Computational Social Sciences

Sorin Adam Matei
Martha G. Russell
Elisa Bertino *Editors*

Transparency in Social Media

Tools, Methods and Algorithms for Mediating Online Interactions

Springer

Computational Social Sciences

A series of authored and edited monographs that utilize quantitative and computational methods to model, analyze and interpret large-scale social phenomena. Titles within the series contain methods and practices that test and develop theories of complex social processes through bottom-up modeling of social interactions. Of particular interest is the study of the co-evolution of modern communication technology and social behavior and norms, in connection with emerging issues such as trust, risk, security and privacy in novel socio-technical environments.

Computational Social Sciences is explicitly transdisciplinary: quantitative methods from fields such as dynamical systems, artificial intelligence, network theory, agent-based modeling, and statistical mechanics are invoked and combined with state-of-the-art mining and analysis of large data sets to help us understand social agents, their interactions on and offline, and the effect of these interactions at the macro level. Topics include, but are not limited to social networks and media, dynamics of opinions, cultures and conflicts, socio-technical co-evolution and social psychology. Computational Social Sciences will also publish monographs and selected edited contributions from specialized conferences and workshops specifically aimed at communicating new findings to a large transdisciplinary audience. A fundamental goal of the series is to provide a single forum within which commonalities and differences in the workings of this field may be discerned, hence leading to deeper insight and understanding.

Series Editors

Elisa Bertino
Purdue University,West Lafayette, IN, USA

Jacob Foster
University of California, Los Angeles, CA, USA

Nigel Gilbert
University of Surrey, Guildford, UK

Jennifer Golbeck
University of Maryland, College Park, MD, USA

James A. Kitts
University of Massachusetts, Amherst, MA, USA

Larry Liebovitch
Queens College, City University of New York, Flushing, NY, USA

Sorin A. Matei
Purdue University, West Lafayette, IN, USA

Anton Nijholt
University of Twente, Enschede, The Netherlands

Robert Savit
University of Michigan, Ann Arbor, MI, USA

Alessandro Vinciarelli
University of Glasgow, Scotland

More information about this series at http://www.springer.com/series/11784

Sorin Adam Matei • Martha G. Russell
Elisa Bertino
Editors

Transparency in Social Media

Tools, Methods and Algorithms for Mediating Online Interactions

Springer

Editors
Sorin Adam Matei
Brian Lamb School of Communication
Purdue University
West Lafayette, IN, USA

Martha G. Russell
mediaX, Stanford University
Stanford, CA, USA

Elisa Bertino
Computer Science
Purdue University
West Lafayette, IN, USA

Computational Social Sciences
ISBN 978-3-319-18551-4 ISBN 978-3-319-18552-1 (eBook)
DOI 10.1007/978-3-319-18552-1

Library of Congress Control Number: 2015943271

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Contents

Part I
Overtures to Transparency in Social Media

Introduction

Sorin Adam Matei, Elisa Bertino, and Martha Russell

As engagement with social media has become a dominant information acquisition and dissemination experience, the nature of the collection, production, and consumption of information has also changed. One of the most significant changes is the lowering of cost and technological barriers for sharing knowledge or opinions. User generated content dominates social media. This challenges traditional methods of collecting, disseminating and evaluating information. As much of the information exchanged on social media is often created or vetted by individuals or corporations whose identities, motives, or abilities are poorly or often simply unknown, we need new tools, theories, and practical strategies for evaluating the quality of the content and the credibility of its authors. Modelling the provenance and impact of authorship on social media is of crucial importance for explaining the emergence and impact of human motivations on social media content generation. Research on presenting, visualizing and explaining the social context of any given user in a social medium information exchange is equally important. In brief, researchers and practitioners need to create theories, methods and tools that make the authorship and dissemination process more transparent. We need new ways to understand at a glance, who, in what context, and if possible why creates or disseminates specific units of content.

The significance of this task cannot be underestimated, especially we consider the following facts:

S.A. Matei (✉)
Brian Lamb School of Communication, Purdue University, West Lafayette, IN, USA
e-mail: smatei@purdue.edu

E. Bertino
Purdue University, West Lafayette, IN, USA
e-mail: bertino@purdue.edu

M. Russell
Stanford University, Stanford, CA, USA
e-mail: martha.russel@stanford.edu

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_1

1. Between seven to nine Google searches for common terms return as one of the first page results a Wikipedia page (Miller, 2012). As a consequence, Wikipedia is the sixth most visited site in the world, attracting about 300 million readers every day (Alexa.com). This is almost 15 % of the entire world Internet population, estimated at around two billion people.
2. Facebook is the most visited site in the world, having more than a billion users, or half or the world Internet population (Alexa.com). Facebook is not, however, just for amusement. One third of US adults get their daily news from Facebook, either as links to articles or as original content (Pew Internet).
3. Twitter, on the other hand, is becoming a leading source for breaking news on a par with traditional newswire services. A study indicated that of 27 seven major breaking stories within a 6 week period in 2011, Twitter beat the newswires eight times and was just as fast in four other instances, showing that it has become just as powerful a source of information as traditional media (Petrovic et al., 2013).

The contributors to this volume present in a synergistic manner some of the significant theoretical and practical contributions in the area of social media reputation and authorship measurement, visualization, and modelling. The book justifies and propose several significant contributions to a future agenda for understanding the requirements for making more social media authorship more transparent. Building on work presented in a previous volume, *Roles, Trust, and Reputation in Social Media Knowledge Markets* (Bertino & Matei, 2014), the present volume discusses new tools, applications, services, and algorithms that are needed for the future of content in a real-time publishing world. These insights may help people who interact with or create content through social media or may assist in analyzing audience attitudes, perceptions and behavior in informal social media or in formal organizational structures. The contributors propose and answer forward looking questions about tools, applications, and services that visually or editorially mediate social interactions in social media. In addition, the volume includes several chapters that analyze the higher order ethical, critical thinking, and philosophical principles that may be used to ground social media authorship or use on valid moral and socially consequential foundations. Together, the perspectives presented in this volume may help us understand how social media content is created and how its impact can be evaluated. The chapters demonstrate thought leadership through new ways of constructing social media experiences and making traces of social interaction visible.

The substantive goal of this synergetic volume is to help researchers and practitioners design services, tools, or methods of analysis that encourage a more transparent process of interaction and communication on social media. Knowing who has added what content and with what authority to a specific online social media project can help the user community better understand, evaluate and make decisions and, ultimately, act on the basis of such information. As mentioned, research and development that support this goals need not ignore the ethical dimension of the problem, which is discussed in the last section of the volume.

The chapters are both retrospective and prospective. Scholars and practitioners look back at the work they have conducted so far, sharing with the readers some of the lessons they have learned from their own work. They also discuss the areas that they find understudied or that promise the greatest intellectual or practical payoff in the future. Many of the contributions place the discussion in the context of social network analysis or "big data" research.

Synthesized thematically, the present volume explores the following core issues:

1. How do author feedback and incentive structures influence participation, value creation and reputation of social communities and social media content creation in various contexts—commerce, education, entertainment, government?
2. In what ways does curation infrastructure influence content creation—e.g., crowdsourcing—and sharing. How do the participants' perception of the factors influence content credibility, risk and trustworthiness? What types of statistical strategies or procedures are needed to better understand how social media roles emerge, function, generate valuable content, accrue trust and inspire credibility?
3. What kinds of tools, especially net-centric statistical analysis aimed at large social media datasets, can be adapted to make social media interactions more transparent to social science researchers or avid content authors?
4. What new approaches are needed to explore security and user identity in social media contexts?
5. What ethical and philosophical dimensions are involved in social media authorship and analysis processes?

The volume is structured in four sections. The first section introduces the main themes of the volume by reviewing the KredibleNet workshop on Reputation, Trust, and Authority funded by a National Science Foundation grant and held at Stanford University in 2013. The workshop proposed an agenda for future research and development related to social media knowledge production. Researchers and practitioners discussed the ways in which online information is transformed into "truth"—validated, repudiated, credentialed, measured, weighed and processed—as well as influenced, manipulated and controlled. The participants examined online behavior in order to reveal the organic emergence and evolution of social roles, hierarchy and elites online. Participants also identified procedures and features instrumental for stimulating, managing and otherwise controlling this behavior. Researchers discussed ways by which social media content is leveraged for insight into public opinion and sentiment, the flow of information, and the rise to prominence of relevant topics and issues. Analysis of social media also exposed the vulnerability and malleability of information and facts. Presentations investigated the potential of online crowdsourcing for accomplishing challenging tasks—including visualizing innovation, organizing people, aggregating information and data, sourcing high quality content, managing complex projects—as well as for seeding misinformation.

The second section proposes a frameworks for understanding the various paths by which authors emerge and content is created on social media, and how the trustworthiness of this content can be assessed. A set of case studies, methodological chapters, and tool presentations discuss the theoretical principles and methodological approaches needed for explaining authoring and trust mechanisms involved in social media. Aleksandar Ignajatovic, Mohsen Rezvani, and Elisa Bertino discuss the principles required by aggregating robust data from inconsistent information. The chapter also introduces a methodology to protect against the malicious information sources that collude in order to perform information deception attacks. James Cavarlee proposes a new methodology for detecting crowdsourcing of unprincipled or socially deleterious tasks, such as spreading malicious URLs in social media, deploying artificial grassroots campaigns (astroturf), spreading rumors and misinformation, and manipulating search engines. His chapter is an important contribution to better understanding the authenticity and validity of user generated content.

User generated social media content represents, at the same time, a very valuable source of information, especially in time of emergencies. Social media flows can be used as sui generis social monitoring tools and just in time information dissemination channels. The chapter authored by the VACCINE team (Jiawei Zhang, Shehzad Afzal, Junghoon Chae, Guizhen Wang, Dennis Thom, Sorin A. Matei, Niklas Elmqvist, and David S. Ebert) is dedicated to a tool suite that identifies influential users in social networks, detects anomalous information diffusion patterns, and locates their geo-spatial coverage or impact. The chapter describes a visual analytics framework that can handle these issues based on dynamic social network analysis. It also discusses a visual analytics approach that allows decision makers to analyze large volumes of social media data to detect and examine abnormal events within location-based social networks.

The next two chapters are dedicated to alternative methodologies and tools for detecting specific social structures of interaction on social media. Sorin Adam Matei, Robert Bruno, and Pamela Morris discuss a methodology derived from social entropy theory and a visualization tool for wiki spaces that can detect optimal social structuration in online collaborative groups. The tool can be used to create self-monitoring, motivation and moderation mechanisms to enhance collaboration and learning in online spaces. Marc Smith, Itai Himmelboim, Ben Shneiderman, and Lee Rainie discuss the applicability of network analysis and of a dedicated tool, accessible to the lay public, NodeXL (a spreadsheet plugin), to categorize the basic structures of communication and interaction on Twitter. The six patterns they uncover reveal the important role played by self-expression, on the one hand, and polarization, on the other, in shaping Twitter interactions. Their research can also start a broader conversation about what is and what is not possible in such social interactional spaces.

Faisal Ahmed and Marina Gavrilova discuss innovative methods for identifying social media authors in conditions of poor provenance documentation. Their methodologies, relying on socio-behavioral metrics captured in time and space, is an important contribution to making social authorship online more transparent and accountable.

Finally, Brian Britt presents a methodology for detecting the emergence of social structures on massive social media content projects, such as Wikipedia. His approach reduces a complex problem to a manageable process, utilizing longitudinal analysis of presence or absence of certain actors in a leading elite of contributors.

The third section continues the conversation about making social media interactions and authorship more transparent, this time through new methods of curation, documentation, and segmentation. Two chapters, one discussing the open collaborative science education platform NanoHub (Michael G. Zentner, Lynn K. Zentner, Dwight McKay, Swaroop Samek, Nathan Denny, Sabine Brunswicker, and Gerhard Klimeck) and Ostinato, a data curation and visualization technique (Jukka Huhtamäki, Martha Russell, Kaisa Still, Neil Rubens), offer two best use scenarios for social media curation and content management. Sorin Adam Matei and Jeremy Foote, on the other hand, explore the manner in which Wikipedia avoids to visualize the social nature of the authorship process on the front page of each article in the name of a minimalist design, which in fact hides a social compact that privileges long tenure users and editors.

Mazdak Nik Bakht and Tamer E. El-Diraby discuss the applicability of social media mining strategies to urban development and planning. They propose analyzing conversation networks on social media as a source of creative ideas regarding project scope, funding, design and operations plans. Such analysis should be influenced by two major factors: domain-relevancy and benchmarking best practices.

The fourth and last section elevates the discussion about authorship, transparency, and credibility to the ethical and philosophical level. Robert Laughlin speaks about the relationship between authorship, credibility, and identity in social media. He suggests that social media anonymity, which is considered by many a given, can lead to social dilemmas, favoring amoral or unethical activities online. He proposes that mandating using real names is not always functional, since the requirement can be so easily circumvented. Rather, credibility can be enhanced online by a method of close reciprocal monitoring and "witnessing."

Michael Steinmann, Julia Shuster, Jeff Collmann, Sorin Adam Matei, Rochelle Tractenberg, Kevin FitzGerald, Greg Morgan, Douglas Richardson use the chapters of the book as a testing ground for a method of exploring ethical concerns in big data analysis. Focusing on privacy, they highlight the fact that all contributions raise and need to tackle privacy concerns, which suggests that big data research should have a more explicit method for dealing with ethical issues. To understand the ethical challenges that can arise from privacy concerns in Big Data, the authors elucidate how privacy in Big Data can be analyzed using two dimensions: (1) different *contexts* in which privacy is relevant and (2) different *principles* that specify the ethical meaning of privacy.

The final chapter of the book is an interview with Howard Rheingold, a pioneer of the idea of "virtual community" and of sociability enhanced by communication technology. The interview starts with a review of the promises, achieved or missed, of the cyberculture and of social media. It continues with a discussion of Rheingold's philosophy and methodology for facilitating credibility and trust online, in which he emphasizes the importance of critical thinking and of developing a system of intellectual and cultural "checks and balances."

The present volume hopes to create a bridgehead in an underexplored territory. Knowledge generating social media is here to stay, yet the manner in which it changes authorship, credibility, and trust is at best imperfectly known. Setting principles, proposing algorithms or tools, and creating social methods for clarifying or addressing them should be a priority. We hope that the volume may bring some clarity to this process of setting priorities, in addition to being a substantive contributions to the emerging field of computational social science.

Acknowledgments Part of the research reported in this book and the editorial work were supported by the NSF award No. 1244708, KredibleNet.

References

Bertino, E., & Matei, S. (2014). *Roles, trust, and reputation in Social Media Knowledge Markets: Theories and Methods*. New York: Springer Publishing House.

Miller, M. (2012, March 22). 3 more studies examine Wikipedia's page 1. Google rankings [Blog]. Retrieved January 14, 2015, from http://searchenginewatch.com/sew/study/2163432/studies-examine-wikipedia-s-page-google-rankings.

Petrovic, S., Osborne, M., McCreadie, R., Macdonald, C., Ounis, I., & Shrimpton, L. (2013). Can Twitter replace Newswire for breaking news? In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media ICWSM*.

Socio-Computational Frameworks, Tools and Algorithms for Supporting Transparent Authorship in Social Media Knowledge Markets

Karina Alexanyan, Sorin Adam Matei, and Martha Russell

1 Introduction

The Internet has enabled collaboration on an unprecedented scale, exhibited in a diverse array of information aggregators, interest based communities, crowdsourcing platforms, open source tools, educational venues and organizations. The low barriers to entry, wide-spread access and distributed nature of these systems have disrupted established notions of authority, reputation, influence, credibility and trust—concepts that are at the root of traditional forms of knowledge production. At the same time, the magnitude of recorded data generated online presents an opportunity and a challenge to scholars of all stripes—producing research fodder not only in the form of content, but also in the form of conversations, interactions, relationships and other networked connections.

A core contribution to meet this challenge and guide future research in this direction was the Kredible.net workshop on Reputation, Trust and Authority, held on October 18, 2013 at Stanford University's Institute for Research in the Social Sciences. Co-sponsored by the Purdue University Discover Park CyberCenter[1] and mediaX at Stanford University[2] and the Social Media Research Foundation,[3]

[1] PI: Sorin Adam Matei, CoPIs: Elisa Bertino, Michael Zhu, Chuanhai Liu

[2] Martha Russell, Executive Director

[3] Marc Smith

K. Alexanyan
Post Doctoral Scholar/Project Manager MediaX, Stanford University, Palo Alto, CA, USA

S.A. Matei (✉)
Brian Lamb School of Communication, Purdue University, West Lafayette, IN, USA
e-mail: smatei@purdue.edu

M. Russell
Stanford University, Stanford, CA, USA
e-mail: martha.russel@stanford.edu

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_2

the workshop brought together scholars from a variety of fields, combining perspectives from social sciences and computer science, from academia and from business. From this interdisciplinary perspective, scholars explored how—amidst the Internet's enormous volume of content and relationships—certain topics, concepts and individuals rise to prominence, develop strong reputations, gain followers and establish credibility and trust. The workshop explored both the opportunities and vulnerabilities of online knowledge creation, presenting methodologies, models and tools for analyzing the production of knowledge, as well as influence and power. They introduced new theoretical and intellectual perspectives and featured advances in mathematical modeling, social network analysis, text extraction and analysis, natural language processing and machine learning. The workshop was an agenda setting event. Its conclusions, summarized thematically here, represent a core contribution to the present volume. They delineate two main directions of research, one directed at the latest methods for detecting social interactions and roles online, the other at applying these methods to core social media platforms, especially those dedicated to crowdsourcing. Throughout, social network analysis and big data were considered essential methods and domains of investigation. We consider the summary of the presentations, discussed below thematically, not only as a record of one isolated conversation, but a springboard for further research and conversation. The case studies can also be used as further inspiration for new and at times normative approaches to the issues discussed.

Overall, researchers discussed the ways in which online information is transformed into "truth"—validated, repudiated, credentialed, measured, weighted and processed—as well as influenced, manipulated and controlled. They examined online behavior in order to reveal the organic emergence and evolution of social roles, hierarchy and elites online. They also identified elements and features instrumental for stimulating, managing and otherwise controlling this behavior. They leveraged social media content for insight into public opinion and sentiment, the flow of information, and the rise to prominence of relevant topics and issues. Analysis of social media also exposed the vulnerability and malleability of information and facts. Presentations investigated the potential of online crowdsourcing for accomplishing challenging tasks—including organizing people, aggregating information and data, sourcing high quality content, managing complex projects—as well as seeding misinformation.

Randy Farmer[4] **and Phil Gomes**[5] discussed the key features of reputation systems and trust online. **Ed Chi**,[6] **Sorin Adam Matei**,[7] **Jeremy Foote**,[8] **Howard**

[4] Suddenly Social, CEO

[5] Edelman, Senior Vice President

[6] Google

[7] Purdue University

[8] Purdue University

Welser[9] **and Jure Leskovec**[10] offered models for understanding social interaction online, exploring the emergence of functional roles in collaborative communities, examining networks of user engagement, and investigating the control and manipulation of users and data. The websites, collaborative communities and social media platforms that these scholars described ranged from leading exemplars such as Twitter, Google Plus and Wikipedia to more specific, yet still data intensive, sites such the geneology community WeRelate, Beer Communities, Breast Cancer Support communities, and the Stock Overflow question and answer website. **Jana Diesner**[11] **and Luo Si**[12] examined online discourse using natural language processing techniques and text analysis. **Itai Himelbolm**,[13] **Katy Pearce**[14] **and Adrian Albert**[15] extracted relational data from Twitter, using social network analysis tools such as NodeXL to explore the diffusion of information in social media, how topics and individuals rise to prominence, and the relationship between online conversations and offline realities. **Martha Russell**, **Kaisa Still**[16] **and Jukka Huhtamaki**[17] also explored relational networks, applying data intensive graph visualization techniques to manage large scale data sets and investigating value creation in innovation and collaboration ecosystems. Other presentations, such as those from **Larry Sanger**,[18] **Michael Bernstein**[19] **and Gerhard Klimeck**,[20] featured online resources and tools developed by the researchers themselves, leveraging the power of online crowds for completing complex creative tasks and producing high quality educational and journalistic content. In contrast, **James Caverlee**[21] investigated how crowdsourcing can be used for the inverse effect of misdirection and misinformation.

Taken together, the projects in this workshop explored the emergence of social roles, the creation of value, and the perception of credibility and trustworthiness in online information. Their approaches combined social science insights into the structure and nature of online interaction (exploring the influence of author feedback, curation infrastructure, and participation incentives) with advances in computational science, data visualization, graph analysis and natural language processing. The methods and results offered innovative statistical strategies, models and methodologies for navigating the large and complex data sets produced by online content.

[9] University of Ohio
[10] Stanford University
[11] University of Illinois
[12] Purdue University
[13] University of Georgia
[14] University of Washington
[15] Stanford University
[16] VTT, Finland
[17] Tempere University of Technology
[18] Infobitt
[19] Stanford University
[20] Purdue University
[21] Texas A&M University

2 Models and Methods for Measuring, Analyzing and Influencing Social Interaction, Functional Roles, and Behavior Online

Social media data provide an extensive research resource, an enormous and complex collection of trackable information on online interactions, and the evolution of roles, status and behaviour over time. At the same time, the concept of reputation is by definition a relational construct, an attribute of an achieved position in a network of interactions. Within the context of social media, reputation is predicated on the multiplicity, intensity and diversity of implicit ties that individuals establish by sharing or contributing content online. Hence, online reputation is a function of the amount and frequency of contributions, multiplied by the velocity at which the content is disseminated.[22] People are influential and important based on the amount of content they share, contribute to, or manipulate. Furthermore, the number of relationships, their direction, intensity, diversity, and the specific locations of the individuals connected by those relationships within the broader network topology illuminate the specific roles individuals play in the network. In other words, knowing the topology of a specific network of social media interaction allows us to derive the functional role and reputation of each node or individual.[23]

The data-intensive nature of social media research presents numerous challenges including data heterogeneity; entity and network discovery; and data size. This resource calls for data management and analytic tools that can address graph and network data on a massive scale, and support timely, effective, and efficient knowledge extraction processes. A new understanding of reputation needs to be incorporated into tools that measure and visualize its magnitude for each social media platform. Such an understanding would include guidelines for generating tools and services that measure reputation relationally, making reputation measurements and visualization an integral and essential part of ordinary individuals' online knowledge production and consumption. Understanding which new tools are needed, and how to design and build these tools, requires input from a broad multi-disciplinary community of scholars, transcending the boundaries of social science, computer science, and statistics.[24]

A number of the studies presented at the Kredible.net Workshop explored online social interaction. These studies investigated a range of online communities to examine the emergence and evolution of social roles and user behavior. In the

[22] A research agenda for the study of entropic social structural evolution, functional roles, adhocratic leadership styles, and credibility in online organizations and knowledge markets by Sorin Adam Matei, Elisa Bertino, Luo Si, Michael Zhu, Chuanhai Liu, Brian Britt. In Matei, S. A. and Bertino, E., editors (2014) Roles, trust, and reputation in Social Media Knowledge Markets: Theories and Methods. New York: Springer Publishing House.

[23] ibid

[24] ibid

process, they introduced new computational models and methods for tracking, measuring, influencing and directing user engagement.

Randy Farmer's[25] presentation articulated the concepts and terminology of reputation systems and defined their mechanisms. Farmer's presentation also provided tools with which one can analyze existing models of reputation systems, as well as design, deploy, and operate online reputation systems. Reputations, Farmer explained, aren't "things," they are "systems." Information can be conceptualized into small, discrete units such as "the reputation statement," which can be understood as the building block of a reputation system. These systems are critical tools for making decisions, rooted in the information we use to make value judgments about people or things. The information individuals use as a basis for these decisions is often externally produced. When people don't have firsthand knowledge of the object being evaluated, they tend to rely more heavily on reputation, and the experiences of others can be an invaluable aid in their decision. Context is also important, as reputation is earned within a particular context, or multiple contexts, and can extend outside context boundaries, and differ across contexts. According to Farmer, as individuals turn to online sources for data, sorting through trillions of pages in search of accurate and valid information, reputations become even more significant. Without reputation systems for features such as search rankings, ratings and reviews, and spam filters, Farmer argued, the Web would have become unusable years ago.

Phil Gome[26] specializes in digital media and reputation management with the global public relations firm Edelman. Gomes' presentation highlighted the key features of trust on the internet, as revealed by Edelman's "Trust Barometer," an annual survey and exploration of issues of trust around the world. Gomes clarified the distinction between reputation, which is based on an aggregate of past experiences, and trust—a forward facing metric of stakeholder expectation. Two things work against trust online, Gomes explained. The first is an understandable, even beneficial, aversion within an anarchic environment to mechanisms of control. The second is the notion that within that environment, reasonable expectations of permanence are extremely low. Based on Edelman's research, Gomes argued that the most trusted online institutions tend to be those that participate in communities in ways that parallel those communities' respective mores and provide a sense of reliability and permanence amid a shifting and amorphous environment.

In his presentation at the Kredible Workshop, **Ed Chi**[27] examined three case studies of social systems at three different stages of development. Chi's research took a model-driven approach to investigating social interactions on the Web. His work addressed one of the key questions of the Workshop, highlighting the statistical strategies or procedures that help researchers understand how social media roles emerge, function, generate valuable content, accrue trust and inspire

[25] Suddenly Social, CEO "Reputation Systems are Everywhere"

[26] Edelman, Senior Vice President "Trust in the Online Environment"

[27] Google "The Science of Social Interactions on the Web"

credibility. Chi applied a variety of models and analytical approaches to gain insight into the evolution of three different online social systems, with a focus on building and maintaining networks of trust, and the differences between offline and online behavior. In particular, Chi studied:

1. A successful social system by looking at Wikipedia's growth,
2. The start of a new social system, with a focus on building trust, by looking at privacy and sharing on Google Plus,
3. How to improve a running social system, with a focus on maintaining trust, by exploring information transmission across linguistic boundaries on Twitter and Google Plus.

One of Chi's insights was that people trust Wikipedia not because of the consensus necessary to create its online content, but because of the transparency of the conflicts that lead up to this consensus. Wikipedia, Chi demonstrated, required a critical mass of trust before experiencing exponential growth. For his research on Google Plus, Chi described the challenges of launching a new social system, with a focus on building trust through the use of privacy controls such as those enabled by "Circles" and the ways in which these controls attempted to delineate the diffusion of information across interpersonal social networks online. Chi also explored information transmission across linguistic boundaries on Twitter and Google Plus, investigating the different approaches and styles used by country, language and culture.

Chi's research methods and systems are informed by models such as information scent, sense-making, information theory, probabilistic models and evolutionary dynamic models. These models are used to understand a wide variety of user behaviors, from individuals interacting with social bookmarks in Delicious to groups working on Wikipedia articles. The models range in complexity from a simple set of assumptions to complex equations describing human and group behaviors. A model-driven approach, Chi argued, helps researchers improve their understanding of how knowledge is fundamentally constructed in a social context, and detect a path forward for further social-interaction research.

Sorin Adam Matei, **Wutao Tan**, **Michael Zhu**, **Chuanhai Liu**, **Elisa Bertino and Jeremy Foote**[28] questioned social media changes the way human organizations work. Their research explored changes in large data sets over time, focusing on the emergence and evolution of structure and social roles in voluntary production social media projects. Specifically, their study looked at Wikipedia, examining how, to what degree, and in what way, membership becomes organized into functional roles and elites. Their study examined the top 1 % Wikipedia contributors over time, to determine if the composition of this group is variable, as users join and exit, or if membership in this "elite" group remained stable over the long term. Does the elite group tend to become durable over time, they asked, preserving its membership even as a vast new number of members join the project every day? In

[28] Purdue University "Do Wise Crowds Have 'Sticky' Elites?"

other words, do wise crowds, groups of people who spontaneously get together to work on common projects, have 'sticky' elites? If that is the case, and the contribution or collaborative processes on Wikipedia and similar projects are dominated by a consistent, stable and long term (sticky) group of elites that are responsible for a vast share of the content—will the project exhibit a system-level **structuration process**, with elites developing functional roles and emerging leadership positions? The findings indicate that the top 1 % members of the project are a resilient group. About 30 % of them are present in the elite at least 2 weeks at a time. The slow turnover suggests the emergence of an adhocratic elite, that is both stable and flexible.

At the root of this research was the observation that voluntary and collaborative efforts online display familiar patterns of uneven distribution of contributions and rewards. With that in mind, their research aimed to discover whether these patterns are random, or if some specific factors lead to the dominance and stability of top "elite" online contributors. As part of this effort, their study aimed to identify the synthetic indicators that would enable researchers to place social media projects on a continuum—from changing leadership to stable. Such indicators offer insights into leadership roles in the social media era and their potential impact on human organizational behavior in general. One of these indicators was the social entropy level, which measures the degree of group structuration. When used in the Wikipedia study, the measure indicates that structuration reached a steady state in the last several years of the project. The level of structuration is also relatively high, indicating the presence of functional roles and leadership positions.

Howard Welser's[29] **presentation** also investigated online organizational structures and social roles, with a focus on function, trust and credibility, and the inherent social benefits of collaborative online systems. In particular, Welser argued that "digital institutions" such as online communities and collaborative projects have the potential to overcome a key problem in contemporary society—the inevitable top down corruption of large organizations. Large organizations, Welser explained, follow an "iron law of oligarchy" in that, despite egalitarian and democratic principles, they tend to concentrate organizational power at the top. Digital institutions, Welser asserted, offer the opportunity to overcome usual limitations to create an alternative social structure that provides truly distributed organizational control.

Welser identified a set of key attributes of such systems, including shared mission, flattened organizational structure, participatory democracy, open access to recorded contributions, large scale collaborative project spaces, semi-automated and/or distributed systems for monitoring, evaluation and sanctioning, double blinded peer review, content evaluation and compensation for digital contributions. This set of attributes, Welser argued, helps reveal problems inherent in extant organizations, and highlights the characteristics, features and insights that can be

[29] University of Ohio, "Breaking the Iron Law of Oligarchy: Computational Institutions, Reputation Systems and Distributed Social Control"

integrated from online interaction systems. According to Welser, certain online organizations exhibit some of these characteristics already, and provide examples for the future—including quantified self projects such as Strava and collaborative projects such as Wikipedia, Reddit and CrowdGrader. These types of distributed systems, Welser explained, provide reputation management through evaluation, self monitoring, and achievement oriented gamification. Ultimately, they create more effective organizations, with more meritocratic reward systems and reduced corruption.

In his exploration of user behavior in online communities, **Jure Leskovec**[30] focused on methods for motivating and steering user behavior. Leskovec's research touched on many of the key questions of this Workshop. He addressed the statistical strategies or procedures necessary for insight into how social media roles emerge, function, generate valuable content, accrue trust and inspire credibility. He also demonstrated the approaches needed to address the challenges of large data sets, and their changes over time. In particular, Leskovec's work highlighted the ways in which author feedback and incentive structures influence participation and value creation online.

Leskovec studies how mechanisms for rewarding user achievements based on a system of badges can influence and steer user behavior on a site—leading both to increased participation and to changes in the mix of activities that a user pursues on the site. Several robust design principles emerged from his framework that could serve to advance the design of incentives for a broad range of sites. Leskovec's driving research questions were: *How do people become members of collaborative communities*? *Can you predict later behavior* (*how long they will stay*) *based on early behavior*? *Can one build incentive behaviors* (*badges*) *so that people will behave well and stay longer*? *What is the optimal set of badges for behavior modification and control*? In order to answer these questions, Leskovec modeled and measured the relationship between individual users and the online community itself. He explored the trajectory of user and member evolution, as well as the evolution of the community as a whole. Essentially, his work examined what is going on as a person is becoming active in a community.

Leskovec's research looked at two online communities in particular—a network of Beer aficionados, and a Breast Cancer support network. His work focused on linguistic change as representative of the relation between users and communities, analyzing language practices (norms, etiquette) as measurable indicators of individual expression and group identity. Leskovec presented a framework for tracking linguistic change, measuring user reaction to linguistic change, and eventually predicting when users will leave the community.

Leskovec found that all users go through a similar life cycle, exhibiting repetitive patterns of assimilation to the style of the community and stagnation as the community evolves, which leads to distancing as the community leaves the user behind. The "lifespan" of the user—that is, their length of membership before final

[30] Stanford University, "Steering User Behavior with Badges"

distancing, or exit—is based on how receptive and adaptable they are to community style and behavior. The greater the distance between the user and the community at the beginning of the cycle, the shorter their "lifespan." These findings on the "life cycle" of members enabled Leskovec to predict a member's potential evolution, based on an analysis of their initial behavior and such parameters as initial distance, speed of assimilation/ approach and level of flexibility or adaptation.

Based on this understanding of member behavior, Leskovec investigated whether it is possible to identify users at risk of departing and influence/inspire their behavior using reputation markers and incentives (such as badges). This approach is based on the prevalence of badges in all social mileu's—military, education, online communities and commerce. Badges recognize and validate wide range of activities, serving as both credentials and incentives. For this aspect of his research, Lescovec asked: *How do criteria for a badge translate into effects on user behavior*? *How should site designers place/use badges if they want particular outcomes*? In response, Leskovec introduced a utility based model for reasoning about user behavior in the presence of badges, and in particular for analyzing the ways in which badges can steer users to change their behavior. This approach steers user behavior and user engagement, motivating the user to trade off between a preferred mix of activities in order to reach a badge. To evaluate the main predictions of his model, Leskovec studied the use of badges and their effects on the widely used Stack Overflow question-answering site. The site offers an enormous data set with two million members, five million questions and ten million votes. His model charted action of Type 1 (Question) against Type 2 (Answer), with badges serving as boundaries as each user moves along the chart. He tracked how users change behavior to reach badges, and the tensions between a tendency to resist behavioral change and the drive to attain a badge. His research found evidence that badges steer behavior in ways closely consistent with the predictions of his model.

Finally, he investigated the problem of how to optimally place badges to induce particular user behaviors. Leskovec's model allows for optimizing the badge placement for optimal behavior steering. If attainment is too easy, there will be little or no change in behavior. If attainment is too hard, change will also be deterred. The "sweet spot" of badge placement identified by Leskovec will inspire and motivate the user to change their behavior in order to reach the badge.

In separate presentations, **Jana Diesner**[31] **and Luo Si**[32] both introduced novel computational strategies, tools and algorithms for understanding how social media roles emerge, function, generate valuable content, accrue trust and inspire credibility. Their work provided examples of advances in computer science that enable the statistical analysis of large social media datasets, helping explain the emergence of new functional roles, and detecting credibility or trust online. In particular, both

[31] University of Illinois, "How do Social Roles, Reputation and Authority Emerge on Social-media Knowledge-generation Projects"

[32] Purdue University, "A Learning Approach for Web Social Emotion Detection"

scholars combine natural language processing techniques with methods from other disciplines—Diesner leveraged recent methodological advancements in analytic capabilities to combine NLP with network analysis and machine learning, while Si's research group applied NLP techniques along with information retrieval, machine learning, intelligent tutoring and text/data mining for life science.

Diesner's research explored online social interaction with a focus on how social roles, reputation and authority emerge on social media knowledge generation projects, and how can they can be operationalized, measured and explained. Her work introduced solutions, methods and tools for text mining/ distilling information from text data. In particular, she applied social network analysis to highlight the content of information produced or shared by network participants. According to Diesner, most text mining work focuses on *named entities* for network nodes—i.e., extracting only proper names for people and groups as potential nodes—disregarding the key fact that the vast majority of textual references to social agents is realized via common nouns that refer to social roles or social collectives (e.g., citizens, protestors). Diesner corrects for this oversight, expanding methods and analysis to also include textual references to social roles or social collectives. Diesner's work revealed the effects of language use in networks, including the transformative role that language can play in the evolution of roles, reputation and authority. Her work addresses a common lack in current research, which often focuses on the fact, frequency or likelihood of information flows, without regard for the content of the texts themselves.

Si's research group approaches online conversation as a measure of public and user opinion. Towards that end, their goal is to measure emotions and predict opinions based on the comments to online news stories. In his presentation, Si argued that the emotions contained within the text of online comments offer insight into the preferences and perspectives of individual user. These insights enable content producers to tailor information to the needs of the users and offer more relevant services to readers. Building on this understanding, Si's group developed a unique system of Meta classification that integrate heterogeneous sources related to online news stories—including not only the content of comments, but also user generated emotion tags. Their experiments on datasets from online news services demonstrated the effectiveness of the proposed approach.

3 Exploring Structure and Dynamics of Networks with Social Network Analysis

The presentations above explored issues of credibility and reputation, and the emergence of social roles, within the context of online interpersonal interaction. The three research projects below focused instead on the flow of information in social media, and the relationship between online conversations and offline realities. In particular, they examined the diffusion of information on Twitter, analyzing the emergence and evolution of influential topics, facts, credibility and "truth" in online conversations, and exposing patterns of information seeking, manipulation and verification. These projects leveraged a variety of methods and tools, including discourse analysis, social graphs and, in particular, the open source social network analysis tool NodeXL.

Itai Himelboim[33] explored patterns of information seeking online, with a particular focus on cases where facts were unclear or in dispute. He selected two specific cases which document disputed factual environments on Twitter—the Navy Yard shooting of September, 2013, which served as an example of a breaking news topic in which facts emerged over time, and The Affordable Care Act, a controversial measure surrounded by disputed information manipulated for political reasons. Himelboim collected data from Twitter based on mentions and replies among users who discussed the two topics, identifying popular hashtags, users and keywords. Using graph analysis and the open source tool NodeXL, Himelboim identified nodes, clusters and relationships surrounding these topics. Himelboim's research affirmed "information silo" theories, finding that one's social networks and network clusters influence exposure to and availability of information, particularly when facts are in dispute. In other words, belonging to a cluster influences exposure to information, on Twitter and across the Web. As a result, the degrees of accuracy and completeness of available facts vary across individuals.

In a similar vein, **Adrian Albert** [34] explored the evolution of information online, with a focus on how topics rise in prominence and influence, and how this influence is reflected offline. Albert began with the understanding that opinions, feedback and other rich content that users generate online offer a 'noisy' measurement of public opinion on topics of societal interest. He selected two particular topics for the focus of his research—energy and the environment—and examined the online discourse surrounding environmental legislation and regulation. Specifically, he analyzed the Twitter accounts of various groups in support and sponsorship of Congressional bills. Albert leveraged Twitter data as a tool for identifying the directionality of influence in the emergence of central topics in the public discourse. His research explored how influential topics change over time, and the channels through which they become adopted in public discourse. His work highlighted, in particular, the language surrounding these topics, and the manner in which they rise in prominence and influence, and are eventually adopted into law or incorporated into regulations.

Katy E. Pearce[35] explored the interplay of online information and offline reality within an authoritarian environment, examining the ways in which credibility, authority and validity can be manipulated online, and the real world effects of this behavior. Her research focuses on the use of **Twitter** during a series of protests in 2013 in Azerbaijan, where media and freedom of assembly are under

[33] University of Georgia, "The Affordable Care Act on Twitter"

[34] Stanford University, "How Social Media Reflect Decisions and Outcomes in the Physical World"

[35] University of Washington, "Social Media and Protest"

authoritarian control. In this context, online media become a primary tool for independent and oppositional communication and organization. Pearce combined qualitative data (observations and interviews) with data collected and visualized using the social network analysis tool NodeXL to reveal evidence of online information manipulation by the Azerbaijani government. Pearce's research documented how pro-government forces leveraged the opposition's main tool—social media—to limit its utility for protest and organization. The exposure of this manipulation influenced real world behavior—emboldening the opposition by reinforcing their views of government control and causing pro-government forces to become more savvy in their techniques.

Martha Russell,[36] **Kaisa Still**[37] **and Jukka Huhtamäki**[38] explored the structure and dynamics of innovation ecosystems, social-media platforms and other networked phenomena. Their work addressed several of the key questions of this workshop, including innovations in graph analysis that advance our abilities to explore and analyze the enormous and heterogeneous data sets produced by social media.

Russell, Still and Huhtamaki highlighted the wealth-creating potential residing in a firm's relationships with its stakeholders by exploring value creation in innovation networks, open innovation and co-creation. Their work leveraged the volumes of digital data generated around activities, interactions and collaboration, as company founders, entrepreneurs, investors, journalists, policy makers and customers share information, and communicate about their needs, experiences and opinions using social media. In their research, Russell, Still and Huhtamaki applied data driven visual analytics and social network analysis for insights into relational capital, looking beyond usual metrics such as stakeholders, customer satisfaction and media exposure to analyze relationships, connections and interactions. This work incorporated the heterogeneous nature of context for a set of unique actors and the unique reciprocal links between them, presenting metrics and network visualizations that 'reveal' this context.

In a separate presentation, Huhtamaki discussed the requirements of next-generation analytics tools for networks. Huhtamaki proposed a cloud-based approach for developing the necessary tools and processes. These tools would involve aspects of interactive computing, reproducible analysis, visual analytics, interactive visualization and scientific visualization. In particular, Huhtamaki advocated for data-driven visual network analysis of the very large datasets produced by social-media platforms such as Twitter and Facebook, and argued that adaptive data modeling methods should be developed to support computational open-data ecosystem analysis.

[36] Stanford University, "Understanding the Wealth-Creating Potential of Relationships Beyond Pretty Pictures Based on the 'Fluff' of Social Media"

[37] VTT Technical Research Center, "Understanding the Wealth-Creating Potential of Relationships Beyond Pretty Pictures Based on the 'Fluff' of Social Media"

[38] Tampere University of Technology, "Data-Driven Network Analytics in the Cloud"

4 Crowdsourcing for Education, Creative Production, News and Misinformation

Crowdsourcing is an online process that delivers services, ideas or content via distributed micro contributions from large groups. In crowdsourcing, problems or tasks are broadcast to users (the crowd) who perform tasks or submit solutions that the crowdsourcer then owns. The benefits of the process for crowdsourcers include the economical and rapid acquisition of solutions and information. Users are motivated to contribute by social contact, intellectual stimulus or financial gain. Although 'crowdsourcing' was coined in 2006, it may describe activities that include crowdvoting, crowdfunding, crowdworking and their negative counterpart—crowdturfing, where the "crowd" is used to manipulate social media and search engine results, spreading rumor and misinformation. Three of the projects described below addressed the challenge of crowdsourcing complex and multifaceted tasks. They provided examples of computational methods and techniques for leveraging crowds to produce rapid, efficient and high quality results—in the fields of education, journalism and creative production. The fourth presentation investigated an inverse phenomenon—the intentional use of crowds to generate false information and mislead users.

Larry Sanger[39] shared his experience in building an innovative, high risk, but high payoff crowdsourcing project—InfoBitt News, an online site for crowdsourced news content. Sanger explored the challenges of crowdsourcing consistently high quality content, as well as the benefits. These benefits, Sanger explained, include speed, scope, quick and efficient ranking, extensive summaries, and, potentially, the elimination of editorial bias. For InfoBitt News, Sanger developed a novel crowdsourcing method that combines five key features:

(1) competition, (2) constrained text (minimum/maximum length), (3) content requirements and rules aimed at quality, (4) gamification, to let users compete measurably, and (5) a shared, high-minded goal. His presentation described how these five features will together attract editors that focus on and create high quality content. He also articulated potential problems, and shared predictions for success.

Michael Bernstein's[40] **team**, **including Daniela Retelny**, **Sébastien Robaszkiewicz**, **and Alexandra To**, addressed the challenge of crowdsourcing creative, open ended and complex tasks. Bernstein's presentation described the online authoring platform, *Foundry*, developed by his group, which provides a modular computational crowdsourcing structure to coordinate crowdsourced teams of experts. Foundry's modular computational workflows enable rapidly assembled expert teams to compete complex and interdependent goals. The tool addresses obstacles such as complexity, lack of structure, busy waiting, blurred boundaries etc. with a flexible, composable and replicable user interface that coordinates and

[39] Infobitt, "How to Crowdsource a High-Quality News Site"

[40] Stanford University, "Enabling Expert Crowdsourcing with Flash Teams"

guides expert flash teams through a wide range of complex tasks. Foundry combines the visual language of team workflow environments with the affordances of flash teams, aiding users in composing modular, elastic and pipelined team designs. The goal of Foundry, Bernstein explained, is to become a library of best practices, workflows and team structures, as well as a first-generation IDE for expert crowd computing.

Foundry's workflows are modular in that they are self contained, replicable and able to be built upon. Based on a formalized series of events, input is received and output produced and then handed off to the next group. A DRI—directly responsible individual—serves as the manager or temporary leader for each component. The workflows incorporate elasticity—the ability to grow or reduce team members—and pipelining, which enables simultaneous work. They can be sequential, concurrent or interdependent. A sample task included crowdsourcing the entire software design process from "napkin sketch" to mock up, to heuristic evaluation, to revised mockup, to software prototype, to user test, to revised prototype—all in 1 day. Other tasks involved crowdsourcing educational content, such as an entire MOOC platform, and creating a short animated video. In the process, "expert crowds" served as core components of the crowd sourcing system, which coordinated "ad hoc" teams of experts to accomplish tasks they couldn't do alone.

Foundry's strengths are its scalability, versatility and quick turnaround. The platform provides a step forward in CSCW (computer supported cooperative work), and the dynamic collaboration of diverse and interdependent participants, affording users a novel way to organize and accomplish tasks, going beyond "being there" and working more quickly and effectively than distributed teams. However, as Bernstein explained, recruitment remains a time consuming task, and the approach is challenged to avoid the inevitable tradeoffs between quality, time and cost, as well as conflicts in coordination and team work.

Gerhard Klimeck[41] described his project **nanoHub**—an open source, collaborative effort for improving the functionality of online education. In particular, Klimeck's group is focused on the online delivery of a broad range of nanotechnology simulation tools for use in education, with the aim of bringing the new insights and approaches being developed in nanoscience into the traditional fields of engineering and applied science in a broadly accessible manner. Klimeck's group developed the RAPPTURE toolkit, containing over 300 tools available to students and educators through their educational portal. The toolkit provides the basic infrastructure for a variety of scientific applications, letting scientists focus on their core algorithm when developing new simulators. These simulators offer serious treatments of fundamentals, taught at an advanced undergraduate or beginning-graduate-student level. RAPPTURE is a net-centric tool, which makes massive computation resources readily available to large groups of users, who in turn employ the tool to produce additional content. According to Klimeck, the

[41] Purdue University, "Mythbusting in Research and Educational Networks"

tool's utility and ease of use have greatly reduced production time for scientists and educators, who have used the toolkit to create over 1,400 new versions.

While the RAPPTURE-enabled projects leveraged rapidly assembled crowds for the production of high quality content in various forms, other efforts have exploited crowd-based production capability to produce misinformation and manipulate content. **James Caverlee's**[42] presentation cited a recent Chinese study, which found that 90 % of tasks on many crowdsourcing platforms are for crowdturfing—using crowds for purposes of misinformation.[43] **Caverlee** investigated examples of crowdturfing, such as spreading malicious URLs in social media, forming artificial grassroots campaigns (astroturf), spreading rumor and misinformation and manipulating search engines. His initial research found that most malicious tasks in crowdsourcing systems target either online communities (56 %) or search engines (33 %). Caverlee's lab is pursing a set of related research activities aimed at uncovering the ecosystem of crowdturfers, developing the core algorithmic approaches for detecting crowdsourced manipulation of social media and online communities, and building new preventive frameworks for maintaining the information quality and integrity of online communities in the face of this rising challenge.

4.1 In Sum

The projects reported at the Kredible.net Workshop explored the socio-evolutionary dynamics of online knowledge production from a variety of angles. The Workshop highlighted a diverse arsenal of analytical tools, models and methods for investigating the emergence and rise to prominence of topics, concepts, behaviors, roles and individuals online. This exploration of authority, reputation, credibility and trust online also provided insights into their inverse—manipulation, the conscious and unconscious spread of misinformation, the variability of facts, and the abuse of influence and power. The conceptual approach used by most of these research projects was one of systems and network theory—online communities, social media networks, reputation systems and networks of innovation and collaboration were all explored from the perspective of connections, relationships, links and nodes. Online communities offered insights into online behavior, roles, engagement, motivation, culture and values, with a focus on how these influence, and are influenced by, reputation, authority and trust online. Social media provided content for social network analysis as well as text analysis of public opinion,

[42] Texas A&M University, "Detecting and Preventing Crowdsourced Manipulation of Social Media and Online Communities"

[43] Wang, G., C. Wilson, X. Zhao, Y. Zhu, M. Mohanlal, et al. Serf and turf: Crowdturfing for fun and profit. In Proceedings of the 21st international conference on World Wide Web. 2012. ACM. *As cited in* http://repository.tamu.edu/bitstream/handle/1969.1/151253/TAMILARASAN-THESIS-2013.pdf?sequence=1

sentiment, the flow of information and the emergence of key topics and issues. Crowdsourcing platforms served as examples of online tools for accomplishing complex organizational tasks, sourcing high quality content and managing projects. Research methods traversed disciplines, combining the use of computational and analytical models with natural language processing, machine learning, social network analysis and data visualizations. The ultimate result served to arm contemporary scholars and "information consumers" with a variety of next generation tools, methodologies, strategies and insights that can serve as "information gauges," helping researchers and users navigate the evolving online environment and make better decisions.

The current state of cutting edge research on transparency and credibility in social media requires a clear visualization of the roles and behaviors that "nudge" users toward specific outcomes. Credibility is not an issue of belief, but of evaluating other users' acts on the basis of their outcomes. A key evaluation strategy is to establish a working relationship with the provider of content, and then coordinate actions with them through a variety of means. Providing the necessary online visualization and information affordances to foster co-orientation is crucial. Even more important, is to provide the means to influence other actors actions through your own acts. Transparency on social media is not only a pious desiderate, but a very real means of improving interaction and strengthening credibility. The multiple perspectives offered in the workshop presentations and in the other chapters of this volume make a significant contribution to this end.

4.2 About

KredibleNet is a global community of scholars and practitioners dedicated to examining the emergence of social roles, authority, credibility, and trust online. **KredibleNet** represents a broad multi-disciplinary community effort, defining, measuring, and operationalizing the changing concepts of "reputation" and "expertise" in social media and collaborative online communities, and leveraging insights into online knowledge creation to design and build new large scale data analysis and management infrastructures. KredibleNet strives to shape the next generation of theoretical and analytic strategies for understanding how knowledge markets are influenced by the social interactions and reputations built online. The workshops, papers, conference presentations, educational or mentoring activities generated by KredibleNet aim to change the way in which knowledge generation in social media spaces is understood and utilized. The tools and algorithms prototyped through KredibleNet are developed to provide "information gauges" that help contemporary information consumers make smarter choices.

mediaX at Stanford University and its members and collaborators worldwide create networks of thought leaders whose collective inquiries address problems in ways beyond any individual organization. Their strength lies in the knowledge and expertise they bring together, through discovery collaborations, to address pressing issues and opportunities. Affiliate program to Stanford's Human Sciences Technology Advanced Research Institute, mediaX catalyzes research to explore how information technology can improve the human experience and how fundamentals of human science can inform the information technology products and services of the future.

Part II
Assessing Provenance and Pathways in Social Media: Case Studies, Methods, and Tools

Robust Aggregation of Inconsistent Information: Concepts and Research Directions

Aleksandar Ignjatovic, Mohsen Rezvani, Mohammad Allahbakhsh, and Elisa Bertino

1 Introduction

Today, more than ever, there is a critical need for organizations to share data within and across the organizations so that analysts, decision makers and control systems can make effective decisions. However, in order for analysts and decision makers to produce an accurate analysis and make effective decisions and take actions, data must be trustworthy. Therefore, it is critical that data trustworthiness issues, which also include data quality, provenance and lineage, be investigated for organizational data sharing, situation assessment, multi-sensor data integration and numerous other functions to support decision makers and analysts. Almost all application domains that we may think of require the ability to assess data trustworthiness; notable examples include: sensor networks (Lim, Moon, & Bertino, 2010; Lim, Ghinita, Bertino, & Kantarcioglu, 2012), social networks (Dai, Rao, Truta, & Bertino, 2012), location-based applications (Dai, Rao, Ghinita, & Bertino, 2011) critical infrastructures, e-health, and peer marking for massive open online courses (MOOCs).

