

Reports of the Workshops Held at the 2016 International AAI Conference on Web and Social Media

Jisun An, David J. Crandall, Roman Fedorov, Casey Fiesler, Fabio Giglietto, Bahareh Heravi, Jessica Pater, Konstantinos Pelechrinis, Daniele Quercia, Katrin Weller, Arkaitz Zubiaga

■ *The workshop program of the Association for the Advancement of Artificial Intelligence's International Conference on Web and Social Media (AAAI-16) was held in Cologne, Germany. There were eight workshops in the program: CityLab, Challenges and Futures for Ethical Social Media Research, Social Media and Demographic Research, Wiki, #Fail: Things That Didn't Work Out in Social Media Research — And What We Can Learn from Them, News and Public Opinion, Social Media in the Newsroom, and Social Web for Environmental and Ecological Monitoring. Workshops were held on the first day of the conference, Tuesday, May 17, 2016. 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. Of the eight workshops held at the conference; organizers from only five included papers in the AAI Technical Reports series, and organizers from six workshops submitted reports.*

CityLab

Based on a recent report from the United Nations, more than 50 percent of the world's population currently lives in cities. This percentage is projected to increase to 70 percent by the year 2050. As massive amounts of people move to urban areas there is a need for cities to be run more efficiently, while at the same time improving the quality of life of their dwellers. Nevertheless, the exact same force that sets the above requirement, that is, the proliferation of urbanization levels, makes this task much harder and challenging, especially in megacities. Despite the aforementioned conflicting dynamics, many city management operations can be facilitated by appropriate exploitation of the unprecedented amount of data that can be made available to authorities from a variety of sources. In the era of big data and ubiquitous and pervasive mobile computing, different types of sensors such as parking meters, weather sensors, traffic sensors, pipe sensors, public transportation ticket readers, and even human sensors (for example, through web technologies, social media, or cell phone usage data) can assist in these

efforts. Furthermore, civic applications can exploit web and mobile technologies to deliver a livable, sustainable, and resilient environment to the citizens.

Harnessing these information streams and technologies presents many challenges that were in the epicenter of the CityLab workshop. How representative are sources such as social media for planning operations? How can we integrate traditional web data sources with novel and heterogeneous civic data? What are the sensing technologies that will facilitate civic data collection? How can we exploit web technology to engage citizens and drive bottom-up policies? Can crowdsourcing deliver robust solutions in urban problems? Are web and mobile-based solutions going to increase the digital divide and inequality? The abundance of information can facilitate solutions but can hinder problems as well. For example, how do we assure that citizens' privacy and open data can coexist? Cities can use their understanding of the citizen to foster stronger relationships with the diverse communities in their constituencies. This understanding could be applied to mobilize people on important issues such as education, health care, political engagement, and community awareness. Also, new digital fabrication tools have been recently used to generate adoptable, dynamic, and interactive architecture able to evolve together with urban dwellers, and it has been shown that new Internet-of-Things devices could effectively capture physical observations to understand how cities and urban centers work. As a result, cities now provide a living lab where applied research can be carried out to understand citizen and services with a focus on collaborative, user-centered design and co-creation.

The CityLab workshop was organized by Daniele Quercia (Bell Labs) and Konstantinos Pelechrinis (University of Pittsburgh). The papers from the workshop were published by AAAI Press as technical report WS-16-16, which was included in *The Workshops of the Tenth International AAAI Conference on Web and Social Media*.

Challenges and Futures for Ethical Social Media Research

Ethical research design has been a cornerstone of the human-computer interaction field since its inception. However, online research has increasingly challenged traditional ethical guidelines for human subjects research, especially with the widespread use of social media content and data. Academic and industry researchers alike, along with institutional review boards and other governing bodies, have struggled with how to deal with these new research paradigms in an ethical way. The goal of this workshop was to explore the difficult and unanswered ethical questions surrounding this tension, with an eye toward best practices for social media and big data research.

