# Computational Social Science for the World Wide Web

Strohmaier, Markus; Wagner, Claudia

This is the postprint of the article originally published in
IEEE Intelligent Systems, Vol. 29, 2014, No. 5, pp. 84-88
available online at: https://doi.org/10.1109/MIS.2014.80

# Computational Social Science for the World Wide Web

**Markus Strohmaier and Claudia Wagner,** University of Koblenz-Landau and GESIS-Leibniz Institute for the Social Sciences, Germany

Computational social science develops computational methods and studies new kinds of social data to advance our theoretical and empirical understanding of human social behavior. By combining methods from both computer- and social sciences, the field pursues a series of novel opportunities that arise when human social behavior and computation become increasingly entangled.

With the inception of the World Wide Web 25 years ago, the way humans interact with each other socially has evolved from the physical realm to the virtual, from a local scale to a global scale, from an individual to a network level, and from the ephemeral to the persistent. Today, the World Wide Web represents not only an increasingly useful reflection of human social behavior, but everyday social interactions on the Web are increasingly mediated and shaped by algorithms and computational methods in general. This is the case with systems where, for example, knowledge about human social behavior is used to recommend contacts (such as Facebook friend suggestions), to recommend products (Amazon recommendations), or to filter and retrieve content (Digg.com or Google). Such systems are sometimes referred to as social- computational systems[1-3] or social machines[4] whose emergent properties are co-determined by the social behavior of their users and by algorithmic computation of machines.[5]

What distinguishes social-computational systems from other types of software systems – such as software for cars or airplanes – is the unprecedented need to understand and model human social behavior on a very large scale (millions of users). Without a thorough understanding of human social behavior in these systems, engineers will fail to satisfice non-functional requirements such as sociability,[6] navigability,[7] and other socially co-determined requirements. This makes social-computational systems a novel class of software systems, and unique in a sense that potentially essential system properties and functions are dynamically influenced by the social behavior of a large number of users. Due to its communicative nature, the World Wide Web is particularly conducive to such systems.

Any effective approach to engineering largescale applications on the Web therefore requires a deeper – both theoretical and empirical – understanding of human social behavior, and corresponding ways of modeling it. Yet, current approaches in understanding human social behavior on the World Wide Web are limited, often application driven, and rarely informed by the rich body of existing social science research literature and theories.

## Computational Social Science

Computational social scientists are usually interested in developing computational methods that enable studying how people think/feel/behave in social situations (social psychology), relate to each other (sociology), govern themselves (political science), handle wealth (socioeconomics), and create culture (anthropology). Computational social science[8] can therefore equip engineers of social-computational systems on the World Wide Web with models, methods, and techniques for understanding human social behavior, and provide a basis for engineering systems that establish system properties such as sociability and others.

In this article, we want to introduce the field of computational social science to the intelligent systems community and discuss how this field can help to advance the current state of understanding and engineering social-computational systems on the World Wide Web.

Overall, this article makes an argument that computational social science offers a unique range of challenges as well as methods and techniques that can help understand and engineer systems on the World Wide Web.

## Traces of Human Social Behavior

Historically, the social sciences have pursued a variety of approaches to understand human social behavior. As early as the 1960s, a distinction between unobtrusive and obtrusive research methods has been made.[9] Unobtrusive (sometimes also called non-reactive) research methods in the social sciences refer to methods where the researcher doesn't intrude into the research context. For example, unobtrusive research methods help sociologists learn something about human beings and the social systems they inhabit without interrupting them by asking questions (questionnaires) or directly observing their behavior (ethnography). Traditionally, unobtrusive research methods relied on physical objects that are capable of representing traces of human activity: the wear of floor tiles around museum exhibits as indicators of popular exhibits; the setting of car radio dials as indicators of favorite stations; the wear on library books and rub and fold marks in their pages.[9] These physical traces can be associated with accretion measures (the build-up of physical traces) and erosion measure (the wearing away of materials). In general, this data has sometimes been referred to as found data,[10] process data, or organic data,[11] emphasizing that scientists have little control over the data generation process.

