

# Reports of the 2015 Workshops Held at the International AAAI Conference on Web and Social Media

*David Garcia, Germaine Halegoua, Yelena Mejova, Nicola Perra,  
Jürgen Pfeffer, Derek Ruths, Ingmar Weber, Robert West, Leila Zia*

■ *The 2015 workshops at the International AAAI Conference on Web and Social Media were held on May 26 in Oxford, UK. The workshop program included seven workshops, including Auditing Algorithms from the Outside: Methods and Implications; Digital Placemaking: Augmenting Physical Places with Contextual Social Data; Modeling and Mining Temporal Interactions; Religion on Social Media; Standards and Practices in Large-Scale Social Media Research; the ICWSM Science Slam; and Wikipedia, a Social Media: Research Challenges and Opportunities. This article contains the written reports of six of the workshops.*

## Auditing Algorithms from the Outside: Methods and Implications

The Auditing Algorithms from the Outside workshop was organized by Mike Ananny (University of Southern California), Karrie Karahalios (University of Illinois in Urbana-Champaign), Christian Sandvig (University of Michigan), and Christo Wilson (Northeastern University). The organizers did not submit a report to *AI Magazine*. No technical report was issued.

## Digital Placemaking: Augmenting Physical Places with Contextual Social Data

People have embraced social media as a means to express their experiences within and knowledge about particular places, and researchers have continued to analyze these digital traces in order to better understand social activities within particular places. Geotagged social media data such as photos, tweets, check-ins, audio, video, and status updates have proliferated and reveal individual and collective senses of place and local insights into interactions between people and place. However, these digital traces alone cannot reveal a holistic sense of place and placemaking. The workshop aimed to investigate definitions of placemaking — commonly understood as a collaborative, community, or human-centered approach to the planning, design, social production of public spaces in order to cultivate shared value and recognize specific physical, cultural, and social identities and experiences in a particular place — and how placemaking intersects with digital and social media. One of the key goals of the workshop was to create a hands-on experience and encour-

age participants from a variety of backgrounds to discover, share, and interact with experimental techniques of urban sensing and data gathering about place and location.

The workshop brought together industry professionals and academics from computer science, social science, urban planning, and the humanities in a discussion of how placemaking intersects with social and mobile media. The workshop was divided into two paper presentation and discussion sessions and an afternoon data gathering and analysis activity. During the afternoon activity, participants were equipped with Empatica wristbands designed to capture physiological signals in real time and were asked to document their journeys through different neighborhoods in Oxford. Each group of participants analyzed the experiential and physiological data collected alongside social media data sets and presented preliminary conclusions about the locations they visited and the processes of placemaking within these locations.

A major theme that emerged from the workshop was a discussion of the potential value in combining multiple methods of data gathering and analysis in the study of place and placemaking. The papers and activities during the workshop illustrated how methods and techniques such as ethnography, elicitation techniques, soundwalks and physical/digital tours, audio recording, cartography, and geocoded physiological sensing could be paired with computational and big data analysis in order to gain a more robust sense of the social production and experience of places. Additionally, the workshop papers and activities generated discussion about mobility and time in relationship to space, place, and colocation. For example, papers such as “Digital Rendezvous: Experiences of Students, Staff, and Alumni on a Campus Memory Map” dealt with mobility and time in terms of using digital media to aggregate and display small histories of place by transient campus populations, while “Tweeting Behavior during Train Disruptions within a City” addressed how mobility and disruptions in mobility are expressed through Twitter and can be understood in terms of time. Another theme presented in several workshop papers was the context of play and storytelling in studies and practices of placemaking. Participants discussed and questioned the ways in which playfulness, memory, pervasive games, and narratives could be paired with or considered alongside social media and other data sets in informative ways. Relatedly, these themes lead to discussions about in what ways digital information and digital activities can not only aid researchers in investigating questions about space and place, but also help people gain and share spatial knowledge and become more active in nurturing places and neighborhoods.

The Digital Placemaking workshop was organized by Germaine Halegoua (University of Kansas), Raz

Schwartz (Facebook), and Ed Manley (University College London), with on-site technical assistance from Panagiotis Mavros (University College London). This report was written by Germaine Halegoua. The papers of the workshop were published as AAAI Press Technical Report WS-15-16.

## Modeling and Mining Temporal Interactions

The emergence of the big data paradigm together with the framework of complex networks has played a crucial role in providing the tools and data sets to begin understanding human interactions and the dynamics of social systems with wide applications to informatics, social sciences, information technology, and epidemiology.

The rise of the social web has resulted in the creation of an unprecedented number of social systems than can be analyzed in great detail. This unique opportunity has fostered an interdisciplinary effort to study their structure and dynamics with network theory taking the forefront. As a result, researchers have unveiled a number of surprising properties about the structure and dynamics of large-scale social systems such as online social networks, scientific collaboration networks, or mobile phone communication networks.