The problem of providing trustworthy data to users and applications is an inherently difficult problem that requires articulated solutions combining different methods and techniques, ranging from iterative filtering (IF) algorithms (Laureti, Moret, Zhang, & Yu, 2006) to semantic integrity and ontology-based reasoning to digital signature techniques—just to name a few. It is however important to notice that technology has made possible to collect data from many different, possibly independent, sources. The advent of the Internet of Things (IoT) will further push

A. Ignjatovic • M. Rezvani • M. Allahbakhsh
University of New South Wales, Sydney, NSW, Australia
e-mail: aleksignjat@gmail.com

E. Bertino
Purdue University, West Lafayette, IN, USA

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_3

such capabilities. The availability of multiple observations and data pertaining to the same event or phenomenon in both the cyber space and the physical space represents an important opportunity for methodologies, referred to as data aggregation methodologies, aiming at assessing data trustworthiness by comparing and aggregating such multiple observations. Such methodologies can also include the use of IF algorithms resulting in iterative data aggregation methodologies. However, a major problem of data aggregation methodologies is that data items representing such observations are often inconsistent. Such inconsistencies arise because of errors, such as human and application errors or sensor calibration errors, or may be a result of deliberate attacks by malicious parties aiming at injecting deceiving information.

The use of provenance techniques may help in addressing such a problem. Provenance tracing makes it possible to trace back the source of a data item and the path that the data item followed in a given system in order to reach the intended recipient. Such provenance information can be used as a factor for assessing data trustworthiness in that it allows one to assign different weights to data items based on the source. An approach that combines IF with provenance has been proposed by Lim et al. (2010) in the context of sensor networks. Such approach is efficient and effective and has been widely extended. However, a major drawback of such approach is that it is not robust against collusion attacks. A collusion attack is one by which multiple malicious parties cooperate in order to inject deceiving information. Under such an attack, the data aggregation methodology will assess data as trustworthy whereas the data is not.

The problem of designing data aggregation methodologies that are robust against collusion attacks has been recently addressed by a novel IF methodology by Rezvani, Ignjatovic, Bertino, and Jha (2015). Such methodology is applicable to both numerical and non-numerical data, and, compared with the "classical" IF algorithms of Laureti et al. (2006), Yu, Zhang, Laureti, and Moret (2006) and De Kerchove and Van Dooren (2007, 2008, 2010) greatly improve the numerical stability of data aggregation as well as robustness against the collusion attacks.

In this paper we provide a survey of IF methodologies for assessing data trustworthiness and introduce a research roadmap to guide future research. In what follows, we first survey the methodology by Lim et al. (2010), Laureti et al. (2006), Yu et al. (2006) and De Kerchove and Van Dooren (2007, 2008, 2010) to introduce the basic concepts and IF with provenance. We then show a collusion attack against such methodology and survey the IF methodology by Rezvani et al. (2015). Experimental results show that this methodology is highly effective against collusion attacks. We then discuss relevant research directions and finally outline a few conclusions.

2 Provenance-Based Data Trustworthiness Assessment

A cyclic and provenance-aware trust computation framework was proposed by Lim et al. (2010) in the context of sensor networks. The proposed framework is based on a heuristic that the more trustworthy data a sensor node reports, the higher the node's trust score is. Moreover, the trustworthiness of a data item depends on the trust scores of the nodes which passed it towards the server node. The nodes through which a data item has been passed in the sensor network represent the *provenance* of such data item. By taking into account such interdependency relationship between the trustworthiness of data items and sensor nodes, a cyclic trust computation has been proposed in which the trust scores evolve gradually. This framework which we briefly review now can be employed as an online trust computation method. In what follows, we first introduce the network model underlying this framework, and the relevant notions of provenance. We then describe the cyclic framework, and finally report results from the experimental evaluation in Lim et al. (2010).

2.1 Background Notions

A sensor network is represented by m sensor nodes n_i, $i = 1, \ldots, m$ with identifier i for node n_i. In such a network, all sensor nodes are responsible for monitoring one event (i.e. nodes report multiple independent observations for one event). The sensor network is modeled as a graph $G(N, E)$, where $N = \{n_1, n_2, \ldots, n_m\}$ is the set of nodes and $E\{e_{i,j}\}$ denotes the set of edges, with $e_{i,j}$ an edge connecting nodes n_i and n_j. Figure 1a shows an example of a sensor network. As one can see in

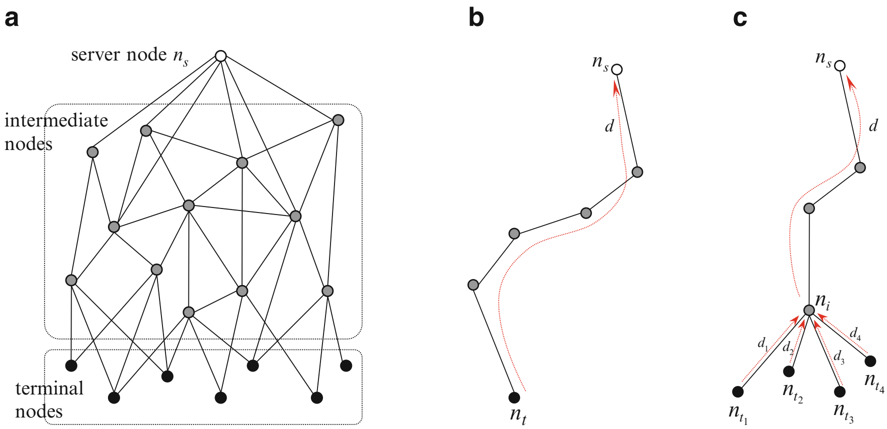


Fig. 1 Sensor network and data provenance examples. (**a**) Sensor network example. (**b**) Simple path example. (**c**) Tree path example

this figure, network nodes in N can be categorized into three types according to their roles in the network: a *terminal*, an *intermediate*, or a *server* node.

Definition 1 (Lim et al. (2010)).
A *terminal node* is a sensing node which generates a data item and sends it to one or more intermediate or server nodes (black filled nodes in Fig. 1a). An *intermediate node* receives data items from one or more terminal or intermediate nodes and passes them to another intermediate or a server node; it may also perform an aggregation function over the received data items and send the aggregate value to an intermediate or a server node (gray filled nodes in Fig. 1a). A *server node* (or base station) receives data items and evaluates continuous queries based on those items (white nodes in Fig. 1a).

Without loss of generality, it is assumed that there is only one server node in the network, denoted by n_s. Moreover, a data item d is represented by a single numeric value v_d.

In data management, the provenance concept represents the path of provisioning a data item. The provenance of a data item d, denoted by p_d, records where and how the data item d has been generated and how it has been passed through the sensor network towards the server n_s.

Definition 2 (Lim et al. (2010)).
The *provenance* p_d of a data item d is a rooted tree satisfying the following properties: (1) p_d is a subgraph of the sensor network $G(N,E)$; (2) the root node of p_d is the server node n_s; and (3) for two nodes n_i and n_j of p_d, n_i is a child of n_j if and only if n_i has passes the data item d to n_j through a direct link.

According to the tree nature of the data provenance, intermediate nodes are categorized into two categories: simple and aggregate.

- A *simple node* is an intermediate node having only one child. For example, in Fig. 1b every intermediate node is a simple node. Accordingly, a data provenance with only simple nodes can be represented by a simple path and this type of provenance is called a *simple provenance*.
- An *aggregate node* is an intermediate node with more than one child nodes. Figure 1c shows an intermediate node n_i which is an aggregate node and generates a new data item d by aggregating multiple data items $[d_1, d_2, d_3, d_4]$ received from nodes $[n_1, n_2, n_3, n_4]$ and passes d to the server n_s. A data provenance with at least one aggregate node is represented as a tree rather than a simple path and this provenance is called an *aggregate provenance*.

As an example of the sensor network, we can assume that a number of different sensors are distributed in a battlefield to collect the enemy locations (Tang et al., 2010). The sensors continuously watch the areas day and night to detect approaching enemies and send alarms to a server node. Moreover, the sensors are using a multihop routing scheme where each sensor may pass through the data of other sensors towards a server node.

2.2 Cyclic Trust Computation Framework

The main idea behind the trust computation approach by Lim et al. (2010) is to model the interdependency relationship between the trustworthiness of data items and their corresponding network nodes (as shown in Fig. 2). As one can see in this figure, the trust scores are assigned to both data items and network nodes, in an interdependent manner. The trust score of a data item is partially measured by the trust scores of the network nodes within its provenance. On the other hand, the trust score of a network node depends on the trustworthiness of data items that are generated by or passed through the node.

Figure 3 shows how the cyclic framework proposed in Lim et al. (2010) uses this interdependency to compute the trust scores of data items and network nodes. As shown in the figure, there are three different types of trust scores, *current*, *intermediate*, and *next*, for every data item and network node. The dashed line has separated the trust computation modules for data items and network nodes; the solid lines are traversed from one computation module to the next one.

For a set of data items received for a same event in the current window, the methodology by Lim et al. (2010) computes the current and intermediate trust scores for each data item in the first and second steps, respectively. The current trust score for a data item depends on the current trust scores of the nodes in its provenance, while its intermediate trust score is computed based on the latest set of

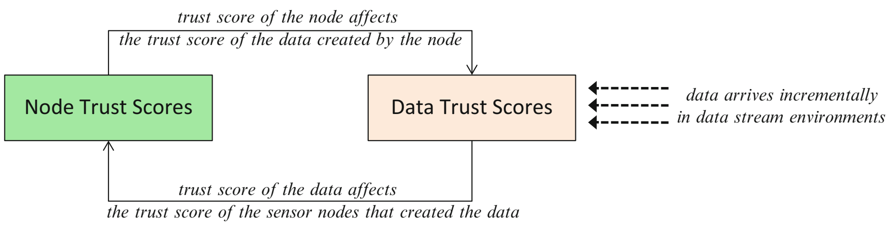


Fig. 2 Interdependency between data and node trust scores

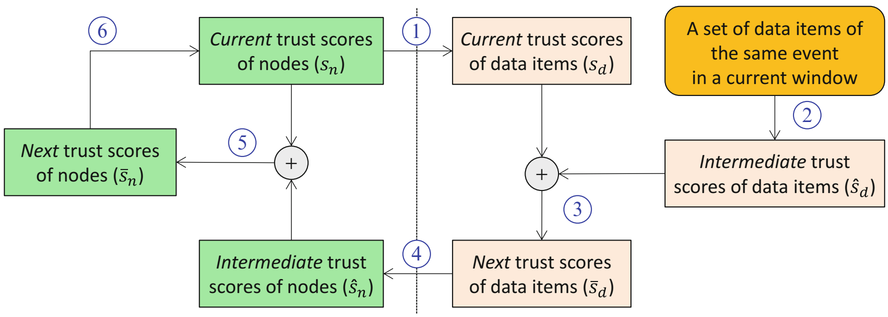


Fig. 3 An cyclic framework for computing trust scores

data items reported for a same event in the current streaming window. In the third step, the next trust score for each data item is computed by aggregating the current and intermediate trust scores of data items.

As shown in left side of Fig. 3, the intermediate trust score for each network node is calculated based on the trust scores of its related data items (step 4). After that, the next trust score for a network node is obtained by combination of its current and intermediate trust scores. Finally, the next trust scores in the current streaming window are copied to the current scores in the next window (step 6). Note that the cyclic trust computation process needs initial trust scores for sensor nodes which are set to one for all nodes at a very beginning of the process.

Computing Node Trustworthiness As we described, the current trust score of a network node n, denoted by s_n, is equal to the next trust score obtained in the previous streaming window for that node. Thus, one needs to compute its intermediate and next trust score in the current window, denoted by $\hat{s}_n$ and $\bar{s}_n$, respectively.

The intermediate trust score of a network node n is computed based on the trustworthiness of its corresponding data items, which is a set of data items that are generated or passed through such a node during the current streaming window, denoted by D_n. The intermediate trust score $\hat{s}_n$ is simply computed as the average of the trustworthiness of its related data items, as follows:

$$\hat{s}_n = \frac{\sum_{d \in D_n} \bar{s}_d}{|D_n|}, \tag{1}$$

where $|D_n|$ is the number of nodes in the set D_n, and the $\bar{s}_d$ indicates the current trust score of data item d obtained in the first step of the proposed trust computation framework (see ① in Fig. 3).

As we described, the next trust score of a network node is computed by the aggregation of its current and intermediate trust scores (see ⑤ in Fig. 3). These trust scores are aggregated using a weighted sum as follows:

$$\bar{s}_n = c_n s_n + (1 - c_n)\hat{s}_n \tag{2}$$

where c_n, $0 \le c_n \le 1$ is a constant which represents the relative impacts of trustworthiness from the current streaming window versus the previous one. In other words, if c_n is small, the trust scores of network nodes can change fast; if c_n is large, the trust scores will change more slowly from one window to the next.

Computing Data Trustworthiness The trustworthiness of a data item d depends on its value v_d and provenance p_d. Moreover, there are three trust scores for a data item d: the current, the intermediate, and the next scores, denoted by s_d, $\hat{s}_d$, and $\bar{s}_d$, respectively.

Current Trust Score s_d The current trust score of a data item d is obtained by aggregating the current trust scores of nodes within its provenance. In the proposed approach, the minimum of the current scores of the nodes in p_d is used as the current trust score. This can be explained by the fact that the trustworthiness of a data item can be dominated by the minimum trustworthy node among all nodes which such a data item has passed through.

If the data item d has a simple provenance, the current trust score s_d is simply computed using the minimum value of current trust scores of nodes in p_d. However, when the data item has an aggregate provenance, it is needed to take into account the nodes with more than one child in p_d. To address this problem, the average of the current trust scores of child nodes is used as their aggregate score. Therefore, these child nodes can be considered as a single child node with a trust score equal to the average of the original child nodes. Using this method, an aggregate provenance is formed as a simple provenance for the trust computation.

Intermediate Trust Score $\hat{s}_d$ An intermediate trust score of data item d, denoted by $\hat{s}_d$ is computed based on the data value similarities and its provenance similarities with other data items reported for the same event. it is assumed that D is the set of data items reported for the same event with d.

In order to compute the value similarity for a data item d with value v_d, the proposed approach uses the assumption that the data values in D are normally distributed and the mean and variance are μ and σ^2, respectively. Therefore, the cumulative probability of the normal distribution is employed to compute the similarity of data value v_d with other values within D. Basically, the computation gives high trust scores to the values close to the mean. Thus, the initial $\hat{s}_d$ is computed as follows:

$$\hat{s}_d = 2 \int_{v_d}^{\infty} f(x)dx \tag{3}$$

As shown in Fig. 4a, the shaded area represents the trust score $\hat{s}_d$ obtained from Eq. (3). Clearly, the intermediate trust score is obtained by considering only the data value similarity. Thus, it is needed to adjust the computation to reflect the provenance similarity of the data item as well. The impact of provenance similarity on the trust score computation is computed based on some intuitive observations, listed in Table 1. For example, it is clear that different provenances

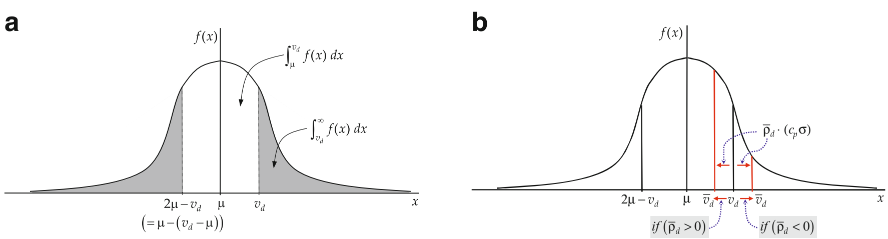


Fig. 4 Computing the intermediate trust score $\hat{s}_d$. (**a**) Intermediate trust score. (**b**) Intermediate trust score adjusted with provenance

Table 1 Impact of provenance similarity on adjusting $\hat{s}_d$

	Similar Data Value	Different Data Value
Similar Provenance	score ↑	score ↓↓↓
	small positive effect	large negative effect
Different Provenance	score ↑↑↑	score ↓
	large positive effect	small negative effect

of similar data values may increase the trustworthiness of data items. Accordingly, a normalized adjustable similarity value is defined for the similarities of the provenance of a data item d with all other data items in D, denoted by $\bar{\rho}_d$. More details can be found in a previous work on provenance-based trustworthiness assessment (Lim et al., 2010).

The adjusted similarity value $\bar{\rho}_d$ reflects the impact of the provenance p_d on the trust computation of the data item d. Thus, it is used to adjust the data value v_d to a new value $\bar{v}_d$ as follows:

$$\bar{v}_d = \min\{v_d - \bar{\rho}_d(c_p.\sigma), \mu\} \tag{4}$$

where c_p is a constant value greater than 0.

Now, the data value v_d in the Eq. (3) is replaced by the $\bar{v}_d$ to adjust the intermediate trust computation for data item d. Thus,

$$\hat{s}_d = 2\int_{\bar{v}_d}^{\infty} f(x)dx = 1 - \int_{2\mu-\bar{v}_d}^{\bar{v}_d} f(x)dx \tag{5}$$

Figure 4b shows how the adjusted similarity value $\bar{\rho}_d$ reflects the value similarity computation.

Next Trust Score $\bar{s}_d$ After computing the current and intermediate trust scores for a data item d, a weighted summation of these two trust values is used to compute the next trust score of data items, denoted by $\bar{s}_d$ (see ② in Fig. 3), Thus,

$$\bar{s}_d = c_d s_d + (1 - c_d)\hat{s}_d \tag{6}$$

where c_d is a constant, $0 \leq c_d \leq 1$, which defines how fast the data trustworthiness evolves as the cycle is repeated.

2.3 Experimental Evaluation

In this section, we briefly summarize the evaluation results from Lim et al. (2010) concerning the effectiveness of the proposed trust computation approach. The experiments were conducted by simulating the sensor networks and generating

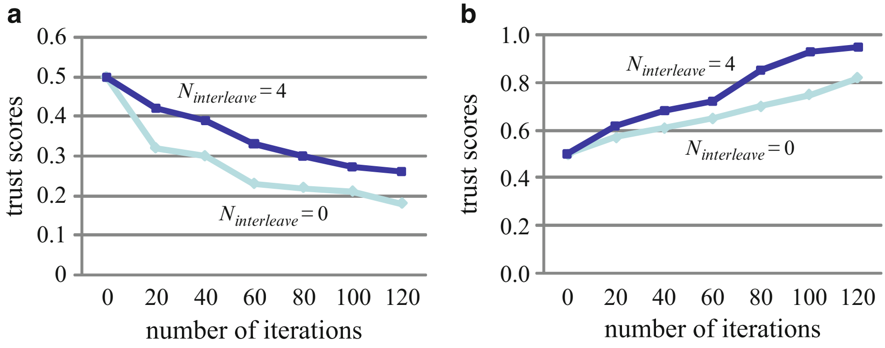


Fig. 5 Change of the trust scores for false data items. (**a**) With false data items, (**b**) with trustworthy data items

synthetic data. For observing the impact of provenance similarity, an interleaving factor was defined which means the interval between the assigned leaf nodes for generating data items in the simulated sensor network. In order to evaluate the effectiveness of the proposed solution, Lim et al. (2010) simulated the injection of false data items into the network and investigated how the proposed cyclic approach reflects this situation in the computation of the trust scores.

Figure 5a (from Lim et al., 2010) shows that when the false data items are injected, the trust scores change rapidly for smaller interleaving factors. This can be explained by the principle that different values with similar provenances rapidly reduce the trust scores (see Table 1). On the other hand, one can see in Fig. 5b that when the correct data items are injected again, the trust scores are increased more rapidly for larger interleaving factors. The reason is that similar values with different provenances result in a large positive effect (see Table 1).

2.4 Summary

This concludes our brief summary of the cyclic trust computation framework proposed in Lim et al. (2010). In Lim et al. (2012) the authors have proposed a game-theoretical defence strategy to protect sensor nodes from attacks and to guarantee a higher level of trustworthiness for sensed data. However, such approach can be compromised with collusive (collaborative) attacks which target the sample mean and variance of the data. In Sect. 4 we demonstrate this and then propose a safer solution based on Iterative Filtering algorithms.

3 IF Algorithms of Laureti et al. and De Kerchove et al.

A relevant class of algorithms for the assessment of information trustworthiness is presented by the *iterative filtering (IF) algorithms*. Pioneering algorithms of such kind were first proposed by Laureti, Moret, Zhang and Yu in their papers appearing in 2006 (Laureti et al., 2006; Yu et al., 2006). Their work was a motivation for the subsequent work of C. De Kerchove and P. Van Dooren in 2007 de Kerchove and Van Dooren (2007) and later in De Kerchove and Van Dooren (2008, 2010). Independently Ignjatovic rediscovered IF algorithms in 2007 (published in 2008, Ignjatovic, Foo, & Lee, 2008) and later introduced other novel algorithms in Lee et al. (2009), Lee, Rodrigues, Kazai, Ignjatovic, and Milic-Frayling (2009), Ignjatovic, Lee, Compton, Cutay, and Guo (2009), Chou, Ignjatovic, and Hu (2013).

The aims of IF-based data aggregation methodologies should be

1. to provide an aggregate value with a provably minimal variance due to stochastic errors of the sources;
2. to insure robustness against non-stochastic errors ranging from hardware faults to collusion attacks from some of the sources, with provable estimates of the level of robustness in terms of the fraction of misbehaving sources.

Moreover, such methodologies should be applicable to both numerical and non-numerical data.

We now explain the essence of IF algorithms using an example of a conference Chair. While such a problem is clearly not among the most pressing ones in the area of data aggregation, its familiarity to the reader makes it a very convenient example to explain both the challenges and our methods.

Let us assume that you are the Chair of a conference, and your referees have done their job: each paper has been reviewed by several referees and every referee has reviewed several papers and you got the scores. However, you suspect that some of the referees might have been unreasonably harsh with their marks; some others might have been sloppy, barely having looked at the papers and thus likely to have made large random errors. Worse, you are worried that some of your referees might have colluded in order to promote the papers of their friends and trash the papers of those against whom they might hold grudges. How should you aggregate the referee's scores and decide which papers to accept in the fairest possible way?

To analyze such a problem, let us assume that there are R referees marking P submitted papers, and, for the sake of simplicity of formulate, let us assume an unusual situation in which each referee marks every single paper. We denote by $M(r,p)$ the mark given by a referee r to a paper p. The main feature shared by most of IF algorithms is that they simultaneously produce approximations of the final aggregate values $\vec{\mu} = \langle \mu(p) : 1 \leq p \leq P \rangle$ (in the present case marks of papers) as well as trustworthiness ranks for the sources $\vec{\tau} = \langle \tau(r) : 1 \leq r \leq R \rangle$ (in this case referees), in a single iterative procedure.

An IF algorithm would typically start by giving all referees the same initial trustworthiness $\tau^{(0)}(r) = 1$ and obtain the initial approximation of the aggregate mark for each paper p as the simple mean of the marks of all referees, $\mu^{(0)}(p) = \sum_{r=1}^{R} M(r, p)/R$. Now, in turn, each referee can be judged on how accurate her marks are, by computing how close her marks are to such an initial approximation of the aggregate marks $\vec{\mu}^{(0)}$. Thus, we compute for each referee r the Euclidean distance $d^{(0)}(r) = \sqrt{\sum_{p=1}^{P} (M(r, p) - \mu^{(0)}(p))^2}$ between her marks $\langle M(r, p) \ : \ 1 \le p \le P \rangle$ and the aggregate values $\vec{\mu}^{(0)} = \langle \mu^{(0)}(p) : 1 \le p \le P \rangle$.

Since the trustworthiness of each referee should be inversely related to her distance (or deviation) $d^{(0)}(r)$, we pick a monotonically decreasing *penalty function* $F(d)$ and define the new estimate of trustworthiness of referee r as $\tau^{(1)}(r) = F(d^{(0)}(r))$. In the next round of iteration we obtain a new estimate $\vec{\mu}^{(1)}$ of the marks of papers as a weighted average of the marks of all referees, with the marks of a referee r taken with a weight $w^{(1)}(r)$ proportional to a referee's trustworthiness $\tau^{(1)}(r)$. In this way the outliers will be penalized, because their distance to the coarse, initial approximation $\vec{\mu}^{(0)}$ of the aggregate marks will be the largest and thus their trustworthiness and corresponding weight the smallest (but no outlier is ever completely excluded!). This process is iterated until it has, hopefully, converged, i.e., for a given precision threshold ε,

while $\sqrt{\sum_{1 \le p \le P} (\mu^{(n+1)}(p) - \mu^{(n)}(p))^2} > \varepsilon$ **repeat:**

$$d^{(n)}(r) = \sqrt{\sum_{1 \le p \le P} (M(r, p) - \mu^{(n)}(p))^2}; \quad (7)$$

$-$computing the distance between $r's$ marks and estimate $\vec{\mu}^{(n)}$

$$\tau^{(n+1)}(r) = F(d^{(n)}(r)); \qquad - \text{ computing the new trustworthiness of r} \quad (8)$$

$$w^{(n+1)}(r) = \frac{\tau^{(n+1)}(r)}{\sum_{1 \le r' \le R} \tau^{(n+1)}(r')}; \quad (9)$$

$-$ computing $r's$ weight by normalising $r's$ trustworthiness

$$\mu^{(n+1)}(p) = \sum_{1 \le r \le R} w^{(n+1)}(r)\, M(r, p), \quad - \text{ computing new estimate of the marks } \vec{\mu} \tag{10}$$

When such iteration terminates after, say, t many rounds of iteration, we get not only the aggregate values of marks of papers $\mu(p) = \mu^{(t)}(p)$ but also an estimate of the trustworthiness of the referees $\tau(p) = \tau^{(t)}(r)$. As we will see, choosing "the best" function $F(x)$ which provides an inverse relationship between distances and trustworthiness ranks is a tricky problem; the most commonly used functions are:

$$(i) \quad F(d(r)) = \frac{1}{d^2(r)}; \quad (ii) \quad F(d(r)) = e^{-d(r)}; \quad (iii) \quad F(d(r)) = 1 - k \cdot d(r),$$

where k appearing in the third function is allowed to be different for each round of iteration, and is chosen so that if r' is the referee with the largest (square of the) distance $d^{(n)}(r')$, then $F(d^{(n)}(r')) = 0$. We now briefly discuss the performance of the above algorithm with the first, reciprocal penalty function; other choices suffer from their own problems.

If (in a simulation experiment) each referee produces true marks plus some independent Gaussian noise with no bias and with variance v_r, then the performance of the above algorithm depends on the distribution of the variances v_r of the referees. For some distributions the algorithm produces an unbiased estimate of the true values with a variance which is remarkably low and essentially equal to the lowest possible variance as dictated by Information Theory, reaching the Cramer-Rao lower bound (CRLB). Note that in such a case the Maximum Likelihood Estimator (MLE) also reaches the CRLB; however, unlike the MLE, the above algorithm **does not** require prior knowledge of the variances of the referees; in fact, this particular form of the algorithm with the reciprocal function can be seen as alternating between estimations of variances of the referees (step 7) and applications of MLE with such estimated approximate variances (step 10).

4 Collusion Attacks

Although the above IF algorithm exhibits better robustness compared to the simple averaging techniques, for some distributions of variances the performance of this algorithm is very bad, with the algorithm producing an estimate of the true marks equal to the marks assigned by one of the referees. The reason for such a behavior is that the penalty function $F(d) = 1/d^2$ has a pole at $d = 0$, and thus the marks of referees act as *attractors* for the iterative procedure: if in the process of iteration the estimated marks get sufficiently close to the marks of any particular referee, the iterative procedure converges in only a few additional steps to the marks provided by that particular referee.

Worse, we have shown Rezvani, Ignjatovic, Bertino, and Jha (2013), Rezvani et al. (2015), such behavior makes the algorithm extremely vulnerable to a collusion attack. Assume that there are R referees among whom C are colluders. The colluders first do their best to estimate the true marks t_p; then $C-1$ of them report heavily skewed marks s_p while the last colluder reports values $((R-C+1)t_p + (C-1)\ s_p)/(R-1)$ as his marks. In such a case the first iteration of the procedure, which takes the mean of all marks, is very likely to produce aggregate marks very close to the marks proposed by the last attacker, causing the algorithm to quickly converge to the marks of the last attacker whose marks are still considerably skewed.

5 Data Aggregation with Protection from Collusions

In order to overcome such instability of the above IF algorithm and make it applicable to compressive sensing in wireless sensor networks in the presence of sensor faults, Chou et al. proposed Chou et al. (2013) to modify the penalty function by adding a small regularisation constant $a > 0$ and define $F(d) = 1/(d^2 + a)$. While this does make the algorithm more robust, it also has a serious drawback: if a is sufficiently large to make the algorithm stable, then the values returned by the algorithm might not differ significantly from the simple mean of the marks of all sources.

In trying to solve this problem in a more satisfactory manner, Rezvani et al. have proposed Rezvani et al. (2015) a better way to provide an initial approximation $\vec{\mu}^{(0)}$. Clearly, without knowing the true values, the algorithm cannot determine the error of each source; however, denoting again the true value of item p (in our example the true mark of a paper p) as t_p, we have that for every pair of sources r_1, r_2 (in the above example referees),

$$\begin{aligned}\sum_{1\leq p\leq P}\frac{(M(r_1,p)-M(r_2,p))^2}{P} &= \sum_{1\leq p\leq P}\frac{((M(r_1,p)-t_p)-(M(r_2,p)-t_p))^2}{P}\\ &= \sum_{1\leq p\leq P}\frac{(M(r_1,p)-t_p)^2}{P}+\sum_{1\leq p\leq P}\frac{(M(r_2,p)-t_p)^2}{P}\\ &\quad +2\sum_{1\leq p\leq P}\frac{(M(r_1,p)-t_p)(M(r_2,p)-t_p)}{P}.\end{aligned} \tag{11}$$

The first two terms on the second line are estimators for the variances v_{r_1} and v_{r_2}, and, assuming that the errors of the sources are reasonably uncorrelated, the last term on the second line should be small. In this way we obtain $\sum_{1\leq p\leq P}(M(r_1,p)-M(r_2,p))^2 \approx v_{r_1}+v_{r_2}$, which results in $R(R-1)/2$ equations in R variables $v_1, v_2, \ldots, v_R$, that can be solved in the sense of the Least Squares. We can now take as the initial approximation $\vec{\mu}^{(0)}$ of the marks the MLE estimation with the obtained approximations of the variances v_r, i.e.,

$$\mu^{(0)}(p) = \frac{\sum\limits_{1\leq r\leq R} \frac{M(r,p)}{v_r}}{\sum\limits_{1\leq r\leq R} \frac{1}{v_r}}. \tag{12}$$

Remarkably, experiments have demonstrated that, even when the errors are significantly correlated, such initial value dramatically improves the stability of the algorithm without any sacrifice in performance. It also improves its robustness against a collusion attack, because the attackers have no way of estimating the variances of other referees (Rezvani et al., 2015). However, in general, the above algorithm can have several fixed points (de Kerchove & Van Dooren, 2010); for that reason, since it does not provide a unique solution, it is not suitable for a real life deployment. Moreover, the algorithm has another serious drawback: it is not applicable to non-numerical data because it crucially depends on using a distance function, $d(r)$.

For that reason the present authors have looked for IF algorithms which are both provably convergent and also applicable to non-numerical data. This was partly addressed by Allahbakhsh and Ignjatovic (2015), Allahbakhsh et al. (2015), Allahbakhsh, Ignjatovic, Benatallah, and Motahari-Nezhad (2013) by altering the main feature of the previously introduced IF algorithms, namely by separating the process of assessment of the trustworthiness of the sources from the actual data aggregation process. We explain the main idea using a Q&A website example.

At a typical *Q & A* website each question is open for new answers for a certain period of time, say 30 days, before the question is closed; users are allowed to vote for the best answer to a particular question for an additional period of time, say 10 days, before the votes are counted and the best answer is declared. In general, there are other, concurrently open questions on the same topic and, as it can be easily observed on such websites, users with the same interest tend to vote for the best answer to a number of questions in the same field, open during the past 30 days or so. For that reason, the following policy of such a social website would not be very restrictive: only the votes of members who are "active" at the time are taken into account, and a member is considered active if he or she has cast her vote for the best answer to a certain number of questions $Q > 1$ which were recently closed. This gives an opportunity to make vote aggregation significantly more robust by deciding **simultaneously** which are the best answers to all questions which have been recently closed, using the following algorithm proposed in

Allahbakhsh and Ignjatovic (2015), Allahbakhsh et al. (2015), Allahbakhsh et al. (2013) by the present CI and his student.

Assume that there are Q recently closed questions; for each question q_i we have a corresponding list Λ_i of n_i answers, $\Lambda_i = \langle a(i,1), a(i,2), \ldots, a(i,n_i)\rangle$. We also assume that there are V voters $v_1, v_2, \ldots, v_V$. Again, for the simplicity of presentation, we assume that each voter has chosen her best answer for every question; for a sparse pattern of votes all quantities involved can be appropriately normalized, according to the total number of questions each voter has participated in choosing the best answer for, see Allahbakhsh and Ignjatovic (2015), Allahbakhsh et al. (2013), Allahbakhsh et al. (2015). The algorithm for vote aggregation is again iterative, and it simultaneously evaluates the ratings $\rho(i,k)$ of all answers to each question posed in the given interval of time as well as the trustworthiness $\tau(m)$ of each voter v_m who participated in voting during that period of time, in the following manner:

Let p be a real number, $p \geq 1$, and let us denote by $m \rightarrow i,k$ the fact that voter v_m has voted for the answer $a(i,k)$ as the best answer to question q_i. In the initial round of iteration, for each question q_i and all of its answers $a(i,k)$, $1 \leq k \leq n_i$, we simply count the number $\nu(i,k)$ of votes which $a(i,k)$ has received. We now obtain the initial ranks of answers as the normalized number of votes, $\rho^{(0)}(i,k) = \nu(i,k)/\sqrt{\sum_{1\leq j\leq n_i} \nu(i,j)^2}$; thus, for all answers $a(i,k)$ to a question q_i we have $\sum_{1\leq k\leq n_i} \rho^{(0)}(i,k)^2 = 1$. We are now again in a position to judge for every voter v_m how good his choices are, namely, to what degree their voting is in agreement with the community sentiment, and assign to them his initial trustworthiness $\tau^{(0)}(m) = \sum_{i=1}^{Q} \{\rho^{(0)}(i,k) : m \rightarrow i,k\}$, which is simply a sum of the normalized number of votes received by all the answers which he voted for. Clearly, a voter v_m will get a large initial trustworthiness only if he has chosen answers which many other community members have also chosen. In the next round of iteration of our vote aggregation procedure not every vote has an equal value, but its value depends on the trustworthiness of the voter. Thus, at every consecutive stage of iteration $n + 1$ we have:

$$\tau^{(n+1)}(m) = \sum_{1\leq i\leq Q} \{\rho^{(n)}(i,k) : m \rightarrow i,k\}; - \text{ computing the trustworthiness of voter } v_m$$

$$\rho^{(n+1)}(i,k) = \frac{\sum_{m\,:\,m\rightarrow ik} \left(\tau^{(n+1)}(m)\right)^p}{\sqrt{\sum_{1\leq j\leq n_i} \left(\sum_{m\,:\,m\rightarrow ik} \left(\tau^{(n+1)}(m)\right)^p\right)^2}};$$

$- $ computing the new rank of answer $a(i,k)$ to question q_i

iterating until $\sum_{1\leq m\leq V} \left(\tau^{(n+1)}(m) - \tau^{(n)}(m)\right)^2 < \varepsilon$. We note that the purpose of the denominator in the expression for $\rho^{(n+1)}(i,k)$ is a normalization which keeps the iteration stable and allows an elegant convergence proof by ensuring that at every stage of iteration $\sum_{1 \leq k \leq n_i} \rho^{(n)}(i,k)^2 = 1$, see Allahbakhsh and Ignjatovic (2015). The parameter p controls filtering; the larger the value of p the more the algorithm is robust against collusion attacks, but larger values also increasingly marginalize honest voters who do not vote entirely in accordance with the prevailing sentiment of the community.

With such a vote aggregation procedure the colluding voters must vote for the best answer for a significant number of other questions posed during the same period of time, and they cannot vote randomly, but must vote in accordance with the prevailing sentiment of the community, in order to receive sufficient trustworthiness. Only then can they vote differently from other voters for the answer to the question they are attacking, and hope that they can prevail over the honest voters. While this does not preclude entirely collusion attacks, it obviously makes them harder to execute.

Also note that in this case the data (the choice of the best answer) is not only non-numerical but also does not have any natural ordering. However, the same algorithm is applicable to numerical choices with values which are integers in a limited range as well as ordered choices. For example, customer feedback is usually in the range of one to five "stars" and the same applies to movie ranking. Market analyst's recommendations are an example of non-numerical but ordered choices (strong_buy < buy < neutral < sell < strong_sell). After such an iterative procedure has converged and ranks $\rho(i,k)$ of all choices have been determined, in case of numerical data one can form a weighted average of such numerical choices, with weights obtained from the ranks; in case of ordered choices it can be left to the user to choose the particular numerical values for the ordered alternatives to reflect user's preferences, and then obtain the aggregate value as a corresponding weighted average.

Allahbakhsh at al. proved that the above algorithm always converges, and extensive tests not only on simulated data but also on real data, such as the publicly available movie rating dataset *MovieLens*, have shown that in terms of robustness against large collusion attacks such an algorithm outperforms the previous IF algorithms, see Allahbakhsh et al. (2015), Allahbakhsh et al. (2013).

Moreover, for cases where we can also rely on historic data, or in a case of a refereeing process where each referee can declare his level of competence for each paper, such additional information can be included into the iterative procedure of such an algorithm in a way that preserves the proof of convergence (Allahbakhsh et al., 2013).

The continuous case, such as aggregation of measurements of sensors, appears to be a significantly harder problem. An aggregation algorithm must be robust against collusion attacks without sacrificing its performance when the sources have only stochastic errors. In fact, even in the presence of a collusion attack, if the fraction of the colluding sources is reasonably small, the algorithms should provide output values which are close to the optimal, MLE estimate based on the data obtained from the sources with stochastic errors only. Rezvani et al. have designed an algorithm which, in extensive tests, appears to meet these requirements (Rezvani, Ignjatovic, Bertino, & Jha, 2014a). This algorithm is based on an idea of propagation of credibility $cr(r)$ of one source to another source. It again takes the simple mean as the initial approximation of the aggregate values $\mu^{(0)}(p)$ and assigns equal initial variance estimates $v^{(0)}(r) = \frac{1}{(P-1)R}\sum_{s=1}^{R}\sum_{p=1}^{P}(M(s,p)-\mu^{(0)}(p))^2$ to all sources; we then repeat until convergence:

$$\mathrm{cr}^{(n+1)}(r) = \left(\prod_{j=1}^{R}\frac{\exp\left(-\frac{\frac{1}{P-1}\sum_{1\le p\le P}(M(r,p)-\mu^{(n)}(p))^2}{2v^{(n)}(j)}\right)}{\sqrt{2\pi v^{(n)}(j)}}\right)^{\frac{1}{R}}; \quad (13)$$

– computing the credibility of source r

$$\mu^{(n+1)}(p) = \sum_{i=1}^{R}\frac{\mathrm{cr}^{(n+1)}(i)}{\sum_{k=1}^{R}\mathrm{cr}^{(n+1)}(k)}M(i,p); \qquad \text{– computing the new aggregate values} \quad (14)$$

$$\mathrm{var}^{(n+1)}(r) = \frac{1}{P-1}\sum_{k=1}^{P}(M(i,k)-\mu^{(n+1)}(k))^2 \qquad \text{– computing the new variance of source} r \quad (15)$$

Thus, at each stage of the iteration, the credibility of the values supplied by a source r is assessed by estimating the likelihood that the values supplied by r might have been obtained by every other source. The credibility is defined as the geometric mean of all of these likelihoods; see Eq. (13). The heuristic underlying such methodology is that the stability of such algorithm should come from the smoothing property of taking a mean of all of these likelihoods. The geometric mean was chosen with a hope that to be able to rigorously prove that, in case of purely stochastic normally distributed unbiased errors, the algorithm converges to the MLE estimation which could have been obtained if the non-colluding sources and their exact variances were a priori known; this would clearly ensure that our algorithm has the minimal possible variance, equal to the CRLB. Figure 6 shows a typical result obtained with 25 sources; 20 sources are "honest" providing the true mark t_p of item p plus a normally and independently distributed unbiased noise with randomly chosen variances between 1 and 5. The remaining 5 sources collude, with the first 4 sources reporting skewed values $s_p = 3t_p$ and the fifth colluder the mean $((R-C+1)t_p + (C-1)\,s_p)/(R-1)$.

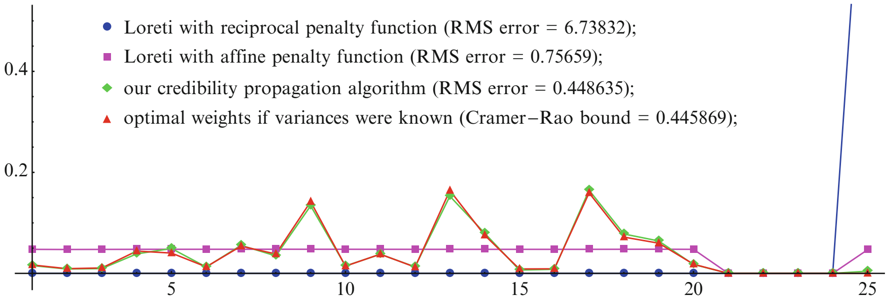


Fig. 6 Reciprocals of normalized variances of sources, estimated using: IF with $F(d) = 1d^2$ (*filled circle*), IF with $F(d) = 1 - k\ d$ (*filled square*), credibility propagation (*filled diamond*), normalized reciprocals of the true variances (*filled triangle*). Also shown are the corresponding RMS value of errors of the aggregate values (discrete values are joined by *lines* for better visual representation)

As it can be seen from Fig. 6, the weights obtained by the IF algorithm with the reciprocal penalty function $1/d^2$, (filled circle), are all essentially zero except for the weight of the last attacker which is 1 (out of range on the graph); the weights obtained by IF algorithm with the affine penalty function $F(d) = 1 - k\ d$, (filled square), are 0 for all attackers except the last one, but all other, non zero weights are essentially equal thus resulting in the simple mean of all honest sources and the last attacker. Finally, the weights produced by the algorithm based on the credibility propagation (filled diamond) are almost indistinguishable from the (normalized) reciprocals of the true variances of the "honest" sources (filled triangle), which in this case represent the optimal weights resulting in an estimator with the smallest possible variance. The RMS values of errors shown on the legend of Fig. 6 demonstrate the superiority of the credibility propagation algorithm. In fact, several IF algorithms—more than a dozen of them—were implemented and test and in all cases the algorithm by Rezvani et al. had the lowest RMS error, only slightly higher than the CRLB, even in the presence of a collusion attack. A *Mathematica* code which produced the above results is available online at http://www.cse.unsw.edu.au/~ignjat/IF.nb.

In addition, Rezvani et al. have applied ideas of the provenance of data (Lim et al., 2010) to design an iterative algorithm for computing the risk of flows and hosts in a computer network (Rezvani, Ignjatovic, Bertino, & Jha, 2014b; Rezvani, Ignjatovic, & Jha, 2013; Rezvani, Sekulic, Ignjatovic, Bertino, & Jha, 2014). For such iterative risk assessment algorithm as introduced in Rezvani et al. (2014b), Rezvani et al. were able to prove its convergence and also obtain sharp analytic estimates for its performance (Rezvani et al., 2014). Future research will aim to integrate the idea of provenance of data with IF algorithms in a single (possibly nested) iterative procedure. Such an integration should be done in a way which preserves the convergence proof of the resulting algorithm

6 Research Roadmap

In many real-life distributed systems such as social networks, rating system, participatory sensing networks and WSNs, the trustworthiness of participants has a significant role in the decision-making processes. While we believe that past results have demonstrated the potential of our IF algorithms as a robust trust framework for these distributed systems, achieving the objective requires much wider research efforts.

Most IF algorithms are still mostly “ad hoc” solutions which do not have a unified mathematical foundation. For example, in the discrete case we still lack an algorithm which, in case of domains which are integers (for example one to five star ratings) takes into account the proximity of votes, rather than just the coincidence of votes. This is clearly unsatisfactory: if a number of voters give a five star ranking to a movie, then a voter which gives it four stars should get some credit from them, and certainly more credit than a voter which gives the same movie only three stars. However, in algorithms by Allahbakhsh and Ignjatovic (2015), Allahbakhsh et al. (2015) both such dissent voters get no credit from the voters giving the movie five stars. Moreover, the degree of such credibility propagation from a voter to the voters who propose similar but not equal scores should depend on the estimated variances of the voters. It is also crucial that domain knowledge be incorporated into the data trustworthiness methodologies. For example, in a sensor network, a sensor that has been deployed for a long time may be considered less trustworthy than recently deployed sensors. Also metrics and methodologies from the area of data quality should be considered here (Reznik & Bertino, 2013).

In some distributed systems such as participatory sensing networks, preserving the privacy and anonymity of participants is mandatory (Wang, Cheng, Mohapatra, & Abdelzaher, 2013). Clearly, if the participatory networks fully anonymize the reported data, it is difficult to accurately estimate the trustworthiness of participants using the current state of our IF algorithms. Decentralization of our trust computation approach could improve the privacy of participant (Hasan, Brunie, Bertino, & Shang, 2013). Thus, proposing a decentralized privacy preserving IF algorithm for robust trust computation is an interesting open research area.

A tremendous volume of data generated by recent technological advances, referred to as *Big Data* can be used to provide data-driven decision-making. Moreover, the interconnected Big Data forms a large data redundancy which can be used to validate data trustworthiness (Labrinidis & Jagadish, 2012). An interesting research direction is to scale the IF algorithms to Big Data in order to extract hidden relationships within the data redundancy.

We will investigate applications of our IF algorithms other than just data aggregation or ranking. One such application was already implemented and tested as a part of an Honors Thesis project (D’Souza, 2011), where it was used to produce a novel recommender system. Taking as an example movie ranking, our algorithm aggregates ratings of movies provided by users, and, as we have explained, besides producing robust ratings of movies it also produces weights for users which reflect to what degree their ratings agree with the prevailing "community sentiment" ranks, as produced by our IF algorithm. We now use the observation that if two users have similar tastes, their weights must also be similar, because their movie ratings, being close to each other, must also be at a similar distance to the community sentiment ranks. Thus, to make recommendations for a particular user, we can restrict our attention only to users whose weights are close to the weight of that particular user.

In conclusion, we believe that the IF algorithms have demonstrated a promising potential for providing robust trust assessment methods for inconsistent information. Moreover, such algorithms provide a robust aggregate of such inconsistent information and can thus play a critical role in WSNs as a method of resolving a number of important problems, such as secure routing, fault tolerance, false data detection, compromised node detection, cluster head election, and outlier detection. They are also applicable to social networks, web services, and many other fields which involve inconsistent information.

References

Allahbakhsh, M., & Ignjatovic, A. (2015). An iterative method for calculating robust rating scores. *IEEE Transactions on Parallel and Distributed Systems, 26*(2), 340–350.

Allahbakhsh, M., Ignjatovic, A., Motahari-Nezhad H.R., Benatallah, B (2015). Robust evaluation of products and reviewers in social rating systems. *World Wide Web* 18(1):73–109

Allahbakhsh, M., Ignjatovic, A., Benatallah, B., & Motahari-Nezhad, H. R. (2013). Robust evaluation of products and reviewers in social rating systems. *WorldWide Web*.

Chou, C.-T., Ignjatovic, A., & Hu, W. (2013). Efficient computation of robust average of compressive sensing data in wireless sensor networks in the presence of sensor faults. *IEEE Transactions on Parallel and Distributed Systems, 24*(8), 1525–1534.

Dai, C., Rao, F.-Y., Ghinita, G., & Bertino, E. (2011). Privacy-preserving assessment of location data trustworthiness. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11 (pp. 231–240).

Dai, C., Rao, F.-Y., Truta, T. M., & Bertino, E. (2012). Privacy-preserving assessment of social network data trustworthiness. In *CollaborateCom* (pp. 97–106).

de Kerchove, C., & Van Dooren, P. (2007). Iterative filtering for a dynamical reputation system. *CoRR* [abs/0711.3964].

De Kerchove, C., & Van Dooren, P. (2008). Reputation systems and optimization. *Siam News, 41*(2), 2008.

de Kerchove, C., & Van Dooren, P. (2010). Iterative filtering in reputation systems. *The SIAM Journal on Matrix Analysis and Applications, 31*(4), 1812–1834.

D'Souza, N. (2011). Applications of adaptive averages: Recommendation systems. Honours Thesis, UNSW.

Hasan, O., Brunie, L., Bertino, E., & Shang, N. (2013). A decentralized privacy preserving reputation protocol for the malicious adversarial model. *IEEE Transactions on Information Forensics and Security, 8*(6), 949–962.

Ignjatovic, A., Foo, N., & Lee, C. T. (2008). An analytic approach to reputation ranking of participants in online transactions. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI)*.

Ignjatovic, A., Lee, C. T., Compton, P., Cutay, C., & Guo, H. (2009). Computing marks from multiple assessors using adaptive averaging. In *International Conference on Engineering Education (ICEE)*.

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *The Proceedings of the VLDB Endowment, 5*(12), 2032–2033.

Laureti, P., Moret, L., Zhang, Y.-C., & Yu, Y.-K. (2006). Information filtering via Iterative Refinement. *EPL (Europhysics Letters), 75*, 1006–1012.

Lee, C.-T., Milic-Frayling, N., Vinay, V., Rodrigues, E.M., Ignjatovic, A., & Kazai, G. (2009). Measuring system performance and topic effectiveness using generalized means with adaptive weights. In *ACM Conference on Information and Knowledge Management (CIKM)* (pp. 2033–2036). New York, NY, USA: ACM.

Lee, C.-T., Rodrigues, E.M., Kazai, G., Ignjatovic, A., & Milic-Frayling, N. (2009). Model for voter scoring and best answer selection in community q & a services. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI)* (pp. 116–123). New York, NY, USA: ACM.

Lim, H.-S., Ghinita, G., Bertino, E., & Kantarcioglu, M. (2012). A game-theoretic approach for high-assurance of data trustworthiness in sensor networks. In *2012 I.E. 28th International Conference on Data Engineering (ICDE)* (pp. 1192–1203).

Lim, H.-S., Moon, Y.-S., & Bertino, E. (2010). Provenance-based trustworthiness assessment in sensor networks. In *Proceedings of the Seventh International Workshop on Data Management for Sensor Networks*, DMSN ’10 (pp. 2–7).

Reznik, L., & Bertino, E. (2013). Poster: Data quality evaluation: Integrating security and accuracy. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS ’13 (pp. 1367–1370). New York, NY, USA: ACM.

Rezvani, M., Ignjatovic, A., Bertino, E., & Jha, S. (2013). Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks. Technical Report UNSW-CSE-TR-201319, School of Computer Science and Engineering, UNSW.

Rezvani, M., Ignjatovic, A., Bertino, E., & Jha, S. (2015). Secure Data Aggregation Technique for Wireless Sensor Networks in the Presence of Collusion Attacks. IEEE Trans. Dependable Sec. Comput. 12(1):98–110.

Rezvani, M., Ignjatovic, A., Bertino, E., & Jha, S. (2014a). Credibility propagation for robust data aggregation in WSNs. Technical Report UNSW-CSE-TR-201414, School of Computer Science and Engineering, UNSW.

Rezvani, M., Ignjatovic, A., Bertino, E., & Jha, S. (2014b). Provenance-aware security risk analysis for hosts and network flows. In *IEEE/IFIP Network Operations and Management Symposium (NOMS)*.

Rezvani, M., Ignjatovic, A., & Jha, S. (2013). Iterative security risk analysis for network flows based on provenance and interdependency. In *IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS 13)* (pp. 286–288).

Rezvani, M., Ignjatovic, A., Bertino, E., & Jha, S. (2015). Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks. *IEEE Transactions on Dependable and Secure Computing, 12*(1), 98–110.

Tang, L.-A., Yu, X., Kim, S., Han, J., Hung, C.-C., & Peng, W.-C. (2010). Tru-Alarm: Trustworthiness analysis of sensor networks in cyber-physical systems. In *Proceedings of the 2010 I.E. International Conference on Data Mining*, ICDM ’10 (pp. 1079–1084).

Wang, X., Cheng, W., Mohapatra, P., & Abdelzaher, T. F. (2013). Artsense: Anonymous reputation and trust in participatory sensing. In *INFOCOM, 2013 Proceedings IEEE* (pp. 2517–2525).

Yu, Y.-K., Zhang, Y.-C., Laureti, P., & Moret, L. (2006). Decoding information from noisy, redundant, and intentionally distorted sources. *Physica A Statistical Mechanics and its Applications, 371*, 732–744.