### Advantages of Found Data

An advantage of unobtrusive research methods is that they allow us to study evidence of actual human social behavior rather than self-reported ones. For example, by studying used containers of alcoholic beverages in a community, one can get a more comprehensive picture of alcohol consumption than from surveys, which might suffer from underreporting. Thereby, we can learn about what people do rather than what they say they do. Unobtrusive methods can also provide a signal where surveys aren't economical or reasonable to obtain. For example, they can provide information on a global scale in near real-time by analyzing social media messages in case of events or disasters, and they can be obtained without burdening subjects. Found data might also help to address the well-studied Hawthorne effect[12] – where people adapt their behavior in response to being directly observed, for example, by trying to conform with perceived social norms and expectations. In addition, found data can alleviate problems of traditional research methods such as non-response biases in surveys. In that sense, found data on the World Wide Web has the potential to not only inform us about user behavior online, but about human social behavior in general (offline).

### Challenges of Found Data

Social scientists have historically also pointed out the many challenges and limitations of unobtrusive methods. For example, internal states of humans such as motivations or concerns can't be easily studied because they may not always cause observable outcomes, or confounding factors may distort dependent variables. Found data often represents a conservative estimate that takes time to accumulate and is inferentially weak. Unknown data aggregation, biases, and privacy concerns represent other common problems of such research methods. Found data on the World Wide Web is particularly prone to these limitations and pitfalls.

Found Data on the Web

Digital traces can also be divided into accretion traces (the buildup of digital traces – for example, user logs, Facebook “likes,” or the collection of tweets produced by a user) and erosion traces (the wearing away of digital traces – for example, removal of content in Wikipedia articles, deletion of Web content, or unfollowing behavior on Twitter). These digital traces provide similar opportunities and limitations to the physical traces studied by social scientists historically. Applying the rigor of social science research to the study of human social behavior via the World Wide Web is likely to yield fruitful insights for engineering social-computational systems.

**Examples**

Understanding the opportunities, challenges, and limitations that come with such research methods is an important prerequisite for computational social scientists. In the following, we want to discuss two approaches for studying human social behavior on the World Wide Web via examples, in particular social Web data and Web experiments.

Social Web Data

Facebook adoption represents an interesting example of a social process. In Johan Ugander and his colleagues' work,[13] they aim to reveal how Facebook adoption materialized by analyzing observational data from email invitations which existing Facebook users could send to their friends. For each user who has received an email invitation, they find his or her potential friends on Facebook by identifying who imported the email address of this user in the past. This information gives the researcher a rough estimate of how the future ego network of the user will look like if he or she should decide to join Facebook. Interestingly, the researchers found that the number of future friends seems to be less important in predicting the probability of a user to join Facebook than the diversity of future friends. That means, the more distinct and unconnected the friends of a user who are already on Facebook, the more likely the user will join as well. This suggests that online recruitment tools may benefit from emphasizing the diversity of users who have already joined a platform, adopted a behavior, or bought a product. Another example is the work by Jure Leskovec and his colleagues,[14] where they aim to predict the sign (positive or negative) of social links in various social media platforms. The authors transform social-psychological theories like balance and status into quantifiable heuristics, which they then use as features to train a prediction model. In three different social Web datasets, the researchers found evidence for global status ordering of nodes. In contrast, they found no evidence for a global organization of these networks into opposing factions, which suggests that balance is operating more strongly at a local level than at a global one.

Web Experiments

In addition to social Web data, Web experiments open up new ways of shaping and influencing social parameters of user groups and the behavior of larger populations. In previous work,[15] Markus Strohmaier utilized randomized experiments to identify the extent to which product adoptions in social media applications are driven by peer influence (friends convincing friends) versus correlated factors (adoption that's independent of peer influence). In other work,[16] Robert Bond and his colleagues conducted a Web experiment on voter turnout, finding that Facebook messages directly influenced the real- world voting behavior of millions of people. In another experiment, Facebook suppressed emotionally-laden messages and showed that they influence the messages a user authors subsequently, thereby hinting at the possibility of emotion contagion in social networks.[17] This has sparked a discussion about the ethics of Web experiments in general.[18] An experiment by Tim Hwang and his colleagues[19] demonstrated the possibility of social bots influencing users on Twitter,[20] showing that users react to automated messages sent by bots. Experiments by Luca Aiello showed that automated bots can gain influential positions in online social networks.[21,22] Damon Centola analyzed the impact of the network topology on the adoption of behavior in online health communities.[23] Participants were embedded in different network topologies and received messages from their health buddies, which informed them about their buddies' activities. Centola found that redundant ties are more important for the adoption of behavior than weak ties, for example, since individual adoption was much more likely when participants received social reinforcement from multiple neighbors in the social network. Social Web data usually doesn't allow us to observe how users behave when they're embedded in different networks, and therefore experiments that manipulate the network structure are necessary. Online platforms for conducting these and other social science experiments are currently under development (see for example www.volunteerscience.com). While these examples highlight some of the potentials of applying computational social science for the World Wide Web, there are a number of unique and pressing challenges that need to be tackled.

