fact checking that spread online; and Botometer, a machine-learning algorithm for detecting social bots on Twitter. Theoretical and empirical studies were presented, showing for example that low-quality information often goes more viral than high-quality information on social media. Among the factors that help interpret these findings are the structural segregation and polarization in online social networks. The resulting echo chambers are exacerbated by algorithms that personalize online experiences and hinder exposure to ideologically diverse sources of information. Other factors include information overload, limited attention, popularity bias, and manipulation through social bots.

Speakers from journalism, social media platforms, government, and communication, computer, and information science described misinformation as a global problem intertwined with increasing distrust of media and public institutions. They all agreed that the problem is complex and supported greater coordination between researchers, journalists, and social media platforms, as well as efforts to improve public media literacy. Their other proposals varied from reputation-based filtering and ranking mechanisms (to prioritize sharers with a track record of reliability rather than simple popularity), to providing link-optimized fact pages to disprove disinformation (to be shared organically in disagreements), to specific digital products for flagging stories as lacking credibility.

A lively discussion ensued about the commitment of platforms to curbing the spread of misinformation. It was pointed out that more should be done to deal with abuses that exploit social bots, fake accounts, and pages, and verified account badges. It was also suggested that an API for access to public Facebook page data would be a great boon to the research community. Workshop participants from Facebook, Google, and Twitter highlighted various efforts. For example, Facebook is pursuing four approaches: (1) disrupting the financial incentives for fake news; (2) developing new products to curb the spread of fake news, such as allowing users or third-party fact checkers to flag posted stories as untrue or unverified; (3) helping people make informed decisions by educating them on how to spot fake news; and (4) launching a partnership between industry and nongovernmental organizations to promote media literacy. Platforms regularly engage with the research community through collaborative programs and grants, but acknowledged the growing demand from third-party researchers for data to tackle the spread of misinformation.

A first panel of five computer and information scientists discussed the research challenge of defining and detecting misinformation in online settings. They highlighted patterns of attack, such as leveraging fake accounts to target specific groups and spread links to fake news stories. The panelists delineated different types of rumor, different motives for creating and spreading inaccurate information (financial, political, and ideological and cultural), and different methods for detecting misinformation based on linguistic, network, account, URL, and crowdsourced features. They lamented the limits of simple heuristics that can only address a fraction of the issue, and the lack of ground-truth data to train sophisticated machine-learning methods, including deep learning; fact checkers cannot cope with the sheer volume and variety of misinformation. On the other hand, crowdsourced information can be easily manipulated; for example, it is easy to recruit workers for astroturfing. Aggregated signals to infer the reliability or controversy of a given piece of content might be helpful. Panelists called for more research to understand why people reshare rumors (which is less well understood than the incentives for creating rumors), identify the intent behind social media posts, and predict the impact of rumors and corrections. Finally, they urged social media platforms to highlight provenance information about sources to help users determine their intents and trustworthiness, and to provide researchers with data about how recipients of misinformation engage with it.

Through an online Wiki survey, workshop participants identified five major research challenges related to the definition and detection of misinformation: (1) Identifying and promoting reliable information instead of focusing on disinformation; (2) Tracking variants of debunked claims; (3) Developing reputation scores for publishers (4) Creating an automated trustmark to promote journalistic integrity, and (5) Collecting reliable crowdsourcing signals.

A second panel included a historian, a psychologist, a physicist, an economist, and a computer scientist who addressed the research challenges in studying cognitive, social, and technological biases that create vulnerabilities to misinformation. They cited studies showing that people have difficulty distinguishing the reliability of news stories to argue that the fight against misinformation is one in which we all must take part, not just big tech companies. One panelist noted that online partisan news leads people to reject evidence, due to these outlets' emotional pull. While social media increase the profile of misinformation, it remains unclear how much this actually shifts public opinions. Panelists emphasized that human limits of information processing make it impossible to keep up with the growing volume of information, resulting in reliance on simple cognitive heuristics to manage the load. These heuristics may in turn amplify cer-

tain cognitive biases. A simple model demonstrated that online echo chambers are inevitable within the current social media mechanisms for content sharing, which tend to cluster individuals into segregated and polarized groups. These network biases distort our perceptions. Finally, panelists noted that research in the field tends to focus on what content people consume rather than the more difficult matter of how they actually absorb that information. Is mass ignorance on a particular issue more or less harmful than the consumption of fake news about that issue? Research may also be distorted by overreliance on Twitter data, which may not be as representative of news consumption by the general population as Facebook or television.