This workshop brought together researchers from

both industry and academia in a multitude of sub-fields such as data security, behavioral health, gender, identity, law, and media studies. The overarching theme of presentations and discussions was that of the ethical challenges inherent in participants' work around social media or big data, regardless of discipline. A number of short talks from participants framed the discussion. One theme was that of vulnerable populations: David Myles (University of Montreal) focused on privacy and data sensitivity while studying online mourning practices, and Jessica Pater (Georgia Tech) talked about issues associated with studying presentations of behavioral health issues like eating disorders and self-harm across social media platforms. Several participants focused on the ethics of new movements within our research paradigm, including Anna Lauren Hoffmann's (University of California, Berkeley) talk on justice and research ethics, Nicholas Proferes' (University of Maryland College Park) discussion of frameworks for open data, Yew Chuan Ong's (University of Sheffield) work on ethics and social bots, and a Twitter research case study from Sara O'Sullivan (University College Dublin). Finally, there were several talks regarding policies and standards for data ethics. Mutlu Binark (Hacettepe University) spoke about the movement in Turkey to establish data ethics policies, Michael Zimmer (University of Wisconsin Milwaukee) presented on conceptual gaps in research ethics policies, and Lauri Kanerva (Facebook) discussed how Facebook has been meeting these gaps and challenges with its newly established review policies. Casey Fiesler (University of Colorado Boulder) closed out the talks with a discussion of ACM SIGCHI proposals for reviewing procedures regarding ethics.

The group collected these insights, organized them, and synthesized commonalities across all genres of research and study that were represented in the workshop. Seven key open questions resulted from this discussion. First, do we need to redefine or make clear what we define as "public" as it relates to data? Second, what do we define as a human subject? We know that people are human subjects, information they share with us is human subject data, but what about the information they share publicly online? Third, obtaining consent: if information shared online is human subject data, are there situations where obtaining informed consent is appropriate or not? Fourth, should relational and associational harms associated with online data be a concern of our community? Fifth, is it acceptable for researchers to violate terms of service (TOS)? If research necessitates breaking TOS, what are the additional research design oversights needed from an ethical standpoint to mitigate risk? Sixth, how do we account for users' expectations in these discussions? Research has shown that tech savviness and digital literacy levels are inconsistent across populations, making this a very complex issue in and of itself. Finally, seventh,

how do we address the use of stolen, leaked, or deleted data in our research? The participants were in agreement that this was the beginning of a much longer, nuanced, and needed conversation that we hope to see manifest within more thoughtful research designs, the addressing of ethical challenges and barriers within papers, and more open discussions within the social media research community and in future workshops.

Casey Fiesler and Jessica Pater served as cochairs for the workshop, along with Anna Lauren Hoffmann (University of California, Berkeley) and Nicholas Proferes (University of Maryland College Park). This report was written by Casey Fiesler and Jessica Pater. This workshop did not issue a technical 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 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. The organizers of the workshop were Bogdan State (Facebook Research and Stanford University), Emilio Zagheni (University of Washington), and Francesco Billari (Oxford University). They planned no proceedings and did not submit a report to *AI Magazine*.

Wiki

The goal of the Wikipedia workshop was to bring together researchers exploring all aspects of Wikimedia websites such as Wikipedia, Wikidata, and Commons. With members of the Wikimedia Foundation's Research team on the organizing committee and with the experience of a successful workshop in 2015, the aim was to continue facilitating a direct pathway for exchanging ideas between the organization that operates Wikimedia websites and the researchers interested in studying them.

The workshop organizers were Robert Wes (Stanford University), Leila Zia (Wikimedia Foundation), Dario Taraborelli (Wikimedia Foundation), and Jure Leskovec (Stanford University). The papers from the workshop were published by AAAI Press as technical report WS-16-17, which was included in *The Work-*

shops of the Tenth International AAAI Conference on Web and Social Media. The organizers did not submit a report for *AI Magazine*.

#FAIL! Things That Didn't Work Out in Social Media Research — And What We Can Learn from Them

Studying social media is challenging for a variety of reasons; for example, data from social media platforms may not easily be available to researchers or it may be difficult to deal with the ever-changing landscape of social media platforms and likewise the ever-changing practices of how users interact with these platforms. As researchers from different disciplines become interested in working with social media data, many individuals or groups of researchers try out new approaches that may be suitable for their particular research questions. Oftentimes they build up expertise in this field through trying several approaches first, before one of them is successful — for example when finding ways for collecting data through the application programming interfaces (APIs) or for setting up privacy protection strategies for social media users. And oftentimes this expertise is not being shared alongside with the actual research results as part of the papers published in the scattered domain of social media research. This workshop series therefore aims at collecting cases in which approaches for social media studies did not work out as expected, for example, due to problems with data collection tools, lack of knowledge or tools to analyze large data sets, or legal and ethical frameworks. The workshop provides a platform to share experiences and lessons learned that may not easily be published in classical formats such as journals or conference papers and it brings together researchers across disciplines to discuss solutions to current problems in social media research.