## Challenges

The entanglement of social behavior and computation has created a vast, uncharted research territory, requiring the attention of computer scientists, social scientists, and Web scientists alike. In the following, we'll highlight three challenges, particularly important to social-computational systems on the World Wide Web.

### Social-Focused Algorithms and Computational Methods

While in computer science, algorithm design has focused on optimizing space and time complexity, in computational social science this focus needs to be expanded to include the social complexity of algorithms. To what extent does an algorithm capture different facets of social theories or human social behavior? What are the hard problems for algorithms that have a high social complexity? And can we quantify or formalize the social complexity of algorithms? At present, there are no methods, techniques, or notations that would help to reason about the social complexity of algorithms in a way that's analogous to existing notations in computer science, such as the Big O notation. The idea of social audits of algorithms[24] might represent a first step in such a direction.

### Computation-Focused Social Theories and Constructs

Although the social sciences have developed a rich body of literature and theories to understand social behavior and social systems, much existing research doesn't lend itself naturally to computation, or doesn't view social-computer interaction as an inherent component of human social behavior. As a consequence, social theories need to expand their focus from studying social interactions to studying social-computer interactions and need to be formalized to an extent that enables computational implementation. Increasing formalization of social theories would aid understanding, improve comparability, facilitate application, and enable empirical validation or simulations. An example of work in this direction is the computational translation of an existing – mostly qualitative – social theory to assess online conversational practices of political parties on Twitter.[25]

Ethics of Engineering Social
Interactions and Systems

When designing social-computational systems on the World Wide Web, engineers need to have tools and research methods available that help them gauge the social impact of engineering decisions. To what extent is it ethical/responsible/necessary to influence social behavior on the Web? What kind of social rules and norms do we want to support and implement? How do user interfaces adapt to users' cultural differences? How can intercultural understanding be facilitated? What kind of experiments require informed consent and what kinds of experiments don't? Katharina Reinecke and Abraham Bernstein offer an interesting example of a Web- based system that adapts to the cultural backgrounds of different groups of users.26 An example of the ethical discussion that we need to lead can be found in the conversations erupting from the Facebook emotion contagion study.[17,18]

We need to study social phenomena on the World Wide Web such as online social networks, not only because they're a (biased) representation of social realities, but also because they can shape social reality via the systems that we build. Understanding human social behavior therefore will become a key prerequisite for designing effective social-computational systems on the World Wide Web. Yet, the broad variety of methods, traditions, and ways of thinking about the problem has the potential to both catalyze and hinder progress. To avoid stagnancy, ideas about the validity of data and methods must be negotiated and allowed to coexist between computer and social scientists. Computational methods for analyzing, modeling, and shaping human social behavior, and the corresponding questions related to ethics, norms, and society in general will represent open research problems for decades to come. The emergence of computational social science is a first step in this direction.

**References**

1 F.-Y. Wang et al., “Social Computing: From Social Informatics to Social Intelligence,”IEEE Intelligent Systems, vol. 22, no. 2, 2007, pp. 79-83; doi: 10.1109/MIS.2007.41.
2 D. Zeng, F-Y. Wang, and K.M. Carley, “Guest Editors' Introduction: Social Computing,” IEEE Intelligent Systems, vol. 22, no. 5, 2007, pp. 20-22; http://dx.doi.org/10.1109/MIS.2007.4338490.
3 D. Zeng, “Social Computing: An AI Perspective,” IEEE Intelligent Systems, vol. 28, no. 3, 2013, pp. 2-5; http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6588861.
4 N. Shadbolt et al., “Towards a Classification Framework for Social Machines,” Proc. Workshop on Theory and Practice of Social Machines, 2013; http://eprints.soton.ac.uk/350513/1/socm2013_submission_9.pdf.
5 M. Strohmaier, “A Few Thoughts on Engineering Social Machines,” Proc. 2013 World Wide Web Conf., 2013;_www.markusstrohmaier.info/documents/a_few_thoughts_on_engineering_social_machines.pdf.
6 J. Preece, “Sociability and Usability in Online Communities: Determining and Measuring Success,” Behaviour and Information Technology, vol. 20, no. 5, 2001, pp. 347-356; http://dx.doi.org/10.1080/01449290110084683.
7 D. Helic and M. Strohmaier, “Building Directories for Social Tagging Systems,” Proc. 20th ACM Conf. Information and Knowledge Management, 2011, pp. 525-534.
8 D. Lazer et al., “Life in the Network : The Coming Age of Computational Social Science,” Science, vol. 323, no. 5915, 2009, pp. 721-723.