Workshop participants identified five major research challenges related to the study of biases that make us vulnerable to misinformation: (1) Investigating the use of language, images, and design in misinformation persuasiveness; (2) Validating model predictions through field experiments; (3) Studying the roles of algorithmic mechanisms in the spread of misinformation; (4) Translating research findings into policy recommendations; and (5) Accessing behavioral data from social media platforms.

The final panel offered perspectives of computational journalists, fact checkers, and computer and communication scientists about countermeasures against misinformation, as well as who can best deliver them. Panelists described efforts to categorize news sources and noted that poor journalism helps spread fake news, sometimes from the very outlets that are supposed to question and disprove misinformation. They argued for the need of partnerships between civil society and platforms to combat fake news effectively. Platforms have a responsibility to promote information that builds trust and to assist journalists in their work. Google is partnering with fact-checking organizations and leveraging semantic web annotations to facilitate the labeling of fake news. Artificial intelligence can be a powerful tool to promote quality and trust in information. However, users play a role in the spread of misinformation, which may be the most challenging problem to address. Some platforms, like Reddit, can be easily gamed by coordinated manipulation. Finally, panelists denounced political attacks that, ironically, leverage fake news to undermine research about digital misinformation.

Workshop participants identified five major research challenges related to feasible/effective countermeasures: (1) Support and scaffold critical thinking; (2) Increase prominence and availability of fact-checking information; (3) Design trust and reputation standards for news sources and social media users; (4) Build tools to track the provenance of digital content; and (5) Develop computational tools to support fact checking.

The ultimate goal of the workshop was to bootstrap a long-lasting initiative between various sectors

(industry, academe, journalism, civil society) for building a reliable web. Follow-up discussion and further collaborative activity is currently under way through online community spaces. We believe that support from both private foundations and federal agencies will be a key ingredient for the success of future collaborative activities, the scope of which must include research, education, and policymaking.

The workshop was supported by the Network Science Institute and the School of Informatics and Computing at Indiana University. Alessandro Flammini served as cochair.

Event Analytics Using Social Media Data

The purpose of the Event Analytics Using Social Media Data workshop was to bring together researchers that are working in a variety of areas that are all related to the larger problem of analyzing and understanding events using social media responses, to discuss what are the recently developed machine-learning and data-mining techniques that can be leveraged to address challenges in analyzing events using social media data, and, from challenges in analyzing events, what are the practical research directions in the machine learning and data mining community. The workshop was organized by Yuheng Hu (University of Illinois at Chicago) and Yu-Ru Lin (University of Pittsburgh). No report was submitted by the organizers.

News and Public Opinion

Media is known to influence our lives, shaping our attitudes and behaviors on different matters. The role of media has become more important in this new digital era, where anyone can play a role in media. Media scholars and social scientists have long been studying the role of media, and they have established theories for understanding the way in which news is produced, distributed, and consumed, as well as its influence. In the digital era, people can take different actions with news content, from reading to sharing to discussing, and all of these actions are digitally recorded. This availability of data gives scholars the opportunity to study whether the existing media theories still hold and to investigate changes in media systems and audience behavior. The goal of this workshop was to bring two disciplines (social sciences and computer sciences) together to seek different methodologies and better approaches to tackling various issues in media studies using large-scale data.

The workshop brought together researchers in diverse domains, such as computer science, social science, and political science, with a variety of methodologies, such as statistical inferences, network analysis, machine learning, and natural lan-

guage processing, to unveil questions in media studies. One major theme of the papers presented at the workshop was data-driven studies on news coverage and their applications. News coverage determines which issue or actor would be visible to people. To have a fair view of news coverage within a country or across countries, it requires news data sets that contain content from all media sources. In this regard, two global news data sets were introduced, The Global Database of Events, Language, and Tone (GDELT) and the Unfiltered News data set. Using such data sets, news coverage of a particular issue (such as refugees) was analyzed. Also, a method to measure the level of press freedom was proposed by exploiting the diversity of news topics for different countries.

