Reports of the Workshops Held at the 2019 International AAAI Conference on Web and Social Media

Diego Alburez-Gutierrez, Eshwar Chandrasekharan, Rumi Chunara, Sofia Gil-Clavel, Aniko Hannak, Roberto Interdonato, Kenneth Joseph, Kyriaki Kalimeri, Sanjay Kairam, Momin M. Malik, Katja Mayer, Yelena Mejova, Daniela Paolotti, Emilio Zagheni

■ *The workshop program of the Association for the Advancement of Artificial Intelligence's 13th International Conference on Web and Social Media was held at the Bavarian School of Public Policy in Munich, Germany on June 11, 2019. There were five full-day workshops, one half-day workshop, and the annual evening Science Slam in the program. The proceedings of the workshops were published in* Research Topic of the Frontiers in Big Data. *This report contains summaries of those workshops.*

Complex Systems Perspectives on Algorithmic Bias

The goal of the Complex Systems Perspectives on Algorithmic Bias workshop was to investigate how the quickly evolving field of algorithmic bias can benefit from theory and methods used in complex systems research.

From search result ranking to friend recommendation, algorithms used in today's online platforms exist within a coevolving universe of model parameters, institutional constraints, and the whims of platform users. In this workshop, we focused on the development and exposition of research that seeks to model, quantify, or theorize the complex system of algorithms, platforms, and users. Through a mix of presentations of accepted abstracts, presentations from keynote speakers, and panel discussions, the workshop brought to the forefront empirical, theoretical, and simulation-based research that can help clarify this interplay using a complex systems perspective. We took a broad view of what a complex system perspective entails, emphasizing only that there is a goal of understanding how these various pieces fit and evolve together.

 ISSN 0738-4602

The workshop brought together researchers from a variety of fields; attendants included scholars in the fields of sociology, computer science, philosophy, and law, and members of both academe and industry. Our first keynote speaker, Krishna Gummadi (head of the Networked Systems Research Group at the Max Planck Institute for Software Systems) structured his talk around algorithmically assisted decision-making in bail sentencing decisions, touching on both ethical and legal sides of automated decision-making. Following him, our first panel focused on the topic of algorithmic bias, debating the true source of bias in algorithmically aided systems and the future of algorithmic bias research.

Our second keynote speaker, Seda Gürses, assistant professor in the school of Technology, Policy, and Management at the Delft University of Technology, presented thoughts on the limits of the idea of fairness and algorithmic solutions to addressing issues of social inequality in technology. Our second panel consisted of social and data scientists who investigate how real-world problems manifest on online platforms. The speakers discussed the role of AI and data science in solving the issues around bias in online communities. Finally, our third panel consisted of social scientists, who study social behavior through simulations or experiments. The panelist argued for including social theory at early stages of any research or solution around issues of algorithmic bias. The panel and audience also discussed difficulties in the interdisciplinary communication that seems crucial to solve the problems the workshop focused on.

Kenneth Joseph and Aniko Hannak served as cochairs of this workshop.

Data for the Well-Being of Most Vulnerable

The Data for the Well-Being of Most Vulnerable workshop was held to highlight and brainstorm the latest developments in the use of new sources of data, including web and social media, in efforts to address the health and other needs of vulnerable populations, such as children, families, marginalized groups, and those at the threat of poverty, conflict, natural disaster, or epidemic risk.

The scale, reach, and real-time nature of the Internet is opening new frontiers for understanding the vulnerabilities in our societies, including inequalities and fragility in the face of a changing world. Thus, the aim of this workshop was to encourage the community to use new sources of data to study the well-being of vulnerable populations including children, the elderly, racial or ethnic minorities, the socioeconomically disadvantaged, the underinsured, or those with certain medical conditions. The selection of appropriate data sources, identification of vulnerable groups, and ethical considerations in the subsequent analysis are of great importance in the extension the benefits of big data revolution bring to these populations. As such, the topic is highly multidisciplinary, and the workshop attracted a diverse audience, including academia across computer science, medicine, and social sciences, as well as representation from industry (Twitter and Microsoft), and government.