Weaponized Crowdsourcing: An Emerging Threat and Potential Countermeasures

James Caverlee and Kyumin Lee

1 Introduction

The crowdsourcing movement has spawned a host of successful efforts that organize large numbers of globally-distributed participants to tackle a range of tasks, including crisis mapping (e.g., Ushahidi), translation (e.g., Duolingo), and protein folding (e.g., Foldit). Alongside these specialized systems, we have seen the rise of general-purpose crowdsourcing marketplaces like Amazon Mechanical Turk and Crowdflower that aim to connect task requesters with task workers, toward creating new crowdsourcing systems that can intelligently organize large numbers of people. However, these positive opportunities have a sinister counterpart: what we dub "Weaponized Crowdsourcing". Already we have seen the first glimmers of this ominous new trend—including large-scale "crowdturfing", wherein masses of cheaply paid shills can be organized to spread malicious URLs in social media (Grier, Thomas, Paxson, & Zhang, 2010; Lee & Kim, 2012), form artificial grass-roots campaigns ("astroturf") (Gao et al., 2010; Lee, Caverlee, Cheng, & Sui, 2013), spread rumor and misinformation (Castillo, Mendoza, & Poblete, 2011; Gupta, Lamba, Kumaraguru, & Joshi, 2013), and manipulate search engines. A recent study finds that 90 % of tasks on many crowdsourcing platforms are for crowdturfing (Wang et al., 2012), and our initial research (Lee, Tamilarasan, & Caverlee, 2013) shows that most malicious tasks in crowdsourcing systems target

J. Caverlee (✉)
Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843, USA
e-mail: caverlee@cse.tamu.edu

K. Lee
Department of Computer Science, Utah State University, Logan, UT 84322, USA
e-mail: kyumin.lee@usu.edu

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_4

either online communities (56 %) or search engines (33 %). Unfortunately, little is known about Weaponized Crowdsourcing as it manifests in existing systems, nor what are the ramifications on the design and operation of emerging socio-technical systems. Hence, this chapter shall focus on key research questions related to Weaponized Crowdsourcing as well as outline the potential of building new preventative frameworks for maintaining the information quality and integrity of online communities in the face of this rising challenge.

2 Background

In a crowdsourcing marketplace like Amazon Mechanical Turk, a participant can be a requester: one who posts a task description and recruits workers to solve this task; a worker: one who performs a task and is typically compensated for this work; or both a requester and a worker. These tasks are usually difficult or computationally expensive for computers to solve, but relatively easy for humans. In many crowdsourcing marketplaces, complex tasks are typically broken down into simpler tasks that can be completed by an individual worker in a reasonable amount of time. For example, validating the quality of a transcribed script from an audio source (as in the case of using crowd workers to construct subtitles for a previously un-subtitled video) may be assigned to multiple, overlapping workers who tackle parts of the task: an individual worker may transcribe a 10-second clip; other workers may repeat this work or verify the quality of this work; eventually, the full-time transcription may then be completed and given to a final worker (or collection of workers) to validate. Workers in these crowdsourcing marketplaces are often cheaply paid and treated as interchangeable by requesters; and since workers are often drawn from the entire world, tasks may be completed at any time by a distributed workforce.

In light of these perceived benefits, we should note that a crowdsourcing marketplace is itself a social system that provides many of the advantages of social systems. That is, the reliance on users themselves to "maintain the community" can lead to many positive effects, including growth in the size and capabilities of the system, the emergence of recognized experts within the system (e.g., workers who are especially fast or precise, or have other desirable qualities), and the flexibility to tackle problems beyond the scope of the original system designers. And yet this relative openness and reliance on users to drive the system may lead to new risks and growing concerns. In particular, we highlight the challenge of *weaponized crowdsourcing*, in which malicious requesters misuse this openness to post tasks that spread malicious URLs in social media, form artificial grassroots campaigns, spread rumor and misinformation, and manipulate search engines. In the same vein, unethical workers will perform these tasks, often by propagating manipulated content to target sites such as social media sites, search engines, and review sites, resulting in the degradation of information quality and the integrity of these online communities.

To illustrate, Fig. 1 shows a typical workflow, wherein (1) a requester first posts one of these tasks (here, a "crowdturfing" task), (2) identifies the appropriate workers to complete this task, and (3) finally, these workers spread their misinformation in a target venue like a social network, a forum/review site, a search engine, or blog. Figure 2 shows an example of a crowdturfing task description that we sampled from the crowdsourcing platform Microworkers.com. This task requires workers to have at least 50 Twitter followers, search for a certain keyword on Google, and then click on a website in the search results. In addition, it requires the workers to retweet an article in the website to Twitter. This task targets not only a search engine but also a social media site, hoping to boost the target website's rank by artificially manipulating both a search engine and a social network. At the time of our collection, 222 workers had completed this task for $0.60 per task completion.

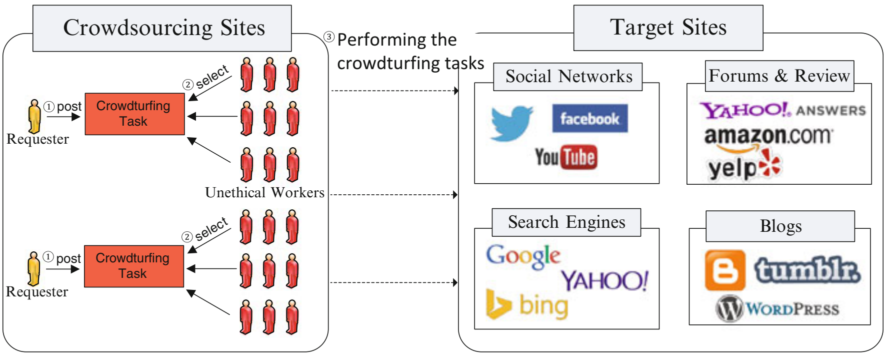


Fig. 1 The interactions between malicious requesters and unethical workers

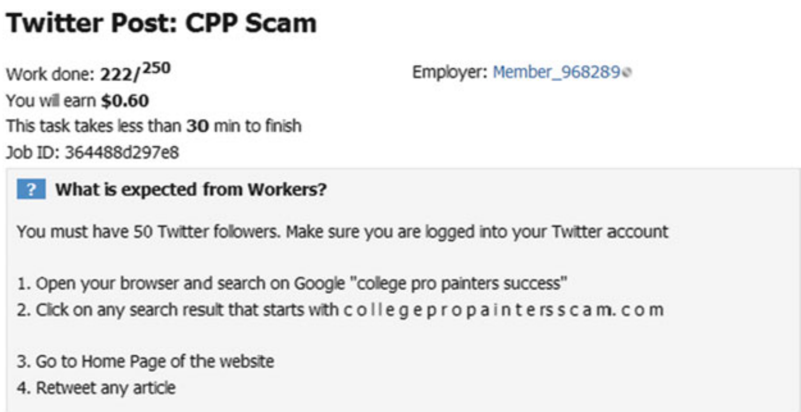


Fig. 2 A crowdturfing task description posted to Microworkers.com

3 Weaponized Crowdsourcing: An Investigation

In this section, we investigate the emerging threat of weaponized crowdsourcing through a multi-part analysis. We sample and report on tasks from existing crowdsourcing marketplaces, characterize the market size, present an initial categorization of the types of campaigns, and investigate the demographics of both malicious requesters and unethical workers.

3.1 Datasets Collected from Crowdsourcing Sites

In order to conduct our analysis about weaponized crowdsourcing, we collected 505 campaigns totaling 63,042 tasks by crawling three popular Western crowdsourcing sites that host clear examples of crowdturfing campaigns: Microworkers.com, ShortTask.com, and Rapidworkers.com during a span of 2 months in 2012. Almost all of the campaigns in these sites are crowdturfing campaigns, and these sites are active in terms of the number of new campaigns being posted. Note that even though Amazon Mechanical Turk is one of the most popular crowdsourcing sites, we excluded it in our study because it has only a small number of crowdturfing campaigns and its terms of service officially prohibits the posting of crowdturfing campaigns. Perhaps surprisingly, Microworkers.com is ranked by Alexa.com at the 4,699th most popular website while Amazon Mechanical Turk is ranked 7,173. We additionally collected 89,667 campaign descriptions and 31,021 corresponding user profiles between July and August 2013 from Fiverr.com, a global microtask marketplace that as of April 2014 is the 130th most visited site in the world according to Alexa (2014).

3.2 Market Size of Weaponized Crowdsourcing

To analyze the market size of crowdturfing campaigns in Microworkers.com, We collected 144 requesters' profiles and 4,012 workers' profiles—where all campaigns in our sample data are crowdturfing tasks and other researchers have found that 89 % of campaigns hosted at Microworkers.com are indeed crowdturfing tasks (Wang et al., 2012).

The 4,012 workers have completed 2,962,897 tasks and earned $467,453 so far, which suggests the entirety of the crowdturfing market is substantial. Interestingly, the average price per task is higher on a crowdturfing site (for Microworkers.com, the average is $0.51) than on the legitimate Amazon Mechanical Turk where 90 percent of all tasks pay less than $0.10 (Ipeirotis, 2010).

Table 1 presents the maximum, average, median and minimum number of tasks done, how much they have earned, and the account longevity for the sampled

Table 1 Characteristics of Crowdturfing Workers in Microworkers.com

	# Of tasks	Total earned ($)	Longevity (day)
Max	24,016	3,699	1,215
Avg	738	117	368
Median	166	23	320
Min	10	1	5

Table 2 Characteristics of Crowdturfing Requesters in Microworkers.com

	# Of campaigns	# Of paid tasks	Longevity (day)
Max	4,137	455,994	1,091
Avg	68	7,030	329
Median	7	306	259
Min	1	0	3

workers. We observe that there are professional workers who have earned reasonable money from the site to survive. For example, a user who earned $3,699 for slightly more than 3 years (1,215 days) lives in Bangladesh where the GNI (Gross National Income) per capita is $770 in 2011 as estimated by the World Bank TradingEconomics (2011). Surprisingly, she has earned even more money per year ($1,120) than the average income per year ($770) of a person in Bangladesh.

The requesters' profile information reveals their account longevity, number of paid tasks and expense/cost for campaigns. As shown in Table 2, many workers have created multiple campaigns with lots of tasks (on an average—68 campaigns and 7,030 paid tasks). The most active requester in our dataset initiated 4,137 campaigns associated with 455,994 paid tasks. In other words, he has spent a quarter million dollar ($232,557)—again a task costs $0.51 on an average. In total, 144 requesters have created 9720 campaigns with 1,012,333 tasks and have paid a half million dollars ($516,289). This sample analysis shows us how the dark market is big enough to tempt users from developing countries to become workers.

3.3 Types of Campaigns

We next analyze types of crowdturfing campaigns to understand the tactics of the requesters. Hence, we first manually grouped the 505 campaigns collected from Microworkers.com, ShortTask.com, and Rapidworkers.com into the following five categories:

- **Social Media Manipulation [56 %]:** The most popular campaigns target social media. Example campaigns request workers to spread a meme through social media sites such as Twitter, click the "like" button of a specific Facebook profile/ product page, bookmark a webpage on Stumbleupon, answer a question with a

Fig. 3 An example social media manipulation campaign

Twitter Post: Getmine

1. Go to http://getminecraftforfree.org
2. Click on the tweet button on the left side
3. Tweet something like "how to play minecraft for free" or "check this site out"
4. Include link to the site

link on Yahoo! Answers, write a review for a product at Amazon.com, or write an article on a personal blog. An example campaign is shown in Fig. 3, where workers are requested to post a tweet including a specific URL.

- **Sign Up [26 %]:** Requesters ask workers to sign up on a website for several reasons, for example to increase the user pool, to harvest user information like name and email, and to promote advertisements.
- **Search Engine Spamming [7 %]:** For this type of campaign, requesters seek to increase the visibility of a particular web page by creating artificial clicks, which are typically interpreted by major search engines as a signal of page quality. A typical task requires a worker to search for a specified keyword on a major search engine (like Google or Bing). The workers should then scan through the search engine results and click on the specified link (which is affiliated with the campaign's requester), towards increasing the number of clicks on the page and ultimately increasing the rank of the page in future searches, as shown in Fig. 2.
- **Vote Stuffing [4 %]:** Requesters ask workers to cast votes. In one example, the requester asked workers to vote for "Tommy Marsh and Bad Dog" to get the best blue band award in the Ventura County Music Awards (which the band ended up winning!).
- **Miscellany [7 %]:** Finally, a number of campaigns engaged in some other activity: for example, some requested workers to download, install, and rate a particular software package; others requested workers to participate in a survey or join an online game.

Next, we also analyzed 121 crowdturfing campaigns randomly sampled from Fiverr.com, and manually grouped them into three categories:

- **Social Media Targeting Campaigns [54 %]:** These crowdturfing campaigns targeted social media sites such as Facebook, Twitter and Youtube. The main purpose of these campaigns are to artificially increase number of friends or followers on these sites, promote pre-selected messages or URLs, and increase the number of views associated with requesters' videos. The requesters expect these manipulations to result in more effective information propagation, higher conversion rates, and positive social signals for their web pages and products.
- **Search Engine Targeting Campaigns [38 %]:** These campaigns targeted search engines by artificially creating backlinks for a targeted site. This is a traditional attack against search engines. However, instead of creating backlinks on their own, the requesters take advantage of workers to create a large number of

backlinks so that the targeted page will receive a higher PageRank score (and have a better chance of ranking at the top of search results). Interestingly, a worker (also called a seller in Fiverr.com) has earned $3 million for helping running search engine targeting campaigns with 100 % positive ratings and more than 47,000 positive comments from requesters who hired the worker. This fact indicates that the search engine targeting campaigns are popular and profitable.

- **User Traffic Targeting Campaigns [8 %]:** The last campaigns aimed to get user traffic to a targeted site. Workers generated user traffic (visitors) for a pre-selected website or web page. With higher traffic, the requesters hope to abuse Google AdSense, which provides advertisements on each requester's web page, when the visitors click the advertisements. Another goal of these campaigns is for the visitors to purchase products from the pre-selected page.

From the analysis of types of crowdturfing campaigns, we can see that most existing crowdturfing campaigns have targeted social media sites and search engines, which raises natural concerns about the information quality and community trust of these systems.

3.4 Countries of Requesters and Workers

Next we analyze where requesters and workers were from in Microworkers.com and Fiverr.com. Do workers and requesters have different country distributions? Can we observe different country distributions of requesters and workers who were involved in crowdturfing campaigns in Microworkers.com and Fiverr.com?

To answer these research questions, we first analyze the countries of workers and requesters in Microworkers.com. From the 4,012 workers' profile information in Microworkers.com, we found that they are from 75 countries. Especially, 83 % of the workers are from the top-10 countries as shown in Fig. 4a. An interesting

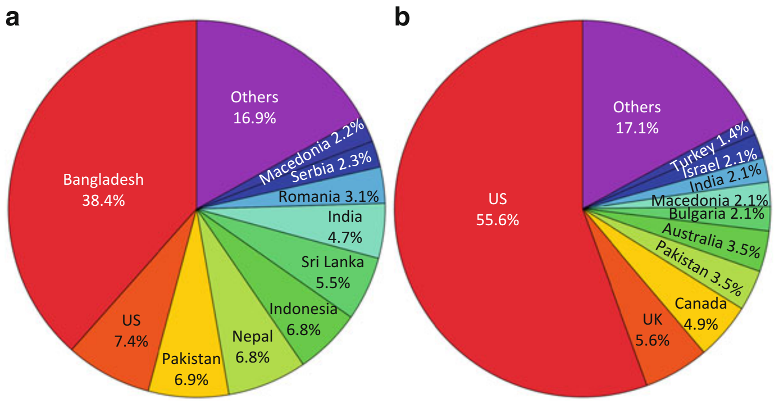


Fig. 4 Top 10 countries of workers and requesters of crowdturfing campaigns in Microworkers.com. (**a**) Workers, (**b**) requesters

observation is that a major portion of the workers in Microworkers.com are from Bangladesh—where 38 % workers (1,539 workers) come from—whereas in Amazon Mechanical Turk over 90 % workers are from the United States and India Ross, Irani, Silberman, Zaldivar, and Tomlinson (2010).

However, requesters in Microworkers.com have a different country distribution. We found that the requesters are from 31 countries. Interestingly, 55 % of the requesters are from the United States, and 70 % of the requesters are from the English-speaking countries: United States, UK, Canada, and Australia. Figure 4b shows the top-10 countries which have the highest portion of requesters. We can see an imbalance between the country of origin of requesters and of the workers, but that the ultimate goal is to propagate artificial content through the English-speaking web.

Next, we analyze countries of workers and requesters in Fiverr.com, and compare their country distribution with country distribution of workers and requesters in Microworkers.com. Interestingly, the most frequent workers who performed crowdturfing tasks were from the United States (35.8 %) as shown in Fig. 5a. The next largest group of workers is from India (10.5 %), followed by Bangladesh (6.5 %) and the United Kingdom (5.9 %). Overall, the majority of workers (52 %) were from western countries. This distribution is very different from country distribution of workers in Microworkers.com in which the most frequent workers were from Bangladesh. This observation might imply that Fiverr.com is more attractive than Microworkers.com for U.S. residents since a worker in Fiverr.com earns higher income (at least $5 per task) than a worker in Microworkers.com (average $0.50 per task). The country distribution of requesters in Fiverr.com (as shown in Fig. 5b) is similar with a country distribution of requesters in Microworkers.com, in which the majority of requesters were from English-speaking countries.

So far, we have investigated the weaponized crowdsourcing market size, examined the distribution of tasks on two platforms, and seen how these platforms attract both workers and requesters from around the world to target successful social and web communities.

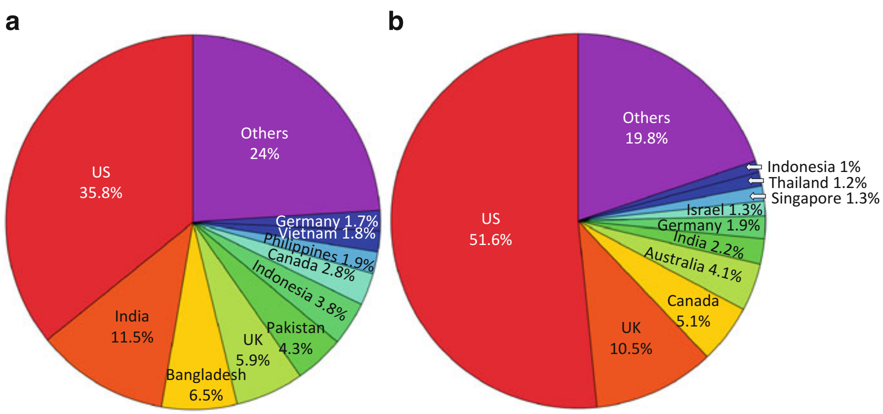


Fig. 5 Top 10 countries of workers and requesters of crowdturfing campaigns in Fiverr.com. (**a**) Workers, (**b**) requesters

4 Preventive Approaches

Given the scale and reach of existing weaponized crowdsourcing marketplaces and concerns over how they may grow in the future, we turn in this section to a discussion of possible preventative approaches for mitigating their impact on socio-technical systems. Our goal is to highlight approaches to detect and prevent the weaponized crowdsourcing problem. Specifically, we highlight three approaches: (i) an approach to detect crowdturfing tasks at the source (in the crowdsourcing platform itself); (ii) an approach to detect accounts of crowd workers who performed crowdturfing tasks in a target site (by looking at the impacts of these tasks in their target); and (iii) a crowdsourcing approach that aims to use the crowd itself to monitor and police itself. Then we turn to a discussion of future steps toward improving our defenses against weaponized crowdsourcing.

4.1 Automatic Crowdturfing Task Detection

One way to solve the crowdturfing problem is to automatically detect and delete crowdturfing tasks to prevent workers from performing the crowdturfing tasks. To measure whether automatically detecting crowdturfing tasks is possible, we present here a prototype crowdturfing task detection classifier.

First, we randomly sampled 1,550 distinct tasks from Fiverr.com and manually labeled them as either a legitimate task or a crowdturfing task. As we described in Sect. 3.3, we found that 121 out of 1,550 tasks were crowdturfing tasks. This labeled dataset was converted to feature values to train and test our SVM-based classifier. Our feature set consists of the title of a task, the task's description, a top level category, a second level category (each task at Fiverr.com is categorized to a top level and then a second level—e.g., "online marketing" as the top level and "social marketing" as the second level), ratings associated with a task, the number of votes for a task, a task's longevity and so on (detailed information can be found in Lee, Webb, & Ge, 2014). For the title and job description of a task, we converted these texts into bag-of-word models in which each distinct word becomes a feature. We also used *tf-idf* to measure values for these text features.

Then, we trained and tested our SVM-based classifier with tenfold cross-validation. Its classification result is shown in Table 3, where we can see a 97.35 % accuracy, 0.974 F_1, 0.008 false positive rate (FPR), and 0.248 false negative rate (FNR). This positive result shows that our classification approach works well.

Table 3 SVM-based classification result

Accuracy	F_1	FPR	FNR
97.35 %	0.974	0.008	0.248

Fig. 6 Word cloud of crowdturfing tasks

We also applied this classifier to a larger testing set containing 87,818 tasks. In this experiment, the 1,550 tasks were used as a training. We built the SVM-based classifier with the training set and predicted class labels of the tasks in the testing set. 19,904 of the 87,818 tasks were predicted as crowdturfing tasks. To verify whether the predicted 19,904 crowdturfing tasks are real crowdturfing tasks, we manually scanned the titles of all of these tasks and confirmed that our approach worked well.

To understand and visualize what terms crowdturfing tasks often contain, we generated a word cloud of titles for these 19,904 crowdturfing tasks. First, we extracted the titles of the tasks and tokenized them to generate unigrams. Then, we removed stop words. Figure 6 shows the word cloud of crowdturfing tasks. The most popular terms are online social network names (e.g., Facebook, Twitter, and YouTube), targeted goals for the online social networks (e.g., likes and followers), and search engine related terms (e.g., backlinks, website, and Google). This word cloud also helps confirm that our classifier accurately identified crowdturfing tasks.

The experimental results confirm that automatically detecting crowdturfing tasks are possible. We expect this approach would filter crowdturfing tasks before workers take the jobs.

4.2 Tracking Manipulated Content and Detecting Workers in Social Media

Another way to solve the crowdturfing problem is to detect crowd workers' accounts in target sites. By linking manipulated content such as URLs and message templates to a target site, we would identify crowd workers' accounts in the target site. By learning these crowd workers' behaviors in the target site, we may automatically detect accounts of crowd workers who have performed crowdturfing tasks. To test this possibility, we selected 65 campaigns, which targeted Twitter, from 505 campaigns collected from Microworkers.com, ShortTask.com, and Rapidworkers.com. There were two types of Twitter related crowdturfing campaigns—campaigns which ask to post a tweet and the ones which ask to follow a user.

- **Tweeting about a link:** These tasks ask the Twitter workers to post a tweet including a specific URL (as in the example in Fig. 3). The objective is to spread a URL to other Twitter users, and thereby increase the number of clicks on the URL.
- **Following a Twitter user:** The second task type requires a Twitter worker to follow a requester's Twitter account. These campaigns can increase the visibility of the requester's account (for targeting larger future audiences) as well as impacting link analysis algorithms (like PageRank and HITS) used in Twitter search or in general Web search engines that incorporate linkage relationships in social media.

Next we tracked the Twitter accounts who participated in these campaigns. For campaigns of the first type, we used the Twitter search API to find all Twitter users who had posted the URL. For campaigns of the second type, we identified all users who had followed the requester's Twitter account. In total, we identified 2,864 Twitter workers. For these workers, we additionally collected their Twitter profile information, most recent 200 tweets, and social relationships (followings and followers).

In order to compare how these workers' properties are different from non-workers, we randomly sampled 10,000 Twitter users. Since we have no guarantees that these sampled users are indeed non-workers, we monitored the accounts for 1 month to see if they were still active and not suspended by Twitter. After 1 month, we found that 9,878 users were still active. In addition, we randomly selected 200 users out of the 9,878 users and manually checked their profiles, and found that only 6 out of 200 users seemed suspicious. Based on these verifications, we labeled the 9,878 users as non-workers. Even though there is a chance of a false positive in the non-worker set, the results of any analysis should give us, at worst, a lower bound since the introduction of possible noise would only degrade our results.

To build a crowd worker detection classifier, we converted the dataset containing information of 2,864 workers and 9,878 non-workers to feature values. Our features consist of four feature groups:

- **User Demographics (UD):** features extracted from descriptive information about a user and his account.
- **User Friendship Networks (UFN):** features extracted from friendship information such as the number of followings and followers.
- **User Activity (UA):** features representing posting activities.
- **User Content (UC):** features extracted from posted tweets.

Table 4 Features

Group	Feature
UD	The length of the screen name
UD	The length of description
UD	The longevity of the account
UD	Has description in profile
UD	Has URL in profile
UFN	The number of followings
UFN	The number of followers
UFN	The ratio of the number of followings and followers
UFN	The percentage of bidirectional friends: $\frac{\lvert followings \cap followers \rvert}{\lvert followings \rvert}$ and $\frac{\lvert followings \cap followers \rvert}{\lvert followers \rvert}$
UA	The number of posted tweets
UA	The number of posted tweets per day
UA	| links | in tweets / | tweets |
UA	| hashtags | in tweets / | tweets |
UA	| @username | in tweets / | tweets |
UA	| rt | in tweets / | tweets |
UA	| tweets | / | recent days |
UA	| links | in tweets / | recent days |
UA	| hashtags | in tweets / | recent days |
UA	| @username | in tweets / | recent days |
UA	| rt | in tweets in tweets / | recent days |
UA	| links | in RT tweets / | RT tweets |
UC	The average content similarity over all pairs of tweets posted: $\frac{\sum similarity(a,b)}{\lvert \text{set of pairs in tweets} \rvert}$, where $a, b \in$ set of pairs in tweets
UC	The ZIP compression ratio of posted tweets: $\frac{uncompressed\ size\ of\ tweets}{compressed\ size\ of\ tweets}$
UC	68 LIWC features (Pennebaker, Francis, & Booth, 2001) which are Total Pronouns, 1st Person Singular, 1st Person Plural, 1st Person, 2nd Person, 3rd Person, Negation, Assent, Articles, Prepositions, Numbers, Affect, Positive Emotions, Positive Feelings, Optimism, Negative Emotions, Anxiety, Anger, Sadness, Cognitive Processes, Causation, Insight, Discrepancy, Inhibition, Tentative, Certainty, Sensory Processes, Seeing, Hearing, Touch, Social Processes, Communication, Other References to People, Friends, Family, Humans, Time, Past Tense Verb, Present Tense Verb, Future, Space, Up, Down, Inclusive, Exclusive, Motion, Occupation, School, Job/Work, Achievement, Leisure, Home, Sports, TV/Movies, Music, Money, Metaphysical States, Religion, Death, Physical States, Body States, Sexual, Eating, Sleeping, Grooming, Swearing, Nonfluencies, and Fillers

From the four groups, we generated a total 92 features as shown in Table 4 (detailed information can be found in Lee, Tamilarasan, et al., 2013).

Using tenfold cross-validation approach and these feature groups, we tested 30 classification algorithms using the Weka machine learning toolkit (Witten & Frank, 2005). To test which classification algorithm returns the highest accuracy, we ran over 30 classification algorithms such as Naive Bayes, Logistic Regression and SMO (SVM) with the default setting. Their accuracies ranges from 86 to 91 %.

Table 5 Worker detection: results

Classifier	Accuracy	F_1	AUC	FPR	FNR
Random Forest	93.26 %	0.966	0.955	0.036	0.174

Tree-based classifiers showed the highest accuracy results. In particular, Random Forest produced the highest accuracy which was 91.85 %. By changing input parameter values of Random Forest, we achieved 93.26 % accuracy and 0.932 F_1 as shown in Table 5.

The experimental results confirm that we can automatically detect accounts of crowd workers who performed crowdturfing tasks.

4.3 Crowdsourced Mitigation

Another possible approach is to mobilize the crowd itself to mitigate the threat of weaponized crowdsourcing. But how can a crowd be organized to police itself? In one direction, we could hire crowd workers whose job is to verify whether a task is crowdturfing or not. This approach can be combined with the above approaches. For example, a crowdturfing task detector could give us a probabilistic assessment of each task (e.g., task A would be a crowdturfing task with 80 % probability). Since sometimes the detector may give us some false negatives, predicted crowdturfing tasks with a low probability would be passed to crowd workers and verified to build a more accurate crowdturfing detection system. A similar work to detect social spammers by crowd workers was studied by Wang et al. (2013).

4.4 Discussion

So far, we have introduced several *algorithmic* approaches for maintaining the information quality and integrity of online communities in the face of weaponized crowdsourcing. We now turn to a forward-looking discussion of other socio-technical approaches including collaboration among crowdsourcing service providers, target companies (e.g., social media and search engine companies), and the government.

First, we suggest creating and maintaining a common repository where employees of crowdsourcing sites and researchers store crowdturfing task descriptions containing manipulated content (e.g., URLs, template messages). Email and web service providers already maintain blacklists to store malicious web page URLs for spam, phishing and malware software distribution. A new repository for crowdturfing tasks would be helpful for employees of crowdsourcing and target sites, and for researchers to actively detect and prevent crowdturfing tasks, manipulated content, and participants.

Second, we have to think of how to increase the cost of running crowdturfing campaigns and how to discourage workers from participating in these campaigns. For example, we could imagine forfeiting malicious requesters' credits and blocking their IP addresses, as well as suspending unethical workers' accounts and blocking their IP addresses. When a user creates an account in a crowdsourcing site, we could require providing an email account and passing a Captcha so that we can delay these malicious requesters and workers from creating accounts and discourage running and participating in crowdturfing campaigns.

An interesting observation that we learned from this work is there are several crowdsourcing sites where almost all tasks are crowdturfing tasks. These crowdsourcing site providers intentionally do not prohibit posting crowdturfing tasks because these providers earn commission (about 20 %) from requesters. In addition, as we mentioned in Sect. 3.2, a crowdturfing task is five times more expensive than a legitimate task, further encouraging these crowdsourcing platforms to allow these crowdturfing tasks. To solve this problem, another potential effort is for governments or specialized organizations to start monitoring crowdsourcing sites for these weaponized crowdturfing tasks, and then to advocate for a strong response (e.g., bringing public pressure to shut down these sites).

5 Conclusion

In this chapter, we have highlighted the challenges presented by weaponized crowdsourcing and begun a discussion of potential countermeasures. As crowdsourcing platforms and systems continue to grow in complexity, variety, and reach, we can naturally anticipate the continued challenge and maturation of threats posed by weaponized crowdsourcing. Moving forward, we believe that weaponized crowdsourcing research is poised to make major breakthroughs in the years to come due to the growing interest and collaboration of researchers and practitioners across disciplines toward improving the transparency and trust of social media and online interactions.

Contributions Portions of this chapter are based on work that appeared in the 2013 and 2014 International AAAI Conference on Weblogs and Social Media (ICWSM) (Lee, Tamilarasan, et al., 2013; Lee et al., 2014).

References

Alexa (2014). Fiverr.com site info—alexa. http://www.alexa.com/siteinfo/fiverr.com.
Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. In *WWW*.
Gao, H., Hu, J., Wilson, C., Li, Z., Chen, Y., & Zhao, B. Y. (2010). Detecting and characterizing social spam campaigns. In *Proceedings of the 10th annual conference on Internet measurement (IMC)*.
Grier, C., Thomas, K., Paxson, V., & Zhang, M. (2010). @spam: The underground on 140 characters or less. In *CCS*.

Gupta, A., Lamba, H., Kumaraguru, P., & Joshi, A. (2013). Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In *WWW Companion*.

Ipeirotis, P. G. (2010). Analyzing the amazon mechanical turk marketplace. In *XRDS*, (Vol. 17, pp. 16–21).

Lee, K., Caverlee, J., Cheng, Z., & Sui, D. Z. (2013). Campaign extraction from social media. *ACM Transactions on Intelligent Systems and Technology, 5*, 9:1–9:28.

Lee, K., Tamilarasan, P., & Caverlee, J. (2013). Crowdturfers, campaigns, and social media: Tracking and revealing crowdsourced manipulation of social media. In *ICWSM*.

Lee, K., Webb, S., & Ge, H. (2014).The dark side of micro-task marketplaces: Characterizing fiverr and automatically detecting crowdturfing. In *ICWSM*.

Lee, S., & Kim, J. (2012). Warningbird: Detecting suspicious urls in twitter stream. In *NDSS*.

Pennebaker, J., Francis, M., & Booth, R. (2001). *Linguistic inquiry and word count*. Mahwah: Erlbaum Publishers.

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI Extended Abstracts on Human Factors in Computing Systems*.

TradingEconomics (2011). Gni per capita; atlas method (us dollar) in bangladesh. http://www.tradingeconomics.com/bangladesh/gni-per-capita-atlas-method-us-dollar-wb-data.html.

Wang, G., Mohanlal, M., Wilson, C., Wang, X., Metzger, M. J., Zheng, H., et al. (2013). Social turing tests: Crowdsourcing sybil detection. In *NDSS*.

Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H., et al. (2012). Serf and turf: crowdturfing for fun and profit. In *WWW*.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*, 2nd ed. New York: Morgan Kaufmann.

The Structures of Twitter Crowds and Conversations

Marc A. Smith, Itai Himelboim, Lee Rainie, and Ben Shneiderman

1 Summary of Findings

Social media promises to provide access to a vast variety of human interactions, important and trivial. More than traditional electronic media or interpersonal contact, social media allows people to find and interact based on common interests rather than physical proximity. Billions of people have embraced these tools, entering social media spaces to exchange trillions of messages. Social media interactions may not be as rich as face-to-face interactions, but they offer access to a wide range of people and topics. Success has led to new problems, as social media offers too many contacts, too many interactions, and poor tools for filtering and gaining an overview of the larger landscape of communication. Social media is created and consumed through tools that limit the observer's view to individual messages or short chains of messages and replies. The leaf and the branch of social media is visible, but not the tree or the forest. The result is an information and interaction deluge. The overwhelming amount of data and the limited ways to understand it can be seen as a negative consequence of social media. For many ordinary users social media is an incomprehensible torrent. Proposed solutions, such as automatic filters that select relevant information for us, are often seen as worse than the problem it is meant to solve. "Filter bubbles" can trap users in homogeneous

M.A. Smith (✉)
Social Media Research Foundation, Belmont, CA, USA
e-mail: marc@connectedaction.net

I. Himelboim
University of Georgia, Athens, GA, USA

L. Rainie
Pew Research Internet Project, Washington, DC, USA

B. Shneiderman
University of Maryland, College Park, MD, USA

© Springer International Publishing Switzerland 2015
S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_5

collections of information, losing sight of the larger range of discussions and content. Social media is inherently a social network, meaning that people use it to create collections of connections that have an emergent form, structure and shape. Interfaces to social media, however, lack insights into the nature, topology, and size of the networks they present. Access to social media network information is of academic and practical interest. Social Network Analysis (SNA) offers a powerful method to conceptualize, analyze and visualize social media—leading to new applications and user interfaces that help users make their own decisions about content relevance and the credibility of other users. Social media can be much more useful for users, and the information in it can be more easily evaluated, if its underlying network structure is made more visible and comprehensible.

In this chapter we demonstrate that mapping the structure of social media is practical and provides parsimonious answers for basic questions like "what type of social interaction am I a part of", "what topics are discussed here", "who are the most active and/or central players in this conversation". Mapping social media networks can reveal the landscape of discussions, highlighting areas of community, fragmentation, division, and celebrity.

2 Twitter Social Media Networks

As Twitter users share information and talk about a wide range of topics they form social media networks. There are intimacies, shouting matches, soapbox barkers, commercial come-ons, cliques, mobs, congregations, and everything in-between. People reply-to and mention one another, forming links that aggregate into various shapes and structures. Popular topics attract crowds of people who form a range of network structures. Twitter data can be collected and analyzed to reveal and visualize the shape of these crowds, summarize their topics of discussion, and figure out which people are at the center of them. In effect, we can create an aerial photo of the social media crowd as it forms while listening carefully to the banter taking place on the ground. The resulting maps and reports can inform users, giving them a better chance to manage the flow of information, make their voices heard, and help them detect untrustworthy communicators or messages.

The Pew Research Center's Internet & American Life Project collaborated with the Social Media Research Foundation to gain insights into the way people use Twitter. The goal was to detect the simplest and most common ways in which Twitter social media conversations take shape. Using a social media network analysis tool called NodeXL[1] the report analyzed data collected from Twitter

[1] NodeXL is a free and open tool for network analysis that provides special support for collecting and visualizing social media network data. The download and support site for "NodeXL"—the network overview, discovery and exploration add-in for Excel is located at: http://nodexl.codeplex.com. The NodeXL Graph Gallery website hosts a collection of social media network visualizations, descriptions, and data sets for download: http://nodexlgraphgallery.org/. NodeXL is created by the Social Media Research Foundation which fosters the creation of *open tools*, *open data*, and *open scholarship* related to social media: http://www.smrfoundation.org/.

conversations and communities related to a range of topics and then generated network visualization maps and reports that highlighted key people, groups, and topics. The goal was to create a taxonomy of basic interactions, at the group level. The report was intend to reveal the basic building blocks of social interactions in social media, starting with the most prevalent social interaction patterns on Twitter. These patterns are important not only as visualizations. They can reveal who drives the conversation, what type of energies are spent on it, and who are the most central nodes (users). These reports allow us to draw some conclusions about the roles different Twitter users play, based on the positions they hold in the networks in which they are a part.

Social media network maps are created by drawing lines between Twitter users for each connection they formed as they follow, reply to, or mention one another. By reviewing many thousands of these social media network visualizations we found six distinctive social media network patterns, which illustrate the range of social typologies and roles that can occur.

We identified six different kinds of social media networks:

Polarized: Polarized discussions feature *two big and dense groups that have little connection between them*. The topics being discussed are often the most divisive and heated social and political subjects. In fact, there is usually little conversation between these groups despite being focused on a shared topic. Polarized groups are not arguing, they are ignoring one another while pointing to different web resources and using different hashtags.

In-group: In-group networks are *tight communities*. They are characterized by smaller groups of highly interconnected people with few disconnected, isolated participants. Many conferences, professional topics, hobby groups, and other subjects that attract communities have an in-group form.

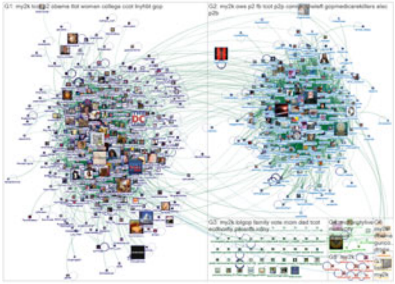
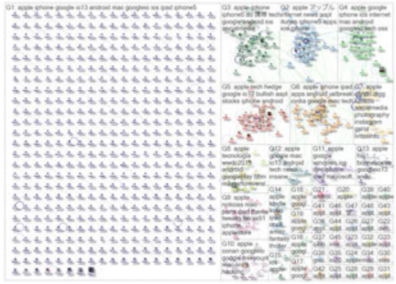
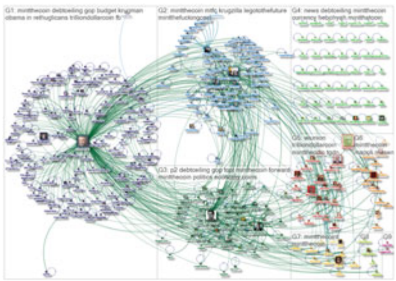
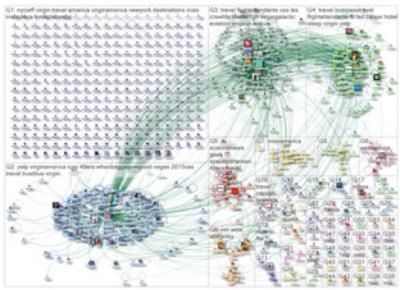
Brands: Products, services, and events often get discussed in Twitter, attracting wide comment from large populations of *disconnected participants*. The better known a brand or the bigger an event, the larger the population talking about it. As the Twitter population around a subject grows, the less likely are the people talking about it to be connected to one another. Therefore, brands networks in social media often have a low density of connections with many "isolated" people who have no connections at all to others talking about the same brand.

Clustered: Some popular topics may attract many smaller groups, which often form around a *few hubs each with its own audience, influencers, and sources of information*. Global news stories often attract coverage from many outlets each with its own following, creating a collection of medium-sized groups.

Broadcast: Conversations around major news stories and the output of media outlets have a distinctive hub and spoke pattern in which *many people repeat what the news organization tweets*, forming a disparate "audience" group. The audience is often connected only to the hub news source, without connecting to one another. In some cases there are smaller subgroups of densely connected people—think of them as subject groupies—who regularly discuss the news with one another.

Support: Customer complaints for a major business are often handled by a Twitter service account that attempts to resolve and manage customer issues around their

Table 1 Six forms of social media network found in Twitter, each form has distinct structural properties that are associated with different kinds of discussions and social processes

	
Polarized: two dense groups with little interconnection	*In-group*: few disconnected isolates, many connections
	
Brand/Public Topic: many disconnected isolates, some small groups	*Clustered*: many medium sized groups, some isolates
	
Broadcast: a hub which is retweeted by many disconnected users	*Support*: a hub which replies to many disconnected users

products and services. This produces a hub and spoke pattern that is different from the Broadcast pattern. In the Support network pattern, the hub account replies-to many otherwise disconnected users. In contrast, in the Broadcast pattern, the hub gets replied to (re-tweeted) by many disconnected people.

Below is an expanded and annotated version of the polarized crowd map from Table 1, which highlights key features illustrated by this “aerial view” of social media crowds (Fig. 1).

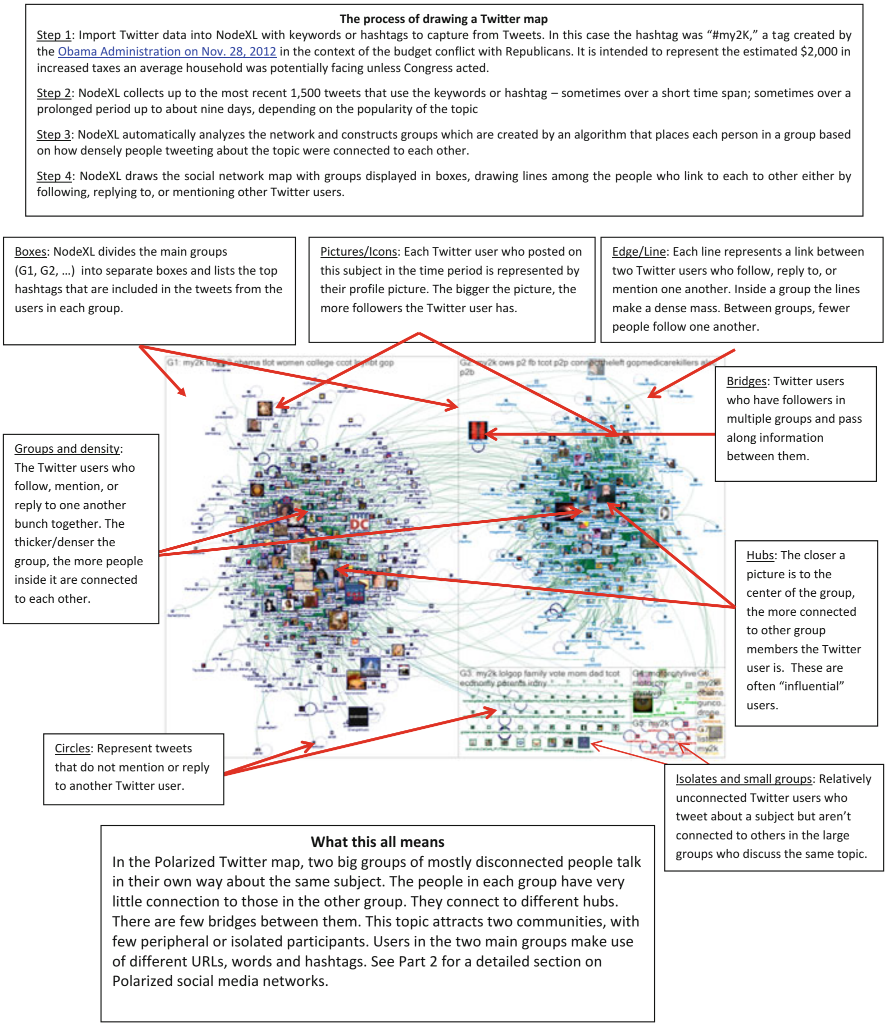


Fig. 1 The process of drawing a Twitter map

3 Influencers: Hubs and Bridges in Networks

These maps also highlight key participants in Twitter conversations in several ways. First, in the visualizations, each user who sends tweets is represented by her/his profile photo. The larger images indicate people who have larger numbers of other users following them. Those are the people who start with the largest audience for what they have to say.

Second, some users in these conversations link to and receive links from more Twitter users than others. Our network maps call out these key people at the center of their conversational networks—they are "hubs" and they are notable because they have the most engaged followers.

Third, the maps point out the people who have links to several groups—they are called "bridges." They are important because they pass along information from one group to another and often play a role in causing a message to "go viral."

The maps show that each kind of crowd has its own pattern of influence with key users occupying strategic locations, like hubs and bridges, in the network.

Our research also assess the content created by the people within each subgroup. Content is analyzed by examining the words, URLs, and hashtags that are most common in each subgroup. Different types of social media network crowds have distinct content structures with varying levels of overlap and diversity.

In the following we document in detail what happens in each kind of social media network crowd, highlighting the information that gained the greatest attention in the population, and the kinds of people and institutions that shape the conversion.

4 A Taxonomy of Network Types: Purpose, Division, Density

Our initial six forms of social media networks can be more precisely defined in quantitative terms as relationships between different network measures.

Structure	Group count and group size	Level of group interconnectivity	Density	Examples
Polarized	2 Large	Disconnected	Few	Political controversy: Divisive topics display separated "echo chamber" pattern
In-Group	2–6 Medium	Connected	Few	Hobbies, professional topics, conferences. No "outsiders," all participants are "members"
Brand/ Public Topic	Many small	Few connections	Many	Brands, public events, popular subjects
Clustered	Many small	No connections	Many	Global media topics
Broadcast	1 large, some secondary	Inbound connections	Moderate	News and media outlets, famous individuals
Support	1 large, some secondary	Outbound connections	Moderate	Companies and services with customer support

The first two networks are opposites of one another in terms of division; the polarized network type is divided while the in-group network is unified. The next pair of networks, the brand and the clustered community, shares a large population of isolates, but they vary in terms of their density of connections; the brand network pattern has small disconnected groups with many isolated participants, while the clustered community network pattern has larger, more connected groups. The last two networks are inversions of one another; the broadcast model features many spokes pointing inward to a hub while the support pattern features a hub linking outward to many spokes. Each of these network types is described in detail below.

5 Part 1: Research Method and Strategy

To understand the nature of Twitter conversations, the Pew Internet Project solicited help from researchers at the Social Media Research Foundation, a group of scholars whose mission is to support the creation and application of open tools, open data, and open scholarship related to social media. The foundation maintains a software project called NodeXL, a plug-in extension to Microsoft Excel spreadsheets that enables **n**etwork **o**verview, **d**iscovery, and **e**xploration. NodeXL allows users to import network data and perform analysis and visualization of networks. NodeXL permits anyone to connect to social media services (including Twitter, Facebook, YouTube, flickr, Wikis, email, blogs and websites) and retrieve public data about the connections among users, pages, and documents. In the specific case of Twitter, the tool captures information about the content of each message (the "tweet"), which may contain usernames, hyperlinks and hashtags, along with information about each author's connections to other Twitter users. In Twitter, these connections include relationships among users who follow one another, who mention one another, and reply to one another.

Our Twitter datasets start with keyword searches that return a set of tweets from the Twitter Search service. Maps are then constructed by examining the content of each tweet that is returned in the Twitter Search results for the query. In addition, NodeXL captures information about the Twitter user's connections to other Twitter members. Data are also retrieved from each user's public Twitter profile, which includes the number of tweets the user has posted, the number of other users that the user follows, and the number of persons who follow that user, among other things. Author statistics are combined with information about the connections among the people who shared the use of the same word, phrase, or term. For example, if Alice and Betty both post a message that includes the term "Obama" and Alice follows Betty on Twitter, our data captures this relationship.

Only publicly available messages are analyzed in our studies. No direct messages or other private content are collected or analyzed. Any message defined by its author as private (from, for example, "protected accounts") is excluded from analysis.

There are clear limits to any dataset captured by NodeXL. The tweets we collect are snapshots of finite periods of conversation around a topic or phrase. The data here do not represent the sentiments of the full population of Twitter users or the larger period of discussion beyond the data collection window. Further, Twitter users are not representative of the full range of the population of the United States or even the population of the Internet or even of social media users generally. Thus, we are not arguing that this analysis represents all that happens on Twitter or that it is a proxy for national sentiment on these topics. However, we believe these data sets contain useful snapshots of the structure of social media networks around topics that matter.

6 Taking "Aerial Photographs" of Twitter Crowds

Our method is similar to taking aerial photographs or short videos of crowds in public spaces, particularly pictures of rallies, protests, political events, and culturally interesting phenomena. No one snapshot or video clip of a crowd completely captures the event, but our method has the benefit of producing crowd photos from social media spaces, a domain that has not been widely pictured before. Like aerial crowd photographs, social media network maps show the size and shape of the crowd along with the key actors in that crowd.

These network maps can reveal information at the level of both individuals and groups. Social media networks often have just a few people who stand out in terms of the unique ways they connect to others. Some networks are composed of just a single group, while others are divided into sub-groups. Groups can be more or less connected to other groups. These shapes tell a story about the kinds of interactions that take place in Twitter.

7 Group Density

Each group can be measured in terms of the density of its connections. A group of people with many connections among its members is more "dense" than a group that has few connections. Our maps allow groups to be compared both in terms of content members of the group share and in terms of how strong the linkage is among members of that group. Density is measured as the ratio of the number of existing relationships among nodes over the total number of possible relationships. The density can vary between zero (i.e., no connections among nodes) and 1 (i.e., all nodes in a network are connected to all other nodes). As groups grow in size it is harder to interact with all other participants, so as a rule, the larger the numbers of people in a social network the lower the density of their connections. As a result, no one value is a specific threshold for high or low density but networks are considered loosely-knit, low density networks when few of the participants are connected to one another.

Twitter social media network maps show how interconnected people are when they engage in conversations. People often "clump" into groups. Some groups have high levels of internal connection and limited connectivity to people outside their group. The amount of internal and external connection is an important indicator of how exposed people are to differing points of view from people in different groups. If there are few ties between groups, people may not be exposed to content from users in other groups.

8 More on Hubs and Bridge

Social network maps created from collections of Twitter relationships often highlight a few individual users who occupy key positions in the network. We refer to the few highly connected users as “hubs.” Many other users follow these hubs; far more than follow the majority of people in the network. Hubs are important because they have large audiences. Some people have fewer connections but are equally important because they have the rare trait of having connections that link across the network to other groups, acting as “bridges.” While big hubs can also occupy the important position of “bridge,” a user with just a few relatively unique connections may also be an important bridge.

9 Part 2: Extracting the Six Conversation and Group Network Structures in Twitter

After examining hundreds of maps of thousands of subjects and events we have found six distinct network structures in Twitter social networks. Each is profiled below. There is no doubt there are other styles and structures of social media networks remaining to discover, as the taxonomy above in fact suggests. The landscape of social media remains a partially undiscovered and poorly mapped terrain. The six network types we describe are intended as initial examples of distinct forms not as an exhaustive list of all possible forms. It is also important to note that these maps only cover Twitter. Different kinds of social media services may generate different patterns of networks. Yet, we have reasons to believe, again, based on the taxonomy listed above, that the six forms are representative for the sets of motivations and constraints that are most commonly encountered on social media.

Social media researchers, managers and participants who wish to extend the typology may want to ask a series of questions related to the social media network maps of their own topics and related discussions. What kind of social media network is formed by the people talking about the topics that matter most to you? How does your topic’s network compare to competing topics? Who are the key people and groups in these networks? How do these networks change over time, particularly as events and engagements occur?

10 Group Type 1: Polarized

Polarized social media networks feature at least two large dense groups that have little inter-connection or bridge between them. These networks are often focused on divisive topics and are especially likely to involve political content.

The “#My2K” hashtag is a good example of this type of network structure. The data set for this visualization is available here: https://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=2272.