This year's workshop added a new focus on groups of users that are particularly vulnerable — a topic that came up in two out of the three invited talks and that also resonated well with the general audience. Munmun de Choudhury (Georgia Tech) gave the opening keynote and provided valuable insights into her broad expertise in studying sensitive topics (such as health issues) combining behavioral data gathered from social media and self-reported data collected through in-depth interviews and focus groups. Yenn Lee (SOAS University of London) presented insights from her recent work on a peculiar practice of creating shared Twitter accounts to publicly express anonymous opinion. She approached social media data with the lens of an ethnographer dealing with very small case phenomena. Finally, Isabella Peters (ZBW Leibniz Information Center for Economics) contributed to the broader collection of lessons learned when working with specific tools in social media research. Her work is based in the field of

studying new types of scholarly citations through so-called altmetrics, and she reported on a comparison of different tools that aggregate such altmetric data for scholarly articles.

This workshop was organized by Katrin Weller (GESIS Leibniz-Institute for the Social Sciences), Luca Rossi (IT University of Copenhagen), and Fabio Giglietto (University of Urbino). This report was written by Katrin Weller and Fabio Giglietto.

News and Public Opinion

Computational journalism has been an emerging discipline in recent years. A lot of inspiring data-driven studies on news production and consumption, news audience, news bias, and tools have been published and discussed. Social media has received a lot of attention from researchers who study public opinion, as their data is a valuable asset for understanding humans' opinions, attitudes, and behaviors. The goal of this workshop was to make these two disciplines meet and to investigate the interplay between news and public opinion.

The workshop brought together researchers in different domains such as computer science, social science, and political science. One major theme of the papers presented at the workshop was data-driven studies on news production and consumption. A talk given by Haewoon Kwak (Qatar Computing Research Institute) focused on gender and political bias hidden in the choice of news photos through the analysis of millions of news photos powered by deep learning-based vision APIs. A talk given by Augustin Chaintreau (Columbia University) focused on estimating the clicks of news articles from the shares of the news articles.

Another major theme was the measurement of public opinion. Researchers are interested in how people express their opinion in social media and opinion can be measured by various actions, such as tweets, following, and others. A talk given by Yu Wang (University of Rochester) focused on how the politician's actions, particularly Donald Trump's actions, are related to the behavior of online users, in particular following the Twitter account of the politician. The third theme of the workshop was the interplay between news and public opinion, focusing on the influence of the media on the topical interests of the public. A talk given by Matthew J. Williams (University of Birmingham) focused on political news-sharing communities on social media. The results revealed that the political segregation in news sharing and the topics mentioned in social media closely align with what people say on the media.

The workshop participants shared their experiences on different methods and data sources that can be used for this line of research. The participants also discussed the practical implications of the research (for example, for newsroom) and showed their inter-

est in attending a future workshop with the same focus as this one.

Jisun An, Haewoon Kwak, and Fabrício Benevenuto served as cochairs of this workshop. This report was written by Jisun An. The papers from the workshop were published by AAAI Press as technical report WS-16-18, which was included in *The Workshops of the Tenth International AAAI Conference on Web and Social Media*.

Social Media in the Newsroom

The Social Media in the Newsroom workshop took place at the Maternushaus in Cologne, Germany, with a turnout of approximately 25 attendees. The workshop had as its main objective gathering researchers and practitioners with expertise and interest in the intersection of social media and journalism, as a subset of computational and data journalism. The main focus of the workshop was on the ever-growing ability of social media to capture and announce breaking news.

The participants of the workshop were keen on furthering the development of tools to facilitate access to social media for news and media professionals, as well as on gaining insights into social media's benefits from a journalistic perspective.

The workshop kicked off with an insightful keynote speech, delivered by Wilfried Runde, the head of innovation projects at Deutsche Welle. In his talk, Runde outlined the main challenges that newsrooms need to deal with when gathering news from social media. The talk largely focused on social media verification practices at Deutsche Welle, which he illustrated with numerous examples of fake images and videos that made it easier for everyone to understand the challenges. Runde also showcased some of the work Deutsche Welle has conducted within the Reveal EC-funded project, as well as some of the forthcoming plans for a new project that focuses on video verification, InVID. The audience was curious about these verification challenges and raised a number of interesting questions, including questions about the ability to perform thorough verification tasks at scale.