Elad Yom-Tov, a principal researcher at Microsoft Research and the workshop's invited keynote speaker, set the tone for the conference by bridging the divide between large-scale data availability and research of vulnerable populations, and emphasizing the measures that need to be taken beyond institutional review board approval to make sure to do no harm. In particular, he focused on vulnerable populations in the United States, and topics of importance in health. Further contributed talks spanned a wide variety of topics, including assessment of ailments around the world, emotional health of students, and well-being perception of populations in dangerous areas, using a variety of sources including microblogging and networking services, as well as mainstream news media. Looking at a longer research view, Latifa F. Jackson and Caitlin Kuhlman of Howard University and Worcester Polytechnic Institute, US, presented efforts they and others have made to include under-represented and vulnerable populations in data science— both datasets and information analytics. In particular, they touched on the inclusion of minority groups in DNA mapping and engaging with Pacific Islander populations in the study of obesity. They also described data science bias training to raise awareness of data bias that could result in algorithmic bias.

An additional talk was given by Mo Chen, a researcher from the local Technical University of Munich, who has given an overview of the Chinese social credit system, illustrating the kinds of red- and blacklists that list people displaying favorable or unfavorable behaviors, respectively. The presentation sparked an active discussion about data availability, how the credit score is computed and to whom it is available, and what its uses may be.

A closing discussion among the participants raised insights that there is a wide interest in the topic of vulnerable populations, there are many possible facets in defining vulnerability, and social media is a potentially valuable resource for tracking and addressing

such properties. A suggestion for the next event was brought up to include an incentive to either make datasets available or encourage participants to declare willingness to collaborate in the future.

Yelena Mejova, Rumi Chunara, Kyriaki Kalimeri, and Daniela Paolotti served as co-chairs of this workshop.

Managing and Designing for Norms in Online Communities

Community norms describe the implicit rules that govern the behavior of participants within a community and arise from complex interactions between stakeholders. Studying, managing, and designing for online communities requires understanding which types of norms are present, which social and technical interventions are available to influence these norms, and when these interventions may be successfully applied. This workshop brought together an interdisciplinary group of researchers and practitioners at the International Conference on Web and Social Media, to identify practical challenges with establishing community norms and to explore how these challenges might be addressed through design.

In the morning session, workshop participants broke into two groups, each focusing on one set of challenges that communities might experience when designing for certain norms. Using the how-might-we ideation approach, each group focused on expressing their challenge as a design question, identifying relevant context and constraints, and developing prototype research or design ideas. The first group explored strategies for detecting and measuring at-scale norms and their propagation through a community. The second group focused on the challenge of defining positive norms; although many communities may agree about definitions of negative behaviors (for example, harassment and abuse), these communities may differ in how they define good behavior. A key takeaway from these discussions was that while existing conversations around online governance tend to gravitate toward the detection and prevention of bad behavior, we can't holistically improve community health by focusing only on the bad. As a first step toward efforts to promote good online behavior, our solution involved providing a blueprint for articulating the values and opportunities an online community provides: Instead of a negative articulation — an ever-growing list of prohibited behaviors — we should use a positive phrasing (for example, you-will-be-able-to) of capabilities in an online community.

In the afternoon session, workshop participants engaged in a role-playing exercise concerning a fictional online social platform. In this exercise, participants considered how we might define community health from the perspectives of various potential stakeholders, including platform owners, moderators, end-users, and third-party developers and researchers. The discussion raised interesting questions about how community health metrics might intersect with the values or labor expectations for each group of stakeholders. For example, from the platform owners' perspective, participants explored the importance of meaningful engagement and sustainability (that is, a community that maintains and propagates itself) to be important markers of community health. Another aspect that participants raised was the concept of symbiosis: How do all the actions of stakeholders meet? How do we avoid meeting the needs of some and not the others? The discussion around symbiosis revolved around how platforms should work to balance the needs of different entities they cater to. In addition to discussions around the need to balance the expectations and needs of different stakeholders, participants debated about who gets to define what types of online behavior are considered to be appropriate, and ways to distinguish good from bad behavior.

The discussion concluded with how potential definitions for community health might handle specific scenarios, such as in specialized knowledge communities, which do not expect to cater to all potential participants, or communities that are subject to coordinated misinformation attacks. One challenge that the participants noticed was the lack of inputs from other stakeholders involved in the process, in particular the absence of inputs from policymakers and researchers. This lack of discussion among entities became especially problematic when dealing with challenges like misinformation. Misinformation is not only a platform-level problem, but also a societal problem. The discussions highlighted the importance of including the voices and perspectives of diverse individuals and different stakeholders during the process of defining platform policies to deal with social problems.

Sanjay Kairam, Eshwar Chandrasekharan, Stevie Chancellor, and Joseph Seering served as co-chairs for this workshop.