Another major theme, following the theme of last year's workshop, was the measurement of public opinion by examining various online user reactions, such as likes, shares, and discussion, and their relations to news popularity. The papers in this theme provided analyses of what led a positive public opinion, using social media reactions as popularity measures on different platforms. Social networking sites such as Twitter and Facebook were the major sources for conducting opinion research. Also, we saw growing interests in online communities such as Reddit to understand public responses on news events. All of the papers on this theme confirmed that the sentiment and the content of the title closely relate to the popularity of news items and highlighted that different aspects of news such as news values and style need more attention in this research. One of the papers studied what makes a successful political campaign and suggested actively using the feedback from social media reactions when running a political campaign.

The third theme of the workshop was fake news, which was a special theme this year. A paper on this theme investigated the linguistic differences in titles between real news, satire news, and fake news and found that fake news is more similar to satire than real news. This paper served to bring the discussion to the responsibility of the media — the changing media landscape is forcing media organizations to generate more and more shocking titles, and that may cause people to be more vulnerable to fake news.

The workshop participants actively shared their expertise on different methods and knowledge of data sources that can be utilized for this line of research. They also discussed the theoretical and practical implications of the research and strongly expressed their interests in attending a future NECO workshop.

Jisun An, Haewoon Kwak, and Fabrício Benevenuto served as cochairs of this workshop. This report was written by Jisun An. Papers from the workshop were published as AAAI : News and Public Opinion — Technical Report WS-17-17.

Observational Studies through Social Media

People's usage of mobile devices, Internet services, and applications creates a rich repository of data across many domains. As explorations and applications by the ICWSM community become more focused on understanding mechanisms and addressing societal and individual-level problems through policy- and individual-level interventions, the importance of careful studies and causal reasoning methods is becoming more critical. The ICWSM workshop on Observational Studies through Social Media and other human-generated content brought together social scientists, computer scientists, and others who are investigating observational studies of these interactions and data across many areas, including public health, medicine, sociology, education and others. The goal of the workshop was to foster discussion and brainstorming in this area. Selected papers from the workshop were included in AAAI Technical Report WS-17-16: Observational Studies through Social Media. The workshop was organized by Elad Yom-Tov (Microsoft Research), Munmun De Choudhury (Georgia Institute of Technology), Emre Kiciman (Microsoft Research), and Tim Althoff (Stanford University). No report was submitted by the organizers.

Perceptual Biases and Social Media

Recent demonstrations of racial and gender bias in the United States have drawn public attention to stereotyping and discrimination in American society and its institutions. However, stereotyping and prejudice are only two examples of how our perceptual biases affect behavior. Confirmation bias, bias in media coverage of events, and preferential attachment are all examples of perceptual biases that shape social processes like the creation of echo chambers and discriminatory behaviors. The emergence of social media as a prominent medium for human communication has the potential to provide a new lens for studying the relationship between perceptual biases and social processes, and the role social media plays in affecting this relationship.

The first part of the workshop began with a keynote talk by Alice Marwick (Data & Society Research Institute and University of North Carolina at Chapel Hill) who spoke about mechanisms through which far-right political organizations use social media to amplify messaging and spread disinformation online. Following the talk, participants broke into small groups to discuss various perceptual biases that impart social media platforms. Perhaps most prominent was the reference to echo chambers and confirmation bias, which were identified as the biggest threats to open democratic societies brought about by social media platforms. Other points were made about how groupthink in competitive social

settings can lead to unethical behavior; and how properties that are external to the content, such as its presentation or recency, effect the perceived credibility of communicated information or affect its impact on other individuals.

The second half of the workshop began with a keynote by Christo Wilson (Northeastern University), who talked about methodologies for auditing biases in social systems and algorithms. Participants then discussed potential ways to train algorithms that are more robust to perceptual biases present in data and provide tools that would better inform end users. For example, a corrective mechanism was proposed to de-bias data before training machine-learning models on it. Other examples include tools like BuzzFeed's "outside your bubble," meant to provide people with the perspective of others, outside their siloed conversation on social media.

We intend to follow up on these ideas with participants and pursue a publication in the form of an opinion piece, summarizing current knowledge and possible future research direction for the study of perceptual biases and social media.

Nir Grinberg, Kenneth Joseph, and Brooke Foucault Welles organized this workshop and wrote this report.

Social Media and Demographic Research

Demography has been a data-driven discipline since its birth. Data collection and the development of formal methods have sustained most of the major advances in our understanding of population processes. The global spread of social media has generated new opportunities for demographic research, as individuals leave an increasing quantity of traces online that can be aggregated and mined for population research. At the same time, the use of social media and the Internet are affecting people's daily activities as well as life planning, with implications for demographic behavior. The goal of this workshop was to favor communication and exchange between the communities of demographers and data scientists. It revolved around the main theme of applications and implications of social media and online data for demographic research.