9 E.J. Webb et al., Unobtrusive Measures: Nonreactive Research in the Social Sciences, Rand McNally, 1966.
10 R.M. Lee, Unobtrusive Methods in Social Research, Open Univ. Press, 2000.
11 R. Groves, "'Designed Data' and 'Organic Data'," Director's Blog, US Census Bureau, 31 May 2011;_http://directorsblog.blogs.census.gov/2011/05/31/designed-data-and-organic-data.
12 H.A. Landsberger, Hawthorne Revisited, Cornell Univ., 1958.
13 J. Ugander et al., "Structural Diversity in Social Contagion," Proc. Nat'l Academy of Science of the United States of Am., vol. 109, no. 16, 2012, pp. 5962-5966; www.pnas.org/content/109/16/5962.full.
14 J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting Positive and Negative Links in Online Social Networks," Proc. 19th Int'l Conf. World Wide Web, 2010, pp. 641-650.
15 S. Aral and D. Walker, "Identifying Social Cnfluence in Networks Using Randomized Experiments," IEEE Intelligent Systems, vol. 26, no. 5, 2011, pp. 91-96.
16 R.M. Bond et al., "A 61-Million-Person Experiment in Social Influence and Political Mobilization," Nature, vol. 489, no. 7415, 2012, pp. 295-298; doi:10.1038/nature11421.
17 A.D.I. Kramer, J.E. Guillory, and J.T. Hancock, "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks," Proc. Nat'l Academy of Science of the United States of America, vol. 111, no. 24, 2014, pp. 8788-8790; www.pnas.org/content/111/24/8788.full.
18 M.N. Meyer, "Misjudgements Will Drive Social Trials Underground," Nature, vol. 511, no. 7509, 2014, pp. 265-265; doi:10.1038/511265a.
19 T. Hwang, I. Pearce, and M. Nanis, "Socialbots: Voices from the Fronts," interactions, vol. 19, no. 2, 2012, pp. 38-45; http://doi.acm.org/10.1145/2090150.2090161.
20 C. Wagner et al., "When Social Bots Attack: Modeling Susceptibility of Users in Online Social Networks," Making Sense of Microposts, 2012; http://markusstrohmaier.info/documents/2012_MSM12_socialbots. pdf.
21 L.M. Aiello et al., "People are Strange When You're a Stranger: Impact and Influence of Bots on Social Networks," Proc. 6th AAAI Int'l Conf. Weblogs and Social Media, 2012, pp. 10-17.
22 Emerging Technology From the arXiv, "How a Simple Spambot Became the Second Most Powerful Member of an Italian Social Network," MIT Technology Rev., 5 Aug. 2014; www. technologyreview.com/view/529696/ how-a-simple-spambot-became-the-second-most-powerful-member-of-an-italian-social-network.
23 D. Centola, "The Spread of Behavior in an Online Social Network Experiment," Science, vol. 329, no. 5996, 2010, pp. 1194-1197.
24 J.N. Matias, "Uncovering Algorithms: Looking Inside the Facebook News Feed," blog, 22 July 2014;_http://civic.mit.edu/blog/natematias/uncovering- algorithms-looking-inside-the- facebook-news-feed.
25 H. Lietz et al., "When Politicians Talk: Assessing Online Conversational Practices of Political Parties on Twitter," The Int'l AAAI Conf. Weblogs and Social Media, 2014, pp. 132-161.
26 K. Reinecke and A. Bernstein, "Knowing What a User Likes: A Design Science Approach to Interfaces that Automatically Adapt to Culture," MIS Quarterly, vol. 37, no. 2, 2013, pp. 427-453; www.misq.org/skin/frontend/default/misq/pdf/V37I2/Reinecke- Bernstein.pdf.

**Markus Strohmaier** is a full professor of Web science at the Faculty of Computer Science at University of Koblenz-Landau (Germany) and Scientific Director for Computational Social Science (CSS) at GESIS, the Leibniz Institute for the Social Sciences (Germany). Contact him at Markus. Strohmaier@gesis.org.

**Claudia Wagner** is a postdoctoral researcher at the Computational Social Science Department at GESIS, the Leibniz Institute for the Social Sciences, and an adjunct lecturer at the University of Koblenz-Landau. Contact her at Claudia.Wagner@gesis.org.