Demographic Research with Web and Social Media Data

The spread of the internet and online communities provide unprecedented opportunities for studying global population dynamics such as fertility, mortality, migration, and health. Internet users do not only leave digital traces of their existence — the online world also influences their behavior, from daily decisions (like commuting patterns and kin interactions) to major life events (like childbirth and migration). Computational social scientists have long used innovative methodologies and data sources to study social phenomena but demography has been slow in incorporating digital approaches, in spite of being a historically data-driven discipline. Demographers have only recently moved into the digital realm, but collaboration between demographers and computationally oriented scientists remains limited. The Demographic Research with Web and Social Media Data workshop aimed to foster dialogue and knowledge exchange between these two communities

by focusing on the applications and implications of web and social media data for demographic research.

The workshop brought together researchers from different disciplines studying digital data for different purposes. Workshop participants interacted with each other through a combination of presentations and interactive activities.

This workshop was organized by the Max Planck Institute for Demographic Research in collaboration with the scientific panel on digital demography of the International Union for the Scientific Study of Population. It was the fourth workshop held at the International Conference on Web and Social Media during the last 4 years revolving around the theme of demography and social media. The continuous presence of demographers at the International Conference on Web and Social Media is a sign of the importance of social media studies for the demographic community as well as the commitment of demographers to advance research on web and social media and their societal implications.

Presentations were either short (10 minutes for presenting plus 5 minutes of discussion) or long (15 minutes for presenting plus 5 minutes for discussion). This maximized the exposure to a range of topics in digital trace research. It also increases the exchange of ideas between different research fields. Keynote speeches were delivered by Yelena Mejova (research leader at the Institute for Scientific Interchange in Turin, Italy) and Maria Sironi (associate professor in the Department of Social Science at University College London).

The workshop was also interactive and fun, thanks to a series of group activities. These included speed dating for ideas (an ice-breaker); online data access, privacy, and use (a series of minipanels); name-dropping (paper titles for potential collaborations); and working with your academic match (a closing activity).

An online evaluation conducted after the workshop showed that participants found the event enjoyable and helpful for networking. In particular, they appreciated the interactive activities but would have preferred fewer presentations. These recommendations will be considered for a future edition of the workshop.

Sofia Gil-Clavel, Diego Alburez-Gutierrez, and Emilio Zagheni were the co-chairs of the workshop.

Modeling and Mining Social-Media-Driven Complex Networks

The growing availability of multifaceted social media data gives rise to unprecedented opportunities for unveiling complex real-world online behaviors. This also supports the proliferation of complex network models where the expressive power of the graph-based relational structure is enhanced through exposing several types of features that are peculiar to the social media platforms.

The aim of the Modeling and Mining Social-Media-Driven Complex Networks workshop was to get an insight on current trends regarding both modeling and mining aspects concerning any type of data-driven network that can be inferred from social media data contexts, such as heterogeneous, multilayer, temporal, location-aware, and probabilistic networks.

The workshop consisted of four sessions spanning a full day, and included nine presentations of accepted papers and an invited talk. During this time, researchers and practitioners discussed the relation between social media and complex networks from several points of view, including both data-driven modeling (for example, modeling of location-based, mobility-based, and topic-based networks) and tasks focusing on the analysis and manipulation of network structures (for example, network anonymization, link prediction, influence, and information spread).

Several discussions brought up the possibility of inferring the topology of a network from features other than the presence of explicit relations between entities, for example, geolocation of tweets, mobility of users, and topical information. This indicated that while in most cases it may be intuitive to represent social media data by the use of features that natively define a topology among entities (for example, friendships, interactions, ratings, and so on), in several contexts it may be more fruitful to remodel information network data on data-driven features that characterize latent relations between entities tailored to a given problem.

Another recurring discussion was over how to conveniently exploit textual content issued by social media. It has been shown how textual content can be exploited to define topic-driven interactions and communities that can be used to characterize a specific phenomenon (for example, the political scenario of a country), but also how graph structures contained in text flows (such as chat logs) can be leveraged to characterize the text itself (for example, by identifying abusive language).

Misinformation spread, a major research topic in recent years, has continued to be an interesting research avenue in social media analysis. This is not surprising, given the impact that this phenomenon has had in major events like political elections.

In recognition of this topic, the workshop hosted an invited talk by Robert West (Ecole Polytechnique Fédérale de Lausanne) on Message Distortion in Information Cascades. During his talk West discussed how information diffusion processes based on iterative summarization can strongly alter the original content of a message (for example, a scientific article), even in the absence of a malicious intent.