The workshop was organized by Emilio Zagheni (University of Washington, Seattle), Ingmar Weber (Qatar Computing Research Institute, HBKU), and Thomas LeGrand (Montréal University, Canada). No report was submitted by the organizers.

Studying User Perceptions and Experiences with Algorithms

From Facebook's News Feed algorithm that shapes the posts and updates we see, to Spotify's recommenda-

tion service that introduces us to new music that we might love, to dating site algorithms that attempt to match us with potential romantic partners, algorithms play an increasingly important role in shaping many aspects of our daily lives. The Studying User Perceptions and Experiences with Algorithms workshop brought together a community of researchers interested in taking a human-centered perspective on studying the experience of algorithms.

The objective of this workshop was to articulate the grand challenges of studying the user-algorithm relationship and to bring together participants interested in developing projects to address these grand challenges. During the first breakout session of the workshop, participants identified a number of outstanding research questions in this area. These included questions such as: What do users think is an algorithm? How do users employ their (mis)understandings of how algorithms work to reverse engineer or manipulate them? And does new information about how to manipulate algorithms change users' perceptions of how the algorithm works? How do values and preferences transfer from people to algorithms? How do different degrees of awareness of algorithms change user behavior? How do algorithms obscure themselves? What makes users hostile or positively disposed to an algorithm? What parts of an algorithm "should" users see or not see? And who should be in charge of making these decisions? How can we combine "big data" methods with "small data" methods to discern longer-term effects of information filter algorithms on users' worldviews?

Participants then self-selected groups for the second breakout session based on the tractable entry points they found to be of interest during the first breakout. Groups used this time to incubate research ideas, focusing on how they might take action on the questions. At the end of the session, groups reported out on the projects they envisioned.

Participants were encouraged to develop a short abstract for the project on user-algorithmic interaction they envisioned, using Heilmeier's questions (www.darpa.mil/work-with-us/heilmeier-catechism) as a guide. In addition to reporting back to the larger workshop about their envisioned projects, participants also developed a short recommended reading list (www.studyingusers.org/reading-list) of articles on this subject matter.

The cochairs for this event were Nicholas Proferes (University of Maryland), Alissa Centivany (Western University), Caitlin Lustig (University of California Irvine), and Jed Brubaker (University of Colorado Boulder). Additional organizers included Lala Hajibayova (Kent State University), Marina Kogan (University of Colorado Boulder), Tanushree Mitra (Georgia Institute of Technology), and Nicole Ellison (University of Michigan). This report was written by Nicholas Proferes.

16th International Conference on Principles of Knowledge Representation and Reasoning
29 October - 2 November 2018
Tempe, Arizona

<http://kr2018.org>

Co-located with DL 2018, NMR 2018

abstract submission : 13 May 2018
paper submission : 20 May 2018

The ICWSM Science Slam

A *science slam* is an epic scientific event where scientists compete with short talks on their research. It's just like a poetry slam, but with science instead of poems. Slammers were completely free to do whatever they wanted on stage. Everything was allowed, including slides, games — the more creative, the better! The only two rules were that the topic of the slam had to be related to social media and the presentation could not take more than 8 minutes. The 2017 ICWSM Science Slam was organized by David Garcia, Ingmar Weber, Aniko Hannak, and Robert West.

Jisun An is a social computing scientist at the Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar.

Giovanni Luca Ciampaglia is a research scientist at the Indiana University Network Science Institute.

Nir Grinberg is a postdoc at the Network Science Institute at Northeastern University and a fellow at Harvard's Institute for Quantitative Social Science.

Kenneth Joseph is a postdoc at the Network Science Institute at Northeastern University and a fellow at Harvard's Institute for Quantitative Social Science.

Alexios Mantzarlis leads the International Fact-Checking Network at the Poynter Institute.

Gregory Maus is a PhD student in Informatics at Indiana University, Bloomington.

Filippo Menczer is a professor of Informatics and Computing at Indiana University, Bloomington.

Nicholas Proferes is a postdoctoral scholar at the Ethics and Values in Design Lab at the University of Maryland's College of Information Studies.

Brooke Foucault Welles is an assistant professor in the Department of Communication Studies and a faculty affiliate of the Network Science Institute and NULab for Texts, Maps and Networks at Northeastern University.