After the keynote talk, the workshop featured presentations from authors of five accepted full papers. Markos Zampoglou, from CERTH in Greece, presented the first paper, which was an appropriate follow-up to the keynote talk as it also focused on verification — in this case looking into the development of automated tools. His presentation included a demo that resulted from the research, called Media Verification Assistant. It allows its users to upload a picture, producing an enhanced output that includes a likely veracity value for the picture. The second paper was presented by Konstantina Papanikolaou, from the Institute for Language and Speech Processing in Greece, who introduced PALOMAR, a tool that uses

named-entity recognition and event-extraction techniques for detecting events, aggregating both social media and news media in Greek. The third paper was by Tamer Elsayed from Qatar University, who presented research on using a classifier to determine what types of questions Arab journalists make on Twitter. Elsayed and colleagues developed a scheme of seven types of questions and annotated a collection of Arabic tweets from journalists asking questions. Using a supervised classifier, the authors showed that they can accurately classify questions by type, which they plan to extend in future work with larger collections of annotated tweets. The fourth paper was presented by Laura Tolosi, from Ontotext in Bulgaria, who has been studying rumor-detection techniques within the PHEME EC-funded project. She presented a study of features that characterize rumors across different events, suggesting those that can be exploited to develop an accurate rumor-detection system. To conclude, Cornelius Puschmann, from the Alexander von Humboldt Institute for Internet and Society in Germany, presented his paper looking at the news sources of Islamophobic groups on Twitter. His research revealed findings with implications both for the study of mass media audiences through the lens of social media, and for research on the public sphere and its possible fragmentation in online discourse.

Interestingly, the audience actively engaged in all the presentations and raised questions that ignited discussion. While the half-day workshop came to an end quickly, the good turnout and active discussions posited a collective interest in following up with further workshops and activities.

The workshop was cochaired by and this report was written by Bahareh Heravi and Arkaitz Zubiaga. The papers from the workshop were published by AAAI Press as technical report WS-16-19, which was included in *The Workshops of the Tenth International AAAI Conference on Web and Social Media*.

Social Web for Environmental and Ecological Monitoring

The exponential growth of social media has provided not only novel techniques for public communication and engagement, but has also generated unprecedented volumes of publicly available, user-generated social media content. These trends open new opportunities for ecological and environmental applications, both in terms of alternative data sources and novel approaches for interacting with the public. The goal of this workshop was to bring together a combination of academic and industrial participants to discuss ideas, challenges, and solutions at the intersection of social media and environmental and ecological science.

In total, the program committee accepted nine papers of three different formats (two full papers, five

short papers, and two full papers) covering a broad range of topics, including social media data mining and analysis, social networks, citizen science, and crowdsourcing. Papers also explored engagement, motivation, and gamification techniques applied to environmental and ecological monitoring.

Robert-Jan Sips, who leads IBM's collaborative academic research programs in the Netherlands at the IBM Center for Advanced Studies (CAS), gave the keynote talk, *Listening Umbrellas, Tweeting Plants: Social Environmental Monitoring*.

The workshop concluded with a brainstorming session that engaged participants in an active discussion of the main challenges of this research field and the prospective directions for future work. Participants agreed that social media analysis applied to environmental and ecological monitoring has been underexplored, and that we should work toward building a research community around this promising area. The workshop was a good first step to creating this community, and many participants agreed that they would like to participate in future workshops with similar foci in the future.

David J. Crandall and Roman Fedorov served as cochairs of this workshop and wrote this report. The papers from the workshop were published by AAAI Press as technical report WS-16-20, which was included in *The Workshops of the Tenth International AAAI Conference on Web and Social Media*.

Jisun An is a postdoctoral researcher at Hamad Bin Khalifa University.

David J. Crandall is an associate professor at the Indiana University Bloomington, USA.

Roman Fedorov is a Ph.D. candidate at Politecnico di Milano, Italy.

Casey Fiesler is an assistant professor at the University of Colorado Boulder.

Fabio Giglietto is an assistant professor at Università di Urbino Carlo Bo, Italy.

Bahareh Heravi is a lecturer at University College Dublin.

Jessica Pater is a research scientist at the Georgia Tech Research Institute.

Konstantinos Pelechrinis is an associate professor at the University of Pittsburgh.

Daniele Quercia is a computer scientist at Bell Labs, Cambridge UK.

Katrin Weller is a senior researcher at GESIS – Leibniz Institute for the Social Sciences.

Arkaitz Zubiaga is a postdoctoral research fellow at the University of Warwick.