“#My2K” is a hashtag proposed by the White House on November 28, 2012 in the context of the ongoing budget conflict with congressional Republicans. The hashtag is intended to represent the “2 K” or the estimated $2,000 in increased tax costs that the average U.S. household was facing unless Congress acted to head off an automatic tax increase. The President proposed this hashtag to rally Twitter supporters to press Congress to preserve the tax break.

To understand what kind of crowd gathered around the “My2K” banner, we collected and analyzed a network graph that represents a network of 688 Twitter users who tweeted a message starting January 6th and ending on January 8th, 2013 that mentioned “My2K.” There is a green edge, or connecting line, for each instance when someone in our sample who tweeted about “My2K” was also a follower of another person who used the term. Separately, there is also a blue edge if someone in our sample “replies-to” or “mentions” another Twitter user who has written about “my2K.” There is a self-loop edge for each tweet about “my2K” that is not a “replies-to” or “mentions.” We call these Twitter users “isolates” in these conversations because they are not connected to others in the conversation.

The social media network map for #My2K looks like:

The #My2K hashtag network map features two dense groups of Twitter users with very few connections between them, meaning few people in one group replied to, mentioned, or followed people in the other group (Fig. 2).

Analysis of the content of the tweets created by the people in these groups showed that the words, hashtags, and URLs mentioned by people in each group are very different despite the common topic of their tweets. In the network map each group is labeled with the ten most frequently mentioned hashtags used by the users in that group. The group on the left is a large dense group of 360 people who often added the hashtag “#tcot” (which stands for “Top Conservatives on Twitter”) and is often used by conservative Twitter users to self-identify with conservative politics. The group on the opposite side of the graph is composed of 254 people who often added hashtags like “#ows” (Occupy Wall Street) and “#p2” (Progressives 2.0) to their tweets.

The map illustrates that conservatives discussed the subject of “My2K” with one another and liberals discussed it among themselves, but few spoke to someone from the other group—or heard from someone in the other group.

Outside of these major groups are smaller groups with just 74 people who have few connections to other users. Some 48 of them had no connections at all—we call them “isolates” because they are not connected to anyone else in this particular Twitter conversation. These disconnected people mentioned the “#My2K” hashtag but were not observed to follow, reply or mention anyone else who did so in this dataset. These may be people who are just starting to mention this topic and related political issues, since they lack connections to people who discuss this already.

In the middle of each of the two large groups are “hubs,” people with many connections. However, in a polarized network, these connections rarely span the

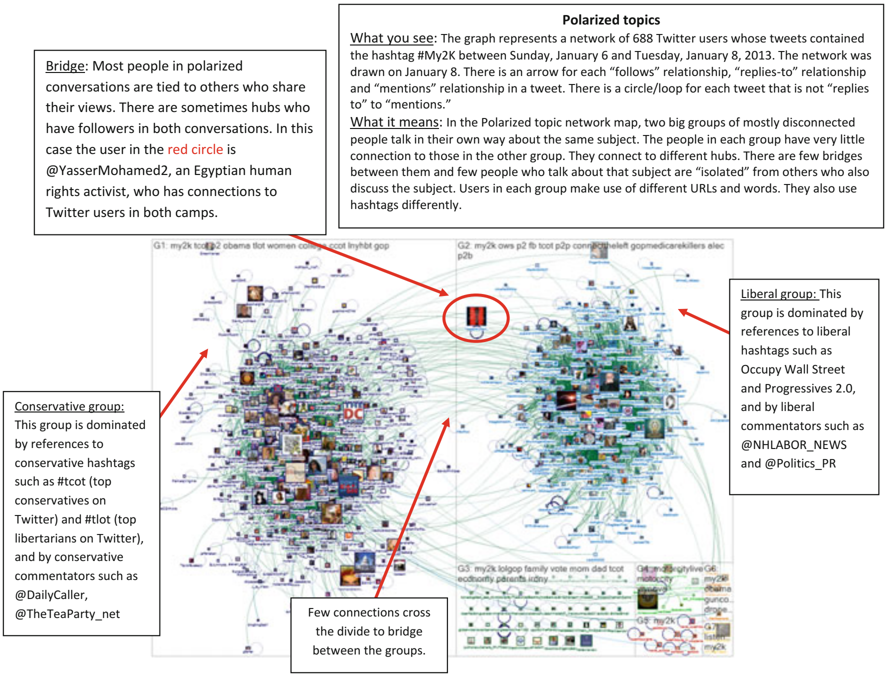


Fig. 2 The network of connections among people who tweeted "#My2K" over the 1-day, 21-h, 39-min period from Sunday, 06 January 2013 at 03:30 UTC to Tuesday, 08 January 2013 at 01:09 UTC

divide to connect to people in the other group. Each group has a small number of highly central core participants. In the conservative-leaning Group 1, the most central people are: @DailyCaller, @TheTeaParty_net, @JC7109, @PeterMAbraham, @saramarietweets—all self-identified conservatives with considerable followings. In the liberal-leaning Group 2, the most central people are: @Politics_PR, @NHLABOR_NEWS, @PaulStewartII, @BODIESOFLIGHT, @CAFalk. The user @YasserMohamed2 stands out as a highly followed user (red icon) who bridges the right wing group and the left wing group.

When the most frequently used hashtags in each group are contrasted, we can get a better sense of the topical focus and orientation each group displays.

Similarly, the most frequently used URLs in the tweets in each group indicate the kinds of web resources each group is interested in sharing. The comparison of the URLs used in Group 1 and Group 2 illustrate the contrast between their political orientations, as seen in Table 2. Group 1 links to partisan news sites devoted to a conservative perspective. Group 2 links to mainstream and liberal news sites and services.

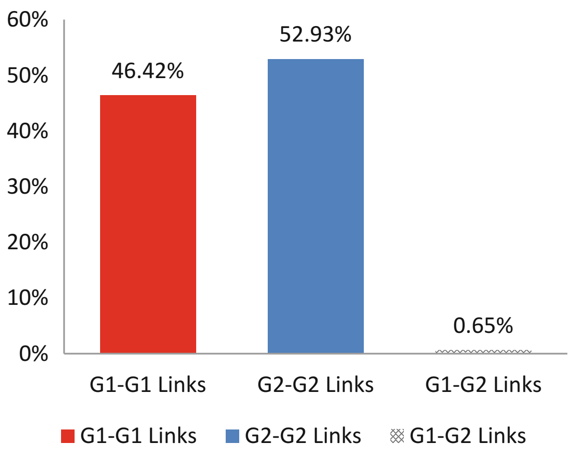
The relative absence of connections between these groups shows that people who tweet about #My2K rarely follow, reply to, or mention anyone who is located in the other group. Indeed, the chart below shows how dense each group is and how

Table 2 Top Hashtags by frequency of mention in Group 1 and Group 2 in the #MY2K Twitter Network

Top Hashtags in the conservative-oriented Group 1[a]	Top Hashtags in the liberal-oriented Group 2
#tcot—top conservatives on Twitter	#ows—Occupy Wall Street
#p2—progressives 2.0	#p2—Progressives 2.0
#obama	#fb—hashtag for posting tweets to Facebook
#tlot—top libertarians on Twitter	#tcot (Top Conservatives on Twitter)
#women	#p2p (peer-to-peer)

[a]Hashtags were identified using the website http://tagdef.com/

Fig. 3 Analysis of links between users in each of the two largest network groups within the hashtag #MY2K showing that very few of the connections among those who used the hashtag crossed group boundaries



few people in each group link to people in the other group (Fig. 3). Some 46 % of all the personal connections in the map are among those in the tight conservative group (G1) and 53 % of the connections are in the tight liberal group (G2). Less than 1 % of the connections are between people in the different groups (Table 3).

There were 13,341 different relationships among those who used the hashtag #My2K from January 6–8, 2013. Figure 3 shows that only .65 % of connections crossed between the two groups.

Most topic networks on Twitter do not look like polarized topics, but many political discussions are structured this way. For instance, similar polarized conversation pattern can be seen in the network of people discussing “Sequester OR Sequestration”: https://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=3441.

Automatic across the board budget cuts, called “sequestration” have been imposed by the US Congress. The topic of “sequestration” is a divisive political issue that generates a clearly polarized pattern in Twitter. The topic attracts a large number of people who appear in Group 1 who share the quality of having no visible connections to others. These “isolates” are an indication of the public quality of the topic—that is, many people have heard of the term even if they do not already

Table 3 Contrasting URLs frequently used in two groups discussing "#My2K"

Top Hashtags in Tweet in Group 1	Top Hashtags in Tweet in Group 2
http://dailycaller.com/2013/01/06/white-house-online-My2K-campaign-fails-as-us-workers-payroll-taxes-increase/	http://ireport.cnn.com/docs/DOC-500857
http://www.breitbart.com/Big-Peace/2013/01/01/Hezbollah-Joining-Cartels-in-Mexico-s-War	http://www.cnn.com/
http://tpnn.com/obama-we-dont-have-a-spending-problem/	http://www.youtube.com/watch?v=SOBsoUZFae8&feature=related
http://www.washingtontimes.com/news/2013/jan/6/obama-supporters-shocked-angry-new-tax-increases/	http://www.huffingtonpost.com/2011/09/15/americas-poorest-states-_n_964058.html
http://mobile.wnd.com/2012/12/the-nazi-roots-of-u-s-gun-control-laws/	http://www.flamethrowermagazine.com/david-koch-secret-right-wing-attack-machine/

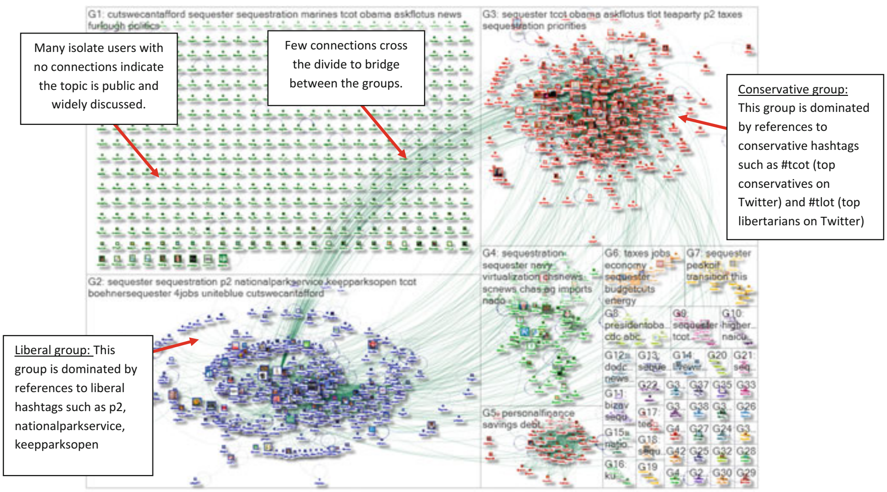


Fig. 4 The graph represents a network of 1,253 Twitter users whose recent tweets contained "sequester OR sequestration" made over the 1-h, 22-min period from Monday, 11 March 2013 at 18:15 UTC to Monday, 11 March 2013 at 19:38 UTC

follow or reply to others who also tweet about the topic. This is a common pattern for brands and well known issues and events (Fig. 4).

What makes this pattern a polarized one is the relationship between groups 2 and 3, two large dense communities of people who have many connections within their group and few to other groups. As seen in Fig. 5, the two groups are linked to one another with only 3 % of links. Contrasted with networks described below, particularly the "in-group" network pattern, this level of inter-group connection is very low. The low level of connection is an indicator that these groups are socially isolated from one another, despite tweeting about the same topic (Fig. 5).

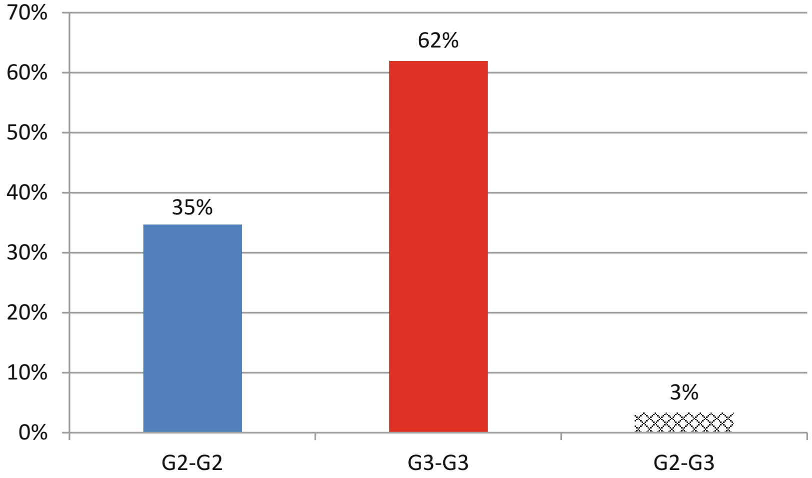


Fig. 5 Analysis of links between users in each of the two largest network groups within the hashtag #sequester showing that only a few of the connections (3 %) among those who used the hashtag crossed group boundaries

Table 4 Contrasting URLs frequently used in two groups discussing "sequester" or "sequestration"

Top URLs in Tweet in Group 2	Top URLs in Tweet in Group 3
http://www.washingtonpost.com/blogs/plum-line/wp/2013/03/11/gop-triumphalism-about-the-sequester-is-premature/	http://www.bernardgoldberg.com/sun-still-rises-after-sequester-so-prez-comes-up-with-plan-b-to-insure-hardship/
http://www.dailykos.com/story/2013/03/11/1193173/-Republicans-are-all-for-sequestration-until-their-something-gets-sequestered-in-their-back-yard	http://thehill.com/blogs/floor-action/house/287371-gop-suggests-dhs-using-sequester-as-excuse-to-weaken-immigration-laws#ixzz2NG7SD4lM
http://www.nps.gov/applications/digest/headline.cfm?type=Announcements&id=13550	http://www.realclearpolitics.com/articles/2013/03/11/obama_flails_as_republicans_stand_firm_on_sequester_117365.html
http://tv.msnbc.com/2013/03/11/obama-jokes-about-sequester-my-joke-writers-have-been-placed-on-furlough/	http://www.youtube.com/watch?v=3gXOV_XWJck&feature=youtu.be
http://www.whitehouse.gov/sites/default/files/omb/assets/legislative_reports/fy13ombjcsequestrationreport.pdf	http://foxnewsinsider.com/2013/03/11/u-s-park-ranger-claims-obama-administration-making-spending-cuts-so-public-feels-pain-from-sequestration/

Another indicator of the divisions between these groups can be seen in a comparison of the URLs user frequently post in tweets in each group as displayed in Table 4. The URLs in Group 2 are critical of conservative positions or reference official documents that describe the impact of budget cuts. In contrast, the URLs mentioned in Group 3 are critical of concerns that the budget cuts will have major

Table 5 Top Hashtags by frequency of mention in Group 2 and Group 3 in the sequestration Twitter Network

Top Hashtags in Tweet in Group 2	Top Hashtags in Tweet in Group 3
Sequester	sequester
Sequestration	tcot
p2	obama
Nationalparkservice	Askflotus
Keepparksopen	Tlot

consequences. They also tie to concerns about immigration politics. And they also cite criticism of the Administration.

The differences between these groups are also reflected in the different hashtags used in the tweets from users in each group.

While both groups used hashtags for "sequester" and "sequestration," they otherwise use different labels in their tweets. Settings these terms aside, in Group 2, the "p2" (Progressives 2.0) hashtag is the most frequently used label while in Group 3 "tcot" (top conservatives on Twitter) is most frequently used. Other Group 2 hashtags ("nationalparkservice" and "keepparksopen") suggest a focus on the negative effects of budget cuts on national parks. In contrast, Group 3 is focused on "Obama," "askflotus" (for questions directed at the First Lady of the United States), and tlot (top libertarians on Twitter) (Table 5).

Many politically controversial topics have this polarized pattern, topics that attract divided populations who converge on the same topic, term or hashtag. For example, discussions about contraception often have a large dense but separate group that is opposed to legal access to birth control. But not all, or even most, topics have this form. There are many topics that have a network that has a pattern that is the opposite of the polarized pattern, the "in-group."

11 Group Type 2: In-Group

Unlike polarized conversations, people in "in-group" conversations have strong connections to one another and significant connections that bridge between sub-groups. These dense networks are often communities of people who are aware of one another and converse often. These networks have many people who follow one another and reply to and mention one another. People who share a common interest and a common orientation to that interest often populate in-groups. These networks are composed of a single group where conversations sometime swirl around, involving different people at different times. In the in-group topic there is no polarized "other" group (Fig. 6).

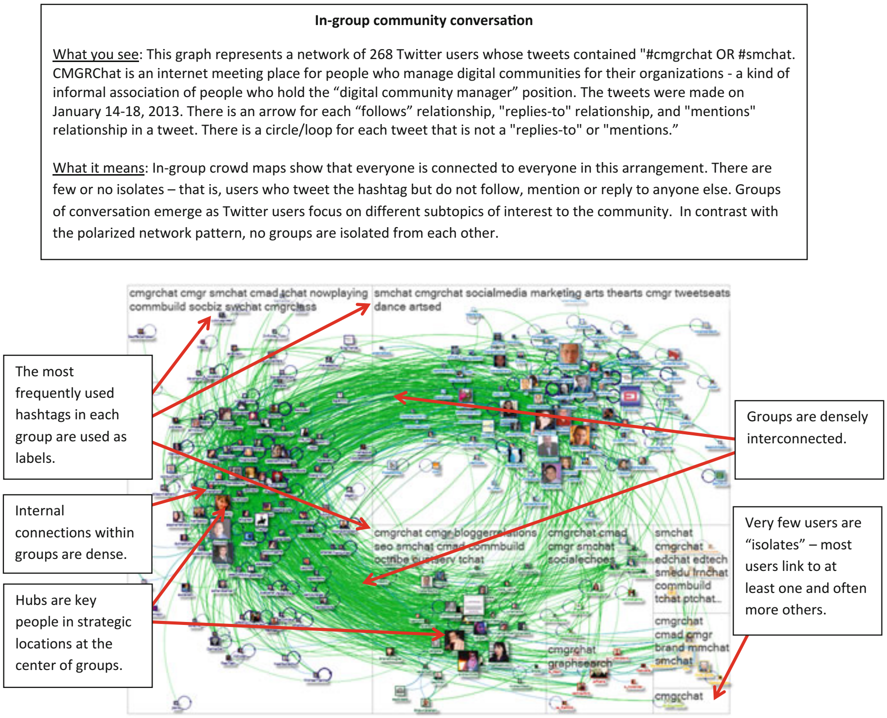


Fig. 6 The graph represents a network of 268 Twitter users whose recent tweets contained "#cmgrchat OR #smchat" made over the 3-day, 21-h, 15-min period from Monday, 14 January 2013 at 18:23 UTC to Friday, 18 January 2013 at 15:38 UTC

In-Group Community Conversation

What you see: This graph represents a network of 268 Twitter users whose tweets contained "#cmgrchat OR #smchat." CMGRChat is an internet meeting place for people who manage digital communities for their organizations—a kind of informal association of people who hold the "digital community manager" position. The tweets were made on January 14–18, 2013. There is an arrow for each "follows" relationship, "replies-to" relationship, and "mentions" relationship in a tweet. There is a circle/loop for each tweet that is not a "replies-to" or "mentions."

What it means: In-group crowd maps show that everyone is connected to everyone in this arrangement. There are few or no isolates—that is, users who tweet the hashtag but do not follow, mention or reply to anyone else. Groups of conversation emerge as Twitter users focus on different subtopics of interest to the community. In contrast with the polarized network pattern, no groups are isolated from each other.

In-group conversations take place in networks that have few if any isolates—people who have no connections to anyone else in the network. In these network maps, isolates are people who use a hashtag or mention a topic, but have not been observed to follow, reply to, or mention *anyone* else who talked about the topic.

The #CMGRChat hashtag is a good example of an in-group topic: https://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=2434. CMGRChat is a social media outlet for people who manage digital communities for their organizations—a kind of informal association of people who hold the "digital community manager" position. Social media professionals who discuss and share resources about the best practices in running message boards, Facebook pages, and Twitter streams populate this network. While some of these contributors have more connections than others, no participant in this discussion has zero connections and most have several.

It is often the case that "everyone knows everyone" in these dense communities of connected participants. The map of tweets using the hashtag #CMGRChat between January 14–18, 2013 shows a conversation divided into sub-groups, but one with many connections that bridge the divisions, suggesting these groups are more closely tied sub-communities rather than divided, insulated and separate communities.

While the frequently mentioned URLs in each of the largest groups in the #CMGRChat networks are different, there is little evidence of a polarized focus. Many of the URLs point to resources related to a diverse range of social media related topics, but the topics are not in conflict with one another (Table 6).

A similar pattern is found in the various hashtags that are more frequently used in each group in the #CMGRChat network. All of the groups mention the common terms #cmgr, #cmgrchat, and #smchat. Group 1 has a focus on the related chat hashtags, Group 2 has a focus on marketing, and Group 3 is focused on bloggers and search engine optimization (SEO) (Table 7).

Examination of the patterns of linkage between groups shows that there is significant cross connection, indicating the presence of a single community, rather than divided polarized groups (Fig. 7).

The #MLA13 hashtag, used in conjunction with the Modern Language Association conference, is another example of an in-group social media network is. The Modern Language Association annual conference attracts many scholars who study culture and language. Like "CMGRChat," the "#MLA13" topic network in Twitter is an in-group with few isolates and just a few small groups with significant interconnections (Fig. 8).

This graph represents a network of 599 Twitter users whose recent tweets contained "mla13": http://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=2274. The network was obtained on Tuesday, 08 January 2013 at 14:54 UTC. There is a green edge for each follows relationship, and a blue edge for each "replies-to" and "mentions" relationship in a tweet. There is a self-loop edge for each tweet that is not a "replies-to" or "mentions." The tweets were made over the 1-day, 19-h, 31-min period from Sunday, 06 January 2013 at 19:05 UTC to Tuesday, 08 January 2013 at 14:36 UTC.

The people who tweeted the hashtag for this conference are highly likely to follow and reply to multiple other people who also mention the name of the conference. A relatively small group of people mentioned the event and had no

Table 6 Contrasting URLs frequently used in three groups discussing #CMGRChat

Top URLs in Tweet in Group 1	Top URLs in Tweet in Group 2	Top URLs in Tweet in Group 3
https://plus.google.com/u/0/events/cc1ho11fo5gopmo-94q5u4bdrtlo	http://socialmediachat.wordpress.com/2013/01/09/arts-diablo/	http://www.buzzstream.com/blog/turning-blogger-relations-into-an-overall-inbound-strategy.html
http://www.womma.org/blog/2013/01/wommachat-on-jan-24-influencers-community-management	http://socialmediachat.wordpress.com/2013/01/09/arts-diablo/#comment-554	http://www.wilhelmus.ca/2013/01/two-facebook-pages-best-practices.html
http://info.socious.com/bid/62373/25-Tweetable-Online-Community-Tips-from-Richard-Millington-s-Book-Buzzing-Communities	http://heidicohen.com/social-media-35-brand-attributes-to-consider/	http://www.feverbee.com/2013/01/meaningful-conversations.html
http://mycmgr.com/community-manager-job-roundup-jan-14/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+mycmgr+(My+Community+Manager)&buffer_share=0d1fa	http://www.huffingtonpost.com/2013/01/09/diablo-ballet-crowdsourcing_n_2443783.html	http://socialmediachat.wordpress.com/2013/01/09/arts-diablo/#comment-554
http://socialmediatoday.com/jd-rucker/1155901/being-bold-social-media-about-risk-versus-reward?utm_source=feedburner&utm_medium=feed&utm_campaign=Social+Media+Today+(all+posts)&buffer_share=dc8aa	http://paper.li/CreativeSage/SMchat	http://mashable.com/2013/01/14/skittles-twitter/

Table 7 Top Hashtags by frequency of mention in groups in the #CMGRChat Twitter Network

Top Hashtags in Tweet in Group 1	Top Hashtags in Tweet in Group 2	Top Hashtags in Tweet in Group 3
cmgrchat	smchat	cmgrchat
cmgr	cmgrchat	cmgr
smchat	socialmedia	bloggerrelations
cmad	marketing	seo
tchat	arts	smchat

connections at all to others. These “isolates” are an indication that news of the event was reaching new communities of people, but the dense connections among most people taking about the “MLA” suggests that this is a community or in-group. The use of hashtags and URLs in the content in each group is another way to contrast these groups. The most frequently mentioned URLS in the largest groups in the

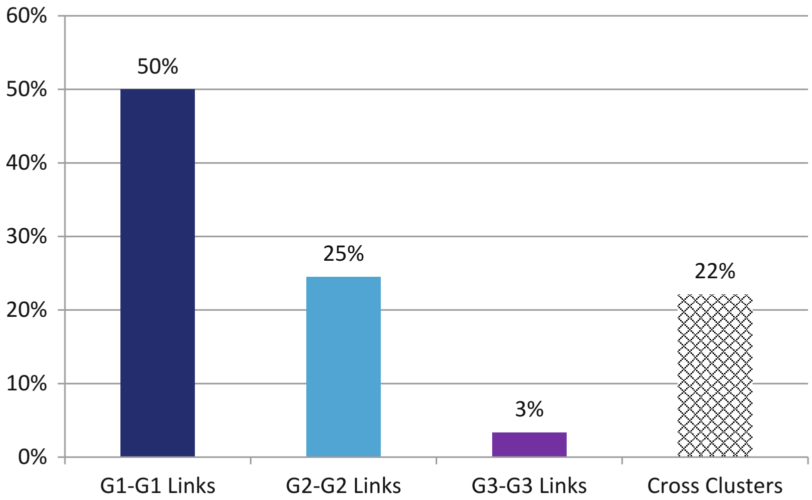


Fig. 7 Analysis of links between users in each of the three largest network groups within the hashtag #CMGRChat. Many of the connections among those who used the hashtag crossed group boundaries

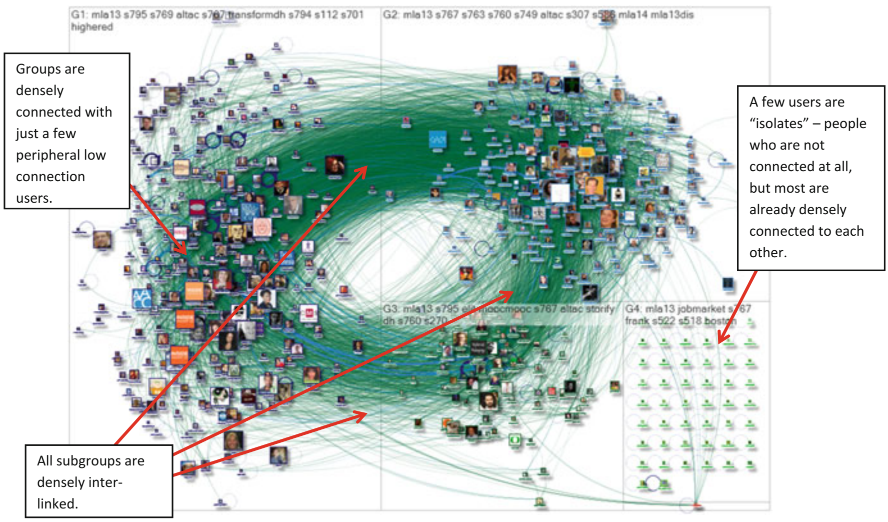


Fig. 8 Network graph of 599 Twitter users whose recent tweets contained "mla13" made over the 1-day, 19-h, 31-min period from Sunday, 06 January 2013 at 19:05 UTC to Tuesday, 08 January 2013 at 14:36 UTC

MLA13 network are displayed in Table 8. The overlap among these lists is an indication that groups shared a common interest and referred to similar content. This is in contrast to networks in which there is little or no overlap in the URLs used in different groups, which would indicate polarization and division. In the MLA network all the sub groups linked to common articles on the "InsideHighEd"

Table 8 Contrasting URLs frequently used in three groups discussing #MLA13

Top URLs in Tweet in Group 1	Top URLs in Tweet in Group 2	Top URLs in Tweet in Group 3
http://www.insidehighered.com/news/2013/01/07/mla-discussions-how-digital-communications-can-help-level-playing-field	http://www.insidehighered.com/news/2013/01/07/mla-discussions-how-digital-communications-can-help-level-playing-field	https://docs.google.com/document/d/1fjtpe3eNUASbilrugEgMKgABcoglwmrwAHJbizv0YUk/edit
http://anitaconchita.wordpress.com/2013/01/07/mla13-presentation/	http://katinarogers.com/2013/01/06/rebooting-graduate-training-mla/	http://storify.com/kathiiberens/the-classroom-as-interface-mla13?utm_content=storify-pingback&utm_campaign=&utm_source=direct-sfy.co&awesm=sfy.co_jD7M&utm_medium=sfy.co-twitter
http://chronicle.com/blogs/conversation/2013/01/06/what-if-the-adjuncts-shrugged/	http://www.uminnpressblog.com/2013/01/from-mla-2013-considering-serial.html	http://chronicle.com/blogs/conversation/2013/01/05/on-the-dark-side-of-the-digital-humanities/
http://storify.com/rogerwhitson/s112	http://nowviskie.org/2013/resistance-in-the-materials/	http://www.insidehighered.com/news/2013/01/07/mla-discussions-how-digital-communications-can-help-level-playing-field
http://www.insidehighered.com/blogs/confessions-community-college-dean/dropping-mla	http://storify.com/rogerwhitson/s112	http://sarahwerner.net/blog/index.php/2013/01/make-your-own-luck/

website and Chronicle.com and Storify websites. The common use of content across these groups suggests that these networks are divided by small differences in social relationships rather than major divisions. These groups are lobes of a common group rather than separate disconnected entities.

The common focus between the groups in the MLA13 network is also reflected in the most frequently used hashtags as displayed in Table 9. The top hashtags in each group refer to the sessions people attended and Tweeted about. The sub-groups represent the sub-populations of people who attended different sessions at the conference. While session 767 was popular in all groups, each group also had at least one term that was unique to it.

The connections people create can stay within their group or cross to end in another group. The measure of these intergroup connections reflects the in-group or polarized character of a network. The rate of internal connection is plotted in Fig. 9. The high level of cross group linkage is a strong indicator that the MLA13 network is an in-group network.

Table 9 Top Hashtags by frequency of mention in groups in the #MLA13 Twitter Network

Top Hashtags in Tweets in Group 1	Top Hashtags in Tweets in Group 2	Top Hashtags in Tweets in Group 3
mla13	mla13	mla13
s795	s767	s795
s769	s763	elit
altac	s760	moocmooc
s767	s749	s767

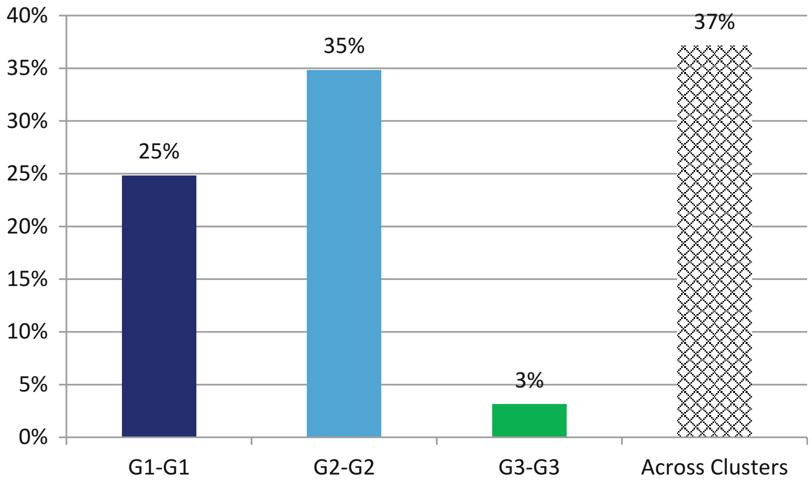


Fig. 9 Analysis of links between users in each of the three largest network groups within the hashtag #MLA13. Many of the connections among those who used the hashtag crossed group boundaries

Groups that use language in unique ways often create in-group networks. These topics share a common quality: People outside the group are unlikely to know or use the term. Technical terms, hobbyist vocabulary, and professional events are all examples of topics that form in-group networks. In-groups often form around topics that have limited general appeal but are topics of great interest to a small minority. People who have a passionate interest in esoteric topics often find one another in social media. These people often form multiple connections to one another as they share information about their niche interest. Therefore, a network map of an in-group community is a useful way to quickly identify the key people, topics and URLS that are central to the discussion of that topic.

12 Group Type 3: Brands, Breaking News, and Big Events

Brands and other public topics are the opposite of in-groups or communities; they have very low density and have many isolated participants. In a brand topic many people are likely to mention the topic without having any connection to one another. Advertised products, public events, and news are likely to have this pattern.

An example of a brand network is the Apple network: http://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=4681.

Apple is a major brand recognized widely around the world. Brands have a distinctive pattern of connection featuring large groups of people who have no links at all to others. In this network most users do not follow, reply or mention any other user who also tweeted about Apple. A large proportion of users share the common attribute of having mentioned the Apple brand name but lack any connection to one another. This pattern is common when a topic or term is widely known. Small groups are present in a brand network, visible in the upper right corner of the network map in Fig. 10. These groups are composed of small collections of users who discuss features and new releases of devices.

The graph represents a network of 834 Twitter users whose recent tweets contained "#apple," taken from a data set limited to a maximum of 1,500 users. The network was obtained on Wednesday, 15 May 2013 at 19:34 UTC. There is a green edge for each follows relationship. There is a blue edge for each "replies-to" or "mentions" relationship in a tweet. There is a self-loop edge for each tweet that is not a "replies-to" or "mentions."

In the groups that formed around the brand, there is limited interaction and little overlap in terms of resources linked to. None of the URLS frequently mentioned in each of the largest groups in the Apple network were mentioned in more than one group. This lack of URL overlap across groups suggests that the groups are distinct and focus on different aspects of the Apple product experience (Table 10).

Users in each group made use of different hashtags as well as URLs. Table 11 displays the frequently mentioned hashtags in the largest groups in the Apple network. The differences in hashtags suggest that each group is devoted to

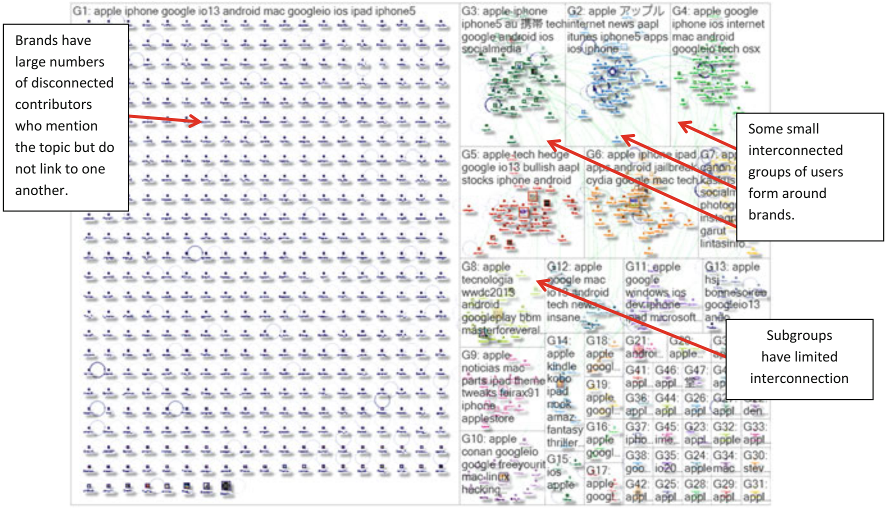


Fig. 10 Network graph of 834 Twitter users whose recent tweets contained "Apple" made over the 1-h, 41-min period from Wednesday, 15 May 2013 at 17:43 UTC to Wednesday, 15 May 2013 at 19:24 UTC

discussion of different Apple products (iTunes, iPhone), investment in Apple, or comparison between Apple and Android mobile devices.

Mentions of brands in Twitter generate networks composed of disconnected individuals and small groups. These groups are relatively interconnected, suggesting that brands are not polarized. The rates of connections between groups discussing Apple in Fig. 11 illustrates the modest levels.

13 Group Type 4: Clustered Community

When groups of people form several evenly sized groups, a network structure different from the Brand structure emerges. We call it a clustered community (or sometimes a "bazaar") because it is a collection of medium sized groups. An example is the discussion of the First Lady Michelle Obama's Twitter username "Flotus": https://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=2440.

While many of these contributors are isolates, demonstrating the brand quality of this topic, there are more connected groups of relatively equal size in a Clustered Community. These types of social media networks have many hubs each with a separate crowd—in a sense, it can be compared to people clustering in different stalls at a bazaar (Fig. 12).

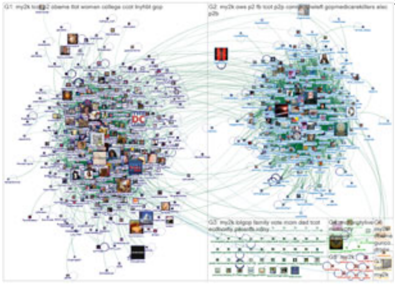
The graph represents a network of 1,260 Twitter users whose recent tweets contained "flotus." The network was obtained on Friday, 18 January 2013 at 18:26 UTC. There is a green edge for each follows relationship. There is a blue

Table 10 Contrasting URLs frequently used in groups discussing #Apple

Top URLs in Group 2	Top URLs in Group 3	Top URLs in Group 4	Top URLs in Group 5	Top URLs in Group 6
http://www.tuaw.com/2013/05/15/google-announces-new-hangout-app-to-hit-ios-today/	http://www.empiremedia.com/what-is-google-play/	http://instagram.com/p/ZV86V8QPdy/	http://finance.yahoo.com/news/hedge-funds-slash-apple-stakes-183323376.html	http://dealspl.us/Cell-Phones_deals/p_roocase-ultra-slim-gloss-black-shell-case-for-apple?r=seanvcxz
http://www.tuaw.com/2013/05/15/belkin-wemo-rolls-out-ifttt-multi-device-control/	http://partners.webmasterplan.com/click.asp?ref=517172&site=2732&type=text&tnb=87	http://mashable.com/2013/05/15/apple-to-samsung-the-s4-infringes-on-our-patents/	http://finance.yahoo.com/news/david-teppers-appaloosa-reduces-apple-175700791.html	http://www.scoop.it/t/future-business-technology/p/4001692666/top-rated-ios-and-android-apps
http://www.macrumors.com/2013/05/15/google-unifies-cross-platform-messaging-services-with-hangouts/	http://dealspl.us/Cell-Phones_deals/p_roocase-ultra-slim-gloss-black-shell-case-for-apple?r=seanvcxz	http://mashable.com/2013/05/14/apple-location-data-stalk-users/?utm_source=twitter&utm_medium=social&utm_content=47853	http://blogs.wsj.com/moneybeat/2013/05/15/time-to-worry-about-apple-again/	http://www.valuewalk.com/2013/05/caller-id-apps-for-iphone-android-blackberry-and-nokia-devices/

http://sportstalkflorida.lockerdome.com/contests/107693493	http://appleinsider.com.feedsportal.com/c/33975/f/616168/s/2bf680a6/l/0Lappleinsider0N0Carticles0C130C0A50C150Cgoogles0Eall0Eaccess0Emusic0Estreaming0Eservice0Eto0Etake0Eon0Espotify0Epandora/story01.htm	http://feeds.feedburner.com/~r/flipboardapple/~3/MFkFfOmb7xE/?utm_source=feedburner&utm_medium=twitter&utm_campaign=flipboardapple	http://www.insidermonkey.com/blog/apple-inc-aapl-billionaire-george-soros-is-also-bullish-on-cupertino-139026	http://www.ft.com/cms/s/0/fef37ffc-bd71-11e2-890a-00144feab7de.html
http://clkuk.tradedoubler.com/click?p=23708&a=1950257&url=https%3A%2F%2Fitunes.apple.com%2Fgb%2Fapp%2F1password-password-manager%2Fid443987910%3Fmt%3D12%26uo%3D2%26partnerId%3D2003&utm_source=dlvr.it&utm_medium=twitter	http://9to5mac.com/2013/05/15/google-maps-coming-to-ipad-this-summer-updated-with-new-design-improved-rating-system-in-app-offers-much-more?utm_source=twitterfeed&utm_medium=twitter	http://feeds.feedburner.com/~r/flipboardapple/~3/bhKgV0Fg--A/?utm_source=feedburner&utm_medium=twitter&utm_campaign=flipboardapple	http://www.businessinsider.com/why-apple-is-unlikely-to-change-its-famous-app-icons-shape-2013-5	http://www.amazon.co.jp/APPLE-mini-2-5GHz-Thunderbolt-MD387J/dp/B009X5EJR8/ref=zg_bs_2151949051_5/375-2400889-4913026?tag=ama012p-22

Table 11 Top Hashtags by frequency of mention in groups in the #Apple Twitter Network

Top Hashtags in Group 2	Top Hashtags in Group 3	Top Hashtags in Group 4	Top Hashtags in Group 5	Top Hashtags in Group 6
 [a]	Iphone	google	tech	iphone
internet	iphone5	iphone	hedge	ipad
news	Au	ios	google	apps
aapl	携帯	internet	io13	android
itunes	Tech	mac	bullish	jailbreak

[a]Apple in Japanese

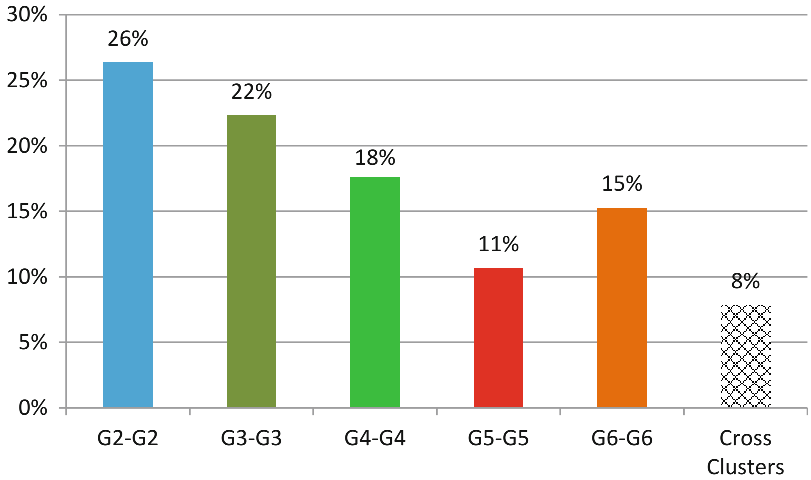


Fig. 11 Analysis of links between users in each of the largest network groups within the hashtag #Apple. Few of the connections among those who used the hashtag crossed group boundaries

edge for each "replies-to" or "mentions" relationship in a tweet. There is a self-loop edge for each tweet that is not a "replies-to" or "mentions."

There was more than the usual amount of social media activity related to Michelle Obama that day because the Administration and activists were eager to tie to her birthday and generate attention for some of their work. At the same time, there were different ways that people made reference to her, the birthday, and the

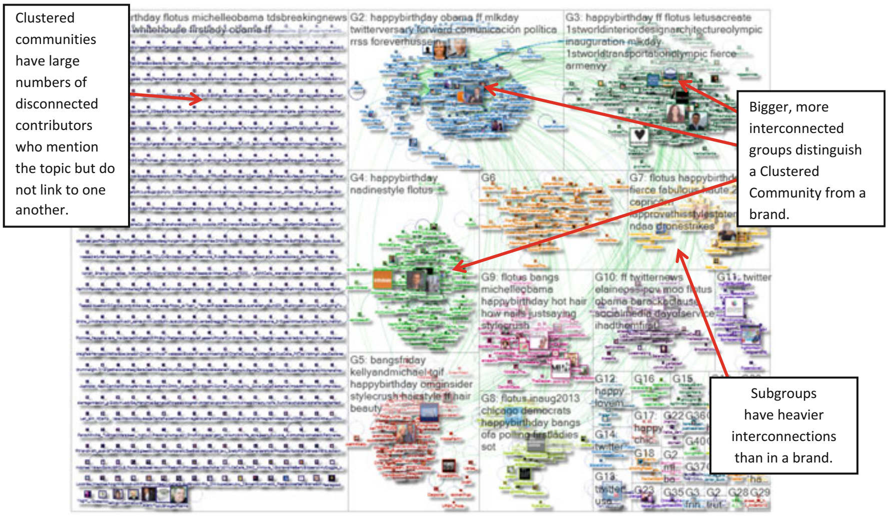


Fig. 12 Network graph of 1,260 Twitter users whose recent tweets contained "Flotus" made over the Friday, 18 January 2013 at 15:16 UTC to Friday, 18 January 2013 at 18:20 UTC

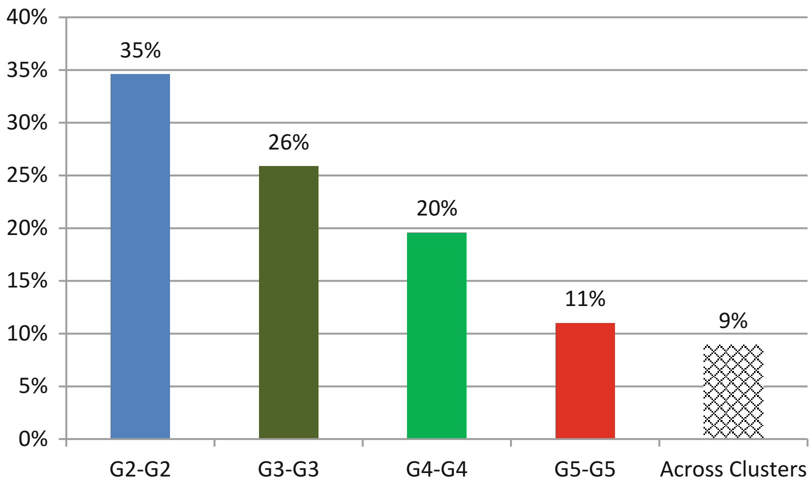


Fig. 13 Analysis of links between users in each of the largest network groups within the Flotus network. Few of the connections among those who used the term Flotus crossed group boundaries

activities of President Obama's allies and that was reflected in the different sources of information and URLs that were cited in each group.

There are 1,608 different relationships among those who used the hashtag #flotus in and across the top four groups (Fig. 13).

Table 12 Contrasting URLs frequently used in groups discussing Flotus

Top URLs in Tweet in Group 2	Top URLs in Tweet in Group 3	Top URLs in Tweet in Group 4	Top URLs in Tweet in Group 5
http://www.whitehouse.gov/blog/2013/01/18/follow-first-lady-michelle-obama-flotus-twitter	http://www.youtube.com/watch?v=YNbAvEe7FbI&feature=youtu.be	http://america.infobae.com/notas/65045-Michelle-Obama-abrio-una-nueva-cuenta-en-Twitter	http://pics.lockerz.com/s/277758941
http://www.youtube.com/watch?v=HYT68Uii1dk&feature=youtu.be	http://www.2013pic.org/service	http://www.2013pic.org/service	https://www.facebook.com/photo.php?fbid=10151341475790480&set=a.389111920479.168476.288878190479&type=1
http://govne.ws/item/Follow-First-Lady-Michelle-Obama-FLOTUS-on-Twitter	https://www.facebook.com/media/set/?set=a.10151436465887994.549161.128463482993&type=1&notif_t=like	http://simpsons.wikia.com/wiki/Michelle_Obama	http://ow.ly/i/1of6A
http://www.whitehouse.gov/blog/2013/01/18/follow-first-lady-michelle-obama-flotus-twitter?utm_source=twitter-feed&utm_medium=twitter	http://www.whitehouse.gov/blog/2013/01/18/follow-first-lady-michelle-obama-flotus-twitter	n/a	http://www.krnb.com/wpblog/?p=32021
http://flic.kr/s/aHsjDE7Xbh	http://obamafoodorama.blogspot.com/2013/01/president-obama-treats-first-lady-to.html	n/a	http://www.whitehouse.gov/blog/2013/01/18/follow-first-lady-michelle-obama-flotus-twitter

At the same time, there was some overlap among the groups because they shared an interest in her. That is evident in the link-count analysis chart below. In Clustered Community conversations many people are in the same conversational "vicinity," but their attention is often focused on separate things. The tone of the shared information in different groups also varies—some is serious, some is funny or wry, some is challenging and skeptical (Tables 12 and 13).

Another example is the network of people who tweeted about the Consumer Electronics Show (CES2013)—a giant trade show aimed at introducing new consumer-focused technology products that occurs every January: https://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=2275.

Table 13 Top Hashtags by frequency of mention in groups in the Flotus Twitter Network

Top Hashtags in Group 2	Top Hashtags in Group 3	Top Hashtags in Group 4	Top Hashtags in Group 5
happybirthday	happybirthday	happybirthday	Bangsfriday
obama	ff	nadinestyle	Kellyandmichael
ff	letusacreate		Tgif
mlkday	1stworldinteriordesignarchitectureolympic		Happybirthday
twitterversary	inauguration		Omginsider

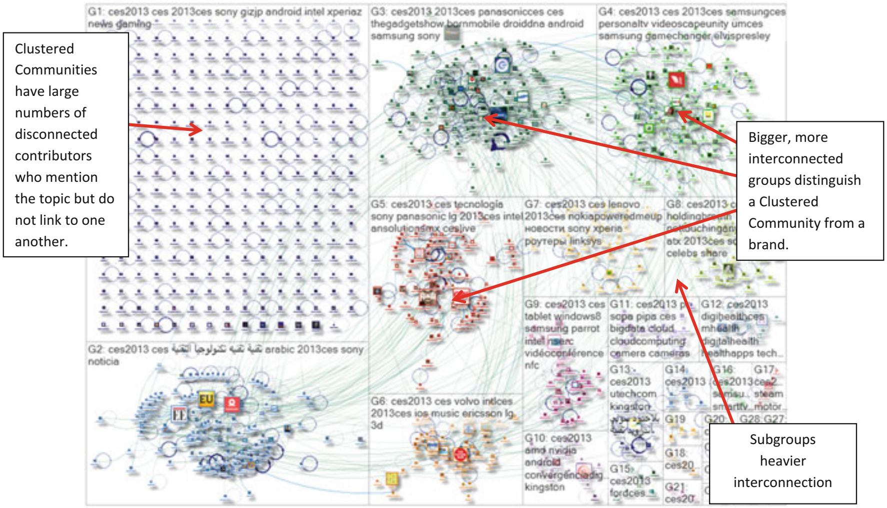


Fig. 14 Network graph of 1,260 Twitter users whose recent tweets contained “ces2013” made over the 47-min period from Tuesday, 08 January 2013 at 16:01 UTC to Tuesday, 08 January 2013 at 16:48 UTC

The graph represents a network of 1,041 Twitter users whose recent tweets contained “CES2013.” The network was obtained on Tuesday, 08 January 2013 at 16:56 UTC. There is a green edge for each follows relationship. There is a blue edge for each “replies-to” and “mentions” relationship in a tweet. There is a self-loop edge for each tweet that is not a “replies-to” or “mentions” (Fig. 14).

Again, the groups are at the same conversational “bazaar” but their interests vary from stall to stall and that is what is reflected in popular URLs in each cluster.

Moreover, there is notable overlap among the groups in a Clustered Community-style Twitter conversation. The people in the cluster don’t just link to and talk to each other, they have relatively overlapping ties to those in other clusters, as shown by the chart below (Tables 14 and 15) (Fig. 15).

There are 1,942 different relationships among those who used the hashtag #Flotus in and across the top four groups.