In this arena, the mining and study of the temporal characteristics of social dynamics, raises new fundamental challenges, both theoretical and computational, with applications to fields such as social sciences, marketing, computer science, and epidemiology. In particular, new tools and frameworks are needed to mine, characterize, and model temporal behaviors as well as the profound consequences that individual characteristics and behavioral correlations have on processes under study.

The workshop aimed at creating a fertile ground for discussing a wide variety of approaches used to study the temporal evolution of online and offline social systems and user behavior. In particular, the workshop focus was to stimulate the cross-pollination between novel data-mining techniques, used to effectively collect, represent, and explore large-scale time-resolved data sets, and new modeling approaches, aimed at understanding and reproducing their relevant features.

The event has been characterized by an interesting set of talks on the development of methods to test different hypotheses about human behavior on the web, characterize collective attention on Twitter, model bursty conversations on social networks, anticipate spikes in social media, and maximize the information value of social media. Furthermore, the workshop hosted two invited speakers: Bruno Ribeiro (Carnegie Mellon University) and Suzy Moat (Warwick University). They gave extremely interesting talks discussing fundamental aspects of big data

analysis and cutting edge methods to study, model, and predict human behavior. The set of speakers and the audience was an interdisciplinary mix of computer scientists, social scientists, and physicists. This was instrumental to spark interesting and productive discussions throughout the event.

Bruno Gonçalves (Aix Marseille), Márton Karsai (ENS Lyon), and Nicola Perra (Northeastern University) organized the event. This report was written by Nicola Perra. The papers of the workshop were published as AAAI Press Technical Report WS-15-17.

## Religion on Social Media

The focus of this first Religion on Social Media workshop was the study of religious communities in social media, providing a venue for social and religious studies scholars to interact with researchers employing web mining, information extraction, network science, and other computational methods. The aim of this workshop was to outline the research questions, methodologies, and collaborations to foster the use of computational methods — especially applied to social media — and to support the study of religious and cultural phenomena. To facilitate knowledge exchange and interaction, the workshop included a variety of activities, including invited talks, poster session, and brainstorming interactive activities.

Religion is one of the major forces in both personal and social life of a great number of people across the world. In a 2012 Gallup study, 59 percent of the sampled world's population said that they think of themselves as religious. As any cultural force of this scale, it has found expression in various forms on social media. Online religious communities may reflect established cultural norms and organizations, while simultaneously presenting a new venue for religious expression and networking. We began discussing these topics with three invited speakers from a wide variety of areas of expertise.

The first speaker was Robin Dunbar of Oxford University, whose work in anthropology and evolutionary psychology has illuminated the cognitive constraints on the quality and quantity of social interactions we are capable of maintaining. In particular, the evolution of religion, he posits, reflects the size of the settlements and their need for cohesion. Doctrinal religions with “high” gods are likely to evolve in large settlements, but are then threatened by cults and fragmentation. Online data shows that the evolutionary constraints on the sizes of our social circles remain despite the promise of higher connectedness through social media. However, the rise of cults around charismatic leaders may be facilitated by the new technology, though these leaders may find it difficult to control their followers.

The issue of collecting social media and Internet data was then discussed by Peter Webster, a historian of contemporary British religious history with partic-

ular interests in the digital turn for history and the archived web. He explained the challenges of archiving social media data, including the fast rate of decay of such content, which disappears quickly. In several case studies, Webster shows the rising engagement with social media, and early adoption of social media by religious organizations.

Finally, the third speaker, Yasmine Hafiz, presented a view from the news media's side, speaking about her experience as an associate editor of Huffington Post Religion. Using real-world examples, she outlined the major trends in the study of religion on social media, including viral multimedia content, instances of social media solidarity, the use of technology to counter stereotypes, and the direct digital transmission of religion by influencers. The challenges such content presents to a news media are many, including the decision on whether to propagate religious material by publishing it, checking for own bias, and determining spokespeople for religious organizations.

The interactive portion of the workshop began with an icebreaking exercise in which all participants paired up such that their expertise was complementary (with color-coded stickers on their name tags signifying expertise in religion, social, or computer sciences). The pairs then brainstormed ideas for hypothetical collaboration. The titles of such works ranged from specific groups, such as women in Egypt, to overarching analyses of religious dynamics of local areas, to delving into particular structures of the religious groups and the development of their hierarchies.