The workshop showed how such tools are extremely flexible and key-enabling for modeling and analyzing phenomena in different domains. At the same time, the discussions brought to light how coping with social-media-driven complex networks can be challenging, due to the high heterogeneity of the data and the absence of a standard way to process them.

The general consensus of participants was that, due to different factors such as the increasing pervasiveness of social media, and the availability of advanced tools and methods to model and analyze complex networks, modeling and mining social-media-driven complex networks is a continuously evolving subject that deserved continued discussion and research.

Roberto Interdonato (The French Agricultural Research Centre for International Development [CIRAD]), Sabrina Gaito (Università degli Studi di Milano), Alessandra Sala (Nokia Bell Labs), and Andrea Tagarelli (University of Calabria) served as co-chairs of the workshop.

Critical Data Science

Critical data science is defined as the practice of working with and modeling data (the data science), combined with identifying and questioning the core assumptions that drive that practice (the critical aspect). It can be regarded as the intersection (or perhaps the union) of data science and critical data or algorithm studies, and an example of a critical technical practice. The objective of the workshop was to create a prime space where scholars from different backgrounds can meet and discuss how to do data science in responsible, sustainable, and interdisciplinary settings.

The workshop included short presentations by participants to support reflection of their own and neighboring scientific practices, and to create opportunities for further cooperation. Participants covered a broad range of backgrounds: industry data science and engineering, computer science, computational social science, linguistics, classics, environmental and human rights activism, social work, digital democracy, and the arts. The workshop was guided by two blocks of questions: politics and practice.

For politics, the questions included: What are our experiences of paradigmatic politics? Who are the insiders, and who are the outsiders for effecting change? Do we feel capable of intervening in curricular decision-making, and can we disrupt dominant narratives of big data hegemony, efficiency, and objectivity? What does it mean to do data science for good, and for whom? What would be my personal priorities: short-term, and long-term?

For practice, the questions included: What concrete actions can we take? How can we create spaces and time for collaboration besides always-hectic, project-based logics? Which incentive and reward structures would we need for that? Which skills do we want to establish in the training of the next generation? How can I/we collaborate? With whom? For what tasks?

Workshop presentations and discussions both delved into how we can change our socio-technical practices. When several computer scientists remembered specific biographical aspects or experiences that led them, unlike many technical practitioners, to be open to nontechnical perspectives, we learned that critical technical practice requires deeply personal involvement with scientific routines. For many, it was personal connections or commitments to political projects that led them to start questioning claims of objectivity, neutrality, and universality, with these aspects commonly leaving little room for reflection. On the other hand, from the broad range of presentations we learned that hybrid approaches — including social scientific analysis, as well as elements of engagement of participation — are indeed possible. Furthermore, there are already communities growing like the Association for Computing Machinery FAT* conference, or critical data and algorithm studies, surveillance studies, science and technology studies, human interface design and data activism; and a number of projects presented at the workshop that remain technical products or analyses perfectly capture these approaches, because they are already overcoming the under-theorized pragmatism that drives so much of software engineering and data analysis today.

The workshop concluded with a set of ideas and priorities on how to design critical technical practice in data science. The ideas and priorities included systematic reflection, participatory action research, team building, publishing venues, linking sectors, education, institution building, documentation, ethical principles, and funding. Katja Mayer and Momin M. Malik served as co-chairs of this workshop.

Diego Alburez-Gutierrez is a research scientist at the Max Planck Institute for Demographic Research.

Eshwar Chandrasekharan is a PhD candidate in computer science at the Georgia Institute of Technology.

Rumi Chunara is an assistant professor at New York University.

Sofia Gil-Clavel is a PhD student at the Max Planck Institute for Demographic Research.

Aniko Hannak is an assistant professor at the Vienna University of Economics and Business.

Roberto Interdonato is a research scientist at The French Agricultural Research Centre for International Development (CIRAD), Unit of Mixed Research TETIS, Montpellier, France.

Kenneth Joseph is an assistant professor at the University at Buffalo, State University of New York.

Kyriaki Kalimeri is a researcher at the Institute for Scientific Interchange (ISI), Turin, Italy.

Sanjay Kairam is a research scientist at Twitch Interactive.

Momin M. Malik is a postdoctoral fellow at Harvard University.

Katja Mayer is the Elise Richter fellow at the University of Vienna.

Yelena Mejova is a research leader at the Institute for Scientific Interchange (ISI), Turin, Italy.

Daniela Paolotti is a research leader at the Institute for Scientific Interchange (ISI), Turin, Italy.

Emilio Zagheni is director of the Max Planck Institute for Demographic Research.