Table 14 Contrasting URLs frequently used in groups discussing CES2013

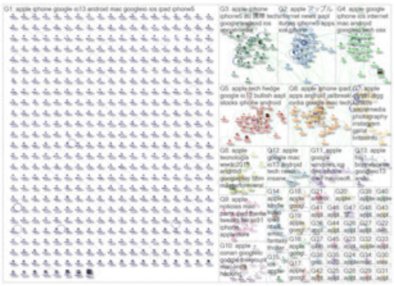
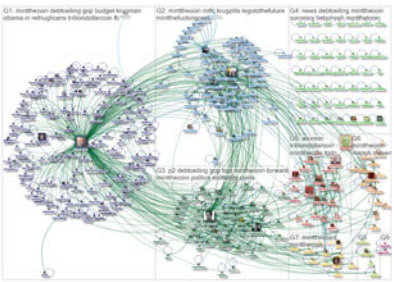
Top URLs in Tweet in Group 2	Top URLs in Tweet in Group 3	Top URLs in Tweet in Group 4	Top URLs in Tweet in Group 5
http://feedproxy.google.com/~r/enterCO/~3/dRBWva8m7MM/?utm_source=twitterfeed&utm_medium=twitter	http://panasonic.com/ces	http://ces.massrelevance.com/	http://www.cnnexpansion.com/tecnologia/2013/01/08/ultrad-le-quita-los-lentes-al-3d
http://www.elespectador.com/tecnologia/articulo-395530-xperia-z-el-telefono-resistente-al-agua-y-al-polvo	http://www.panasonic.com/promos/ces/2013/?cm_mmc=PNA-Web-_-Alias-_-Panasonic-_-CES2013-Microsite-Alias-10003-12212012	http://www.ipglab.com/2013/01/08/the-trigger-lexus-autonomous-driving/	http://www.revistasumma.com/tecnologia/33622-los-5-gadgets-mas-curiosos-del-ces.html
http://www.eluniversal.com/vida/130108/los-gigantes-de-la-tecnologia-dejan-ver-sus-nuevos-productos	http://www.qualcomm.com/sweepstakes/ces2013	http://www.youtube.com/watch?v=-pdOCi-83Fc&feature=youtu.be	http://www.cnnexpansion.com/tecnologia/2013/01/07/3m-touch-systems
http://www.elespectador.com/especiales/articulo-395516-tecnologia-se-toma-vegas	http://gadgetshow.channel5.com/gadget-show/gadget-news/sony-xperia-z-first-full-hd-phone-heads-to-the-uk	http://instagram.com/p/UOq3OLSdUT/	http://conecti.ca/2013/01/08/video-en-vivo-keynote-de-apertura-ces2013-a-cargo-de-panasonic/?utm_campaign=[VIDEO]%20En%20Vivo:%20Keynote%20de%20apertura%20#CES2013%20a%20cargo%20de%20Panasonic&utm_medium=twitter&utm_source=twitter
http://feedproxy.google.com/~r/enterCO/~3/dRBWva8m7MM/?utm_medium=twitter&utm_source=twitter-feed	http://www.ilounge.com/index.php/ces2013/report/incipio/	http://www.flickr.com/photo.gne?short=dJQ4pZ	http://rubiko.mx/lo-mas-relevante-del-ces2013-dia-uno/

14 Group Type 5: Broadcast

The broadcast pattern is dominated by a hub and spoke structure, with the hub often being a media outlet, surrounded by spokes of people who repeat the messages generated by the news organization or personality.

An example is the conversation about New York Times columnist Paul Krugman's article that appeared on January 11, 2013. The NodeXL map contains

Table 15 Top Hashtags by frequency of mention in groups in the CES2013 Twitter Network

Top Hashtags in Tweet in Group 2	Top Hashtags in Tweet in Group 3	Top Hashtags in Tweet in Group 4	Top Hashtags in Tweet in Group 5
 [a]	panasonicces	samsungces	Tecnología
	thegadgetshow	personaltv	Sony
	bornmobile	videoscapeunity	Panasonic
	droiddna	umces	Lg
arabic	android	samsung	Intel

[a]Technology, Tech, Technology, Technical, respectively

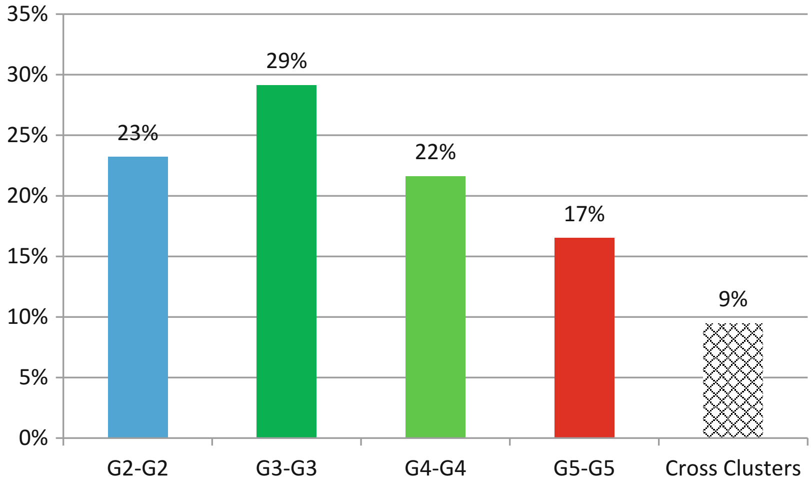


Fig. 15 Analysis of links between users in each of the three largest network groups within the hashtag #CES2013. Few of the connections among those who used the hashtag crossed group boundaries

people who linked to this column on Twitter: https://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=2313.

The graph represents a network of 399 Twitter users whose recent tweets contained "http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html." The network was obtained on Friday, 11 January 2013 at 14:27 UTC. There is a green edge for each follows relationship. There is a blue edge for each "replies-to" or "mentions" relationship in a tweet. There is a self-loop edge for each tweet that is not a "replies-to" or "mentions" (Fig. 16).

In this conversational structure, the "audience" of people who linked connect only to the Paul Krugman account are visible in Group 1, while Group 2 and Group 3 contain denser collections of people who could be considered part of the community interested in discussing Krugman's article. A collection of isolates suggests that the article was visible to many people, even if they weren't discussing it in Krugman discussion communities (Tables 16 and 17) (Fig. 17).

Advocacy organizations also often generate a broadcast pattern. For example, the "KilltheTrade" discussion focuses on the restriction of trade in endangered animal products. https://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=2483.

The graph represents a network of 1,196 Twitter users whose recent tweets contained "killthetrade," taken from a data set limited to a maximum of 1,500 users. The network was obtained on Monday, 21 January 2013 at 19:24 UTC. There is a green edge for each follows relationship. There is a blue edge for each "replies-to" or "mentions" relationship in a tweet. There is a self-loop edge for each tweet that is not a "replies-to" or "mentions." The tweets were made over the 2-day, 4-h, 8-min period from Saturday, 19 January 2013 at 15:07 UTC to Monday, 21 January 2013 at 19:15 UTC (Fig. 18).

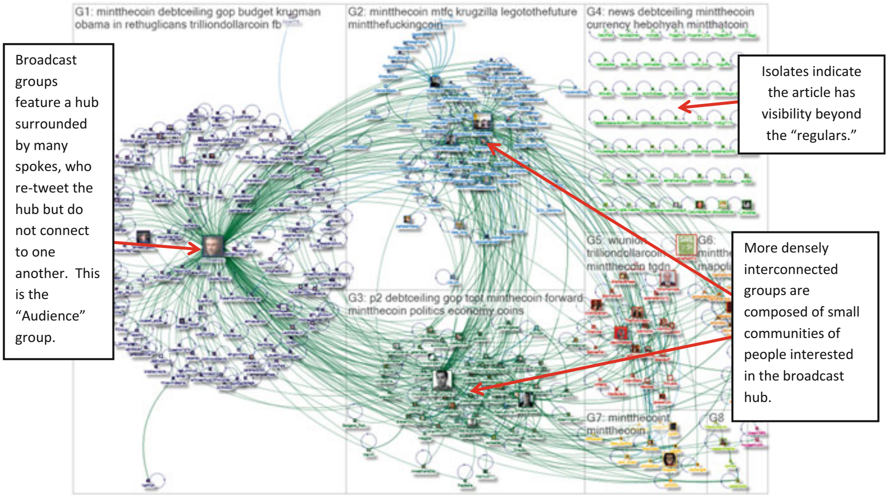


Fig. 16 Network graph of 399 Twitter users whose recent tweets contained a URL to a NYTimes article by Paul Krugman made over the 12-h, 32-min period from Friday, 11 January 2013 at 01:52 UTC to Friday, 11 January 2013 at 14:24 UTC

Table 16 Contrasting URLs frequently used in groups discussing a New York Times article by Paul Krugman

Top URLs in Tweet Group 1[a]	Top URLs in Tweet Group 2	Top URLs in Tweet Group 3
http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html?smid=tw-NytimesKrugman&seid=auto	http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html?smid=tw-NytimesKrugman&seid=auto	http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html?smid=tw-share
http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html?smid=tw-share	http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html?smid=tw-share	http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html?smid=tw-NytimesKrugman&seid=auto
http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html	http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html	http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html
	http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html?smid=tw-NytimesKrugman&seid=auto&_r=0	http://www.nytimes.com/2013/01/11/opinion/krugman-coins-against-crazies.html?hp
		http://www.nytimes.com/

[a] Links appearing only once were removed

Table 17 Top Hashtags by frequency of mention in groups in the Paul Krugman Twitter Network

Top Words in Tweet in Group 1	Top Words in Tweet in Group 2	Top Words in Tweet in Group 3
coins	rt	Coins
against	crazies	Against
crazies	against	Crazies
nytimeskrugman	coin	Rt
rt	coins	Krugman

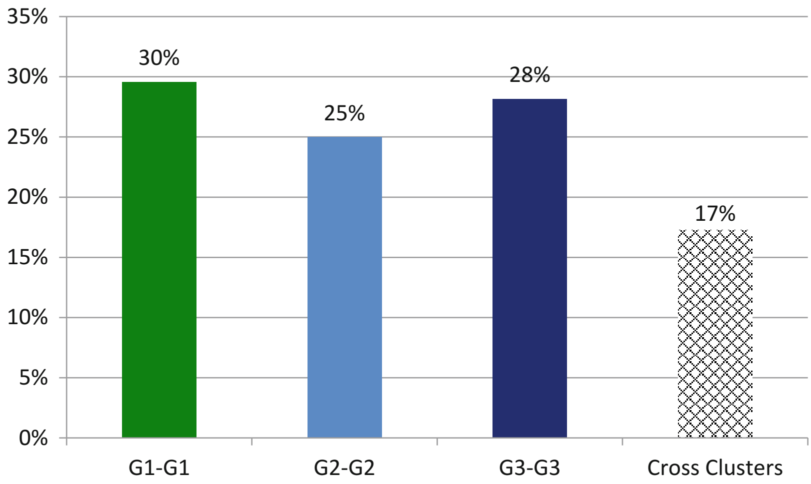


Fig. 17 Analysis of links between users in each of the three largest network groups within the Paul Krugman article network. There are 1,072 different relationships among those who used the URL for Krugman's column in and across the top four groups. Many of the connections among those who used the hashtag crossed group boundaries

At the center of the largest group is the account for the World Wildlife Foundation surrounded by a large number of participants who connect only to the WWF account. This is a low density hub-and-spoke group that contains the audience for the WWF. In contrast, Group 2 and Group 3 are communities composed of densely connected participants who all have many links to one another.

A broadcast network often has one or two large hubs with many spokes while the other groups are relatively small (Tables 18 and 19) (Fig. 19).

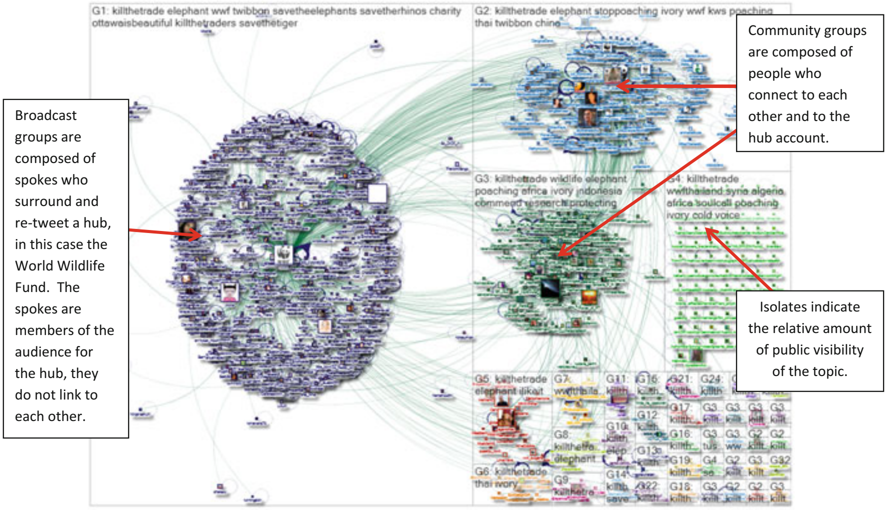


Fig. 18 Network graph of 1,196 Twitter users whose recent tweets contained killthetrade made over the 2-day, 4-h, 8-min period from Saturday, 19 January 2013 at 15:07 UTC to Monday, 21 January 2013 at 19:15 UTC

Table 18 Contrasting URLs frequently used in groups discussing "KillTheTrade"

Top URLs in Tweet in Group 1	Top URLs in Tweet in Group 2	Top URLs in Tweet in Group 3
http://wwf.panda.org/ban?utm_source=socialmedia&utm_medium=twitter&utm_content=thaipetition&utm_campaign=iwtc	http://wwf.panda.org/ban?utm_source=socialmedia&utm_medium=twitter&utm_content=thaipetition&utm_campaign=iwtc	http://wwf.panda.org/ban?utm_source=socialmedia&utm_medium=twitter&utm_content=thaipetition&utm_campaign=iwtc
http://wwf.panda.org/ban	http://wwf.panda.org/ban	http://wwf.panda.org/ban
http://wwf.panda.org/elephants	http://wwf.panda.org/elephants	http://wwf.panda.org/elephants
http://www.youtube.com/watch?v=MFdfocXRCT0	http://www.youtube.com/watch?v=MFdfocXRCT0	http://forcechange.com/53815/urge-indonesia-to-crack-down-on-illegal-ivory-imports/
http://ow.ly/gP2OE	http://www.africam.com/wildlife/tembe_webcam	http://forcechange.com/52018/commend-research-protecting-wildlife-against-illegal-poaching-in-africa/

Table 19 Top Hashtags by frequency of mention in groups in the "KillTheTrade" Twitter Network

Top Hashtags in Tweet in Group 1	Top Hashtags in Tweet in Group 2	Top Hashtags in Tweet in Group 3
elephant	elephant	wildlife
wwf	stoppoaching	elephant
twibbon	ivory	poaching
savetheelephants	wwf	africa
savetherhinos	kws	ivory

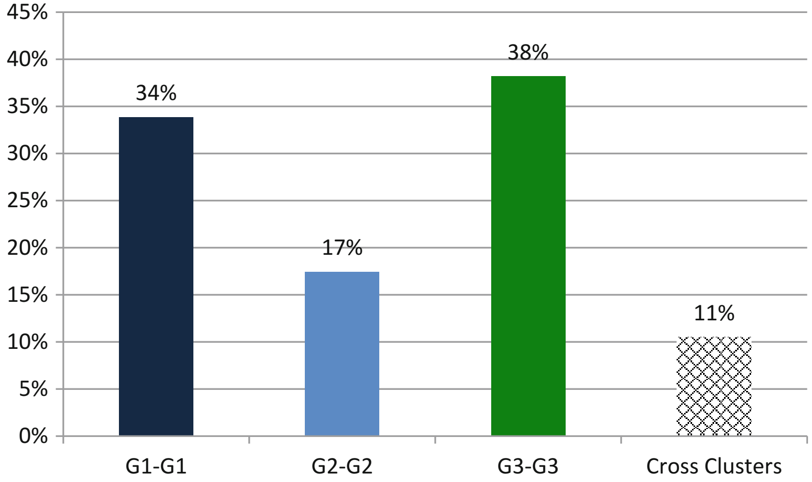


Fig. 19 Analysis of links between users in each of the three largest network groups within the "KillTheTrade" network. There are 4,314 different relationships among those who used the hashtag #killthetrade in and across the top four groups. Many of the connections among those who used the hashtag crossed group boundaries

15 Group Type 5: Support

Many large companies provide customer support via Twitter, maintaining a user account to listen to and reply to user complaints and issues. This account replies to many other accounts, which are not linked to one another directly. https://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=2956.

The graph represents a network of 388 Twitter users whose recent tweets contained "delllistens OR dellcares." The network was obtained on Tuesday, 19 February 2013 at 17:44 UTC. There is a green edge for each follows relationship. There is a blue edge for each "replies-to" or "mentions" relationship in a tweet. There is a self-loop edge for each tweet that is not a "replies-to" or "mentions." The tweets were made over the 6-day, 21-h, 58-min period from Tuesday, 12 February 2013 at 19:34 UTC to Tuesday, 19 February 2013 at 17:33 UTC (Tables 20 and 21) (Figs. 20 and 21).

The support pattern is also visible in the Virgin America network: https://nodexlgraphgallery.org/Pages/Graph.aspx?graphID=2414.

The graph represents a network of 1,040 Twitter users whose recent tweets contained "Virgin America." The network was obtained on Wednesday, 16 January 2013 at 22:48 UTC. There is a green edge for each follows relationship. There is a blue edge for each "replies-to" or "mentions" relationship in a tweet. There is a self-loop edge for each tweet that is not a "replies-to" or "mentions." The tweets were made over the 7-day, 18-h, 19-min period from Wednesday, 09 January 2013 at 04:18 UTC to Wednesday, 16 January 2013 at 22:38 UTC (Fig. 22).

Table 20 Contrasting URLs frequently used in groups discussing DellCares OR DellListens

Top URLs in Tweet in Group 1	Top URLs in Tweet in Group 2	Top URLs in Tweet in Group 3
http://dell.to/OqhRhj	http://www.youtube.com/watch?v=A-qq2gOLIOg&feature=share&list=PLmbFlhPb2qyWJ330CTZBEmPUqYpYKRXlK	https://pbs.twimg.com/media/BBT3RrHCEAA6CXp.jpg
http://dell.to/XazIZH	http://www.youtube.com/watch?v=g1OOq1t8Ybk&list=PLmbFlhPb2qyWJ330CTZBEmPUqYpYKRXlK&index=17	http://lt.dell.com/lt/lt.aspx?CID=68634&LID=4675173&DGC=SM&DGSeg=CBG&RED=301&DURL=http://en.community.dell.com/support-forums/customercare/f/4674/p/19491559/20299447.aspx&buffer_share=96621&utm_source=buffer
http://dell.to/XaPGD9	http://mashable.com/2013/02/18/ubuntu-tablet/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Mashable+%28Mashable%29	
http://del.ly/l/6015nMIV	http://www.dell.com/content/topics/topic.aspx/global/products/pedge/topics/en/config_calculator?c=us&l=en&s=gen	
http://news.cnet.com/8301-1009_3-57569018-83/microsoft-delivers-fixes-for-windows-8-windows-rt/	http://www.dell.com/support/drivers/us/en/19/DriverDetails?driverId=KT6P7	

Table 21 Top Hashtags by frequency of mention in groups in the DellCares OR DellListens Twitter Network

Top Hashtags in Tweet Group 1	Top Hashtags in Tweet Group 2	Top Hashtags in Tweet Group 3
dellcares	dellcares	nevahold
windows8	delllistens	whatawaste
xps	windows8	nosolutions
csrblast	dell	
frustrated	supportquality	

Note: URLs mentioned only once were removed

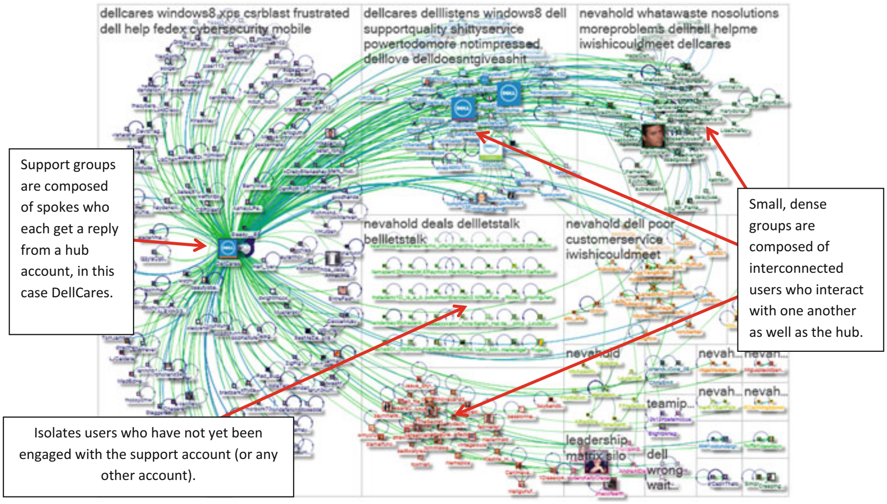


Fig. 20 Network graph of 388 Twitter users whose recent tweets contained delllistens OR dellcares made over the 6-day, 21-h, 58-min period from Tuesday, 12 February 2013 at 19:34 UTC to Tuesday, 19 February 2013 at 17:33 UTC

This map illustrates a hybrid that has brand features and a hub-and-spoke structure that is an indicator of a customer service account along with smaller community groups of densely connected industry analysts and journalists. This pattern resembles the broadcast pattern discussed below but is distinguished by the high rates of mutual interactions between the hub account and the disconnected spokes of customers seeking travel assistance (Tables 22 and 23) (Fig. 23).

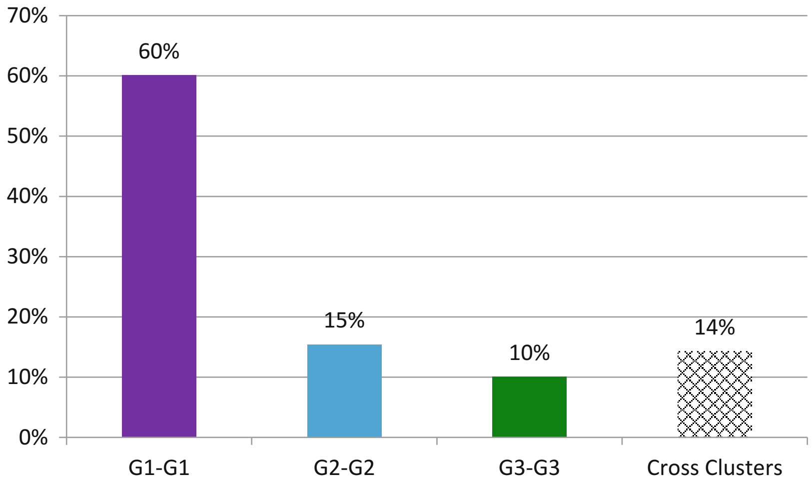


Fig. 21 Analysis of links between users in each of the three largest network groups within the delllistens OR dellcares networks. There are 1,445 different relationships among those who used the words delllistens OR dellcares in and across the top four groups. Many of the connections among those who used the hashtag crossed group boundaries

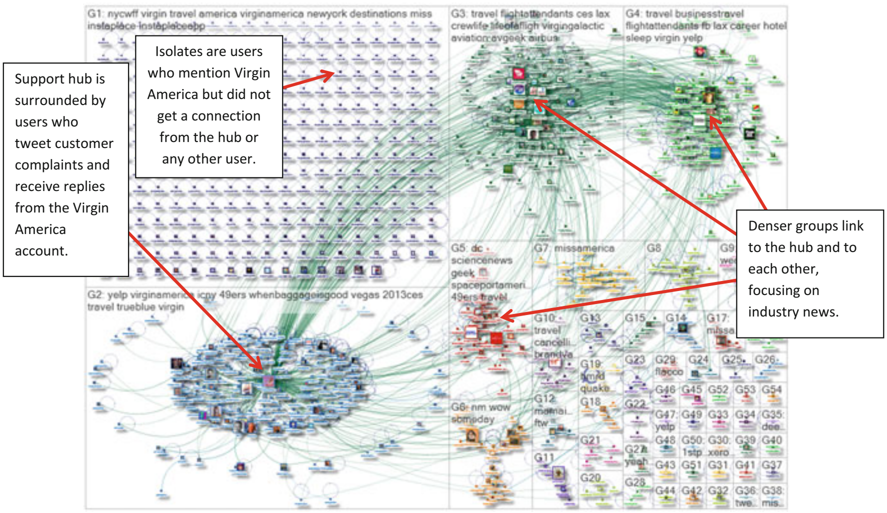


Fig. 22 Network graph of 1,040 Twitter users whose recent tweets contained "Virgin America" made over the 7-day, 18-h, 19-min period from Wednesday, 09 January 2013 at 04:18 UTC to Wednesday, 16 January 2013 at 22:38 UTC

Table 22 Contrasting URLs frequently used in groups discussing "Virgin America"

Top URLs in Tweet in Group 2	Top URLs in Tweet in Group 3	Top URLs in Tweet in Group 4
http://www.virginamerica.com/vx/lax-loft?cid=sm_social_FRI_118_2p_engagement_TW&stop_mobi=yes	http://www.cntraveler.com/daily-traveler/2013/01/first-class-cabins-singapore-airlines-emirate-etihad-cathay-asiana-ana-virgin-america-atlantic-el-al?MBID=twitter_#slide=1	http://www.cntraveler.com/daily-traveler/2013/01/first-class-cabins-singapore-airlines-emirate-etihad-cathay-asiana-ana-virgin-america-atlantic-el-al?MBID=twitter_#slide=1
http://www.yelp.com/biz/virgin-america-san-francisco-4	http://boardingarea.com/blogs/dealswelike/2013/01/15/virgin-america-matching-united-and-american-airlines-status/?utm_source=twitterfeed&utm_medium=twitter	http://fb.me/2aKzbWNeq
https://foursquare.com/nik_nik/checkin/50f470cde4b09661797ef01a?s=_zls9TjyRheB4G7dE1KTb-20Hfc&ref=tw	http://www.usatoday.com/story/todayinthesky/2013/01/16/virgin-america-adds-las-vegas-lax-nonstop/1840285/	http://www.smartertravel.com/airfare/virgin-america-sale-ends-Monday.html?id=13687378&source=rss&utm_source=twitterfeed&utm_medium=twitter
http://www.fastcompany.com/1675455/why-tech-nerds-love-flying-virgin-america	http://www.ausbt.com.au/virgin-america-s-lax-loft-lounge-rules-allow-only-australian-kids	http://Jump.priceline.com/pricebreakers/deal/PB_AIRVirginAmerica59_01152013.html?refid=PMSOCIAL&refclickid=TWITTER_PBI01152013-0200
http://www.yelp.com/biz/virgin-america-westchester	http://www.prnewswire.com/news-releases/virgin-america-launches-new-route--daily-flights-from-los-angeles-to-las-vegas-187146971.html	http://www.sun-sentinel.com/business/consumer-talk-blog/sfl-virgin-america-fll-route-sale-20130115,0,7873138.story

Table 23 Top Hashtags by frequency of mention in groups in the Virgin America Twitter Network

Top Hashtags in Tweet in Group 2	Top Hashtags in Tweet in Group 3	Top Hashtags in Tweet in Group 4
yelp	travel	travel
icny	flightattendants	businesstravel
49ers	ces	flightattendants
whenbaggageisgood	lax	fb
vegas	crewlife	lax

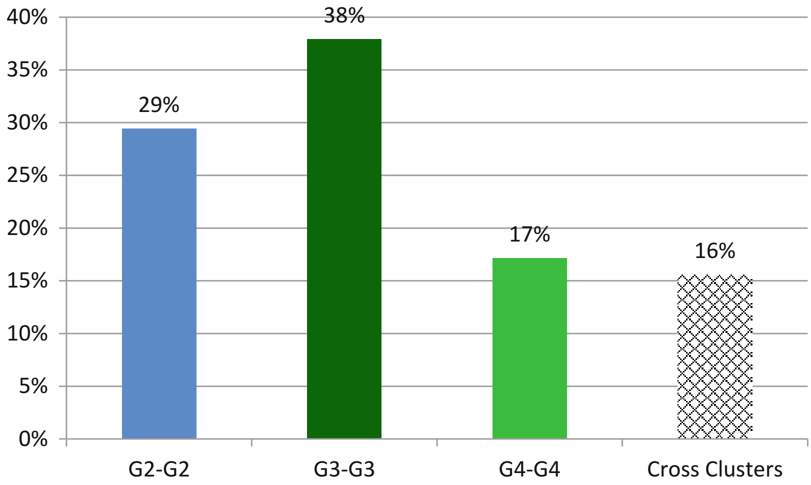


Fig. 23 Analysis of links between users in each of the three largest network groups within the "Virgin America" network. There are 2,353 different relationships among those who used the term Virgin America in and across the top four groups. Many of the connections among those who used the hashtag crossed group boundaries

16 Conclusions

Social media comes in different forms and structures. Mapping social media networks can enable a better understanding of the variety of ways individuals form groups and organize online. Social media network maps of Twitter have illustrated six different structures of connection around different kinds of topics.

It is possible to imagine any number of ways that these insights could find application for those interested in using social media to promote causes, to engage the stakeholders who are interested in their organizations or missions, and to start or enter social media conversations that matter to them.

For instance, those who run social media accounts for their organizations can explore how some of the conversational "styles" might be most applicable and useful to their work. Additionally, they might see how the "natural" structure of a conversation around their core topics could profit from adjustment. For example, a brand may want to cultivate community, or an in-group might want to open up to outsiders. Using these maps, participants can assess the type of social media network in which they participate and set a target for what they want their group to be like.

Social media is used by millions of individuals who collectively generate an array of social forms from their interactions. Social media network maps can be useful in understanding the variety of social structures that emerge. Network maps can reveal the structures of the crowd and highlight strategic locations or roles in these webs of connection. By mapping social media network spaces, researchers and practitioners can learn about the most common and best uses for these communication services.

Additionally, network analysis provides insights into social media that can help individuals and organizations make informed decisions about online conversations. An organization may have a goal to create a discussion with a particular kind of social structure, like a community or a brand. Creating social media network maps of these topic spaces can be a useful way to track progress. Social media managers, for example, have many topics of interest, including brand names, events, products, services, companies, and candidates. Managers may want to ask themselves "Which kind of social media network is my topic most like?" Further, they may want to select a network type as their desired goal for their topic discussion. With a goal in mind, additional maps can be created over time to measure the difference between the current state of the topic network and the desired one. As experiments with various social media engagement strategies are performed, social media network maps can track the impact on the structure of social media spaces.

Resources

Herbert-Read, J., Perna, A., Mann, R., Schaerf, T., Sumpter, D., & Ward, A. (2011). Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences, 108*(46), 18726–18731. doi:10.1073/pnas.1109355108.

Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication, 18*(2), 40–60. doi:10.1111/jcc4.12001.

Himelboim, I., Smith, M., & Shneiderman, B. (2013). Tweeting apart: Applying networks analysis to explore selective exposure on Twitter. *Communication Methods and Measures, 7*(3), 169–197.

Katz, Y., Tunstrøm, K., Ioannou, C., Huepe, C., & Couzin, I. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences, 108*(46), 18720–18725. doi:10.1073/pnas.1107583108.

NodeXL Graph Gallery. http://nodexlgraphgallery.org.

NodeXL. http://nodexl.codeplex.com.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York: Penguin Press.

Smith, M., Rainie, L., Sheiderman, B., & Himelboim, I. (2014). Mapping Twitter topic networks: From polarized crowds to community clusters, Pew Internet and American Life. http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/.

Social Media Research Foundation. http://www.smrfoundation.org.

Welser, H., Gleave, E., Fisher, D., & Smith, M. (2007). Visualizing the Signatures of Social Roles in Online Discussion Groups. *The Journal of Social Structure, 8*(2), 564–586. http://www.cmu.edu/joss/content/articles/volume8/Welser/.

Visible Effort: Visualizing and Measuring Group Structuration Through Social Entropy

Sorin Adam Matei, Robert Bruno, and Pamela L. Morris

1 Introduction

A large amount of research supports the benefits of group collaboration in terms of positive outcomes, individual satisfaction, and powerful cognitive effects (Johnson & Johnson, 1999; Slavin, 1996). The practice of computer-mediated collaboration (CMC) comes in many forms and many definitions for its meaning have been proposed. However, much research still needs to be done to understand the nature of the processes that take place during CMC. For example, despite recurring claims that online collaboration is innately egalitarian (either in terms of access or outcomes) and potentially superior due to some form of "collective intelligence" that spontaneously emerges without much coordination (Kelly, 1995; Rheingold, 2002), there is mounting evidence that online interaction follows traditional patterns of human interaction (Lampe, Ellison, & Steinfield, 2006; Matei & Ball-Rokeach, 2001).

We hold that effective group collaboration using CMC needs division of labor, coordination, and clear goals. Moreover, CMC groups that are rooted in norms or

A version of this paper has been presented at the Global Communication Forum, Jiao Tong University, Shanghai, September 29–30, 2010.

S.A. Matei (✉)
Brian Lamb School of Communication, Purdue University, West Lafayette, IN 47907, USA
e-mail: smatei@purdue.edu

R. Bruno
Communication Studies Department, City University of New York-LaGuardia, 718-704-7780, Long Island City, NY 11101, USA
e-mail: rbruno@lagcc.cuny.edu

P.L. Morris
Department of Communication Studies, University of Wisconsin-La Crosse, 1725 State St., La Crosse, WI 54601, USA
e-mail: pmorris@uwlax.edu

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_6

local cultures and that foster specific ethical guidelines are more likely to be productive. Conversely—and quite significantly—individual effort, inputs, and outputs are regularly observed to be *unevenly distributed* with naturally-occurring coordination and/or power hierarchies accompanying these uneven distributions. Barabási (2003) and Huberman (2001) have documented this uneven distribution for linkages between websites while Anderson (2006) and Shirky (2008) have done the same thing for online interactions related to e-commerce and online content consumption.

It is therefore of great importance that online collaboration be supported by new tools and be studied with appropriate methodologies that determine in what manner such uneven distribution of effort functions or how it can be modeled to facilitate maximum individual and group effectiveness. At the same time, egalitarian work paradigms can and should be employed in an informed, measured and intelligent manner. This is especially important in view of numerous claims that egalitarian collaborative systems are the preferred future organizational form (Brafman & Beckstrom, 2006), which would foster some form of "wisdom of crowds" (Lease, 2007; Powazek, 2009; Tapscott & Williams, 2006).

Some practitioners speculate that online groups are particularly adept at solving large problems by breaking them down into smaller and roughly similarly sized tasks to be allocated to many uncoordinated participants (Tapscott & Williams, 2006). A related expectation is that the larger the group and the more equitable the social structure, the more likely the problem will be solved effectively (Brafman & Beckstrom, 2006). As an example, an often invoked broadly-distributed process such as open source software development has been labeled by Raymond (2001) as the "bazaar" process. Accordingly, he notes that the hugely successful Linux operating system is the product of "bazaar" style micro-negotiation and collaboration between unknown and equally qualified programmers who take turns in fixing each other's mistakes. Illustrating the power of distributed open source programming, he states, "Given enough eyeballs, all bugs are shallow" (p. 30).

The egalitarian assumption that surrounds online interaction can be interpreted in many ways. One could be that equality of access should not be confounded with that of outcome or consumption. This distinction could be very important if the undeniable fact that the Internet gave more people more access to educational, business, or entertainment resources than previous media is to be reconciled with the body of observable evidence, supported by sociological theory, which suggests that collaboration online is in fact highly structured, that the Web has leaders and followers, and that equality of contributions and consumption is rarely if ever present in spontaneously emerging online groups (Kuk, 2006; Shirky, 2008). In opposition to Raymond's perspective, Kuk found a correlation between structuring, participation inequalities and the most productive processes of open source software development.

Taking a cue from this evidence, we propose a method for measuring the amount of equality and the emergence of social structure in groups that participate in CMC. The method relies on measuring the level of social "entropy" of an online environment. Social entropy, which will be discussed at length below, captures the degree of equality, evenness, and diversity of collaboration in any given system or group. The measure is visualized within the wiki environment "Visible Effort" (Veffort.us) with color-coded page frames and graphs, which can be used by learning groups for self-monitoring their collaborative progress. You can visit the experimental site at Veffort.us.

The measure and visualization method proposed serve two goals. First, they are used for measuring and visualizing the degree of collaborative evenness and emergence of social structure in a collaborative online wiki environment. Second, they can be used for steering the collaborative processes to attain specific goals (Matei & Bruno, 2015; Matei, Oh, & Bruno, 2006). This can be accomplished either passively or actively. It can passively provide users feedback on the processes that take place in their online space or can actively provide site administrators, project leaders or instructors the information necessary to intervene and moderate collaborative efforts. The present paper will illustrate these capabilities by describing a specific quasi-experimental teaching activity in tandem with a detailed discussion of theoretical justification, methodological underpinning, and technological capabilities of the Visible Effort approach.

2 CMC and Uneven Online Interaction

A significant amount of empirical evidence indicates that CMC in online environments tends to be distributed in the shape of a highly skewed curve (Anderson, 2006; Huberman, 2001; Kittur, Chi, Pendleton, Suh, & Mytkowycz, 2007; Ortega, Gonzalez, & Robles, 2008). Examples include the well-known metric of 10 % of Wikipedia editors contributing almost 90 % of the online encyclopedia's articles (Matei & Bertino, 2014; Ortega et al., 2008); which use similar inequities of production along the lines of 20–80 % that occur within the practice of the open source software (OSS) and Linux movement (Matei & Bruno, 2015); and multiple manifestations of uneven social distributions on Yahoo user groups, assorted emailing lists, user-generated "question & answer" forums, and so on (Matei & Bruno, 2015). Although utilizing different measurement techniques and theoretical perspectives, other terms that have cropped up in recent years to describe this extreme inequality are "Zipf's Law," "Power Law," or "long tail" distributions (Anderson, 2006; Barabási, 2003; Huberman, 2001). These terms all point to the fact that online phenomena, be it amount of contributions to a user-generated site, traffic, overall attention or usage share are highly skewed (Huberman, 2001). Some may say that the figures are nominal, and this would not be untrue. Nielsen (2006) proposed for the online environment a so-called "90/10/1" rule, which is probably closer to the truth, as Wikipedia research confirmed (Matei & Bruno, 2015): 90 % of users are mere consumers of content, 10 % contribute some time, while 1 % are responsible for the bulk of the contributions. Yet, even this radically skewed distribution would not alter the core idea, namely, online interaction can and is skewed.

However, this phenomenon is not native to computer-mediated environments. Seminal studies of small discussion groups ranging in size showed that top contributors dominate the conversation to the tune of 40–50 % of the time, with the next participator coming in at a percentage in the teens, and all those that follow generally registering below 10 % of the total (Bales, 1950; Stephan & Mishler, 1952). This suggests that human interactions tend to follow a skewed output and input allocation curve. While part of such skewness can be tracked to power, privilege, and control issues, much of it can be put under the rubric of functional differentiation of roles and tasks (Bailey, 1990; Matei & Bruno, 2015). Any task-oriented group needs to allocate roles, rewards, responsibilities, and workloads. Allocation involves a coordination mechanism, attendant communication processes, implementation schedules, and so on. These work best when redundancies are minimized and activities are distributed according to the nature of the task and to individual qualifications. These processes result in uneven distribution of individual input and output. Thus, a significant part of group inequalities can be tracked down to the functional requirements of forming human groups.

While the reality of uneven online collaboration and its impact is an undeniable fact, its ultimate theoretical explanation is still insufficiently understood. To some online activists and media observers, who for the past decades have promoted the idea of cyberspace as a liberating and equalizing force (Barlow, 1994; Benkler, 2007; Hiltz & Turoff, 1978; Raymond, 2001; Tapscott & Williams, 2006), these findings might appear as phenomena of less importance than purported peer-production processes that encourage egalitarian participation (Benkler, 2007). Yet, this opinion might ignore an important argument. As groups increase in size, they meet the hard barriers of mounting transaction costs. When narrowly defined, such costs are the financial expenditures associated with social and economic exchanges. When broadly understood, transaction costs are the energy, time, or financial resources spent on maintaining a group's coordination and communication mechanisms (Coase, 1937; Surowiecki, 2004). In the absence of hierarchies and division of labor, group members need to constantly survey all the other members and communicate with them to keep the project going. This takes more and more attention and resources, which as the group increases in size can undermine its ability to subsist as a whole. The typical solution to this problem is to create specialized roles and coordination mechanisms, which allow some of the members to work on the intended group goal, while other members manage the collaboration process. It is also only fair to note that highly hierarchical and strictly compartmentalized groups, with tightly defined divisions of labor, can run into problems of their own. The most prominent is that of inefficient utilization of resources, poor allocation of effort, and inability to fully capture and redistribute local or tacit knowledge throughout the organization (Coase, 1937).

The dilemmas of human collaboration were neatly captured in the seminal work "Wisdom of Crowds" (Surowiecki, 2004). Although sometimes understood as an argument for flat organizations and egalitarian collaboration, the book makes a more complex point. It highlights the fact that task-oriented social groups work optimally when combined with a high degree of autonomous decision supported by flexible methods of aggregating and communicating information about group processes. Groups are, according to Surowiecki, more likely to come to right solutions when sufficient diversity of opinion, expertise, and interest is combined with social structures and communication tools that can aggregate these opinions and experiences and make them visible to the group in an effective way. Extending Surowiecki's phrase, we propose that for groups to be wise, they need division a labor, role allocations, and the communication tools and channels that allow them to become aware of their own inner working. Furthermore, self-awareness can be enhanced if information refers not only to the task and its completion rate, but also to the manner in which its outcome is produced. Given the uneven and socially structured nature of human tasks already discussed, it is especially important that information aggregation systems communicate in an effective manner how effort has been allocated, who has done what and to what effect. While this can be accomplished in many ways, the ideal situation would be one where such information reflects both global and individual facets of collaboration. In what follows we will present a methodological approach and online tool for monitoring and fostering group collaboration, especially in a learning environment. The tool provides information about the level of collaborative evenness and group structure through charts and colors that reflect group entropy levels. In addition, the tool is meant to facilitate our understanding of how uneven collaboration influences group effectiveness especially in a learning environment.

3 Measuring Collaborative Unevenness

3.1 Shannon's Entropy Theory

In previous work (Matei et al., 2006) we have proposed Shannon's Theory of Communication (Shannon & Weaver, 1949) as an approach and its companion measure, social entropy, as a possible measure for understanding collaboration within online and/or technological systems, especially wikis. Shannon used the social entropy index to capture the degree to which a communication system contains information (Shannon & Weaver, 1949). To accomplish this, Shannon employed a well-known physics measure, entropy, which is connected to the second law of thermodynamics, that states that all physical systems have a tendency to devolve to the point where the level of energy is zero and all their elements are equally likely to be in a random state. Shannon took the entropy measure from the physical to the communicative and as we will show below, to the social realm. His novel proposition was that communication can be conceived in terms similar to those of a physical system. In nature, when all elements of a system (e.g., atoms) occur randomly, their prevalence is approximately equal. The system is in a state of chaos and entropy is at a maximum. When physical particles get organized in more and more complex compounds, which privilege some elements at the expense of others, entropy decreases.

Communication can be seen as a system as well. Symbols, similar to atoms in the physical world, are the basic units. A communication system will probably contain no information and its entropy will be at a maximum when symbols are equally likely to occur. In other words, when the order of the symbols is decided by chance alone, there is no information (Shannon & Weaver, 1949). On the other hand, information-laden communication will utilize specific units of meaning more often than others, and entropy will decrease as symbols, just like physical particles, occur in a biased manner (Seife, 2007). Thus, if applying the entropy formula to a communicative system, the less organized it is, the higher the entropy and the less likely to contain information. The opposite is also true—the more organized the system, the higher the amount of information, and the lower the entropy.

3.2 Social Entropy Theory

Shannon's theory can be extended further, from communicative to social interaction. If we consider communication broadly, as the main mechanism by which social interaction takes place, all human affairs can be understood through the exchange processes that make them possible. Social interaction can be seen as an extended process of communication reliant upon a system of symbols and can be studied through the lens proposed by Shannon. Social systems whose members interact with each other in a nearly random manner, quasi-egalitarian, are more likely to lack a definite structure. Social systems that form a specific structure of interaction, where symbols are exchanged according to specific rules and patterns possess a more definite, structured form. Moreover, while in the first situation the exchanges will be completely even in terms of output/input ratios (everyone is equally likely to send symbols to everyone else), in the second case there will be a definite bias in terms of who will send information to whom.

From a mathematical or statistical perspective, social entropy measures to what degree specific system units (individuals) are more likely to contribute to or in the workings of the system than what chance alone would predict. The social entropy of a group is maximized when a group member is just as likely to communicate, share the effort or contribute an output unit as any other member. In statistical terms, for each of them, contribution would not be greater than what chance alone would predict. It would be purely random. On the other hand, as members take upon themselves or are assigned specific tasks and communicate in a patterned way by interacting in a preferential manner with other members, frequency and amount of output or contribution become non-random. Chance alone cannot predict these outcomes. Entropy, when measured as likelihood of individuals to contribute randomly, starts to decrease. When non-random behavior emerges, however, we have more than simple unevenness and deviation from what chance alone would dictate. Patterned interaction goes hand in hand with roles, rules and division of labor or functional differentiation. The group has become, in fact, structured. More concisely, a social group is more structured when its members are organized in a specific chain of communication and coordination, where some interact more than others, and less structured when members interact randomly (thus, theoretically, equally) to each other. Calculating the entropy of each social situation reveals in fact how structured the group is. Structure is inversely proportional to entropy.

4 Entropy: A Higher Level Structural Indicator

As previously mentioned, groups that are dominated by some of their members are also more likely to have a given structure. This structural characteristic can be captured in a direct way by social entropy: top heavy groups have lower, while egalitarian groups have higher, entropy levels. In this we take a cue from Shannon's original intent in proposing social entropy as a measure for how "informed" (organized) a social (communicative) reality is.

In extending Shannon's theory from information to other realms of inquiry, we continue a line of work with a distinguished past. For example, social and communication scientists, such as Hiltz and Turoff (1978), Schramm (1955) or Bailey (1990) have applied entropy theory and its attendant methodologies to specific social scientific problems, such as small group structuring, system theory, media landscape organization, diversity of media production, and so on. Economists, environmental scientists, or human geographers have also used entropy to characterize the social structure and diversity of industries, occupations, species, or populations (Bailey, 1990; Matei et al., 2006).

In our own work we have analyzed the emergence of social structures on Wikipedia utilizing articles as systems, contributors as system units and their amount of contribution as means for characterizing "system states" (Matei, Braun, & Petrache, 2009; Matei & Bruno, 2015). Calculating the degree to which contributions to Wikipedia articles are random or not, we observed that such contributions tend to be generated by a relatively small group of logged in contributors. Using entropy as a synthetic measure of contribution bias we found that article specific entropic contribution values tend to decrease and to reach a plateau after the 500th editorial intervention. Furthermore, even after this point, entropy keeps decreasing steadily, although at a slower pace. In other words, after the 500th editorial intervention the structure of collaboration within an average Wikipedia article is dominated by a relatively small number of users whose influence keeps increasing at small but steady pace. At the level of the entire Wikipedia space, entropy varies widely at the beginning of the project (2001–2002), reached a peak in 2005, and reached a steady state by 2006 (Matei & Bruno, 2015). Overall, 1 % of Wikipedia users generate 77 % of content. This reflects findings of similar research, such as of Ortega et al. (2008), who found that less than 10 % of Wikipedia members contribute up to 90 % of content, a trend that has dominated Wikipedia for the last several years.

Rooted in this scholarly tradition and building upon our own research, we propose that social entropy could be used to measure how structured or unstructured a group is. More specifically, we reformulate Shannon's theory of information to suggest that:

1. Information and "structure" go in the opposite direction of entropy;
2. Information and structure, especially in the social realm, are intrinsically connected; and,
3. Structure (of a language, symbol system, or group organization) can be measured with one synthetic indicator, namely entropy.

We emphasize the connections between social entropy and structure because groups are more than mere aggregations of people who share the same space. A group is the structure of ties between its individuals. Individuals that occupy specific roles in this structure communicate, contribute or interact in a specific way. The distribution of outputs in the group will follow the curve of abilities, productivity, task and power allocation specific to each role. Employing Shannon's entropy measure to describe group efforts, communicative patterns and collaborative patterns, we expect that as a group becomes more structured (i.e., roles emerge, tasks are assigned or assumed, power and information starts flowing from specific nodes to other nodes), imbalances in the distribution of communication or work will appear.

In other words, as the group starts to form and its structure to emerge, group units (individuals) start behaving in a predictable and non-random way. This predictable pattern entails a specific amount of unevenness. It is important to mention that "specific" has no normative meaning in our research. We have no a priori preference for any given level of unevenness, nor do we think that unevenness is demanded by "natural," individual characteristics. Rather, we propose that unevenness, while ever present, is a dynamic group process. Any group member can theoretically occupy any level of contribution or interaction. For each group and type of structure, some of which can be flatter while other more hierarchical, there is a "specific" level of unevenness and social entropy that needs to be observed and explained, not predicated.

5 Visible Effort: A Technology for Moderating Wiki Collaboration

In what manner can social entropy be employed for building and employing online collaborative tools? We use entropy in a collaborative tool, built on top of a wiki platform that communicates in a direct and active way, i.e., how even or uneven the collaborative efforts of any given group is at any specific point in time. Specifically, Visible Effort measures and displays entropy levels and, as discussed above, group structure. Entropy and structural information are funneled directly back into the collaborative process, or delivered to the group moderators or administrators (who can monitor and direct the process in a proactive manner).

The Visible Effort tool, used with a wiki, has the ability to measure and monitor on a continuous basis the degree to which a group is structured. If needed, it can also be used to maintain collaborative work within certain levels of equitability and evenness. Thus the tool serves a double purpose. On the one hand, it can be used as a monitoring tool, for understanding how collaboration is structured. On the other, it can be employed for adjusting collaboration along particular parameters desired by the instructor or site administrator.

Visible Effort is powered by a Mediawiki extension. Mediawiki is a content management system, originally designed to power Wikipedia, through which content can be edited by any user, including non-registered ones. All changes are permanently stored, and access to information that was edited or added is instantaneous. In addition, all pages come with "talk" areas, which allow discussions and interactions about the editing process. This makes it well adapted for collaborative work, especially of a textual nature.

The fact that all contributions of all users are preserved, regardless of whether they still exist in the current version of the text or not, facilitates an ongoing analytic process that can tell, for each point in time, how even or structured the process of collaboration is. This is accomplished by counting the number of characters that each user has contributed to the document. This count may also include credit for images or other types of content, depending on the option chosen by the administrator. There are two counts that may be utilized. The *gross* contribution uses the total number of words the user has contributed over the document's entire life, whether those words have survived into the current version or not. The *net* contribution is the count only of contributed words that exist in the current, or latest, version of the document. Once calculated, these values are stored by Visible Effort for each revision of the document, so that users can view the contribution scores for any past version of the document.

To process any particular revision for word counts, Visible Effort retrieves the wiki-markup pages for the current and the immediately preceding revision, converts them to plain text, and stores them in files. A UNIX utility is used to compare the files on a word-by-word basis. A difference value calculated for each specific version is assigned to each user and saved in the wiki database. These values are then used for calculating entropy values. Entropy values are then used to shape the page layout using easily comprehensible conventions. The goal is to provide "at-a-glance" information about the collaborative process. As collaboration becomes more (or less) even, background colors change and the graph indicates the size of the collaborative group and who has done the most work so far. In this way the cognitive effort involved in comprehending the project's collaborative status is dramatically minimized.

Key visual elements of the collaborative space (page) are formatted using visual cues that communicate the project status through a diversity of measures. The visual elements include text frames of specific colors and interactive displays (charts). Of these, the most important is the frame that surrounds the page, which changes colors/shades according to the entropy value of each page version that is displayed at a particular point in time—the colors darken or condense as the level of entropy increases. This communicates, at a glance, to the instructors and to the users how even (or structured) the collaboration process currently is. When the color is the lightest, the collaborative effort should be assigned to only one member of the team, thus entropy is 0. When the color is the darkest, there is perfect equality (evenness/high entropy = 100). In addition, there is a chart that visually reflects the distribution of effort for each collaborator as well as tabular information that reflects the number of words or characters contributed by each individual. The system allows electing to visualize or not visualize the entropy levels of each given page, according to the manager or instructor's preferred strategy. Administrators can use the entropy level as a direct indicator for the users, who would be able to see how even or balanced the collaborative effort is. Or, they can hide the information from the users, who would work blindly. Managers or instructors would only send textual and verbal messages to participants about their level of contribution or, given the data provided by VE, they could alter or improve the assignment while it is underway.

6 Use Scenario

Online collaborative learning is in many situations a very effective educational tool. For example, researchers continue to examine the possibility of how distance-learning within virtual worlds, like Second Life, fosters socialization (Kehrwald, 2008), while providing virtual spaces for exploration and creativity that enhance the collaborative learning experiences. Such community learning spaces foster interaction and intrinsic motivation while discovering new knowledge (Faiola & Smyslova, 2009). Moreover, the notion of intrinsic motivation has significant implications for researchers interested in understanding what occurs when the learning activity and environment elicit motivation in students. This is seen when the goals and rewards of learning are meaningful or when the learning assists the learner in obtaining valued accomplishments (Brandt, 1995; Chance, 1992).