More developed projects were then to follow, with authors presenting lightning talks and posters on their studies. Data-driven studies proposed to use it for insights which can be applied to policy and decision making, such as the detection of cyber hate speech. Others proposed to augment current surveys on religious groups, such as the Guttman Report on the Identity and Beliefs of Jewish Israelis, or demographic segments, such as young adults discussing Christianity but not actively involved in a church, by quantitative and qualitative analysis or social media data such as Instagram, Twitter, and Facebook, and incorporating insights from Google search trends and Wikipedia article edit history. Methodologies from social studies were also introduced, including “auto-netnography” — an ethnographic method to study online and Internet-based materials. Finally, both worlds converged in a mobile application where users can read devotional texts and prayers from multiple churches. It provides a view of the private religious and topical preferences, and possible crossovers from interest in one church to others.

To elaborate on the ideas presented by the keynote speakers and discussed during the poster session, an extended brainstorming session concluded the work-

shop. During this session, much like in the icebreaking exercise, groups were brainstorming an idea for a paper. Five ideas were fleshed out, including the title and brief description of methods and data necessary.

Overall, we were pleased with the level of engagement of the participants and by how smoothly both the icebreaking activity and the later activities went. Given the feedback we received and our own impression we would largely keep the format for future editions. However, we would slightly cut down the later brainstorming session in favor of more time for the icebreaking exercise as these relatively short one-on-one interactions led to more mixing than the more in-depth and longer brainstorming sessions.

Yelena Mejova (Qatar Computing Research Institute), Ingmar Weber (Qatar Computing Research Institute), Nathan Matias (Massachusetts Institute of Technology), Lu Chen (Wright State University), and Adam Okulicz-Kozaryn (Rutgers University) organized the workshop. This report was written by Yelena Mejova and Ingmar Weber. No technical report was published.

## Standards and Practices in Large-Scale Social Media Research

Recent years have witnessed numerous instances in which previously established claims, results, and findings have not held up to various forms of external verification. The increasing attention social media mining methods receive in industry and government adds urgency to the need for social media (and, more broadly, social informatics) research to be open, attentive to empirical sources of bias, and reproducible. The intent of this workshop was to identify the obstacles to achieving these goals and frame constructive recommendations for the ICWSM and broader social media research community to take.

The workshop was organized into three sections, which addressed data sharing, bias, and reproducibility. Each section featured the presentation of peer-reviewed papers followed by extended small-group discussion periods.

The outcome of the workshop was a set of recommendations that reflected the day's discussion. Participants recommended that ICWSM institutionally adopt a data-sharing platform and provide a guidelines document explaining best practices for data sharing, the handling of bias, and approaches to ensuring reproducible research. In order to bring more focus to the problem of availability of data, a data challenge was suggested as well as changes to the reviewer form that would require reviewer comments on data availability and quality. The length of conference papers was agreed to be an issue, leading to the suggestion that ICWSM (and other conferences) support the inclusion of supplementary information with submissions. Finally, the participants

agreed that a recurring workshop on the topic of reproducible science would help continue advancing the principles around which this year's workshop was organized.

The Standards and Practices in Large-Scale Social Media Research workshop was organized by Derek Ruths (McGill University) and Juergen Pfeffer (Carnegie Mellon University) who also wrote this report. The papers of the workshop were published as AAAI Press Technical Report WS-15-18.

## The First ICWSM Science Slam

One of the main purposes of the International Conference on Web and Social Media (ICWSM) is to connect computer scientists with social scientists, providing a venue where collaboration across disciplines is encouraged. But arguably the biggest challenge of interdisciplinary science is communication: Collaborations across disciplines often face the barrier of using different terminologies and methods, which can lead to misunderstandings and frustration. Working in a multidisciplinary environment requires new approaches to overcome the cultural and historical barriers between scientific disciplines, sometimes in a pretty unorthodox way.

This year, on the evening before the main conference started, we hosted the first ICWSM Science Slam at the Oxford Old Fire Station. The format of the event was designed to engage the audience with 10 talks, each up to 10 minutes long. Each participant, called a "slammer," presented his or her latest research in a fun and engaging environment. The event was fully booked and the venue filled up before the first slammers started. The audience voted for their favorite talks based on the intensity of their applause and cheering after the talks. The only rules slammers had to observe were (1) a time limitation of 10 minutes and (2) a topical fit with ICWSM. As a result, there were rapping videos, viral cats pictures, "mean" tweets, and even jokes about the organizers.

To encourage communication across disciplines, we gave a special touch to the event by giving a special topic prize to the best slam related to the topics of inequality, discrimination, and the digital divide. While questions related to inequality and divides are pervasive in the social sciences the topic of measuring discrimination with respect to gender, income, or ethnicity is a relatively new topic in computational social science. Few participants qualified for the special topics prize, but the winner was Jahna Otterbacher from the Open University of Cyprus with her talk on stereotyping through linguistic bias. Otterbacher wonderfully explained the complex message of how in- and out-group biases manifest in language, turning some technical difficulties in her talk into comic examples that carried the message to the audience.