Yet, it is inarguable that within groups some individuals have more to offer, others less, and teachers are intimately aware of this reality. If left to his own devices, Stephen, a motivated academic star, may do more than his fair share in the project. Clearly, though meant to benefit the group, teachers would be misguided in stifling his contributions in an attempt to bring them down to the level of the others. Likewise, Sally, a reserved, shy student, may have something of value to contribute to the group even though her participation efforts might not seem overtly active or significant (Lave & Wenger, 1991). Such a perspective is congruent with constructivist learning theory (Vygotsky, 1978), which emphasizes the social nature of knowledge and learning. It is expected that individuals will learn more when interacting with others, because they will be able to construct knowledge socially. Furthermore, collaboration need not be perfectly egalitarian to be successful.

Groups lead by the best students, who contribute above average, tend to perform better (Webb, Nemer, Chizhik, & Sugrue, 1998).

Our current usage of Visible Effort is situated in this constructivist context. Visible Effort aims to foster smart user choices and interactions, as well as instructor interventions, all guided by knowing how even or uneven the collaborative process is. At present, the extension is used in a number of research activities that aim to better understand the advantages and disadvantages of collaborative learning. Another motive behind our research is that while the positive effects of structured collaborative learning (Johnson & Johnson, 1999; Slavin, 1996) have been well known for some time, previous research seemed to explain this in view of individual attributes (Dillenbourg, Baker, & Blaye, 1996). Synthetic group measures that capture general level of structure in the manner proposed by our interpretation of Shannon's social entropy have rarely been used in research on group learning. In addition, equality of effort seemingly has always been the assumed goal. While our research agenda makes no specific point whether this preference for a normative state of equality arose from value-laden positions or not, we do propose that complete evenness of effort would rarely be an ideal operational state of interaction.

To test this proposition, we have devised a quasi-experimental program, which utilizes VE. Students are tasked to create group reports and term glossaries that are incorporated in class assignments. The main goal is to empirically identify the degree to which collaborative evenness promotes learning or not. Learning is measured as acquisition of knowledge related to specific concepts and theories discussed in the group reports and glossaries. Our main contention is that learning outcomes improve as groups become differentiated. As members start contributing according to their level of knowledge and learning needs, a specific social structure of learning emerges. This structure offers each student a given role and comfort zone. Consequently, students will contribute in different ways, according to their needs, abilities, and motivations. The groups they participate in will be characterized by a specific level of collaborative differentiation and unevenness that will go hand in hand with a specific level of learning effectiveness. We further hypothesize that the relationship between learning outcomes, group structure and collaborative unevenness is curvilinear. If collaborative unevenness and its companion level of group structure reach the level where some of the group members constantly dominate the collaborative process or where too many members "free ride," learning is disrupted. Group processes are increasingly hindered by discussions and conflicts about optimal level of contribution, reward allocation, and equity. Collaboration slows down or even ceases. However, the inverse is also problematic. On the other extreme, collaboration can also become too even, wherein top performers may not be allowed to stand above the others and consequently raise performance of the group whole. In this context, we are interested in finding out to what degree making the level of collaborative evenness and group structure known to the group members through the visual cues provided by Visible Effort can maintain the group within optimal collaboration values.

Our ultimate goals are, thus, three: (1) to determine the range of collaborative unevenness within which collaboration and learning are optimal, (2) to uncover the inflection point where collaborative unevenness and group structure ceases to promote learning in an online environment, and (3) to understand to what degree visual feedback can be used for moderating group CMC behavior.

7 Aligning the Conceptual with the Actual

Bruno explored these issues in a study of learning gain through wiki interaction (2010). He focused mostly on the social dynamics, rather than on the visualization effects. He observed small groups of individuals (<10 members) working on finding answers to questions posited by the researchers. A total of 170 undergraduate students were organized into 23 groups and tasked to answer questions posted on the Visible Effort wiki. The questions were mainly focused on widely available information about university campus buildings, history, or traditions. The answers were purely textual. Respondents could add new data, edit, or replace existing information. Respondents were asked factual questions about the topics both before and after the activity. Comparing answers in pre and posttests, Bruno (2010) calculated a net knowledge gain score, which was averaged for each group. Comparing learning gain with group inequality, quantified as normalized entropy (observed entropy/maximum possible entropy) he detected a curvilinear relationship. The highest and lowest levels of entropy hindered knowledge gain at the group level. Knowledge gain was maximized at a level between these two extremes, although it should be said that it was toward the high end of the spectrum, where interaction was much more even than in non-experimental situations (Bruno, 2010).

8 Significance

Bruno's findings highlight the ability of online interaction spaces to foster learning and the need to monitor the level of structuration/entropy to better understand the optimal levels of interaction. Of course, this is best done in spontaneously occurring situations, rather than in experimental settings. Furthermore, research needs to be conducted on the impact of the interaction visualization of present level of structuration/entropy by the participants in the ongoing social interaction. Would such visualization help interactors better understand their current status in the collaboration process? Would the higher performing actors become more or less motivated by their presence in the contributor elite? Will the other contributors be positively or negatively impacted by their relatively lower position in the interaction hierarchy?

Above and beyond these questions, the Visible Effort wiki can be utilized by teachers or knowledge managers in a unique manner. It offers immediate individual and group-level participation feedback that can be passively or actively utilized. It is not proposed to take the place of other tasks any teacher must undertake in the way of student and group assessment, but is simply another powerful implement in the toolkit. This paper only scratches the surface of what is possible. If proven effective, the theoretical and technical applications of the ideas discussed here could conceivably be applied in countless ways, collaboratively utilizing countless emerging technologies. The tool might also be extended to other platforms, such as online management and writing environments (Google Docs, Microsoft Office Live or Zoho), where collaboration can be supported by group work on free standing documents, not directly connected to a wiki, and for any type of assignment. In identifying and isolating what constitutes optimal student collaboration, many different kinds of group projects with different intended goals and outcomes could be carried out—not only those of a cognitive nature. And of course, in that sense, many different forms of learning could also be conducted and measured through similar means. What is not in doubt is the significant benefit theory-driven technologies, such as the Visible Effort wiki, would offer students, instructors, and business organizations.

Acknowledgements The authors would like to thank the programmers Bill Schroeder and Shammi Diddla, as well as David Eisert and Terri Patterson who worked on the software code of the MediaWiki extension. Visible Effort is supported by a competitive grant awarded by the Teaching and Learning Technologies Unit at Purdue University.

References

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. New York: Hyperion.

Bailey, K. D. (1990). *Social entropy theory*. New York: SUNY Press.

Bales, R. (1950). A set of categories for the analysis of small group interaction. *American Sociological Review, 15*(2), 257–263.

Barabási, A. (2003). *Linked: How everything is connected to everything else and what it means for business, science, and everyday life*. New York: Plume.

Barlow, J. P. (1994, August 11). *Jack in, Young Pioneer*. Retrieved December 12, 2014, from http://w2.eff.org/Misc/Publications/John_Perry_Barlow/HTML/jack_in_young_pioneer.html.

Benkler, Y. (2007). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Brafman, O., & Beckstrom, R. (2006). *The starfish and the spider: The unstoppable power of leaderless organizations*. New York: Portfolio.

Brandt, R. (1995). Punished by rewards? A conversation with Alfie Kohn. *Educational Leadership, 53*(1), 13–16.

Bruno, R. J. (2010). *Social differentiation, participation inequality and optimal collaborative learning online* (Order No. 3444483). Available from Dissertations & Theses @ CIC Institutions; ProQuest Dissertations & Theses A&I. (858608194). Retrieved from http://search.proquest.com/docview/858608194?accountid=13360.

Chance, P. (1992). The rewards of learning. *Phi Delta Kappan, 73*, 200–207.

Coase, R. (1937). The nature of the firm. *Economica, 4*(16), 386–405.

Dillenbourg, P., Baker, M., & Blaye, A. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in humans and machine: Towards an interdisciplinary learning science*. Amsterdam: Pergamon Press.
Draper, J. V., & Blair, L. M. (1996). Workload, flow, and telepresence during teleoperation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1030–1035. Minneapolis, MN: IEEE Society.
Faiola, A. & Smyslova, O. (2009). Flow experience and telepresence in second life: Correlating pleasure, immersion, and interaction in virtual communities. In G. Salvendy and J. Jacko (Ed.), *Proceedings of the 13th International Conference on HCI*. San Diego, CA: Springer.
Hiltz, S. R., & Turoff, M. (1978). *The network nation: Human communication via computer*. Reading, MA: Addison-Wesley.
Huberman, B. (2001). *The laws of the web*. Boston: The MIT Press.
Johnson, D., & Johnson, R. (1999). Making cooperative learning work. *Theory into Practice, 38*(2), 67–73.
Kehrwald, B. (2008). Understanding social presence in text-based online learning environments. *Distance Education, 29*(1), 89–106.
Kelly, K. (1995). *Out of control: The new biology of machines, social systems and the economic world*. Reading, MA: Addison-Wesley.
Kittur, A., Chi, E., Pendleton, B., Suh, B., & Mytkowycz, T. (2007). *The Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie*. Presented at the 25th ACM SIGCHI, San Jose, Retrieved from www.users.cs.umn.edu/~echi/papers/2007-CHI/2007-05-altCHI-Power-Wikipedia.pdf.
Kuk, G. (2006). Strategic interaction and knowledge sharing in the KDE developer mailing list. *Management Science, 52*(7), 1031–1042.
Lampe, C., Ellison, N., & Steinfield, C. (2006). A Face(book) in the crowd: Social searching vs. social browsing. *Proceedings of CSCW-2006*, 167–170. New York: ACM Press.
Lave, J., & Wenger, E. (1991). *Situated learning. Legitimate peripheral participation*. Cambridge, England: Cambridge Press.
Lease, E. (2007, January 22). *Wise crowds with long tails* [Web log message]. Retrieved from http://techessence.info/node/85.
Matei, S. A., & Bertino, E. (Eds.). (2014). *Roles, trust, and reputation in social media knowledge markets: Theories and methods*. New York: Springer.
Matei, S. A., & Bruno, R. (2015). Pareto's 80/20 Law and social differentiation: A social entropy perspective. *Public Relations Review, 41*(2), 178–186. http://doi.org/10.1016/j.pubrev.2014.11.006
Matei, S. A., Braun, D., & Petrache, H. (2009). The structure of social collaboration on Wikipedia. *Paper presented at Wikimania 2009*, Buenos Aires, August 26–28.
Matei, S., & Ball-Rokeach, S. J. (2001). Real and virtual social ties: Connections in the everyday lives of seven ethnic neighborhoods. *American Behavioral Scientist, 45*(3), 550–564.
Matei, S., Oh, K., & Bruno, R. (2006, Nov). Collaboration and communication: A social entropy approach. Paper presented at the *National Communication Association Annual Convention, San Antonio*. Retrieved from http://matei.org/ithink/wp-content/uploads/2008/12/communicationcollaborationentropyfinal.pdf.
Nielsen, J. (2006, October). *Participation inequality: Encouraging more users to contribute* [Web log message]. Retrieved from http://www.useit.com/alertbox/participation_inequality.html.
Ortega, F., Gonzalez, J., & Robles, G. (2008). On the inequality of contributions to Wikipedia. In *Proceedings of the 41st Hawaii International Conf on Systems Science*. Washington, DC: IEEE Computer Society.
Powazek, D. (2009, May). The wisdom of community. *A List Apart Magazine,* 283. Retrieved from www.alistapart.com/articles/.
Raymond, E. (2001). *The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary*. Cambridge, MA: O'Reilly.

Rheingold, H. (2002). *Smart mobs: The next social revolution* (1st ed.). Cambridge, MA: Basic Books.

Schramm, W. (1955). Information theory and mass communication. *Journalism Quarterly, 32*, 131–146.

Seife, C. (2007). *Decoding the universe*. London: Penguin.

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Shirky, C. (2008). *Here comes everybody*. New York: Penguin Press.

Slavin, R. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*(1), 43–69.

Stephan, F. F., & Mishler, E. G. (1952). The distribution of participation in small groups: An exponential approximation. *American Sociological Review, 17*(5), 598–608.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. New York: Doubleday.

Tapscott, D., & Williams, A. D. (2006). *Wikinomics*. New York: Portfolio.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Webb, N., Nemer, K., Chizhik, A., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal, 35*(4), 607–651.

Stepwise Segmented Regression Analysis: An Iterative Statistical Algorithm to Detect and Quantify Evolutionary and Revolutionary Transformations in Longitudinal Data

Brian C. Britt

This text offers a number of new, innovative measures for use in studying human behavior and social systems, particularly in social media and organizational contexts. Briefly, Ignajatovic, Rezvani, and Bertino (2015) discussed assessments of content reliability and contributor trustworthiness, as did Mustaro, Frango, Gobbato, and Kuma (2015). Russell, Still, and Huhtamäki (2015) offered a framework for measuring ecosystemic relational capital. Matei (2015) detailed the Visible Effort tool for measuring system-wide social entropy. And Wei, Zhu, Liu, Matei, and Britt (2015) used activity data to profile Wikipedia users and classify them based upon their elite stature or lack thereof. These metrics serve to build upon more rudimentary figures like the sheer number of followers of a given user or the amount of content contributed in terms of post count or word count, and they join the array of increasingly sophisticated tools developed over the years that we may use to evaluate social processes and products in the offline and online realms alike.

The measures listed above all focus upon human interactions, but they may be implemented in a broad range of settings, from social media sites like Facebook and YouTube to offline interactions like strategic planning meetings within a corporation or community. Further, these tools may be used to repeatedly assess the same content, subjects, or groups over time, which makes them valuable metrics for addressing one of the biggest gaps in the social scientific literature (and, for that matter, the natural sciences): that of longitudinal studies and, more specifically, of change, growth, stagnation, and decline, as we may define them across a wide range of domains, over long periods of time.

Until recently, longitudinal studies were exceedingly rare in many disciplines, as it was difficult and time-consuming to repeatedly and consistently collect data over extremely long periods. Researchers who were determined to collect worthy longitudinal data found their efforts plagued by participant attrition, the loss of

B.C. Britt (✉)
South Dakota State University, Brookings, SD, USA
e-mail: brian.britt@sdstate.edu

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_7

theoretical relevance for their study, changes in the measurements to be applied, and numerous other obstacles (Menard, 2002; Ruspini, 2002). But now, with the massive expansion of electronic communication and records, there are countless long-term data sets on human behavior that are primed to be mined and analyzed immediately (Smith, 2015), representing a veritable gold mine for anyone interested in conducting retrospective longitudinal research.

This, however, brings up a serious problem. Present methodological approaches suffer from several major shortcomings that prevent them from adequately modeling inherently volatile social systems like those described throughout this text, perhaps in large part because longitudinal data within such disciplines was not commonly available until recently. Internal and external forces can fundamentally and permanently transform such communities without warning (see Britt, 2011, 2013), so a simple linear or curvilinear equation may neglect to identify and account for these critical change moments, or "breakpoints," yet even the most sophisticated tools that have been developed to quantify change over time suffer from significant limitations that lead them to miss many of the most important findings in exploratory studies.

There are many long-standing techniques like ARIMA and Markov chain modeling, the latter of which was used in the present volume by Wei et al. (2015), which can provide great insight into more systemic, internally driven evolutionary processes. However, these approaches are unable to explain many revolutionary changes like those described above, especially those that are spurred by external forces. Such methodologies are only suitable if one can safely rule out the likelihood of instantaneous transformative change within a given data set, a requirement which frequently cannot be satisfied in longitudinal studies.

A comparatively small number of other tools exist to pinpoint key change moments, but these few algorithms are generally unsuitable for detecting transformative changes in social systems. They tend to be especially susceptible to the sheer amount of random noise inevitable in studies of human subjects, and the sheer quantity of data available in contexts like the aforementioned social media sites (see Smith, 2015) likewise causes substantial computational problems.

As a response to the analytical and computational limitations of existing methodologies, this chapter offers a versatile, iterative regression-based approach that was specifically designed to be robust enough to overcome these shortcomings. This methodology, known as stepwise segmented regression analysis, combines the principles of segmented and stepwise regression to detect changes as they occur in any given measure over time, making it easy to apply to any set of sequentially ordered data. The algorithm's robustness ultimately makes it suitable for an extremely wide range of online and offline contexts and metrics, from Twitter posts and Wikipedia revisions to individual attitudes and intracorporate discussions. It is likewise well-suited to many research contexts within the natural sciences for which the potential influence of outside forces is being considered (such as, for instance, the longtime debate over the possible impact of humans on the environment). All told, it is an ideal tool for initial explorations of virtually any evolving process or product, regardless of whether the particular unit of analysis is an organization, a sub-group, an individual, a document, the environment, or anything else that one might choose to study.

1 Data Organization

Any attempt to analyze changes over time must first begin by sequentially ordering the data in question. Generally speaking, this consists of organizing the data into equally sized windows, with each data point comprising consecutive minutes, hours, days, weeks, or whatever other unit of time the researcher might choose. For example, we might consider the number of Twitter posts made on a given subject on Jan. 1, Jan. 2, and so forth, which subsequent data points representing the tweets made on subsequent consecutive days. Window sizes may naturally vary depending on the research context and the researcher's particular goals—some may even choose to define "window size" based solely upon the precision of the original data set, using the exact timestamps recorded for particular actions (down to the minute, second, millisecond, etc.). The selection of an appropriate window size and shape is beyond the scope of this chapter; interested readers should consult a time series analysis handbook such as *Time Series Analysis and Its Applications* (Shumway & Stoffer, 2010) for further guidance.

It should be noted that some researchers have suggested using overlapping rather than consecutive time windows. Using our above example, one data point might comprise the tweets made at any time on Jan. 1, while the next could run from noon Jan. 1 to noon Jan. 2, and so forth, with each 24-h window beginning at the midpoint of the preceding window rather than its end. This can help to compensate for the often-arbitrary selection of a particular window size, mitigating concerns about susceptibility to random noise or the loss of information due to a window that is too small or too large. Overlapping windows also have the effect of smoothing the data and stabilizing variance patterns, as they effectively represent a moving average; however, such smoothing may mask the precise locations of breakpoints or entirely hide them from detection. They also introduce multicollinearity that may be undesirable in some settings.

A complete discussion of the advantages and disadvantages of overlapping data windows is likewise beyond the scope of this chapter. With that said, while detecting breakpoints in smoothed data may be more difficult in some cases, as long as the overlapping windows are still equally sized and spaced, the overlap would not necessarily invalidate the analysis—or at the very least, the overlapping windows would be as legitimate as they are when used elsewhere (see Harri & Brorsen, 2009).

2 Methodological Objectives

In order to understand the requirements for an improved methodology to study social systems, human behavior, and natural processes as they evolve over time, it is instructive to consider a simple example. Imagine trying to fit a regression equation, for instance, to the increasing number of news articles published about Barack Obama during the 2008 US presidential campaign. Unlike his closest rival in the

Democratic primaries, Hillary Clinton, many analysts did not consider Obama to be a legitimate contender for the presidency until relatively late in the election cycle. In fact, prominent Republican commentator Rush Limbaugh openly celebrated Obama's rise to prominence for most of the campaign, believing that he would do little more than expose Clinton's weaknesses during the primaries, and that if he did somehow defeat Clinton, that he would be trounced by eventual Republican nominee John McCain (Mooney, 2008).

A careful examination of media patterns could help to reveal when Obama started to be seen as a legitimate contender in the eyes of the press, a moment in time that is not otherwise easy to pinpoint. It is possible, certainly, that Obama's media coverage simply underwent a gradual expansion as coverage of the election itself increased, but it is also possible that one or more polls, debates, primary victories, or other factors served as tipping points for media consciousness. Such tipping points would be suggested by sudden jumps or rapid ascensions in a plot of his press coverage—breakpoints in an otherwise relatively stable curvilinear trend—representing points at which the media recognized a greater need to pay attention to him as a candidate.

Importantly, we might expect similar evolutionary processes and revolutionary breakpoints from examining, for instance, Facebook posts about Obama. Just as it is possible that media recognition of Obama underwent dramatic shifts over time, it may be that public perceptions of him as a legitimate presidential candidate grew by leaps and bounds at particular moments, again in response to polls, debates, primary elections, or any number of other forces such as media reports, campaign advertisements, or the withdrawal of other candidates from the race. The question is very similar regardless of whether the area of emphasis is on the mass media or social media, a fact that further highlights the wide-ranging need and broad potential application for a methodology that identifies such breakpoints.

Either way, this brings to light the first requirement for any such methodology: the identification of both continuous and discontinuous breakpoints. Briefly, these represent "bends" and "jumps" in continuous lines, examples of which are provided in Figs. 1 and 2, respectively. In Fig. 1, the linear trend in the data obviously changes at the instance between the red and blue lines, representing the 100th data point, with a constant slope shifting to a negative one. Continuous breakpoints like this, in which the slope instantly changes but the data values do not undergo an instantaneous shift, may be colloquially considered to resemble an elbow. In the political example above, we might expect to see such a bend in the data from a successful commercial, aired over time, that did not have an immediate impact but that accelerated or decelerated the rate at which media recognition of Obama as a viable candidate grew.

In contrast, Fig. 2 showcases an example of a discontinuous breakpoint, as there is a clear, instant change in the data values. Such a discontinuous breakpoint may be accompanied by a change in slope, which may exacerbate or mitigate the immediate discontinuity, but such a slope change is not necessarily essential. Figure 3, for instance, features a breakpoint that separates two periods of relative stability within the data set. In our political scenario, a major event like a debate result or a surprise

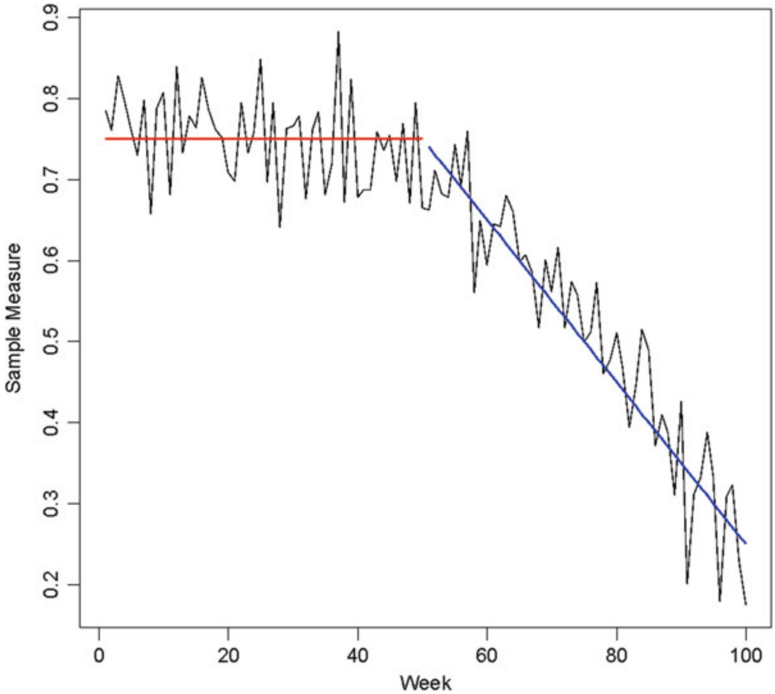


Fig. 1 Fictional data generated from two normally distributed linear trends, split by a continuous breakpoint, with the trends plotted in *red* and *blue*

outcome in a primary election might have caused the sort of immediate impact that we see in Figs. 2 and 3 as opposed to the slope change observed in Fig. 1.

When considering processes that can obviously change in gradual or sudden ways, it would clearly make little sense to focus exclusively on continuous or discontinuous breakpoints, so a methodology designed to detect breakpoints should be able to pinpoint both continuous bends and discontinuous jumps in the data. This includes discontinuous breakpoints like that of Fig. 3, where the breakpoint itself is obviously important even if the periods that bound it lack any significant evolutionary growth or decline within the measure of interest.

Crucially, limiting this search to the identification of a single breakpoint would likewise be insufficient—in our example, one can easily imagine any number of forces and events having their own dramatic effects at multiple points in time, from speeches and commercials to debates and election results, and the same could be said of any other behavioral or attitudinal processes that may be repeatedly altered. While this may sound obvious, it is important to acknowledge, as most prior breakpoint identification studies have been directed toward applications in which only a single breakpoint is expected. For example, toxicologists often want to determine the threshold dose, or the minimum concentration at which a chemical or drug begins to have a noticeable effect on subjects (see Calabrese & Baldwin, 2003;

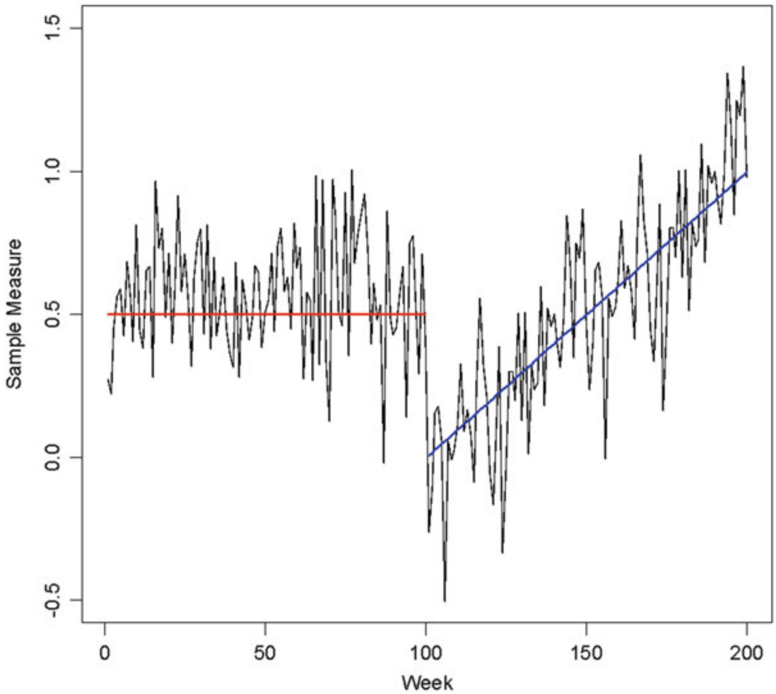


Fig. 2 Fictional data generated from two normally distributed linear trends with different slopes, split by a discontinuous breakpoint, with the trends plotted in *red* and *blue*

and Cox, 1987, for an in-depth review of threshold models in toxicology). Thus, they aim only to find a single breakpoint connecting a line segment with zero intercept and slope—representing no discernible effect from smaller doses—with a second segment that extends from zero, rising in a curvilinear fashion as the applied dosage increases.

In contrast, attitudes and behaviors may fluctuate repeatedly and dramatically over time, as human activities are far less uniform than chemical changes upon dosage increases. The same can be said of many processes in the natural sciences, such as localized weather phenomena or the global ecological landscape. A threshold model may offer useful insights into an isolated event, such as the tipping point at which a riot initiates (Granovetter, 1978), but a single threshold cannot explain the long-term development of more complicated processes, such as public sentiments about race relations over the twentieth century. Any examination of human subjects or natural processes as they develop over an extended period demands that we be prepared to identify multiple potential breakpoints, not just one.

This also implies a third requirement for this methodology: it must be able to pinpoint continuous and discontinuous breakpoints as they occur at unknown dates and times. In other words, the methodology cannot assume in advance the point at which a change "should" occur, as one might in a regression discontinuity study

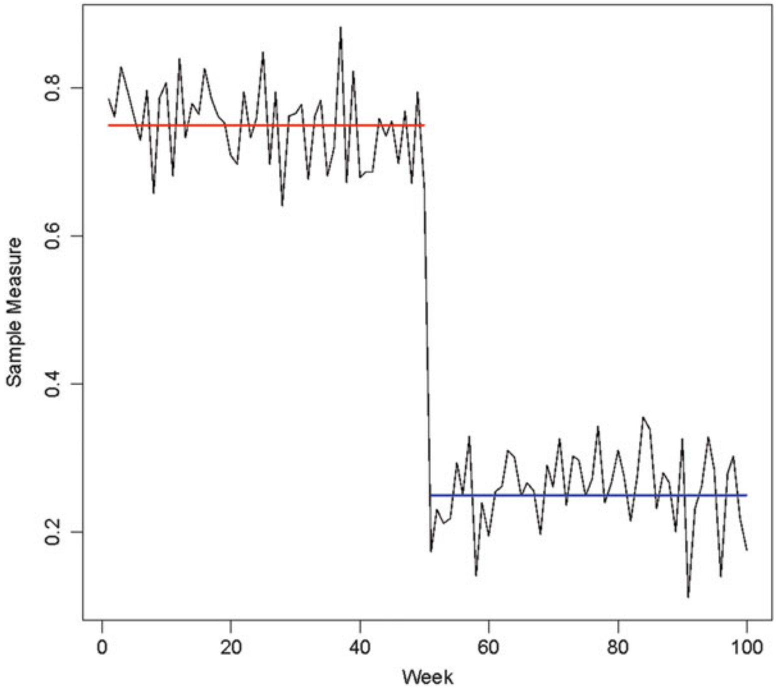


Fig. 3 Fictional data generated from two normally distributed linear trends with identical slopes, split by a discontinuous breakpoint, with the trends plotted in *red* and *blue*

design (see Lee & Lemieux, 2010). After all, the research contexts of interest within this chapter are not studies in which a known condition is applied to subjects at a particular moment in time. Rather, the point is to develop an algorithm for exploratory studies of natural processes and social interactions and products as they change over time in order to detect the emergent transformations that occurred throughout their development. This means that we cannot assume in advance when those changes occurred but must be able to locate them based on the data alone.

Lastly, considering that the primary goal is to evaluate moments of dramatic revolution between periods of stability or gradual evolution, this methodology must be reasonably robust against heteroscedasticity. After all, if we are deliberately looking for fundamental changes in a given data set, then inconsistencies in variance may very well be among those changes. It is to be expected, for example, that if we evaluated a growing corporation, a social media site like Twitter, or an article under development on Wikipedia, that the size and form of the organization or document of interest would change over time, and therefore so would the variance of any measure we might take. The collective behavior of Twitter users, for example, would have been much more susceptible to fluctuation during its earliest days, when only a few dozen or hundred members frequented the site. Under those conditions, so the short-term activity or absence of a single individual would have had a major impact on any organization-level measurement. In contrast, millions upon millions of registered users tweet on a daily basis today and form a relatively stable mass that is hardly affected by individual behavioral changes.

Besides, organizational size aside, if the point is to look for forces or events that fundamentally alter the attitudes or behaviors of people and the forms of groups and documents, then we must take the human element into consideration. If we were to uncover an event that truly shook the foundations of the existing social system to such a degree that it resulted in a revolutionary change within the community, then it is essential to consider the loss of norms that would result.

If, for example, something transpired to alter the manner in which Wikipedia contributors tended to deliver their contributions, with a measurable change in their behavior, then we may infer that their previous, long-standing way of behaving was displaced in some manner. At least in the short term, this could very well leave them adrift, seeking a new "normal" set of activities for themselves and for the organization as a whole, and perhaps experimenting with numerous different ways of participating under the new order before reaching a state of stability.

This scenario naturally lends itself to heteroscedasticity, whether assessed in terms of the organization, the individual, or the documents produced, so this nonconstant variance must be effectively managed by the proposed methodology in order for the fundamental change to be adequately captured and assessed. Importantly, this does not necessarily mean that the methodology should involve transforming the data to remove such heteroscedasticity, as doing so could mask the change of interest just as organizing the data into overlapping time windows would; the only demand is that heteroscedasticity should not unduly violate the core assumptions of the methodology and result in major analytical problems.

In short, an effective methodology to detect and quantify changes in social scientific and natural scientific processes must be able to identify both continuous and discontinuous breakpoints; it must be able to detect an unknown number of breakpoints rather than targeting an arbitrary pre-set number; it must be able to determine the points at which those breakpoints appear throughout the full period of analysis rather than merely checking a limited predetermined subset of moments in time; and it must be able to effectively handle likely heteroscedasticity in the data.

3 Limitations of Existing Methodologies

Now that we have fully established the requirements for a suitable methodology to identify fundamental changes in a given measure over time, it is worthwhile to consider the limitations of other established analytic approaches, highlighting a few of the most common potential alternatives.

First, it would be reasonable to expect sequentially ordered data to exhibit some autocorrelative properties, which might suggest the use of an ARIMA model or a similar statistical approach. The problem is that the key components of an ARIMA model have the potential to mask phase shifts, effectively negating the very trends we hope to observe.

For instance, the entire point of differencing data is to remove linear and higher-order terms, which can make it extremely difficult to detect the locations of otherwise clear changes in the trend over time. We may take, as an example, the fictional data set from Fig. 2, which was generated from two distinct linear trends (given by the red and blue lines) along with normally distributed random noise. In Fig. 2, the trend appears to be fairly constant for the first 100 data points. Around that 100th data point, however, its value plummets, and then the fictional metric appears to increase in a roughly linear fashion for the remainder of the data set. This radical change is easy to observe, even with the naked eye.

Figure 4, however, shows the result of differencing this example data set. From this plot, there does not appear to be any change at all, despite how obvious the transition was in Fig. 2. Now that the data has been differenced, it would be virtually impossible to detect the 100th data point as a key change moment, regardless of the statistical approach employed.

It is clear that while differencing may be necessary to fit autoregressive or moving average terms to certain data sets, it can also mask moments of flux and substantially inhibit any subsequent effort to detect changes. Furthermore, ARMA

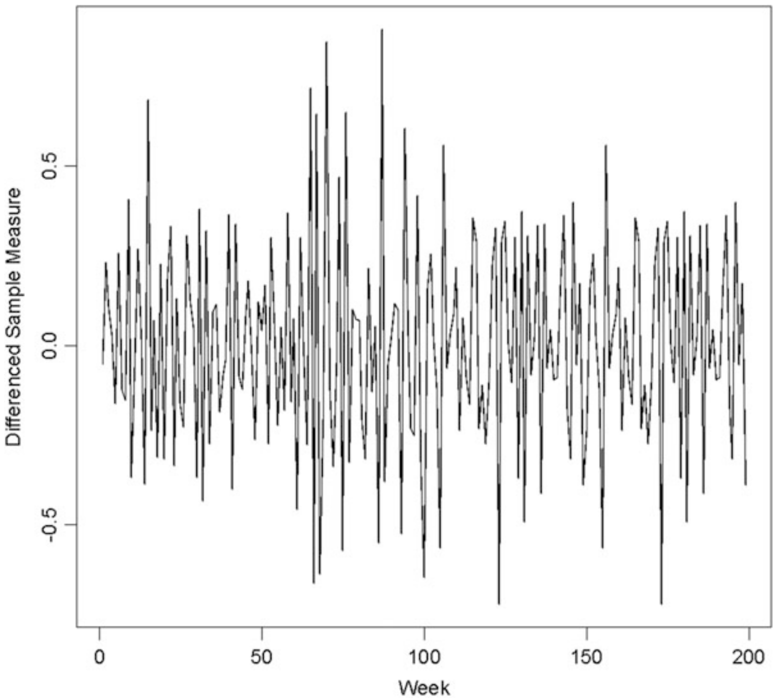


Fig. 4 Fictional data generated from two normally distributed linear trends, split by a discontinuous breakpoint, differenced once

terms alone can obscure otherwise distinct breakpoints. As previously noted, moving averages may help to smooth jagged plots, but in this case such jaggedness may be the most crucial part of the data. Even if a breakpoint is still evident after applying a moving average term, its location may be left in doubt, as the analysis of any given data point will necessarily include information from other data points as well, blending them together and making it difficult to determine at which point a change occurred. Autoregressive terms typically have similar effects, so at best, ARIMA terms could mask the exact location of a given breakpoint, and at worst, they could hide the breakpoint entirely.

A few other techniques exist to handle problems related to change points in data, but each of these suffers from its own critical limitations. Regression discontinuity analyses, for instance, are often used to detect repeated-samples treatment effects in quasi-experiments, but this presupposes that the location of the change of interest is known and only tests the significance of the change at that location, whether in terms of the intercept, the slope, or a higher-order term. In other words, this particular implementation of segmented regression merely assesses the differences between two known line segments. In exploratory studies like those described in this chapter, it is instead necessary to first identify significant revolutionary breakpoints and evolutionary periods whose locations, length, and number are unknown—as has already been established, we cannot assume that we know breakpoint locations in advance, so by extension, we also cannot know the evolutionary periods between those breakpoints.

Several iterative methods have been developed to find the most appropriate breakpoint locations in a data set, given the number of breakpoints for which to search. The SAS statistical package, for instance, uses steepest-descent and gradient descent approaches as well as the Newton, Gauss-Newton, and Marquardt methods of estimating breakpoints. However, these iterative methods can only be used on continuous breakpoints such as those shown in Fig. 1. As such, they are unable to properly handle discontinuous breakpoints like those of Figs. 2 and 3, which we would expect to find in many research contexts. We also cannot presuppose a particular number of breakpoints, so the fact that these methods demand such *a priori* knowledge makes them unsatisfactory for exploratory studies of human attitudes and behaviors.

As a final alternative, a team of statisticians led by Achim Zeileis (Zeileis, Kleiber, Krämer, & Hornik, 2003; Zeileis, Shah, & Patnaik, 2010) devised the "strucchange" algorithm which determines the number and combination of breakpoints in time series data by minimizing the Bayesian Information Criterion (BIC). However, Britt (2013) reported that this approach tends to dramatically overfit data with a great deal of random noise, exhibited by variance that is large compared with movement due to linear and higher-order terms. This is the same sort of short-term volatility and long-term stability that we would expect to see in studies of human behavior or the global climate, with hour-to-hour, day-to-day, and month-to-month variations being far more significant than any year-to-year shifts.

Heteroscedastic variance exacerbates that problem, as Zeileis's algorithm is prone to flooding the more volatile areas of the data set with numerous breakpoints, perfectly fitting random noise and generating comparatively mediocre fits for the more meaningful shifts which occur during periods of lesser variance. Considering the likelihood of observing heteroscedasticity in studies like those described in this chapter and the problems with transforming data in order to remove said heteroscedasticity, Zeileis's approach is not appropriate for such research contexts.

4 Stepwise Segmented Regression

4.1 Underlying Philosophy

With the inadequacy of the above approaches in mind, this chapter proposes an alternative, more broadly applicable methodology based on the principles of stepwise regression model selection. This approach, which was initially devised to examine the development of the Wikipedia co-authorship community (Britt, 2013), expands upon Crawley's (2007, pp. 425–430) method of conducting segmented regression analyses using indicator functions (see also Lemoine, 2012) to quantify the significance of linear evolutionary trends. The present chapter synthesizes and builds upon Crawley's initial ideas, developing a technique that (1) detects and quantifies revolutionary changes to a data set in addition to the significant evolutionary trends between them and (2) seamlessly integrates multiple continuous and discontinuous breakpoints as well as higher-order regression terms into its modeling procedure.

Although we have established that it is important to quantify both evolutionary trends and moments of revolution, the fundamental idea behind this algorithm is to focus almost exclusively on the breakpoints signifying revolutionary change. There are two reasons for this.

First, any methodology that targets the periods of evolution rather than the moments of revolution risks overlooking breakpoints like that given in Fig. 3, where an obviously important revolutionary breakpoint is circumscribed by two periods of stability rather than evolution. If more emphasis is placed on identifying evolution rather than revolution, and an important breakpoint is bounded by two periods of stability rather than change, it is difficult to say that the breakpoint itself will be detected.

Second, the evolutionary periods are inevitably determined by the revolutionary moments that begin and end them. In other words, by definition, instances of revolution change the existing course of evolution, whether they do so via a discontinuous jump in the data (see, for instance, Figs. 2 and 3) or simply by altering its slope (see Fig. 1); consequently, the revolutionary moments, or breakpoints, effectively serve as endpoints bounding each distinct period of evolution or stability. Once the period between two consecutive breakpoints has been identified, it is simple to model its evolutionary trend. In short, pinpointing the breakpoints necessarily means isolating any evolutionary periods as well, so there is no need to specifically target evolution in this analysis—detecting the revolutionary breakpoints results in the same end.

4.2 Procedure

With that in mind, the process of identifying breakpoints proceeds as follows. For a given measure, taken over time, a regression model is fitted with, at minimum, an intercept. We may refer to this as the base model, as this term will never be removed at any stage of the modeling procedure. If the researcher wishes to include higher-order terms in the model, then those should also be added at this time—for instance, the base model might include an intercept, a slope, a quadratic term, and a cubic term. Again, these base terms are permanent fixtures within the regression model and may never be removed.

Next, interaction terms incorporating indicator functions are added and removed from the model based on a stepwise model selection threshold, such as the common $\alpha = .15$ standard. Using this example, at each stage of the process, terms are added to the model if they would be statistically significant at the $\alpha = .15$ level (the forward selection step), then insignificant terms are removed if they no longer meet the $\alpha = .15$ threshold after said addition (the backward selection step). The back-and-forth cycle between the forward selection and backward selection steps repeats until no further interaction terms can be added to the model and no interaction terms already present in the model can be removed.

Each indicator function corresponds to all data points following a given point in the data set. If a given observation of the measure in question occurs before that given point, the indicator function takes on a value of 0. Otherwise, the indicator function yields 1. For instance, if we began a modeling procedure by adding an interaction term corresponding to data point 100 in the model, then the *Intercept** $I(n > 100)$ term would equal 0 for data points 1–100, with n representing the number of a particular data point in its sequential order. For all observations from point 101 to the end of the data set, the $I(n > 100)$ indicator function would be 1, so in the absence of other interaction terms at earlier data points, the sum of the *Intercept**$I(n > 100)$ interaction term plus the global *Intercept* term would yield the estimate of the intercept after data point 100. If we added another interaction term, *Intercept**$I(n > 200)$, then the global *Intercept* term would represent the estimated intercept for data points 1–100; the sum of *Intercept* and *Intercept**$I(n > 100)$ would be the estimated intercept for data points 101–200; and the sum of *Intercept*, *Intercept**$I(n > 100)$, and *Intercept**$I(n > 200)$ would estimate the intercept for all data points from 201 onward. In short, the interaction term represents the change in the intercept that is observed at the breakpoint, regardless of the presence of other terms.

The same approach applies for any higher-order terms in the regression model (linear, quadratic, cubic, etc.). For such higher-order terms, the indicator function is multiplied by the number of the data point, taken to a particular exponent. For instance, the slope change occurring at data point 100 would be given by $n*I(n > 100)$, while the quadratic change at data point 100 would be estimated by $n^2*I(n > 100)$.

A given data point is deemed a statistically significant breakpoint whenever any of the interaction terms including its representative indicator function is significant enough to be added to the model at the chosen stepwise threshold. In other words, if adding a breakpoint to an existing data segment would split it into two segments with significantly different intercepts, slopes, or other terms, then the data point at which the breakpoint is to be added is deemed significant.

Importantly, whenever an interaction term for a particular indicator function is significant at the stepwise threshold, regardless of the order of the term interacting with the indicator function (intercept, slope, quadratic, etc.), all interaction terms involving that indicator function are added to the regression model. For instance, if a given analysis incorporates intercept, slope, and quadratic terms in its core model, and $n*I(n > 100)$ is found to be significant enough to be added to the model, then $Intercept*I(n > 100)$ and $n^2*I(n > 100)$ would also be added. Likewise, none of the three terms would be removed from the model unless all three ceased to be statistically significant. In effect, either the entire breakpoint is included in the model with all of its associated terms, or the entire breakpoint is omitted. Put another way, the line segments immediately preceding and immediately following the breakpoint are treated as wholly distinct, as that is what a breakpoint is intended to signify, so incorporating terms of all orders within each line segment's model ensures that their respective estimates will not overlap.

Terms were also grouped in this way in order to prevent spuriously overfitting the number of breakpoints. Otherwise, one could envision, for instance, a slope change being added to the model in the 100th data point, and a quadratic change later added at data point 101—not because both points served as independent breakpoints modeling distinct changes in the given measure, but merely because fitting the second change to data point 101 happened to explain more variance than adding a second indicator function for data point 100 to the model. Obviously, this would result in the false identification of additional, meaningless breakpoints, making it more difficult to determine where significant changes actually occurred.

Notably, the stepwise segmented regression procedure may result in some higher-order terms being inestimable, even if interaction terms are grouped together to prevent spurious overfitting. It may be possible, for instance, to identify $Intercept*I(n > 100)$ as a significant term, and then to subsequently add $Intercept *I(n > 102)$ to the model as well. This would result in three distinct line segments comprising data points 1–100, data points 101–102, and the data from point 103 onward. Such an outcome is not inherently problematic; however, if the analysis was devised to fit linear terms and quadratic terms, then the linear terms $n*I(n > 100)$ and $n*I(n > 102)$ as well as the quadratic terms $n^2*I(n > 100)$ and $n^2*I(n > 102)$ would also be added to the model. The result is obvious: $n^2*I(n > 100)$ cannot be estimated, as you cannot uniquely estimate a quadratic term for a line segment comprising only two data points (101 and 102). Ultimately, however, this is not a substantial problem for model fitting in this context, as the primary purpose of this methodology is to identify the breakpoints themselves. As long as the breakpoints are accurately identified, it is perfectly acceptable to exclude statistically invalid terms from the modeling procedure and to accept the model given by the lower-order terms that remain.

4.3 Step-by-Step Example

Let us consider how this process might unfold for an example measure, taken 500 times (and therefore resulting in 500 sequentially ordered data points). We may choose to examine this data set in terms of intercept, slope, and quadratic terms. Based on these conditions, we begin with the following core model:

$$Y = Intercept + n + n^2$$

Then, as noted above, the model-building procedure considers the possibility of adding a new intercept, slope, and quadratic term to the model, each of which would be multiplied by a particular indicator function.

$$Y = Intercept + n + n^2 + Intercept{*}I(n > X) + n{*}I(n > X) + n^2{*}I(n > X)$$

The process iteratively builds the regression model with X equal to the number of every data point in the data set for which a breakpoint could potentially be identified. There is no need for it to evaluate the $X = 500$ case, as $I(n > X)$ would be 0 for all points in the data set, so 499 models are developed corresponding to potential breakpoints at data points 1–499. After the procedure builds the 499 models corresponding to the 499 possible values of X, it selects the value of X which would give the regression model with the smallest error sum of squares. For the selected value of X, if $Intercept{*}I(n > X), n{*}I(n > X)$, or $n^2{*}I(n > X)$ is statistically significant at the $\alpha = .15$ level, all three terms are added to the model. For the sake of the example, suppose that the value of X which would yield the smallest error sum of squares is $X = 300$. Then the process would consider the model

$$Y = Intercept + n + n^2 + Intercept{*}I(n > 300) + n{*}I(n > 300) + n^2{*}I(n > 300)$$

and if $Intercept{*}I(n > 300)$, $n{*}I(n > 300)$, or $n^2{*}I(n > 300)$ was statistically significant, this model would be accepted. For instance, if $n{*}I(n > 300)$ was statistically significant, that would mean that the slope of the segment from data points 1–300 was significantly different than the slope of the segment from data points 301–500.

After the three terms corresponding to a particular significant breakpoint are added to the model (the forward selection step), all potential breakpoints that have become statistically insignificant are removed from the model (the backward selection step). For instance, let us say that after several more iterations, at the conclusion of a particular forward selection step, the above model has grown into the following, with breakpoints defined at data points 190, 200, 300, and 310:

$$\begin{aligned} Y = {} & Intercept + n + n^2 + Intercept{*}I(n > 190) + n{*}I(n > 190) + n^2{*}I(n > 190) \\ & + Intercept{*}I(n > 200) + n{*}I(n > 200) + n^2{*}\left(> 200\right) + Intercept{*}I(n > 300) \\ & + n{*}I(n > 300) + n^2{*}I(n > 300) + Intercept{*}I(n > 310) + n{*}I\left(> 310\right) \\ & + n^2{*}I(n > 310) \end{aligned}$$

Since a forward selection step just ended, the process moves into a backward selection step in which all terms are again checked for significance. If all three terms corresponding to a potential breakpoint have become statistically insignificant, those three terms are removed from the model. Suppose that in this case, $Intercept{*}I(n > 200)$, $n{*}I(n > 200)$, and $n^2{*}I(n > 200)$ are all statistically insignificant terms, and $Intercept{*}I(n > 300)$, $n{*}I(n > 300)$, and $n^2{*}I(n > 300)$ are also statistically insignificant. In that case, the two sets of three insignificant terms are removed, and the model becomes:

$$Y = Intercept + n + n^2 + Intercept{*}I(n > 190) + n{*}I(n > 190) + n^2{*}I(n > 190) \\ + Intercept{*}I(n > 310) + n{*}I(n > 310) + n^2{*}I\big(n > 310\big)$$

Since the backward selection step has concluded, the procedure will then move to another forward selection step, evaluating whether the terms for another breakpoint can be added to the model.

The process rotates back and forth between the forward selection and backward selection steps, adding the terms for a significant breakpoint, then removing the terms for any insignificant breakpoints, and then adding the terms for another significant breakpoint, until no further terms could be added or removed from the model. As soon as there are consecutive forward and backward selection steps with the model left unchanged, the procedure terminates, and the resulting model is deemed complete.

It should also be noted that the terms in the model, which we commonly call independent variables in most regression analyses, may be independent from the measure in question but are certainly not independent from one another, as adding or removing the interactions for any particular indicator function inevitably changes the number of data points estimated by the terms defined by the preceding indicator function. For instance, in the final step given in the above example, the removal of all terms that include the $I(n > 200)$ and $I(n > 300)$ indicator functions alters the meaning of the $Intercept{*}I(n > 190)$, $n{*}I(n > 190)$, and $n^2{*}I(n > 190)$ terms—they change from estimating the intercept, slope, and quadratic terms for the line segment from points 191–200 to estimating the same terms for the line segment from points 191–310.

We can therefore see that adding terms to the model or removing terms from it has the potential to profoundly affect the estimation of any other terms. This multicollinearity is inescapable, but unlike traditional model-building, in this case it is not cause for alarm. Recall, once again, that the primary goal is to find breakpoints indicating significant differences between adjacent data segments. If the potential effect of an indicator term is masked due to multicollinearity, that only means that it does not highlight a substantial change from the trends already in the model and does not deserve to be included as a breakpoint in its own right. In much the same way, typical regression diagnostics for supposedly problematic phenomena like outliers are meaningless in this context, as the entire point of this analysis is to detect significant deviations from the existing model.

Once the final model is determined, the respective significance of each term is assessed at a standard significance level for hypothesis testing, such as $\alpha = .05$. Because of the large number of factors that is likely under consideration, the Holm-Bonferroni correction should be applied to the model selection and significance testing thresholds in order to control the experiment-wise Type I error rate. In our example, in addition to the three core model terms included from the beginning, a total of 499 breakpoints were evaluated throughout the analysis. Each of these breakpoints, in turn, was comprised of three terms that could be significant—the intercept, the slope, and the quadratic term—and if any one of these terms was significant, all three were added to the regression model. As such, the Holm-Bonferroni correction in this example should be applied based on the evaluation of 1,500 different regression terms.

Finally, after applying the Holm-Bonferroni correction, we would determine that there is a statistically significant breakpoint present at any particular data point for which any of the interaction terms incorporating its representative indicator function are significant. In our example, if the $Intercept{*}I(n > 190)$ term in our above example is statistically significant, then regardless of the significance (or lack thereof) of the $n{*}I(n > 190)$, and $n^2{*}I(n > 190)$ terms, we would conclude the presence of a significant breakpoint at data point 190, signifying a major revolutionary change that occurred between the 190th and 191st data points. We may then use the parameter estimates given in the regression model to evaluate the evolutionary trend or phase of stability that was active until the 190th data point, and compare it with the unique period of evolution or stability that started at data point 191.

5 Discussion

5.1 Heteroscedasticity and Robustness

One of the most important features of the stepwise segmented regression algorithm is that its basis in regression makes it extremely flexible and robust. For instance, unlike other approaches noted earlier in this chapter, stepwise segmented regression is relatively resistant to random noise and robust against violations of the constant variance assumption. Many of the processes for which this methodology may be applied tend to feature a great deal of random noise compared to relatively slow evolutionary trends, whether the measure of interest targets behavioral activities like social media use or natural scientific phenomena like the global climate, so the fact that this methodology's basis in regression helps it to resist improper fitting due to noise is very important. As for the question of constant variance, the common regression assumption of homoscedasticity ensures that least squares estimators are the best linear unbiased estimators of regression terms, but constant variance is not necessary for the estimates to be unbiased, consistent, and asymptotically normal. We might easily expect a degree of heteroscedasticity to accompany statistically significant changes to an established curvilinear trend, so this algorithm's relative lack of susceptibility to heteroscedasticity compared with other methodologies makes it especially attractive. In fairness, it is possible that extreme heteroscedasticity might still result in some overfitting within the higher-variance portions of the data, but this procedure remains much less susceptible to such modeling problems than others outlined in this chapter.

Notably, if heteroscedasticity in a particular data set was such a major concern that it cast any potential findings in doubt, one could choose to transform the data or run the stepwise segmented regression algorithm using weighted least squares estimation in order to counteract the effects of the nonconstant variance. However, researchers should be aware that heteroscedasticity is rarely a crippling problem for regression-based methodologies like this one, and that both of these solutions has the potential to mask significant changes, either making breakpoint locations difficult to pinpoint or hiding their existence entirely.

As a general rule, this procedure is best conducted on raw, untransformed data, in order to estimate breakpoint locations and effects as precisely as possible. If there do appear to be major modeling problems, it may be worthwhile to apply overlapping windows (i.e., a weighted or unweighted moving average) to smooth the data, as careful implementations may reduce heteroscedasticity while only minimally masking legitimate breakpoints. Beyond that, data transformations and weighted least squares estimation should only be attempted as an absolute last resort.

Additionally, as previously noted, heteroscedasticity itself may be a significant finding. In other words, a substantial variance change may represent a breakpoint of interest, even in the absence of significant changes in the estimated values of regression terms. Such a finding could indicate the stabilization of a process, perhaps among social media users learning and solidifying community norms, or it could instead represent organizational destabilization in response to a crisis. However, there is currently no widely accepted statistical approach to detect and pinpoint a significant change in the variance of a time series process, as most formulae like Bartlett's test are merely designed to detect the presence or absence of heteroscedasticity as a violation of common statistical assumptions. Considering the potential value of pinpointing exactly where or when the variance of a process dramatically grows or declines, it would be worthwhile for scholars and practitioners to have a statistical algorithm to locate significant variance changes beyond an imprecise "eyeball test."

5.2 Extensibility

Another key feature of this methodology is the extent to which it can be adapted based on the particular requirements of the researcher or a given research context. The potential use of weighted least squares estimation is just one example; one could also envision this algorithm being adapted for survival functions, partial least squares regression, non-normal distributions, and any number of other contexts, models, and estimation methods.

Of course, perhaps the most obvious extension of the stepwise segmented regression methodology is the introduction of higher-order terms. Just as the iterative approach allows this procedure to evaluate any number of potential breakpoints within a given data set while maintaining reasonably low computational complexity, its basis in regression makes it easy to include higher-order terms within the modeling process.

With that said, much like the question of transforming a data set or using weighted least squares estimation, researchers should take care when using higher-order terms in a stepwise segmented regression analysis. As a general rule, extremely high-order terms are rarely statistically or practically significant, so there is typically minimal use of higher-order factors in regression equations. This norm is even more important for stepwise segmented regression. Since this approach naturally incorporates substantial multicollinearity among regression terms, the inclusion of excessive higher-order terms has the potential to generate major confounding. There may be conflicts between terms attributed to the same breakpoint, or higher-order terms at consecutive breakpoints may falsely show dramatic shifts from line segment to line segment when a simpler model would reveal no significant change.

As such, terms of a higher order than the linear or quadratic level should only be tested if there is a realistic reason to suspect that they may be important. Just like the dual questions of transformations and weighted least squares estimation, this does not represent an inherent problem with the methodology itself; it is merely a modeling principle for practitioners to keep in mind when implementing it.

6 Conclusions

All told, this chapter offers a new approach to identify and quantify multiple continuous and discontinuous breakpoints within a single data set, along with the periods of evolution or stability between them. Past research has, for the most part, focused more on examining the significance of known breakpoints or on finding just a single breakpoint, or it has been restricted to the identification of continuous breakpoints and the exclusion of discontinuous ones. Further, these approaches have generally focused on finding significant evolutionary trends between breakpoints rather than identifying significant revolutionary moments separating different evolutionary trends or periods of stability, and even innovative algorithms like the strucchange procedure (Zeileis et al., 2003, 2010) appear to overestimate the number of breakpoints and lend undue weight to obviously random fluctuations in data segments with especially high variance.

The approach developed in this chapter, on the other hand, accounts for continuous and discontinuous line segments, is capable of detecting any number of breakpoints, seeks significant breakpoints even if the evolutionary "trends" that bound them are themselves periods of relative stability, and minimizes the false detection of breakpoints that result only from noise or heteroscedasticity in the data. Researchers must still be cautious of creating modeling problems by inserting extraneous higher-order terms into the model, as too many unnecessary terms may introduce undue multicollinearity and confounding effects. Regardless of this isolated cautionary note, however, for the purpose of finding significant breakpoints without overfitting random noise, this approach is a dramatic improvement upon existing procedures.

All told, stepwise segmented regression offers a broad range of prospective applications in the social and natural sciences. It is a robust methodology with limited computational complexity and broad extensibility to a wide array of research needs and domains, particularly given its ability to detect and pinpoint multiple continuous and discontinuous breakpoints, to quantify the significance of the revolutionary changes that they represent, and by extension to provide meaningful estimates of the evolutionary periods between the breakpoints. In sum, this algorithm offers the potential to address some of the most important questions of our time, including the gradual development or decline of human behavior, the transformation or downfall of prominent social communities and organizations, and even the possible impact of industrialization on the world in which we live.

References

Britt, B. C. (2011). *System-level motivating factors for collaboration on Wikipedia: A longitudinal network analysis* (Master's thesis). Retrieved from ProQuest Dissertations and Theses Database. (AAI 1501175).

Britt, B. C. (2013). *Evolution and revolution of organizational configurations on Wikipedia: A longitudinal network analysis* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (AAI 3604712).

Calabrese, E. J., & Baldwin, L. A. (2003). The hormetic dose-response model is more common than the threshold model in toxicology. *Toxicology Sciences, 71*(2), 246–250.

Cox, C. (1987). Threshold dose-response models in toxicology. *Biometrics, 43*(3), 511–523.

Crawley, M. J. (2007). *The R Book*. West Sussex, England: Wiley.

Granovetter, M. (1978). Threshold models of collective behavior. *The American Journal of Sociology, 83*, 1420–1443.

Harri, A., & Brorsen, B. W. (2009). The overlapping data problem. *Quantitative and Qualitative Analysis in Social Sciences, 3*(3), 78–115.

Ignajatovic, A., Rezvani, M., & Bertino, E. (2015). Robust aggregation of inconsistent information—Concepts and research directions. In S. A. Matei, M. Russell, & E. Bertino (Eds.), *Transparency in social media—Tools, methods and algorithms for mediating online interactions*. New York: Springer.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*, 281–355.

Lemoine, N. (2012, August 19). R for ecologists: Putting together a piecewise regression. Retrieved June 6, 2013, from http://climateecology.wordpress.com/2012/08/19/r-for-ecologists-putting-together-a-piecewise-regression/.

Matei, S. A. (2015). Visible effort: A social entropy methodology and research agenda for moderating social dynamics. In S. A. Matei, M. Russell, & E. Bertino (Eds.), *Transparency in social media—Tools, methods and algorithms for mediating online interactions*. New York: Springer.

Menard, S. (2002). *Longitudinal research* (2nd ed.). Thousand Oaks, CA: Sage.

Mooney, A. (2008, May 8). Rush Limbaugh urges vote for Obama. Retrieved June 1, 2014 from, http://www.cnn.com/2008/POLITICS/05/07/limbaugh.obama/.

Mustaro, P., Frango, I., Gobbato, A., & Kuma, T. (2015). Improving search results on open academic content repositories: A co-authorship trustability network-based architecture. In S. A. Matei, M. Russell, & E. Bertino (Eds.), *Transparency in social media—Tools, methods and algorithms for mediating online interactions*. New York: Springer.

Ruspini, E. (2002). *Introduction to longitudinal research*. London, England: Routledge.

Russell, M., Still, K., & Huhtamäki, J. (2015). Measuring relational capital in/for/about ecosystems. In S. A. Matei, M. Russell, & E. Bertino (Eds.), *Transparency in social media—Tools, methods and algorithms for mediating online interactions*. New York: Springer.

Shumway, R. H., & Stoffer, D. S. (2010). *Time series analysis and its applications*. New York: Springer.

Smith, M. (2015). Prospects and challenges for social media research. In S. A. Matei, M. Russell, & E. Bertino (Eds.), *Transparency in social media—Tools, methods and algorithms for mediating online interactions*. New York: Springer.

Wei, W., Zhu, M., Liu, C., Matei, S. A., & Britt, B. C. (2015). When do social media elites become "sticky" and how does this matter: Methods and lessons learned from mining the Wikipedia corpus (2001–2010). In S. A. Matei, M. Russell, & E. Bertino (Eds.), *Transparency in social media—Tools, methods and algorithms for mediating online interactions*. New York: Springer.

Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics and Data Analysis, 44*, 109–123.

Zeileis, A., Shah, A., & Patnaik, I. (2010). Testing, monitoring, and dating structural changes in exchange rate regimes. *Computational Statistics and Data Analysis, 54*, 1696–1706.

Towards Bottom-Up Decision Making and Collaborative Knowledge Generation in Urban Infrastructure Projects Through Online Social Media

Mazdak Nik-Bakht and Tamer E. El-Diraby

Watch a man at play for an hour and you can learn more about him than in talking to him for a year.

—Plato

1 Business of Knowledge and Modern Infrastructure Industry

Evolution of civil infrastructure from a technical artifact into an engineering system and a national asset over the past century has created a new discourse for development, construction, and management of infrastructure, which more and more emphasizes soft and subjective aspects of the system. Modern civil infrastructure is a complex system composed of the physical network of assets together with the social network of actors/users, and their interactions through the operational processes of the system (Lukszo & Bouwmas, 2005). This defines a *sociotechnical* system whose behavior cannot be studied without respect to the associated agents and the related social/institutional infrastructure. This system will be governed by organizational policies as well as social norms and standards. Such a definition for civil infrastructure has improved the role of the society from customers and end users of a service into stakeholders who may influence specifications of the system. This new role introduces new opportunities and challenges to domain decision makers. On one hand, it creates great opportunities for social engagement. Technical and professional decision makers can distill the distributed knowledge

M. Nik-Bakht (✉)
University of Toronto, Toronto, ON, Canada
e-mail: mazdak.nikbakht@mail.utoronto.ca

T.E. El-Diraby
Department of Civil Engineering, University of Toronto, Toronto, ON, Canada

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_8

of public communities (referred to as non-expert or non-mainstream knowledge by Brabham & Sanchez, 2010) to reinforce the decision making procedure. On the other hand, given the diversity of interests and technical sophistications involved, an active participation of the public may result in a chaotic nature for the decision process.

If the 'inter-organizational/networked-based arrangement'—as stated by Keast and Hampson (2007)—is the new measure to facilitate innovation development and diffusion within the construction engineering and management (CEM) industry, expanding such networks to include non-technical stakeholders and end users of the urban infrastructure in the future can expedite the process of innovation in CEM even more. Taking advantage of the '*user innovation*' by involving lower layer nodes of the technical social network in construction projects (such as site workers and technology users) improves the process of innovation in CEM (Sanvido & Paulson, 1992; Slaughter, 1993). This can now be extended to include the knowledge-enabled social communities and end users of the built environment. Active involvement of communities is poised to move sociotechnical infrastructure into an innovative and socially-savvy environment for decision making. Apart from the contribution to the innovation, a mixed network of technical and nontechnical decision makers/decision participants is a response to the demand for more active role of NGOs, political and social groups and communities towards a sustainable development in global construction industry (as addressed by Levitt, 2007). Therefore, the role of Public Involvement (PI) agencies is no longer to promote 'the' best solution, but to empower communities to discover it through democratizing innovation (Von Hippel, 2005).

Public engagement has traditionally aimed at maintaining a balance of power between citizens and their government in the process of decision making. The main role of this mechanism is to minimize the impacts and risks of failure of decisions made in development and construction of infrastructure. Evolving community engagement processes from a passive process of 'public relations' to a process of 'engaged partnership' is one of the requirements for achieving a desired socio-technical model of infrastructure decision making. In the past, community engagement was mostly limited to informing and educating the public with the aim of maintaining the required level of public support. The process is now evolving into other forms of consultations to establish a two-way communication. As Hansen and Jackson suggest (2001), the success of community engagement processes is closely tied into involving the community in a timely manner (from early stages of the project) and continuously (during different phases of the project lifecycle), as well as acknowledging the role of end users (as customers) in shaping decisions related to the infrastructure. At lower levels, static tools such as web pages and open houses are commonly used with objectives such as providing the community with the project specific information or making them aware of decision impacts. However, advances in *Social Web* (Web 2.0) have added a new dimension that includes multi-purpose collaboration between community members for collective deliberation on complex topics.

1.1 Role of Web 2.0 in Public Involvement

As Brabham and Sanchez (2010) indicate, traditional methods of PI are problematic due to the lack of efficiency to engage a fully representative sample of the community, and power dynamics of face-to-face meetings: “citizens may feel their opinions are downed out or feel compelled to self-censor”. Moreover, it is difficult, if not impossible, to ‘educate’ participants of a public meeting with the required level of technical information and details within the limited time frame of a meeting. By analyzing the literature of public engagement best practices in transportation planning, Wagner (2013) suggested three main principles for performance of the engagement process: *accessibility*, *interaction*, and *outcome–orientation*. Web 2.0 can help the online community (e-society) to outperform the offline public in almost all of these dimensions. Epidemiology of knowledge through Social Web is upgrading the e-society into the k-society (knowledge-society). Acquiring the knowledge distributed among the online k-society can result in the development of more robust plans. Moreover, people claim to enjoy participating through online social media and it can increase the level of engagement (Evans-Cowley, 2012).

In short, a meaningful public engagement process must be “open, ongoing, and allow for two-way exchange of information” (Wagner, 2013), and Social Web can help to create all of these features. Many infrastructure planning organizations consider the infrastructure as a marketable product and approach the online public engagement in a similar way to online marketing. Engaging customers in the field of commerce and using reverse marketing mechanisms to collect user innovations are among the best motivating scenarios in this regard.

1.2 Era of Prosumers

Informatics in its modern form does not deal with segregated producers and consumers of knowledge anymore. Today they are both morphed into “*prosumers*”; a portmanteau formed by contracting the word professional, or also producer, with the word consumer to emphasize the active role of consumers in producing the products they use. Circulation of knowledge between producers and prosumers is the hallmark of a big data movement and is a key factor of the evolving knowledge economy. By relying on human intelligence, this movement performs tasks which are impossible to accomplish otherwise. Wikipedia is an example of prosumer culture outcomes. With more than 26 million articles in 286 languages, it is one of the most (if not the most) popular encyclopedias in the world. It has at least 70,000 formal prosumers (editors) among its estimated 365 million readers. YouTube is another ultimate example of *prosumerism*. People on this website create (post), “consume” (watch), re-create (re-mix) and exchange the products (videos) produced by other prosumers.

In a more formal way, this mentality has given rise to *crowdsourcing* which expands beyond marketing and chattering, to enable "knowledge workers" to solve problems. *InnoCentive*, a problem-solving marketplace, has 250,000 "solvers" competing for more than $35 million in prizes. Still in its (perpetual) beta version, *Amazon MTurk* is utilizing hundreds of thousands of workers (500,000 as of Jan 2011) from all around the globe (over 190 countries as of January 2011) for a large number of human intelligence tasks of different types. *Spigit* in the corporate social innovation field, *Covisint* in the automobile manufacturing industry, and *Salesforce* in customer relationship management are among other examples of successful cases in this regard. In business administrations, this mindset paves the way for *reverse marketing* which in turn alleviates harvesting *user innovation* for new designs. As part of a program called "*future by Airbus*", Airbus ran a 2-year global consultation with more than 10,000 of its future passengers, asking them for their requirements, demands, and innovative ideas for a 2050 aircraft. The idea behind this program was to involve people from various backgrounds to shape the future of aerospace industry in a more sustainable way. Since launched in 2008, they have been running a competition every two years with a prize of €30,000, asking students for their innovative solutions (Future by airbus, 2012).

Wisdom of the crowd can particularly be helpful in developing context-sensitive solutions for case-specific issues. "*Online urban guides*" (systems such as *social recommenders*, *rating sites*, and *review services*) are good examples of this type. Today, websites such as *Tripadvisor* and *Yelp*, may be among the most reliable sources of knowledge regarding local services. By adapting the user-generated content and aggregating prosumers' micro knowledge, these websites help decision making in cases where not enough information is documented formally and officially, or where the documented information is in form of a '*negotiable knowledge*'.

1.3 Prosumerism in Infrastructure Industry

In the domain of infrastructure, prosumer culture can be helpful in different stages from detection and adjustment of demand to selection of design and construction alternatives. Knowledge-enabled communities discussing different aspects of their built-environment can reflect demands, interests, and (from time to time) innovative solutions for the infrastructure system. This is an invaluable opportunity for collecting the distributed *micro-knowledge* and shaping more pluralist solutions. Mining online discussions can result in capturing and formalizing the knowledge generated through communications among users who are a part of the socio-technical system and continuously interact with it. On the other hand, analyzing patterns of online social connectivity among such decision participants can foster creation of teams for public consultation in different stages of planning and construction. One major advantage of such a model is its *self-organizing* nature. On one hand, dynamics of discussions will be maintained by participants (which can eliminate concerns regarding outdated information on project websites), addressed by Wagner (2013) among others.

Also, the wisdom of crowds can help to classify and prioritize the project-related content. On the other hand, the users unintentionally evaluate each other through activities such as liking, sharing, following, mentioning, etc. This establishes each individual's influence level in a bottom-up manner.

Governments and other macro-level decision makers in the AEC industry (Architecture, Engineering, and Construction) have recently noticed such advantages and have started to benefit from prosumer culture in the process of engaging public partnerships for infrastructure projects. '*New York city bike share*' program is an example of using prosumers' knowledge in the process of demand detection. In 2011, New York City Department of Transportation (NYC DOT) asked local residents to suggest locations for bike stations through community workshops and meetings and then shared a draft of suggested locations with the users in form of an interactive map. The users were asked to help NYC DOT to refine stations' locations by voting for or against locations suggested by others, offering new stations, starting new arguments, or attending existing discussions and submitting reasons in support or against other people's suggestions. In another experiment, Kansas Department of Transportation (KDOC) launched a program called K-TOC (Kansas Transportation Online Community) in 2009, which was one of the first efforts in creating an infrastructure-specific online community. KTOC aimed to provide a forum for people and policy makers to communicate directly on transportation related issues. Encouraging online discussions, respecting customers by showing them they are being heard, showcasing opinion dynamics under social interactions, direct information exchange between all levels of the transportation industry, and unfiltered access to audience inputs were mentioned among the fundamental benefits of this program (KTOC, 2010).

2 Attempts and Shortcomings

Many studies in the literature such as Chinowsky, Diekmann, and O'Brien (2010) and Levitt (2011) have emphasized the role of managing construction projects through management of social networks involved. Since introducing the social network model of construction by Chinowsky, Diekmann, and Galotti (2008), some researchers have focused on social network analysis (SNA) of construction projects (Di Marco, Taylor, & Alin, 2010, and Wambeke, Liu, & Hsiang, 2012 among others). These studies however, primarily focused on a network of internal stakeholders involved in construction projects (traditionally called '*actors network*'). Moreover, scope of these models was mainly project management, rather than knowledge management. Parallel to these studies, given the diversity and complexity of stakeholders, in infrastructure management, network decision making has been emphasized as a process to reflect interests of all stakeholders in the final solution (Bruijn & Heuvelhof, 2000). In this respect, using online communication channels through the web was suggested as a more direct way to involve participation from external stakeholders of projects. However, scope of many practices in this regard (such as Lorenz, 2011) has been limited to channels run and owned by formal decision makers. Popularity of Social Web has recently drawn the attention of researchers in domain of civil infrastructure to online social media.

Evans-Cowley (2012) refer to public engagement through online social media (and particularly micro-blogging) as 'Micro-participation'. Studies on micro-participation have been mostly centered on the content and sentiment of public inputs and have more or less ignored the social network formed in the background of discussions. Although the domain has generally agreed on micro-participation as a demand for PI, several barriers challenge its efficiency. The following paragraphs briefly discuss these challenges.

First and foremost, involving the wisdom of prosumers is normally associated with some levels of risks. When several parties, with diverse (and in many cases, conflicting) interests are involved in the problem solving, the information provided by different parties takes the form of *negotiated knowledge* (Bruijn & Heuvelhof, 2000). Information covering different angles of the problem and representing different decision makers' perspectives may contradict each other, although all being true and valid. Such an issue is more severe when dealing with problems having a soft and subjective nature. Dichotomy of online and offline identities and attitudes is another challenge in involving online communities. This may result in incorrect or even fake opinion expressions among other issues. Usually forums where opinions are posted publicly suffer an obvious polarization of opinions, indicating that more moderate ideas are either not expressed or are dimmed under more extremist arguments. Moreover, involving public communities will change the nature of decision making from a structured project into a chaotic process (Bruijn & Heuvelhof, 2000). For example, in the case of NYC bike share, in period between September 2011 and April 2012, more than 10,000 locations were suggested and more than 60,000 supporting votes on the suggested locations were collected (NYC Bike Share, 2012).

On the other hand, the rapid growth of online social media and social networks (in form of websites such as Facebook and Twitter) has enabled citizens to express their opinions on various topics in full transparency, using a variety of devices. This helps governments to become independent from their proprietary social communication channels. It also creates a unique opportunity for a pro-active engagement system. As a result, in the beginning of 2012, Only 3 years after launching KTOC, KDOT suspended this website and transferred all activities hosted there to its headquarter Facebook page and Twitter account. In the announcement of this decision, KDOT's social media manager stated:

"*Given the rapid public embrace of the agency's Facebook presence, the fact that Facebook hosts an online audience far larger than any that can be reached by a proprietary online community, and that Facebook is free, it is no longer possible to justify the annual expense[s] associated with the K-TOC software lease*." (Quinn, 2012)

In North America alone, 82 out of the 100 strategic infrastructure projects announced by *North America strategic infrastructure leadership* in 2011, have now active Facebook or Twitter accounts to post news, host public discussions, and collect the feedback. Based on TRB (Transportation Research Board) transit cooperative research program—synthesis 99, major transportation providers who use micro-participation to involve the public in USA and Canada find Twitter and Facebook in many aspects the most convenient communication tools (Bregman & Watkins, 2013).

As an example, the network of Twitter followers of a Light Rail Transit (LRT) project in Toronto is shown in Fig. 1. This network was collected in May 2014, when the project was in the early construction phase. The network has 2,078 nodes and 46,852 edges in total, where node is a twitter ID and an edge connecting node

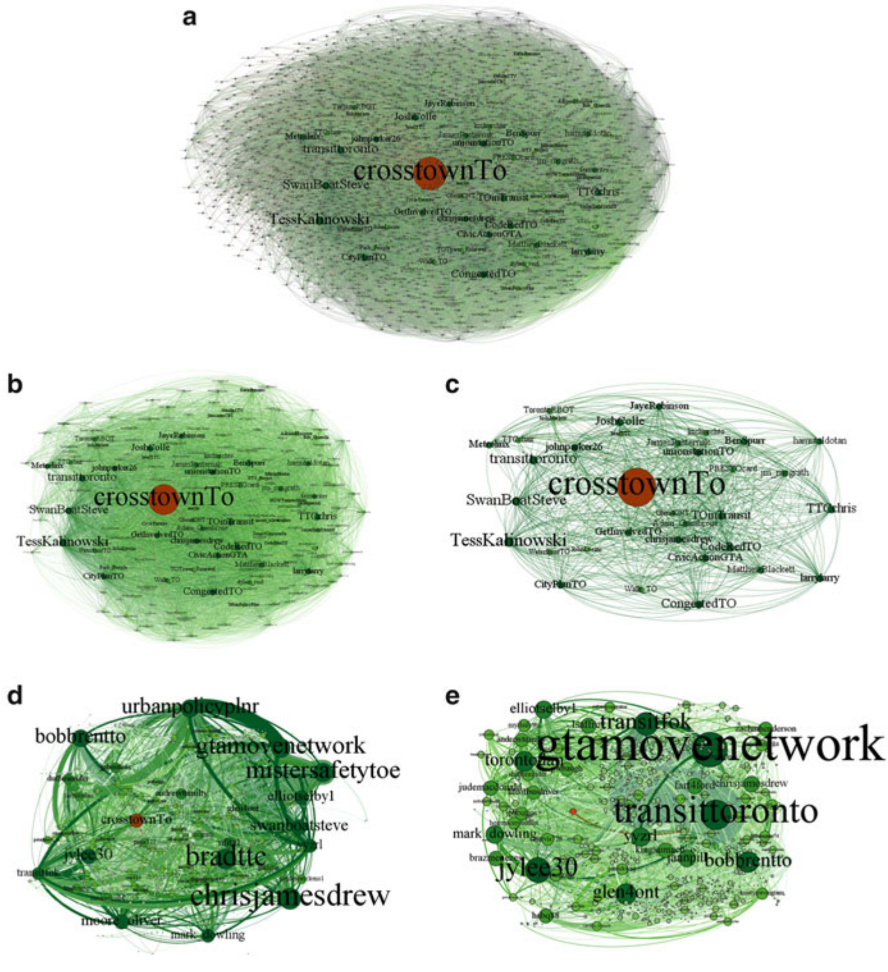


Fig. 1 Network of followers of Eglinton Crosstown LRT project on Twitter (collected on May 2014). (**a**) Complete network of project followers on Twitter. (**b**) Filtering out nodes with under 50 followers in the network. (**c**) Filtering out nodes with under 300 followers in the network. (**d**) Sub-network of mentioning. (**e**) Sub-network of Re-tweeting

A to node B implies that A is following B on Twitter. In order to detect the order beneath the chaos of such a network, we have filtered nodes with in-degree centralities lower than certain thresholds (to remove nodes with low number of followers). If a threshold of 50 followers is used, only 8.5 % of nodes and 17.1 % of edges will remain in the network (Fig. 1b). When we raise the threshold up to 300, as shown in Fig. 1c, only 1.5 % of nodes and 1.3 % of edges (34 nodes and 614 edges) will be kept. Although the lower-sized core can be interpreted easier, it is not yet a meaningful input to the decision making procedure.

Taking advantage of listening to what citizens say about their built environment on online social media can provide planners with an opportunity to engage the public in a very different way (Evans-Cowley & Griffin, 2011). However, harvesting relevant items from the corpus of user generated content on the social media and analyzing them to achieve meaningful results require tools and methods which do not formally exist in the field at this time. As the survey by TRB synthesis 99 indicates, public relation agencies in North America use online social media with goals such as ‘communicating with current customers’, ‘improving customer satisfaction’, and ‘improving agency image’. Online social media plays the role of a communication channel to connect with the customers/community for a real-time communication, advocacy, and feedback collection. Moreover, it can put the customers in power by creating opportunities for user innovation (Bregman & Watkins, 2013).

Evans-Cowley and Griffin focused on project-related ideas discussed over social media and analyzed public discussions from perspectives of content (type and theme) and sentiment. By starting a program called SNAPPatx (Social Networking and Planning Project), they investigated if micro-participation can be analyzed to help understand the public’s views on transportation issues. They used linguistic analysis and word count to assess emotional cognitive and structural components of more than 8,300 relevant tweets collected around transportation related issues in Austin, Texas. Results of this study approved that aggregation of microblogs can create meaning and help to understand perspective of the public community. However, from the aspect of providing decision makers with meaningful inputs, results of this study could not satisfy expectations. Official decision makers expected real-time data analysis with more meaningful results. More importantly, the public officials were generally interested in the identity of the users on top of aggregation of their opinion (Evans-Cowley & Griffin, 2011). The research was using off-the-shelf software for linguistic analysis; while as the researchers admitted, context-specific tools are required for this purpose. The SNAPPatx report ends with an emphasis on the demand for further empirical investigations to find ways in which information extracted through microblogging can be processed and weighted. It also insists on the need for developing a model to use micro-participation in planning effective engagement practices.

Therefore, most of studies in infrastructure PI have tried to evaluate online social media as a communication channel. They originally focus on how to build an online relationship with users to engage them in a dialogue, and more or less ignore the social network formed in the background of micro-participation. The conversational nature of social media however, has a ‘networked’ structure in which expressed ideas are linked to the followers of a project, including end users and public officials. These followers are nodes of the same layer of a network. El-Diraby (2011) refers to the network of people and network of ideas for AEC projects and their interdependency. Analyzing anonymous comments would focus on the ideas and ignore the people. As Evans-Cowley and Griffin (2011) mention, it is only a little more than "a finger in the wind"! Web 2.0 plays the role of a platform which brings people together and connects them to each other in the form of a heterogeneous network. It also documents/showcases their ideas and keeps the flow of project-related discussions among them. Networks which are formed around infrastructure projects are called *Infrastructure Discussion Networks* or IDN for short (Nik Bakht & El-diraby, 2013a, 2013b). The common needs and shortcomings in studying such networks can be summarized as follows:

- *Understanding the contents of* '*chattering*': While there exist many online IDNs, not much work has been done in applying formal socio-semantic analyses to understand the contents of the discussions and the relative importance and connections between ideas exchanged. The few attempts in this regard have used off-the-shelf software tools and were limited to finding keywords and sentiment of discussions.
- *Profiling the Stakeholders*: with the increasing role of society in decision making, we need to create means to profile community members. Understanding citizen attitude and what impacts it is very important to customizing both the project and its communication policy to their needs. No systematic process or protocol has been offered to date for analyzing IDN layout (connections between members) to help in profiling and/or analyzing stakeholders.
- *Distilling crowd knowledge for long-term us*: Context-sensitive solutions for urban infrastructure require local knowledge which is not exclusively owned by internal stakeholders (aka decision makers) anymore. Such knowledge is distributed among users (or future users) of the service and the social organization who interacts (or will interact) with the system. One prerequisite to such a heterogeneous and networked decision making scenario is distilling micro-knowledge of prosumers and aggregating it in a bottom-up fashion. It is important, then, to analyze the links between ideas, stakeholder profiles, project features to extract the basic rules/facts that makes a project more acceptable to the "new" decision makers. Through tracking IDN evolution over time and comparing different IDNs we can create a more suitable and customizable models of what makes a project a successful one. In other words, we should feedback the results of IDN analysis and comparison to reshape our ontologies of infrastructure projects.

In the light of these observations, we argue that effective analysis of the IDN would be the intersection of semantic and social analyses; while the former determines *what* has been uttered, the latter reveals *who* the utterer is. Text mining and natural language processing can handle semantics of discussions over the IDN (similar to Evans-Cowley, 2012, Lorenz, 2011, and Nik Bakht & El-Diraby, 2013b), and tools from SNA can help to uncover the composition of the social network which drives the discussions (similar to Nik Bakht & El-Diraby, 2013b).

In order for micro-participation to be integrated into a comprehensive engagement plan, infrastructure industry needs domain-specific tools to deal with IDNs from the two aspects mentioned above. This from one side requires benchmarking best practices from other domains and evaluating their applicability to the domain of infrastructure, and from the other side involves understanding infrastructure-specific content and trends of community inputs. The lack of tools to analyze seemingly chaotic public discussions and dialogues results in wasting opportunities for collecting prosumers' knowledge and user innovations. This is becoming frustrating to both communities and official decision makers (Evans-Cowley, 2012). Off-the-shelf tools cannot be the perfect answer to this need, unless being validated and customized through adequate empirical analysis and investigation in nature of existing IDNs. Social network analytics must be combined with text mining to classify followers of the project based on their affiliations and interests (stakeholders' typology), rank them based on their influence level (stakeholders' power level), and detect subject and sentiment of their discussions (stakeholders' vested interests and position). Outputs of such a detailed analysis may then provide decision makers with insights regarding stakeholders, which can help the PI program target the right people to be engaged and the right interests to be addressed by the plan.

3 From Big Data to Knowledge

As mentioned, IDNs are networks of 'people' and 'ideas'. Formalizing analysis of online discussions in the process of micro-participation requires evaluating the two aspects and integrating them to create meaningful inputs for decision makers.

Managing distributed knowledge of prosumers through Social Web requires adhering to the social business model. Such a business model is founded based on helping customers to collaborate with stakeholders by sharing and organizing information via Web2.0 technologies. This is a deviation in traditional PI mindset; from a *re-active* approach which *pushes* the public to generate inputs, into a *pro-active* method which *pulls* information required for decision making from users' discussions.[1] An example in using collective intelligence of crowds to detect small magnitude earthquakes can help to highlight the difference between the two approaches. "Did You Feel It?" was the name of a project using tools of Web 2.0 in a re-active fashion to crowdsource detection of small magnitude earthquakes. In this project, participants were asked to fill in online surveys about their experiences during an earthquake (Atkinson & Wald, 2007). Crooks, Croitoru, Stefanidis, and Radzikowski (2013) tried a pro-active alternative for reaching the same goal.

[1] Traditional methods in PI mostly rely on pushing end-users via survey questionnaires and other tools to contribute inputs to the decision making procedure, as well as trying to educate them about the project related decisions and then collect their feedback during public hearings and community meetings. In all these methods, contributors "react" to the initiators within a defined scope and specific framework.

They detected responses to the earthquake over Twitter by following the hashtag *#earthquake* right after the event of US East Coast earthquake in August 2011. They used social media feeds as a sensor system to detect and locate the 5.8 magnitude earthquake, without pushing users to generate any inputs.

In the pro-active school of thought, prosumers routinely express their experiences and opinions in full transparency and in a free format through multiple channels available by online social media. On top of administering these channels, the role of PI officials will be to distill the knowledge through detecting relevant comments, classifying, and evaluating them. As emphasized earlier, penetration of agencies such as Twitter and Facebook leaves no room for official decision makers' propriety social networks to involve the community in a truly active manner. Tools and techniques are required to monitor the general social media and automatically collect the relevant data. The relevance would refer to the semantics of discussions and social value of the supporters. Such data must be then processed into the meaningful knowledge. In more specific words, the relevant content generated by the public over Social Web must be processed to answer the following questions (which are essential to the stakeholder analysis in many domains including civil infrastructure):

- *Who are the project followers*? Answering this question is essential to understanding the typology of stakeholders;
- *What is the relative influence of stakeholders*? Finding the answer to this question can provide decision makers with insights regarding community leaders and level of impact that each project follower can have on others. This can be known as 'network value' of project followers and can be used to evaluate the impact level of ideas discussed;
- *Which topics are being discussed*? Detecting, understanding, and clustering topics discussed over IDNs can help decision makers with identifying needs, vested interests, feedback, and user innovations;
- *What is the sentiment of discussions*? Stakeholders' position in terms of being proponent or opponent to the project and/or decisions is normally reflected in the sentiment of their discussions.

It is noted that answering these questions must be an ongoing process during different phases of project lifecycle. Dynamics of answers to these questions can help to explore patterns of order existing beneath the chaos of public participation.

Analysis of microblogs (mainly, Twitter networks) centered on infrastructure projects can be vital in answering the questions above. On one hand, many infrastructure projects in North America have active Twitter accounts, and on the other hand, the open API (Application Programming Interface) of Twitter provides a great opportunity for researchers as well as field practitioners to pull data from this microblogging website. Moreover, Twitter not only archives social opinion in form of short statements, but also keeps the record of connectivity among individuals in the form of following, mentioning, and re-tweeting. Therefore, the 'networked-ness' of ideas can be tracked by studying Twitter. In the following, we address some of the results from studying IDNs formed in Twitter. Most of these analyses can be directly repeated for similar websites such as Facebook, online forums, and blogs.

3.1 IDN as Network of People

Studies show that IDN as a collection of social entities connected through social relational ties more or less shares the general behavior of general social networks. As an example, in the network of followers of a project over Twitter, different levels of social connections can be defined among the nodes. They range from weak ties such as “following” to stronger ties such as “mentioning” or “direct messaging”. Figure 1c and d respectively show the network of mentioning and re-tweeting among followers of Eglinton Crosstown LRT project in the period of May and June of 2014. This is another way to filter the chaos of IDN and reach into the core of a network. The results in these cases are weighted directed graphs in which node A is connected to B with an edge of weight n if A has mentioned B n times/has re-tweeted n tweets by B. Although as Huberman, Romero, and Fang (2008) suggest, such networks reflect stronger ties and can reveal more meaningful interactions among the nodes; as indicated by Easley and Kleinberg (2010) among many other authors, strength of weak ties cannot be ignored. Focusing on any of these types of connections uncovers a social network at a different level. The network built in this way portrays a layout of social connectivity among followers who have vested interest in the project. Studying IDN as a graph of social identities connected through social linkages reveals topographical and geometrical properties which have roots in formation and evolution of such networks.

It is shown that the geometry of IDN depends on the project’s nature and behavior of the community. Like many other social networks, IDNs follow power law degree distribution. This is a result of ‘popularity’ mechanism among project followers; while there are a few popular nodes with a high number of followers, majority have a low number of followers. As it is shown by Nik Bakht and El-diraby (2014), this behavior becomes more dominant (the rich get richer) as the project proceeds in its lifecycle. Networks with this behavior are called ‘scale-free’; this is not only due to the mathematical scale-free nature of the power function, but also to emphasize that networks related to fundamentally different systems exhibit the same characteristics. Therefore, models and algorithms from other domains, including reverse marketing and the social business model can be benchmarked and directly applied in infrastructure planning to improve the efficiency of the public consultation process.

Moreover, it is shown that IDNs are local and issue-centered ‘small worlds’ with a relatively high clustering, comparatively small diameters, and short average path lengths. Small world phenomenon, which refers to the rich nature of social networks in terms of short paths, facilitates information diffusion and provides a good opportunity for viral marketing around the project. On the other hand, measures and algorithms to evaluate the influence level of nodes can assist to find the high influential nodes among project followers and involve them in the process of consultation with the public. As projects proceed in their life cycle, their IDNs mature and demonstrate a behavior which is closer to more established online social networks. IDN of projects in later phases of their lifecycle engage more followers and grow in size. Moreover, their growth is involved in *triadic closure*; i.e., more connections are formed among friends of friends. The fact that evolution of IDN does not stop as the project progresses into later phases of its lifecycle can be a significant opportunity to extend the online public engagement into the whole project lifecycle as a continuous and self-organizing process.

As suggested in (Nik Bakht & El-diraby, 2014), Topological parameters of the IDN can provide decision makers with indicators of maturity and performance measures for the public outreach programs. Moreover, project followers in the IDN group together and form limited numbers of communities based on different forms of similarities (Nik Bakht & El-diraby, 2013a, 2013b). Project funding sources, various levels of decision makers, geographic similarity and project impacts are among the main criteria which segment the followers into communities. Analysis of influence patterns can detect the leaders of those communities. Detecting communities and their leaders can help the PI practitioners with the team-building process (by engaging community leaders), and classifying cores of public interests. This is particularly value-adding when hidden nodes from the public community are detected. For example, finding influential figures (such as a prominent journalist or an urban activist), who influence public's mind the most is normally not an easy task if ever possible, in traditional PI practices.

Therefore, answering the first two questions in the set of four questions above requires detecting communities in the social graph of the IDN and uncovering patterns of influence among its nodes. Figure 2 illustrates some applications of topology analysis for IDNs.

Detecting patterns of influence—Real world dependencies among people in many cases are reflected in their online social relations. As mentioned, social connections over the web have various types and levels: from loose concurrencies such as subscription to the same group, to more direct relations like 'following' and

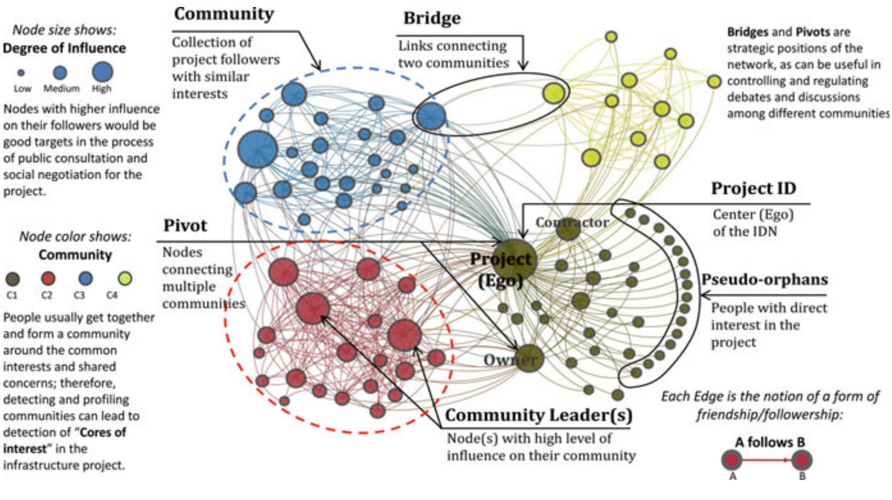


Fig. 2 Analysis of IDN as network of 'people'; social network analytics

'friendship', and stronger ties such as direct messaging, re-tweeting, and mentioning. Although none of these can be a perfect representative of the real/offline relations, some of them can be adequately taken into account as indicators of the social influence. For example, from a 'following' relationship on Twitter or Facebook, it can be inferred that the followee may have a level of influence on the follower (Huberman, Romero, & Wu, 2009) (at least all of the followee's posts can be seen by the follower). After taking such an assumption, detecting top influential nodes of the IDN can help the PI processes to find the right people from the public to be involved. There are several measures and methods for finding influential nodes in a network. Different types of centrality (including degree, betweenness, and eigenvector) are among classical tools for this purpose. In particular, researchers in the domain of construction have widely used these metrics to analyze project networks. In a previous study, we tested the methods normally used in ranking webpages (such as Hyperlink-Induced Topic Search: HITS, and PageRank) for this purpose (Nik Bakht & El-diraby, 2013a). Such methods typically consider not only the number of followers (quantity) for an individual, but also their level of importance (quality). It was shown that given the size and level of complexity of the IDN, such methods provide more precise measures to analyze influence patterns in IDNs. Having a full list of top influential nodes, the team-building processes can target those who represent different vested interest in the project and at the same time have higher levels of influence on other followers of the project.

Studying communities of interest—Social networks are typically composed of clusters of nodes called communities. Nodes within each community are densely connected to each other, and are sparsely connected to nodes from other clusters. Existence of communities has roots in the social behavior of community members. During formation and evolution of a social network, people are interested in joining groups in which not only they have more friends, but also their friends are more closely connected to each other. Therefore, communities of a social network typically form around commonalities and shared interests. Detection of communities, therefore, not only can classify those who have interest in the project, but also can classify vested interests with respect to the project. On top of community leaders, nodes with strategic and inter-disciplinary positions can be great sources of feedback and/or innovation. These nodes can also assist in regulating cross-community relations. Pivot nodes which are at the intersection of multiple communities, and bridges that connect two communities to each other are examples of such nodes (Fig. 3).

3.2 IDN as Network of Ideas

As shown above, an IDN can be modeled and studied as a graph of social identities, connected to each other through direct social linkages. On the other hand, connectivity among people in an IDN can be defined through the ideas they support (or oppose!). Monitoring ideas discussed and analyzing them can also help add

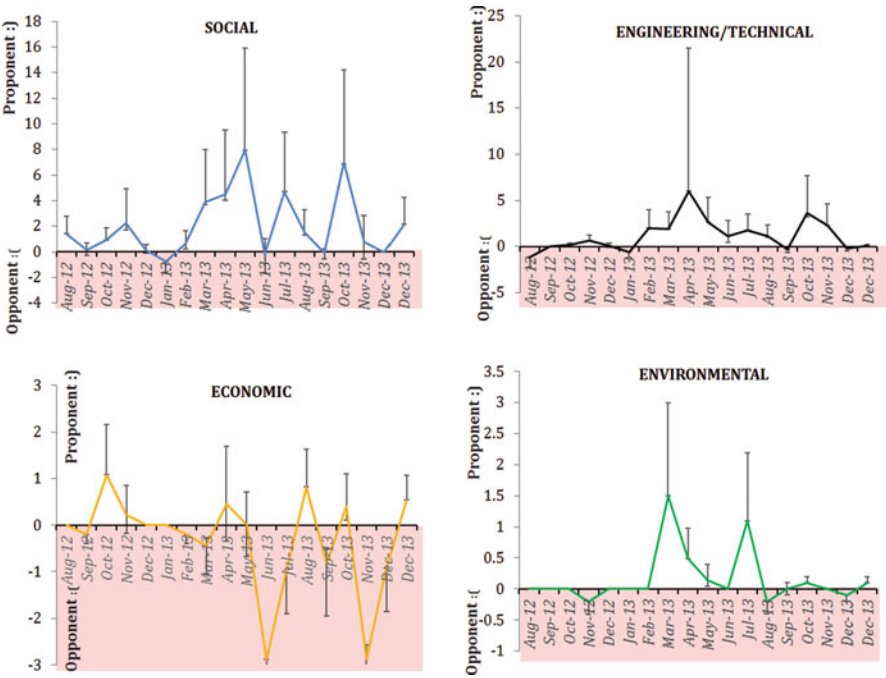


Fig. 3 Eglinton Crosstown LRT project discussion profiles in the four dimensions of sustainability

meaning to network clusters, detect vested interests, and monitor opinion dynamics among projects' followers. Computational Linguistics (CL) can support these aims. In order to 'understand' the content of discussions and to follow their dynamics, domain-specific tools may be required. As mentioned before, complete analysis of the big data over IDN will be the intersection of SNA and CL; the latter can help to interpret results of the former, and also to answer the last two out of the four questions we introduced earlier.

Profiling people and labelling communities—As it was addressed in the previous part, detecting communities of followers in the graph of IDN can provide decision makers with a good insight for team-building and problematization process. To make sure the PI program has involved the right people and to guarantee that all different groups of project followers have their voices in process of decision making; different groups must not only be detected (through community detection), but also they should be 'labeled' based on their backgrounds and roles with respect to the project. The labeling requires a better understanding on the typology of followers. Users normally describe themselves in their online profiles, referred to as 'bio' or 'description'. Text mining can help to classify users' descriptions in each community of the IDN. In our previous study, we used an information retrieval measure called Term Frequency-Inverse Document Frequency (TF-IDF) which is usually used in topic detection to achieve this goal (Nik Bakht & El-Diraby, 2013b). This method scores up terms with high frequency in the set of descriptions for one community, and at the same time scores down terms which are common across multiple communities. By detecting discriminating terms for each community, this method crystallizes the sets of buzzwords for descriptions of each community's members. Semantic clustering of these terms can then automatically label each community and provide decision makers with a layout of the social composition around the project.

Detection of core interests through profiling communities of interests—If labeling communities is clustering 'people', detecting core interests can be known as clustering 'ideas' discussed. As mentioned, shared interests and common values bring people with different backgrounds together in form of a community. Detecting cores of interest among project followers is an important insight that analysis of IDN can provide decision makers with. Any solution offered by decision makers must try to address as many core interests as possible. Shared interests can be detected either by semantic clustering of ideas discussed as suggested by Steinhaeuser and Chawla (2008), or in a reverse format by labeling common opinions expressed within each community, similar to Nik Bakht and El-diraby (2013a, 2013b). In order to achieve this goal, text mining must target discussions supported by members of each community. The result can also detect the topics discussed within and across various communities and portray the '*social dialogue*' around the project (Nik Bakht & El-Diraby 2013b).

Detection and evaluation of opinions—As mentioned, dynamics in structural properties of the IDN can be an important indicator for monitoring the public outreach. Dynamics of the IDN however, is not limited to the followers' interconnectivity, but also at a more sophisticated level, it includes dynamics of opinions they support and express. By defining 'opinion' as a combination of the 'subject' discussed, and 'sentiment' of the discussion, each discussion over social media can be modeled as an instance of opinion expression. Detection of the opinion will consequently require automatic classification of the subject, and the sentiment of the discussion. Such classifiers can be trained using NLP and machine learning techniques among other methods. Detected opinions for one project can then be aggregated and analyzed statistically over the time, to identify patterns of dynamics in the social opinion. Monitoring formation and evolution of opinions enables decision makers to detect bottlenecks in communication with the public. It also allows detection of social alarms from analysis of the opinion dynamics. Since offline social opposition most of the time lags the online declaration of dissatisfaction, detecting online alarms will give the official decision makers enough time to change the decisions appropriately, or to apply timely policies to prevent formation of snowballing social opposition and to reduce the risk of failure in such projects before it is too late.

4 Project Discussion Profile

By selecting a particular context for analysis of discussions over IDN, results of SNA and lexical analysis can be aggregated to form the profile of online discussions for a particular project. This not only classifies major vested interests and highlights followers' positions (in terms of being proponent or opponent) over the time, but also synthesizes them according to the network value of the followers. The context is scoped as a semantic space with a certain number of semantic clusters as dimensions of the space. In this space, each data point (e.g., each tweet in the case of Twitter, or each comment in case of Facebook or forums) is represented as a vector. Every entry of such a vector corresponds to one semantic cluster (topic) and takes numerical values only if that topic is covered by the discussion. Sign of entries highlight sentiment of the discussion; for this purpose, either a binary method (similar to Sousa, 2005: opponent -1, proponent +1, or neutral 0), or a fuzzy system (similar to Olander, 2007 to highlight intensity of support or disapproval) can be used. Signed binary vectors resulted in this way represent opinions as subject-sentiment dyads.

In order to model the network value of discussions, opinion vectors must be connected to the influence level of people who express them. This can be numerically represented as the product of the binary vector of opinion and impact factor of the utterer. The latter can be evaluated as a normalized factor from analysis of influence in IDN through measures such as centrality or PageRank. Weighted vectors in each space can then be combined and the resultant will aggregate all data points for a project in form of the *project discussion profile*.

As an example at the project level, Fig. 3 shows the discussion profile for Eglinton Crosstown LRT project in within a time-span of 18 months. Sustainability was selected as the context of analysis and its components (environmental, social, economic, and engineering) were forming dimensions of the semantic space. As it is seen, public satisfaction in online media reaches its minimum in December of 2012; this is when decision makers revised the environmental assessment study and removed two major stops from the plan. However, after public consultations and listening to community's feedback, the plan was modified and the stops were returned back in May 2013. At this point, the social dimension hits its maximum level. Also, as it is seen, while in many cases engineering, environmental, and social sustainability have similar trends (with different amplitudes), economic-related opinions move in an opposite direction in most of the cases. By the beginning of construction, in summer 2013, negative sentiment with respect to the economic aspect is at its maximum. Figure 3 also gives a range for opponent and proponent discussions at each snapshot.

Project discussion profiles can be known as a formal distillate of online media discussions. This can be used as a measure to evaluate outcomes of different communication strategies and control the public consultation procedure among other applications.

5 Concluding Remarks

Trends of change in profile of decision makers for construction and development of the urban infrastructure improved them from individual decision makers to a hierarchy of technical decision makers, and recently to a network of technical and official decision makers. Such an evolution, along with modern trends of prosumerism and knowledge epidemiology over Web2.0 is now calling for another shift to a heterogeneous network of official decision makers and public decision contributors. Public involvement programs should be upgraded to pave the road for such a shift. Social web can help decision makers of a project to engage important sectors of the community and develop bidirectional communication strategies to not only educate the public regarding the project and related decisions, but also build trust and promote a culture of collaboration through formation of IDNs. At the same time, similar to many other domains, an IDN can be a great source of prosumers' knowledge and innovative ideas. In fact, the advantages of using Web 2.0 in community engagement go beyond improving the quality of communication with local communities; it can build the foundation for reaching solutions which are more robust and more innovative. In addition to harnessing innovative ideas, the core interests extracted from IDN discussions can be embedded in the ultimate solution offered by the decision making process. This can create a true sense of ownership of an urban project by local communities, which is a key factor to the project success from different aspects including stakeholder management and social sustainability.

Some of core contributions of IDNs to the public consultation process were addressed in this chapter. On one hand, structure of the IDN as a social network can provide decision makers with some insight about the performance of their public engagement practices. This structure also includes important information regarding influential nodes among the internal and external stakeholders. On the other hand, semantic analysis of the user-generated content can lead decision makers towards a better understanding of vested interests and social concerns/values. In general, IDNs can help to direct the decision making towards a more pluralistic process, rather than a pre-planned project.

In order to take advantage of IDNs, more research is required in at least two major streams: logistics of the IDN, and behavioral issues. From the logistics point of view, developing context-aware tools and context sensitive mechanisms is necessary to specifically support the realm of infrastructure. This ranges from creating topic/sentiment classifiers for infrastructure-specific discussions, to defining performance measures for assessment of public involvement in a project through its IDN. The latter can facilitate monitoring of bottlenecks in communication with the public, and detection of public dissatisfaction alarms with respect to a project or its certain aspects. At a behavioral level, different phases in the lifecycle of projects and their IDNs must be studied to develop models for predicting behavior of IDNs in different conditions. Such models will be helpful in creating online (and maybe offline) communication strategies with the public in infrastructure projects.