The winner chosen by the audience was Robert

West from Stanford University who presented his work on destroying jokes and the related website unfun.me. In West's entertaining presentation, we learned about the asymmetry of comedy, which makes destroying jokes much easier than creating them. Through data produced in a crowd-sourced platform, West illustrated how sarcastic headlines of satiric news websites can be "unfunned," and the properties of humor that we learn through that process. The response of the audience was unanimous and cheers for him were almost 10dB louder than for his closest competitor!

The engagement of the audience and the speakers in this first ICWSM Science Slam are a positive example of how we can create new institutions in emerging fields like computational social science. In the days following the event, several participants and attendants expressed their wish to attend again next year at ICWSM in Cologne.

Going forward, we plan to largely keep the successful format but try to find a bigger venue. Due to the uncertainty of the popularity of the event we settled for a smaller venue this year but this meant that we could not open up attendance to the general public as we had originally planned. In the future, with a bigger venue, we will advertise locally to give interested residents a chance to attend.

David Garcia, Ingmar Weber, and Aniko Hannak served as cochairs of the ICWSM Science Slam. Taha Yasseri and Jonathan Bright served as local organizers of the workshop. This report was written by David Garcia and Ingmar Weber. The workshop did not issue a technical report.

## Wikipedia, a Social Pedia: Research Challenges and Opportunities

Wikipedia is one of the most popular sites on the web, a main source of knowledge for a large fraction of Internet users, and one of very few projects that make not only their content but also many activity logs available to the public. For these reasons, Wikipedia has become an important object of study for researchers across many sub-

fields of the computational and social sciences, and the goal of this workshop was to create a forum for researchers from all these fields. The workshop succeeded in attracting attendees from across a wide variety of areas, including human-computer interaction, natural language processing, data mining, social network analysis, machine learning, medical informatics, cognitive science, and education.

A unique feature of this workshop was the involvement of the Wikimedia Foundation through a cochair and an invited speaker. In his talk, Dario Taraborelli from the Wikimedia Foundation proposed a research agenda for the Wikimedia community motivated by challenges the project has been facing due to shifts in how the site is used (with increased access from mobile devices, among other factors). Some papers presented by other participants have shed further light on Wikipedia usage in terms of access patterns and pageview statistics.

A second major theme was the cultural diversity inherent in Wikipedia's multilingual nature. Several papers investigated differences and commonalities across language versions in terms of content coverage, with the clear result that the different editions vary widely in the content they focus on. This point was made particularly clear in a talk by Brent Hecht.

The largest portion of Wikipedia content is the natural language text making up its articles, but for many types of analysis and applications it is helpful to have access to a database of facts expressed in tabular form. One of the goals of the relatively new Wikidata project is to provide such a database of Wikipedia facts. Speaker Magnus Manske gave an introduction to Wikidata and showed and demonstrated several applications that leverage it.

Further main topics of interest were the growth and evolution of Wikipedia in terms of both its content and the policies that govern it. The latter aspect was elucidated by Brian Keegan in his invited talk. Other participants presented studies of different aspects of Wikipedia's social dynamics, for which the data from conversations on article and user talk pages consti-

tute an excellent source of information. Cristian Danescu-Niculescu-Mizil exemplified this fact in his talk by showing that Wikipedia discussions in combination with techniques from natural language processing may be leveraged to better understand human dynamics of power, coordination, and politeness.

The workshop concluded with a brainstorming session where participants split into three groups to discuss some of the prominent topics and questions that had emerged from the talks and posters presented earlier. Each group concluded by presenting their main findings to the remaining participants. The final plenary discussion resulting spontaneously from these presentations was inspiring and useful in establishing a direct dialogue among Wikipedia researchers as well as with the Wikimedia Foundation.

Robert West, Leila Zia, and Jure Leskovec served as cochairs of this workshop. This report was written by Robert West, Leila Zia. The papers of the workshop were published as AAAI Press Technical Report WS-15-19.

**David Garcia** is a scientist at ETH Zürich.

**Germaine Halegoua** is an assistant professor at the University of Kansas.

**Nicola Perra** is an associate research scientist at Northeastern University.

**Jürgen Pfeffer** is an assistant professor in the Department of Computer Science, Carnegie Mellon University.

**Yelena Mejova** is a scientist in social computing at the Qatar Computing Research Institute.

**Derek Ruths** is an associate professor at the School of Computer Science, McGill University.

**Ingmar Weber** is a senior scientist in social computing at the Qatar Computing Research Institute.

**Robert West** is a Ph.D. candidate in the Computer Science Department at Stanford University.

**Leila Zia** is a research scientist at the Wikimedia Foundation in San Francisco, California.