There are issues and barriers which may postpone the applicability of IDNs in practice. The difference between offline and online attitudes of many users as well as issues such as multiple identities, security and trust are among other examples of this type. Collaboration between practitioners and researchers will be required to address such issues and solve them in the future.

References

Atkinson, G. M., & Wald, D. J. (2007). Did you feel it? Intensity data: A surprisingly good measure of earthquake ground motion. *Seismological Research Letters, 78*, 362–368.

Brabham, D. C., & Sanchez, T. W. (2010). Crowdsourcing public participation in transit planning: preliminary results from The Next Stop design case. *TRB 89th annual meeting*. Washington, DC: Transportation Research Board of the National Academies.

Bregman, S., & Watkins, K. (2013). *Best practices for transportation agency use of social media*. Boca Raton, FL: CRC Press.

Bruijn, H. D., & Heuvelhof, E. T. (2000). *Networks and decision making*. Utrecht: Lemma Publishers.

Chinowsky, P., Diekmann, J., & Galotti, V. (2008). Social network model of construction. *ASCE Journal of Construction Engineering & Management, 134*, 804–812.

Chinowsky, P., Diekmann, J., & O'Brien, J. (2010). Project organizations as social networks. *Journal of Construction Engineering and Management, 136*(SPECIAL ISSUE: Governance and leadership challenges of global construction), 452–458.

Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). #Earthquake: Twitter as a distributed sensor system. *Transactions in GIS, 17*, 124–147.

Di Marco, M. K., Taylor, J. E., & Alin, P. (2010). Emergence and role of cultural boundary spanners in global engineering project networks. *Journal of Management in engineering, 26*(3), 123–132.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. New York: Cambridge University Press.

El-Diraby, T. E. (2011). Civil infrastructure as a chaotic socio-technical system: How can information systems support collaborative innovation. *CIBW078*. Computer Knowledge Building.

Evans-Cowley, J. S. (2012). Griffin G. *Microparticipation with social media for community engagement in trasportation planning., 2307*, 90–98.

Evans-Cowley, J., & Griffin, G. (2011, Feb 12). *Micro-participation: The role of microblogging in planning*. Retrieved Feb 27, 2014, from SSRN: http://ssrn.com/abstract=1760522 or http://dx.doi.org/10.2139/ssrn.1760522

Future by airbus. (2012). Retrieved February 20, 2014, from Airbus.com: http://www.airbus.com/innovation/future-by-airbus/.

Hansen, J. S., & Jackson, M. C. (2001). St. Louis redefines community engagement. *Land Development and Public Involvement in Transportation, 1780*, 140–144.

Huberman, B. A., Romero, D. M., & Fang, W. (2008). Social networks that matter: Twitter under the microscope. *First Monday, 14*(1), 1–9.

Huberman, B. A., Romero, D. M., & Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday, 14*(1), 8.

Keast, R., & Hampson, K. (2007). Building construction innovation networks: Role of relationship management. *ASCE Journal of Construction Engineering & Management, 133*, 364–373.

KTOC. (2010). Retrieved February 26, 2014, from FHWA Transportation library connectivity: http://libraryconnectivity.org/archive/downloads/KTOC-Poster.pdf.

Levitt, R. E. (2007). CEM research for the next 50 years: Maximizing economic, environmental, and societal value of the built environment. *ASCE Journal of Construction Engineering and Management, 133*, 619–628.
Levitt, R. E. (2011). Towards project management 2.0. In *Engineering project organizations conference*. Denver, CO: EPOC.
Lorenz, J. (2011). Building support for new transportation funding and financing program. *Transportation Research Records, 2245*, 8–16.
Lukszo, Z. V., & Bouwmas, I. (2005). Intelligent complexity in networked infrastructure. *IEEE international conference on Systems, Man and Cybernetics,* (pp. 2378–2383). Waikoloa Village, HI: IEEE.
Nik Bakht, M., & El-diraby, T. (2014). Infrastructure discussion networks: Analyzing social media debates of LRT projects in North American cities. *TRB 3rd annual meeting compendium of papers*. Washington, DC: TRB.
Nik Bakht, M., & El-diraby, T. E. (2013a). Analyzing infrastructure discussion networks: order of 'influence' in chaos of 'followers'. *Csce annual conference-4th construction specialty conference* . Montreal: CSCE.
Nik Bakht, M., & El-Diraby, T. E. (2013b). What does social media say about the infrastructure construction project? Beijing, China: CIB W78.
NYC Bike Share. (2012). Retrieved February 26, 2014, from NYCDOT: http://a841-tfpweb.nyc.gov/bikeshare/suggestion-archive/.
Olander, S. (2007). Stakeholder impact analysis in construction project management. *Construction Management and Exonomicsa, 25*, 277–287.
Quinn, P. (2012, Jan 06). *Kansas department of transportation—Northeast Kansas*. Retrieved February 27, 2014, from Facebook: https://www.facebook.com/NEKansasKDOT/posts/146758275436663.
Sanvido, V. E., & Paulson, B. C. (1992). Site-level construction information system. *ASCE Journal of Construction Engineering and Management, 118*, 701–715.
Slaughter, S. E. (1993). Builders as sources of construction innovation. *ASCE Journal of Construction Engineering and Management, 119*, 532–549.
Sousa, A. O. (2005). Consensus formation on a triad scale-free network. *Physica A, 348*, 701–710.
Steinhaeuser, K., & Chawla, N. V. (2008). Community detection in a large real-world social network. In *International conference on social computing, behavioral modeling and prediction* (pp. 168–175). Phoenix, AZ: Springer.
Von Hippel, E. (2005). *Democratizing innovation*. Cambridge, MA: MIT Press.
Wagner, J. (2013). Measuring the performance of public engagement in transportation planning: Three best principles. *Journal of the Transportation Research Board, 2397*, 38–44.
Wambeke, B. W., Liu, M., & Hsiang, S. M. (2012). Using Pajek and centrality analysis to identify a social network of construction trades. *ASCE Journal of Construction Engineering and Management, 138*(10), 1192–1201.

Biometric-Based User Authentication and Activity Level Detection in a Collaborative Environment

Faisal Ahmed and Marina Gavrilova

1 Introduction

Collaborative and social activities, interactions, and communications are great sources of behavioral biometric information, which has potential applications in user identification in social media (Sultana, Paul, & Gavrilova, 2014a). Recently introduced social behavioral biometric features (Sultana et al., 2014a; Sultana, Paul, & Gavrilova, 2014b) utilize data obtained from different online and offline sources, such as Twitter and Facebook, blogs and discussion forums, online games, face-to-face and virtual meetings, online collaborative activities, etc. Analysis of these data sources can potentially reveal user habits, use of technology, ways users contribute to knowledge generation and collaborative creation of content (Sultana et al., 2014a, 2014b). Having these new types of behavior-mediated biometrics helps not only to identify users in an online environment, but also to quantify their contribution to shared knowledge generated online, such as Wikipedia. This, in turn, allows to study patterns of combined behavior, prevent access to unauthorized content based on individual user behavior, and in some cases, even to prevent unexpected (and potentially harmful) activities.

This approach to transparency and identification of online interactions can be extended to other, cognate realms. Social interaction through virtual spaces is one of them, Meetings, real or especially virtual, are important collaborative activities in any organization, which plays a key role in the dissemination of information and knowledge and thus, provides decision-support (Yu, Ozeki, Fujii, & Nakamura, 2007). However, almost 50 % of the total time spent in meetings is reported to be wasted (Mosvick & Nelson, 1987). The main reasons are loss and distortion of information, sub-optimal decision making, and mismanagement (Mosvick & Nelson, 1987).

F. Ahmed (✉) • M. Gavrilova
Department of Computer Science, University of Calgary, Calgary, AB, Canada
e-mail: faahmed@ucalgary.ca; mgavrilo@ucalgary.ca

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_9

Hence, with the increasing development of computing technologies, there has been a growing demand from industries for new generations of interactive technologies to support high productivity and to optimize time spent during meetings. In addition, secure real-time user authentication and access right management solutions are being actively researched and deployed in both universities and corporate settings (Deutschmann, Nordstrom, & Nilsson, 2013; Penteado & Marana, 2009; Zhou & Bhanu, 2007). In response to these demands, researchers are actively seeking new communication and collaboration technologies, which will reduce meeting attendees' enrollment time and provide the meeting organizer with the tools for efficient real-time attendance tracking and access resource management.

Traditional approaches to the problem involve the use of IDs, smart cards, and user password/logon, coupled with user identity management solutions, all of which require a significant amount of time and additional resources (Jeon, Kim, & Kim, 2011; Kumar, 2010). Rapid development of biometric technologies opened the door to a new class of fast and reliable identity management solutions and changed the research landscape (Down & Sands, 2004; Penteado & Marana, 2009; Yanushkevich, Gavrilova, Wang, & Srihari, 2007). A biometric system can be defined as a pattern-recognition system that can recognize individuals based on the characteristics of their physiology or behavior (Prabhakar, Pankanti, & Jain, 2003). Human biometric traits can roughly be divided into two categories: physiological and behavioral. Physiological biometric systems utilize certain physical characteristics, such as face, iris, ear, fingerprint, palm, etc. for individual recognition. On the other hand, behavioral biometric systems rely on human behavior-mediated activities, such as gait, voice, handwriting, signature, etc. The idea to utilize the state-of-the-art biometric recognition technologies for meeting room management has recently attracted much attention due the efficiency and unobtrusiveness of the technologies. This proposal is based on the premises that physiological and behavioral biometrics can be seamlessly integrated with technologies enabling meeting room setup, and, in addition to individual access management, can provide highly efficient group authentication capabilities. The objective is to enable a meeting organizer to immediately start the meeting, keep track of attendance, and allow all participants to seamlessly access shared resources without compromising secure contents. In addition, the developed system should support archiving meeting statistics, determining input of each of the participants, and recognizing certain activities. This, in turn, will help create a more conductive collaborative environment, use time more effectively, and inform project managers on successful patterns.

In this chapter, we present our research efforts toward developing an interactive and intelligent meeting room system based on multi-modal biometrics, focused not only on user authentication, but also on activity-detection and individual contribution tracking during collaborative activities. A meeting room represents a dynamic collaborative environment, where different individual and group interactions take place. Accurate identification and summarization of these activities and interactions require reliable acquisition, synthesis, analysis, and integration of multiple sources of data. In addition to that, development of intelligent virtual and online meeting room systems demands tools and methods for reliable understanding of the virtual interactions among the participants. This chapter contributes to the emerging domain of user identity, authentication, and security as well as behavior and collaborative interaction analysis via (1) combining behavioral (such as gait) and physiological (such as face, voice, etc.) biometric traits for reliable and efficient user authentication and identity management solution, (2) identifying potential individual and group activities to be tracked in order to analyze the workflow of a collaborative environment, and thus, (3) providing selected content access and enhancing security. The ideas presented here can also be extended to design intelligent online and virtual environments that can monitor and detect user activity and contribution level in an unobtrusive manner. In our discussion, we address the following questions, which are imperative for the realization of an intelligent collaborative environment:

- How can individual identification and authentication be performed in an effective manner? Which biometric traits should be considered for a meeting-room specific intelligent system?—In Sect. 3.2, we address these two questions and propose a multi-modal biometric system that utilizes human gait, face, and voice.
- How can the meeting room workflow be analyzed in an effective manner? How can the system identify the different activities among the participants?—We discuss these issues in Sect. 3.3, where we present a list of individual and group activities along with how to automatically identify these activities. We argue that, identifying these activities plays an important role in analyzing the meeting room workflow.
- How can individual contribution level be evaluated? Which features are required to be considered?—In Sect. 3.4, we propose a list of potential features, which can be used to evaluate individual contribution effectively.
- How to analyze group dynamics and behaviors? Which features are important to be considered for this analysis?—We argue that, analyzing group dynamics and behavior is potentially dependent on the recognition of group activities and workflows, which is presented in Sect. 3.3.

2 Related Work

In recent years, several works (Sultana et al., 2014a, 2014b) introduce the concepts of social media collaborative content analysis through user behavior and social network mining, spatio-temporal analytics, and text authorship recognition. We are particularly interested in analyzing the collaborative activities in a meeting room environment, which can potentially lead to a smart meeting room system. The functionalities of the traditional smart meeting room systems found in the literature are usually limited to only synthesizing and archiving important information presented at the meeting for future analysis (Yu & Nakamura, 2010). One of the earlier efforts toward building smart meeting rooms is the meeting browser system designed by Waibel, Bett, and Finke (1998). Their proposed architecture comprised four major components: (1) a speech transcription engine, (2) a statistics-based summarizer, (3) a speech-event detector, and (4) a non-verbal structure for detecting visual cues and speaker type. The objective was to facilitate human users to quickly review and search records of human interactions (Waibel et al., 1998). Chiu, Kapuskar, and Wilcox (1999) proposed an architecture based on computer controllable video cameras, video conference cameras, and ceiling microphones in order to capture human activities during meeting. Rui, Gupta, and Cadiz (2001) employed omni-directional camera technology in order to construct a meeting capture system that can cover the 360° view of a meeting room. They also addressed some user-interface and client–server performance issues in their work, such as amount of user involvement, camera control rules, identification of meeting context, and view of participants (Rui et al., 2001). A similar meeting recorder was designed by Lee, Erol, Graham, Hull, and Murata (2002). Their system included a user interface that can exploit metadata description in order to facilitate efficient browsing and identification of important events. Jain, Kim, and Li (2003) developed a system that can capture meeting data from multiple cameras and microphones and perform a semantic-based indexing of the captured data. Wellner, Flynn, and Guillemot (2004) proposed a similar method for browsing elements of interest in a recorded meeting. Their system can display both processed interval data and meeting transcripts as specified by the user (Wellner et al., 2004). However, all these systems focus only on finding important information from meeting recordings and do not facilitate any real-time authentication, access control, or individual contribution level tracking.

Realization of an intelligent meeting room system with real-time authentication, activity detection, and active speaker and contribution tracking capabilities requires a combination of image and video processing, audio-speech processing, and multimodal biometric technologies (Mikic, Huang, & Trivedi, 2000). One of the first attempts toward this direction of research was made by Mikic et al. (2000). Their work (Mikic et al., 2000) focused on activity monitoring and summarization in a group-meeting setup. However, they tracked only three types of activities. The tracked activities were: (1) a person standing in front of the whiteboard, (2) a presenter presenting in the meeting, and (3) a non-presenter speaking. Another approach proposed by Stanford, Garofolo, Galibert, Michel, and Larrun (2003) employed different meeting metadata derived from meeting room sensors in order to construct a meeting data corpus aware of the underlying context. Examples of the selected metadata include (1) spoken words, (2) speaker identity, (3) speaker locations, (4) time tags, and (5) sentence-like units.

Apart from constructing the whole intelligent meeting room system, researchers have also investigated specific aspects and components of an intelligent meeting environment. For example, Hornler and Rigoll (2009) modeled activity and dominance in a smart meeting room environment. Low level acoustic and visual features were fused using a Hidden Markov Model (HMM) in order to classify activity levels of meeting room participants. They selected Mel frequency cepstral coefficients (MFCC) as acoustic feature and global motions (GM) as visual features.

However, in their experiments, low-level fusion of these features resulted in a low recognition performance. Hence, they urged researchers to investigate other features such as eye movement, detection of slide change, etc. Nait-Charif and McKenna (2003) developed a head tracker-based activity detection system for smart meeting rooms. The system uses the head position information in order to detect certain activities, such as entering, exiting, going to the whiteboard, getting up, and sitting down (Nait-Charif & McKenna, 2003). However, the recognition ability of the proposed method is heavily dependent on the accurate detection of head position and scene-specific constraints (Nait-Charif & McKenna, 2003). Busso, Hernanz, Chu, Kwon, and Lee (2005) proposed a meeting participant identification method based on video and audio processing. Based on their experimental results, they concluded that, inclusion of complementary modalities can successfully increase the recognition performance. Stiefelhagen (2002) tracked focus of attention in a meeting room setup by using an omni-directional camera to simultaneously track the head poses of all the participants. A probabilistic framework is then used to map the head poses to their corresponding targets.

All these research efforts comprise different aspects of an intelligent meeting room environment. However, they have limited functionalities as stand-alone systems. Hence, a holistic architecture is required to make use of all these components in an effective manner. In this chapter, we borrow different concepts from these existing works and aggregate them with our own ideas in order to construct a complete multi-modal biometric-based collaborative environment.

3 Proposed Multi-Modal System

Our proposed multi-modal system comprises four major components: (1) multi-modal sensing, (2) individual identification and tracking, (3) activity recognition, and (4) individual contribution analysis. Figure 1 shows the architecture, components, and workflow of our proposed system. In this section, we present a detailed discussion on these components.

3.1 Multi-Modal Sensing

Real-world meetings encompass a variety of activities of different modalities, such as speech, presentation, gesture, etc. Hence, our proposed system employs multi-modal sensing in order to collect accurate and complete data under varying and dynamic conditions. We propose to use Microsoft Kinect v2 in order to collect both the video and audio data in a meeting room environment. Microsoft Kinect was originally introduced as an add-on device for the Xbox 360 gaming system, which can detect the physical movement or voice commands of the user and thus enables the user to play games without any physical controller. Kinect is made up of an

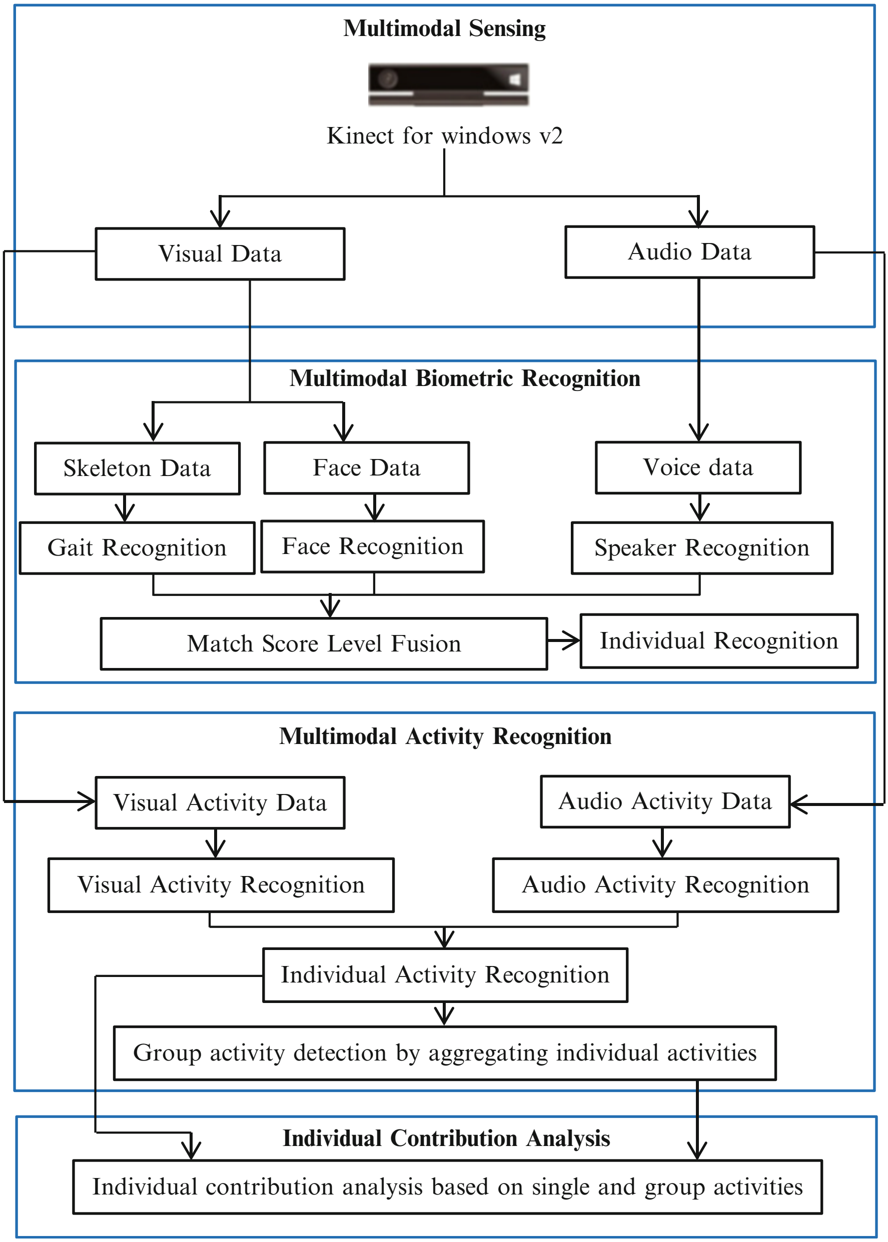


Fig. 1 Components of the proposed multi-modal system

array of sensors, which include (1) a color camera, (2) a depth sensor, and (3) a multi-array microphone setup. The depth sensor comprises a monochrome CMOS camera and an IR emitter. The depth sensor can build a 3D map of the objects by emitting human eye-invisible IR and then analyzing the light and shadow of the image captured by the CMOS camera. The multi-array microphone has an ambient noise cancellation mechanism, which can also be used to detect the source location of voice. In addition, Kinect sensor can construct a 3D virtual skeleton from a human body (Shotton et al., 2011). All these functionalities of Kinect have led to its application in different real-world problems, such as home monitoring (Stone & Skubic, 2011), healthcare (Chang, Chen, & Huang, 2011), surveillance (Popa, Koc, Rothkrantz, Shan, & Wiggers, 2012), etc.

In our multi-modal biometric system, Kinect plays an important role in data collection. Our system uses the Kinect data collected from the color camera, the depth sensor, and the microphone array. These different types of data facilitate multi-modal sensing and recognition of meeting participants as well as their voice, gestures, and activities.

3.2 *Individual Identification and Tracking*

Individual identification and authentication is a common problem for any intelligent environment. Traditional approaches use certain tokens, such as ID card or password verification for secure authentication. However, these solutions have some potential drawbacks. For example, ID cards can be lost, stolen, or forged, while passwords can be forgotten or compromised. Therefore, these approaches are vulnerable to forgery and do not provide sufficient security (Jain, Hong, & Kulkarni, 1999). On the other hand, biometric-based individual identification and authentication presents a convenient, secure, and unobtrusive alternative, which is being widely-deployed in many real-world applications.

A meeting room represents a dynamic and changing environment. As a result, relying on a single biometric trait for individual identification and authentication may not produce satisfactory performance due to several factors, such as noise, occlusion, intra-class variations, inter-class similarities, etc. (Ross & Jain, 2004). Hence, we propose a multi-modal biometric system to overcome this limitation. A multi-modal system exploits biometric data from multiple sources for individual recognition and authentication. Presence of multiple and fairly independent biometric information makes such systems more robust in a changing environment (Monwar & Gavrilova, 2009; Ross & Jain, 2004). We propose to use three different biometric data for individual recognition and authentication: (1) gait, (2) face, and (3) voice. Match score level fusion will be used to combine the recognition results of these three types of biometric.

3.2.1 Gait Recognition

We propose a model-based gait recognition method based on the 3D skeleton data provided by Kinect. A Kinect is required to be placed in a position so that it covers the entry point of the meeting room. Thus, this device can be used to capture both the gait and face data while individuals are entering the meeting room. Using the Kinect v2, we can detect and track 25 different skeletal points of an individual. These points are: (1) SpineBase, (2) SpineMid, (3) Neck, (4) Head, (5) ShoulderLeft, (6) ElbowLeft, (7) WristLeft, (8) HandLeft, (9) ShoulderRight, (10) ElbowRight, (11) WristRight, (12) HandRight, (13) HipLeft, (14) KneeLeft, (15) AnkleLeft, (16) FootLeft, (17) HipRight, (18) KneeRight, (19) AnkleRight, (20) FootRight, (21) SpineShoulder, (22) HandTipLeft, (23) ThumbLeft, (24) HandTipRight, and (25) ThumbRight. Preis, Moritz, Martin, and Claudia (2012) presented a Kinect skeleton-based gait recognition method based on 13 biometric features: height, the length of legs, torso, both lower legs, both thighs, both upper arms, both forearms, step-length, and speed. However, these features are mostly static and represent individual body structure, while gait is considered to be a behavioral biometric, defined as the pattern of the movement of body parts during locomotion. In our opinion, movement of the skeleton points over time is an important gait feature. Gabel, Gilad-Bachrach, Renshaw, and Schuster (2012) used the difference in position of these skeleton points between consecutive frames as their feature. We propose to combine the change information of the skeleton points over time and the skeleton structure features for robust gait recognition.

3.2.2 Face Recognition

Face recognition involves two major tasks: (1) face detection in image scene, and (2) face recognition. Face detection refers to detecting whether there is any face in the image scene, while recognition involves matching the detected face with the templates available in the system database (Luo, Gavrilova, & Wang, 2008). Many successful face detection (Viola & Jones, 2004) and face recognition algorithms (Bashar, Khan, Ahmed, & Kabir, 2014; Jafri & Arabnia, 2009) can be found in the literature. In recent years, local appearance face descriptors based on local binary pattern (LBP) (Ahonen, Hadid, & Pietikainen, 2006) and its variants (Ahmed, Bari, & Hossain, 2014; Guo, Zhang, & Zhang, 2010; Zhou, Wang, & Wang, 2008) have attained much popularity due to their computational efficiency and robustness to challenges, such as illumination and pose variations (Zhao & Pietikainen, 2009). Local binary pattern is a simple, yet effective texture primitive that encodes the local texture information into a binary pattern by thresholding the gray-level values of a local neighborhood with respect to the center. The encoded binary pattern acts as a template for micro-level texture details of an image, such as edges, spots, or corners. We propose to use LBP-based facial feature representation for the face recognition task.

3.2.3 Speaker Recognition

In a typical meeting scenario, there may be multiple competing speakers speaking at the same time. Hence, text-independent speaker recognition in a meeting room is a challenging task, which usually requires a microphone array to facilitate directional discrimination (Moore & McCowan, 2003). A survey of the state-of-the-art speaker recognition methods can be found in (Kinnunen & Li, 2010). Friedland and Vinyals (2008) presented an effective speaker identification system based on Gaussian mixture model (GMM). They reported an accuracy of about 85 % in a small meeting room environment. We propose to use Gaussian mixture probabilities (Narayanaswamy, 2005) for the speaker recognition task in our multi-modal meeting room system.

3.2.4 Match Score Level Fusion

Match score level fusion method combines the match scores obtained for different biometric data and is applicable to a wide variety of multi-biometric scenarios (He et al., 2010). In order to obtain a single match score from multiple sources of biometric data, match score level fusion employs different arithmetic operators, such as addition, subtraction, median, maximum, minimum, etc. on the match scores obtained for individual biometric (Gavrilova & Monwar, 2013). In our proposed multi-modal system, individual match scores generated for gait, face, and voice will be combined using a simple sum rule and the resulted sum will be used as the match score for the final decision (Gavrilova & Monwar, 2013). Match scores from different biometrics may not have the same range of values. Therefore, score normalization must be applied before combining individual match scores (Gavrilova & Monwar, 2013; Ross, NandaKumar, & Jain, 2006). Figure 2 shows the match score level fusion scheme for our proposed multi-modal biometric system.

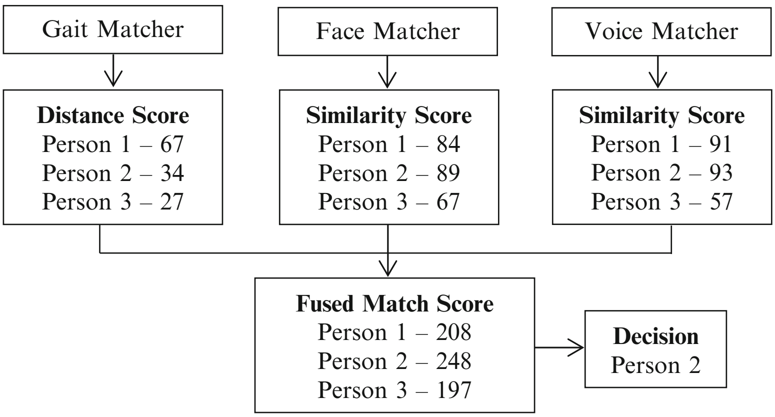


Fig. 2 Match score level fusion for gait, face, and voice-print biometric (adapted from (Gavrilova & Monwar, 2013))

3.3 Activity Recognition

Activity recognition is one of the key components of any intelligent system designed for a collaborative environment. Recognizing activities in a meeting room involves analyzing multi-modal data sources (Yu & Nakamura, 2010). In our proposed system, we analyze both video and audio data in order to recognize two types of activities: (1) individual activities and (2) group activities.

3.3.1 Individual Activity Recognition

Individual activity recognition is a challenging task that has attracted much attention due to its potential applicability in different scenarios, such as video surveillance (Niu, Long, Han, & Wang, 2004), health monitoring (Alhamid, Saboune, Alamri, & Saddik, 2011), etc. However, only a few researchers have addressed the problem of activity recognition in a meeting room context. Nait-Charif and McKenna (Nait-Charif & McKenna, 2003) developed a head tracker system, which tracks the head position of meeting room participants in order to detect different activities, such as entering, exiting, going to the whiteboard, getting up and sitting down. However, this system has two potential limitations, such as (1) it is heavily dependent on scene-specific constraints (Nait-Charif & McKenna, 2003) and (2) the tracking takes place in a 2D space, which limits its applicability greatly. Another approach proposed by Mikic et al. (2000) considered three basic activities: (1) a person located in front of the whiteboard, (2) a lead presenter speaking, and (3) other participants speaking. However, the objective of the activity recognition was only to select the best-view camera.

In our proposed multi-modal system, we aim to detect a total of 11 different activities. Table 1 lists the selected activities. Detection of these activities provides important context information regarding the meeting. In addition, it also facilitates analyzing participant contribution level. We propose to use the Kinect-based 3D skeleton model in order to detect activities like entering the room, exiting the room,

Table 1 List of selected individual activities

Visual activities	Audio-based activities
1. Enter the room	1. Presenter speaking
2. Exit the room	2. Non-presenter speaking
3. Sit down	
4. Get up	
5. Raise hand	
6. Gesture	
7. Change of facial expression	
8. Note-taking	
9. Idle	

sitting down, and getting up. The Kinect skeleton model provides skeleton point positions in a 3D space, which is more robust than the 2D head tracker-based method proposed by Nait-Charif and McKenna (2003). Instead of using only the head position, we can also analyze other skeleton points, such as shoulder center, shoulder left, and shoulder right. These four skeleton points can be used together to model the upper body movement more accurately. In our opinion, analyzing lower body skeleton points is not reliable since in most cases, lower body parts will be occluded by the meeting table or other obstacles. In addition, both gesture and note-taking activities can be recognized by tracking the movements of wrist and hand skeleton points. We propose to use local texture-based feature representation for detecting facial expression change. As mentioned previously, local texture patterns provide an efficient and effective way to represent facial features. These features have been successfully applied in many facial expression recognition applications (Ahmed, 2012; Ahmed & Kabir, 2012).

Detecting audio-based activities such as presenter speaking and non-presenter speaking can be performed using voice-print recognition. We define presenter as any individual who is standing in front of the whiteboard or presentation board in the meeting room. On the other hand, non-presenter is any individual who is not standing in front of the presentation board.

3.3.2 Group Activity Recognition

A significant amount of research in social psychology (McGrath, 1984) has established group activities as one of the most important components of a meeting (Morgan et al., 2001). A group as a whole facilitates disseminating information as well as making important decisions in a meeting scenario. Hence, recognizing group activities in a meeting facilitates understanding the context, detecting the level of collaboration, finding interesting patterns, etc. Zhang, Perez, Bengio, McCowan, and Lathoud (2004) proposed a framework based on a two-layer Hidden Markov Model (HMM) for group activity recognition in a meeting room setup. The key idea was to decompose group actions in a set of individual actions and detect those individual actions in order to determine group activities. In our proposed multi-modal system, we use this idea and define a set of group activities, which can be further decomposed into a set of individual actions. These group activities are:

- **Discussion**: Discussion can be defined as a number of participants speaking simultaneously. This activity can be detected by analyzing the audio data.
- **Monologue**: Monologue can be defined as one participant speaking for a long period of time with very little or no interruption. This activity recognition can be performed by analyzing the audio data.
- **Laughter**: This action can be defined as a number of participants laughing simultaneously. For laughter detection in a meeting room scenario, Kennedy

and Ellis (2004) proposed a method based on mel-frequency cepstral coefficients (MFCCs), delta MFCCs, and modulation spectrum. In their experiments, MFCCs achieved the best recognition rate. Hence, we will use mel-frequency cepstral coefficients for detecting laughter in our proposed system.

- **Voting**: A typical voting scenario can be defined as a number of individuals raising their hands. This activity can be recognized by tracking individual hand positions.
- **Group Note-taking**: This action can be defined as a number of individuals in the meeting room taking notes at the same time.
- **Group Focus of attention**: The group focus of attention can be determined by considering individual focus of attention based on their head orientation, pose, and eye gaze. Waibel et al. (2003) tracked focus of attention based on head orientation estimated using a neural network. A probabilistic model was then employed to map the head orientation to other participant locations.

3.4 Individual Contribution Analysis

An individual can contribute in many different ways in a meeting, which makes real-time automated contribution tracking a very challenging research problem. In addition, collaboration and communication in online environments (Matei, Oh, & Bruno, 2006) should also be considered. In our proposed multi-modal system, we define contribution level as a weighted sum of several features. The selected features for measuring individual contribution in a meeting are listed below:

- **Individual presentation time**: This measure can be defined as the time spent by an individual standing in front of the presentation board and speaking. Presentation is a key component in any meeting which facilitates dissemination of knowledge and thus supports the decision making process. Therefore, a presenter plays a key role in a meeting room environment.
- **Non-presenter speaking time**: This can be defined as the amount of time any non-presenter individual speaks for in a meeting.
- **Difference between meeting start time and individual entering the room**: This measure can be specified as how late a participant is for the meeting. Individuals who attend the meeting late typically contribute less in the meeting.
- **Individual being the focus of attention**: This can be defined as the amount of time for which an individual is the focus of attention in the meeting room.
- **Raise hand**: Raising hand represents a form of contribution via asking questions or voting. Hence, the number of times an individual raises hand should be an important feature.
- **Note-taking trigger**: The note-taking trigger can be defined as how many times an individual's presentation or speaking initiated a note-taking action.

- **Idle time**: This can be defined as how much time a participant was inactive for (for example, sitting idly without speaking or writing anything). Individual idle time should also be considered in contribution level detection.

4 Concluding Remarks

Behavioral biometric recognition is an emerging domain of research, which can potentially be applied to manage user identity and security in collaborative and social activities, interactions, and communications. In addition, utilizing these traits can lead to a smart collaborative environment design that can utilize data obtained from different online and offline sources, such as Twitter, online blogs, discussion forums, online games, meetings, etc. Analysis of these multi-modal data sources can potentially reveal user behavioral traits, habits, patterns of technology usage and thus, enables automated quantization of user contribution in knowledge generation and collaborative content creation. Analysis of behavioral biometrics thus provides a new set of tools that can effectively be used for the security and user identity management in social media contexts. In this chapter, we present our current research efforts toward building a multi-modal biometric-based authentication, activity and behavior analysis tools for collaborative environment. The objective is to automatically manage user authentication and secure access to shared resource and content, analyze meeting workflow and group dynamics, participation of the users in collaborative activities and thus quantify individual contribution. A set of physiological and behavioral biometric traits has been introduced that can potentially be used for this task.

Real-world meetings encompass a variety of activities of different modalities, such as speech, presentation, gesture, etc. Therefore, we propose to integrate complementary modalities, such as audio and video for accurate individual recognition and authentication. In addition, we present a Kinect-based 3D visual activity tracking system that tracks different activities based on Kinect skeleton data. In addition, audio-based activities can be tracked from the data collected by the Kinect microphone-array. These tracked activities are then used to define individual contribution and collaboration level. Our proposed system does not rely on any extensive human-computer interactions, rather it processes multi-modal data sources for automated authentication and activity recognition. However, the latest Kinect camera is able to track at most six skeletons simultaneously, which limits the applicability of the proposed method to a small meeting room setup. The concepts presented here can also be extended to design online and virtual collaborative meeting room environments, which can be used to monitor and track online interactions and contributions of meeting participants. In addition, the concepts of understanding online collaboration, collaborative knowledge generation, and user-influence analysis discussed in Chaps. 6, 7, and 8, respectively can also be applied to an online collaborative environment scenario. We hope that the ideas and questions presented in this chapter will help researchers in this field to advance the current state-of-the-art technologies.

References

Ahmed, F. (2012). Gradient directional pattern: A robust feature descriptor for facial expression recognition. *IET Electronics Letters, 48*(19), 1203–1204.

Ahmed, F., Bari, H., & Hossain, E. (2014). Person-independent facial expression recognition based on compound local binary pattern (clbp). *International Arab Journal of Information Technology, 11*(2), 195–203.

Ahmed, F., & Kabir, M. H. (2012). Directional ternary pattern (dtp) for facial expression recognition. In *Proceedings of IEEE International Conference on Consumer Electronics* (pp. 265–266).

Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(12), 2037–2041.

Alhamid, M. F., Saboune, J., Alamri, A., & Saddik, A. (2011). Hamon: An activity recognition framework for health monitoring support at home. In *Proceedings of IEEE Instrumentation and Measurement Technology* (pp. 1–5).

Bashar, F., Khan, A., Ahmed, F., & Kabir, H. (2014). Face recognition using similarity pattern of image directional edge response. *Advances in Electrical and Computer Engineering, 14*(1), 69–76.

Busso, C., Hernanz, S., Chu, C. W., Kwon, S., & Lee, S. (2005). Smart room: Participant and speaker localization and identification. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2* (pp. 1117–1120).

Chang, Y. J., Chen, S. F., & Huang, J. D. (2011). A kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities, 32*(6), 2566–2570.

Chiu, P., Kapuskar, A., & Wilcox, L. (1999). Meeting capture in a media enriched conference room. In *Proceedings of International Workshop on Cooperative Building*, LNCS, Vol. 1670 (pp. 79–88).

Deutschmann, I., Nordstrom, P., & Nilsson, L. (2013). Continuous authentication using behavioral biometrics. *IT Professional, 15*(4), 12–15.

Down, M., & Sands, R. (2004). Biometrics: An overview of the technology, challenges and control considerations. *Journal of Information Systems Control, 4*, 53–56.

Friedland, G., & Vinyals, O. (2008). Live speaker identification in conversations. In *Proceedings of ACM International Conference on Multi-media* (pp. 1017–1018).

Gabel, M., Gilad-Bachrach, R., Renshaw, E., & Schuster, A. (2012). Full body gait analysis with Kinect. In *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1964–1967).

Gavrilova, M. L., & Monwar, M. (2013). *Multimodal biometrics and intelligent image processing for security systems*. Hershey, PA: IGI Global.

Guo, Z., Zhang, L., & Zhang, D. (2010). Rotation invariant texture classification using lbp variance (lbpv) with global matching. *Pattern Recognition, 43*, 706–719.

He, M., et al. (2010). Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recognition, 43*(5), 1789–1800.

Hornler, B., & Rigoll, G. (2009). Multi-modal activity and dominance detection in smart meeting rooms. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 1777–1780).

Jafri, R., & Arabnia, H. R. (2009). A survey of face recognition techniques. *Journal of Information Processing Systems, 5*(2), 41–68.

Jain, A., Hong, L., & Kulkarni, Y. (1999). A multimodal biometric system using fingerprint, face, and speech. In *Proceedings of International Conference on Audio-and Video-based Biometric Person Authentication* (pp. 182–187).

Jain, R., Kim, P., & Li, Z. (2003) Experiential meeting systems. In *Proceedings of ACM Workshop on Experiential Telepresence* (pp. 1–12).

Jeon, I. S., Kim, H. S., & Kim, M. S. (2011). Enhanced biometrics-based remote user authentication scheme using smart cards. *Journal of Security Engineering, 8*(2), 237–254.
Kennedy, L. S., & Ellis, D. P. W. (2004). Laughter detection in meetings. In *Proceedings of NIST ICASSP 2004 Meeting Recognition Workshop* (pp. 118–121).
Kinnunen, T., & Li, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. *Speech Communication, 52*(1), 12–40.
Kumar, M. (2010). An enhanced remote user authentication scheme with smart card. *International Journal of Network Security, 10*(3), 175–184.
Lee, D., Erol, B., Graham, J., Hull, J. J., & Murata, N. (2002). Portable meeting recorder. In *Proceedings of ACM International Conference on Multimedia* (pp. 493–502).
Luo, Y., Gavrilova, M. L., & Wang, P. S. P. (2008). Facial metamorphosis using geometrical methods for biometric applications. *International Journal of Pattern Recognition and Artificial Intelligence, 22*(3), 555–584.
Matei, S., Oh, K., & Bruno, R. (2006). Collaboration and communication: A social entropy approach. In *Proceedings of National Communication Association Annual Convention*, San Antonio.
McGrath, J. E. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice-Hall.
Mikic, I., Huang, K., & Trivedi, M. (2000). Activity monitoring and summarization for an intelligent meeting room. In *Proceedings of Workshop on Human Motion* (pp. 107–112).
Monwar, M. M., & Gavrilova, M. L. (2009). Multimodal biometric system using rank-level fusion approach. *IEEE Transactions on Systems, Man, and Cybernetics, 39*(4), 867–878.
Moore, D. C., & McCowan, I. A. (2003). Microphone array speech recognition: Experiments on overlapping speech in meetings. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (Vol. 5, pp. 6–10).
Morgan, N., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Janin, A., Pfau, T., Shriberg, E., & Stolcke, A. (2001). The meeting project at ICSI. In *Proceedings of HLT Conference*.
Mosvick, R., & Nelson, R. (1987). *We've got to start meeting like this! A guide to successful business meeting management*. Glenview, IL: Scott, Foresman.
Nait-Charif, H., McKenna, S. J. (2003). Head tracking and action recognition in a smart meeting room. In *Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*.
Narayanaswamy, B. (2005). Improved text-independent speaker recognition using Gaussian mixture probabilities. M.Sc. Thesis, Carnegie Mellon University.
Niu, W., Long, J., Han, D., & Wang, Y. F. (2004). Human activity detection and recognition for video surveillance. In *Proceedings of IEEE International Conference on Multimedia and Expo, Vol. 1* (pp. 719–722).
Penteado, B. E., & Marana, A. N. (2009). *A video-based biometric authentication for e-Learning web applications. Enterprise information systems* (pp. 770–779). Berlin, Germany: Springer.
Popa, M., Koc, A. K., Rothkrantz, L. J. M., Shan, C., & Wiggers, P. (2012). Kinect sensing of shopping related actions. *Communications in Computer and Information Science, 277*, 91–100.
Prabhakar, S., Pankanti, S., & Jain, A. K. (2003). Biometric recognition: Security and privacy concerns. *IEEE Security and Privacy, 1*(2), 33–42.
Preis, J., Moritz, K., Martin, W., & Claudia, L. P. (2012). Gait recognition with kinect. In *Proceedings of International Workshop on Kinect in Pervasive Computing*.
Ross, A., & Jain, A. (2004). Multi-modal biometrics: An overview. In *Proceedings of European Signal Processing Conference* (pp. 1221–1224).
Ross, A., NandaKumar, K., & Jain, A. K. (2006). *Handbook of multibiometrics*. Berlin, Germany: Springer.
Rui, Y., Gupta, A., & Cadiz, J. J. (2001). Viewing meetings captured by an omnidirectional camera. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 450–457).
Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., & Blake, A. (2011). Real-time human pose recognition in parts from single depth image. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1297–1304).

Stanford, V., Garofolo, J., Galibert, O., Michel, M., & Larrun C. (2003). The NIST smart space and meeting room projects: Signals, acquisition, annotation and metrics. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 4* (pp. 6–10).
Stiefelhagen, R. (2002). Tracking focus of attention in meetings. In *Proceedings of IEEE International Conference on Multimodal Interfaces* (pp. 273–280).
Stone, E. E., & Skubic, M. (2011). Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In *Proceedings of Pervasive Health Conference*.
Sultana, M., Paul, P. P., & Gavrilova, M. (2014). A concept of social behavioral biometrics: Motivation, current developments, and future trends. In *Proceedings of International Conference on Cyberworlds* (pp. 271–278).
Sultana, M., Paul, P. P., & Gavrilova, M. (2014). Mining social behavioral biometrics in Twitter. In *Proceedings of International Conference on Cyberworlds* (pp. 293–299).
Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision, 57*(2), 137–154.
Waibel, A., Bett, M., & Finke, M. (1998). Meeting browser: Tracking and summarizing meetings. In *Proceedings of Broadcast News Transcription and Understanding Workshop* (pp. 281–286).
Waibel, A., Schultz, T., Bett, M., Denecke, M., Malkin, R., Rogina, I., Stiefelhagen, R., & Yang, J. (2003). SMaRT: The smart meeting room task at ISL. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 4* (p. 752).
Wellner, P., Flynn, M., & Guillemot, M. (2004). Browsing recorded meetings with Ferret. In *Proceedings of International Workshop on Machine Learning for Multimodal Interaction* (pp. 12–21).
Yanushkevich, S., Gavrilova, M., Wang, P., & Srihari S. (2007). Image pattern recognition: Synthesis & analysis in biometrics. *Series in Machine Perception & Artificial Intelligence, Vol. 67*, p. 423.
Yu, Z., & Nakamura, Y. (2010). Smart meeting systems: A survey of state-of-the-art and open issues. *ACM Computing Surveys, 42*(2), 1–20.
Yu, Z., Ozeki, M., Fujii, Y., & Nakamura, Y. (2007). Towards smart meeting: Enabling technologies and a real-world application. In *Proceedings of ACM International Conference on Multimodal Interfaces* (pp. 86–93).
Zhang, D., Perez, D. G., Bengio, S., McCowan, I., & Lathoud, G. (2004). Modeling individual and group actions in meetings: A two-layer HMM framework. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*.
Zhao, G., & Pietikainen, M. (2009). Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. *Pattern Recognition Letters, 30*, 1117–1127.
Zhou, X., & Bhanu, B. (2007). Integrating face and gait for human recognition at a distance in video. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 37*(5), 1119–1137.
Zhou, H., Wang, R., & Wang, C. (2008). A novel extended local binary pattern operator for texture analysis. *Information Sciences, 178*, 4314–4325.

Part III
Improving Transparency Through Documentation and Curation

In the Flow: Evolving from Utility Based Social Medium to Community Peer

Michael G. Zentner, Lynn K. Zentner, Dwight McKay, Swaroop Samek, Nathan Denny, Sabine Brunswicker, and Gerhard Klimeck

1 Introduction

The word "media" evokes images of lively rooms packed with reporters frenetically covering events. Combined, the words "social media" today make one think of many people interacting online in a way that circumvents conventional media, but essentially accomplishes the same thing: making people rapidly aware of events, whether they are as global as an emerging international conflict or as local as the facial expression of one's cat on a given morning. Scholars on computer-mediated communication and human computer interaction regularly use the term social media in referring to a group of Internet-based technologies that allows users to easily create, edit, evaluate, and/or link to content or to other creators of content (c.f., Kaplan & Haenlein, 2010). In practice, one tends to think of Facebook and Twitter, where the nexus of interaction is a micro-expression of an event or idea. We may think less often of sites such as Wikipedia or LinkedIn, where the nexus of interaction is the more persistent longitudinal development of an article or professional profile, respectively.

M.G. Zentner (✉)
Information Technology, Purdue University, West Lafayette, IN, USA

Network for Computational Nanotechnology, Purdue University, West Lafayette, IN, USA
e-mail: mzentner@purdue.edu

L.K. Zentner • S. Samek • G. Klimeck
Network for Computational Nanotechnology, Purdue University, West Lafayette, IN, USA

D. McKay • N. Denny
Information Technology, Purdue University, West Lafayette, IN, USA

S. Brunswicker
Research Center for Open Digital Innovation, Purdue University, West Lafayette, IN, USA

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*, Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_10

Regardless, the term "media" implies passive instruments that accept input from some people and allow reading of that input by other people. Input and output may undergo cycles of development as people "discuss" input. But when and where will we reach a point where "media" is no longer the right term and we are on the precipice of the medium becoming an active contributor to the body of knowledge that may be developed by it and its interactors? Perhaps "where" is in the online conduct of science and "when" is not so distant in the future.

The discussion in this chapter will use as an example nanoHUB.org, the world's largest online facility for the conduct of science related to nanotechnology. nanoHUB's "in the flow" development philosophy, its current status, its role as a platform in the nanotechnology and social science communities, and forward looking developments will be described in the context of it becoming a peer within its own community.

2 Design for Utility or Design for Social?

As it is the intent to discuss social media evolving into something different, the term "online interaction space" will be used here instead. The biggest upfront considerations when designing an online interaction space are utility margin and uniqueness. Utility margin conceptually is the difference between the benefit users gain from an online interaction space and the effort they must expend in using it. Note that the word "benefit" is important: to simply do something more efficiently is not sufficient if the user perceives little or no value in the outcome. Uniqueness is the degree to which no substitute exists that can satisfy the same needs by the same or different mechanisms. Too often interaction space creators are enamored with large successes and assume that providing the same or marginally more relevant capabilities to a more focused niche will ensure success, only to be disappointed when the intended audience does not adopt their creation. They misestimate their intended user's perception of utility margin (e.g. the benefit of an exclusive membership does not outweigh the work of having to log in to yet one more site) and uniqueness (e.g. a LinkedIn group will accomplish enough of what is needed to serve their community without a whole new site). In other cases, they seek to introduce a new behavior to their intended users without facilitating any of the current activities in which the intended users engage. As a result, the potential user has no frame of reference within which to judge marginal utility even though uniqueness may be high. Particularly when the design involves a new social mode of interaction that has no marginal utility unless a large community participates (the network effect), gaining users will be even more challenging. Designing to be social without designing for an achievable marginal utility is taking a shortcut that is highly likely to lead to failure.

Considering marginal utility and uniqueness, an interaction space must be designed in terms of its nexuses and modes of interaction to create an 'affordance' for everyday activities. The concept of an affordance refers to the action potential that can be taken given a technology to support a particular everyday task (Gibson, 1979; Majchrzak & Markus, 2013). The nexuses and modes can be chosen by looking at the assets and methods of working with those assets that the intended audience employs on a regular basis. In so doing, the interaction space is designed to facilitate some portion of the intended user's everyday workflow.

Without embedding one's creation into the flow of a user's normal activities and thus satisfying the selfish need to accelerate their individual efforts, gaining an appreciably sized audience of intended users is not likely and the network effect cannot be achieved. As pointed out by the psychologist Csikszentmihalyi, the interaction space may afford a flow state, a mental state, in which an individual user is fully immersed into a daily activity, and enjoys the process of the activity (Csikszentmihalyi, 1990).

One tends to attribute a sort of prescience of design to systems like Twitter and Facebook that, when viewed today, appear to have been successfully designed for social interaction without respect to utility. However, the first SMS transmission occurred in 1992 (Snowden, 2006) long before Twitter was conceived in 2006 (Miller, 2010), which simply created the ability to broadcast short messages to more than one follower simultaneously. Facebook also started by fulfilling the selfish need of people wanting to see pictures of other people on the same college campus, and only subsequently grew into the platform for social interaction. Both systems satisfied the selfish needs of the individual before gaining enough mass to realize the network effect. They were designed to be "in the flow" of the user, providing differentiation and beneficial marginal utility.

3 nanoHUB

Today nanoHUB hosts 342 simulation tools and 4,144 online resources in the form of courses, videos, animations, and downloadable documents (Fig. 1). nanoHUB currently serves over 325,000 users annually, over 13,000 of which run in excess of 500,000 simulations that consume more than 14,000 CPU days annually (Fig. 2). This may seem like success by design; however, like the much more successful and popular online interaction spaces discussed above, nanoHUB also achieved its audience by gradual evolution.

When created, nanoHUB was primarily focused on a goal of delivering access to simulation tools and computational resources over the web. The nexus of collaboration was the simulation tool. What previously had been individuals or small groups creating simulation tools for small user audiences changed with the introduction of nanoHUB: the audiences became much larger. As time progressed and audience size grew, additional modes of interaction were introduced, all based upon fitting into different elements of user workflows and in effect "purchasing the right" in the mindshare of the user to engage them in a successively broader social environment.

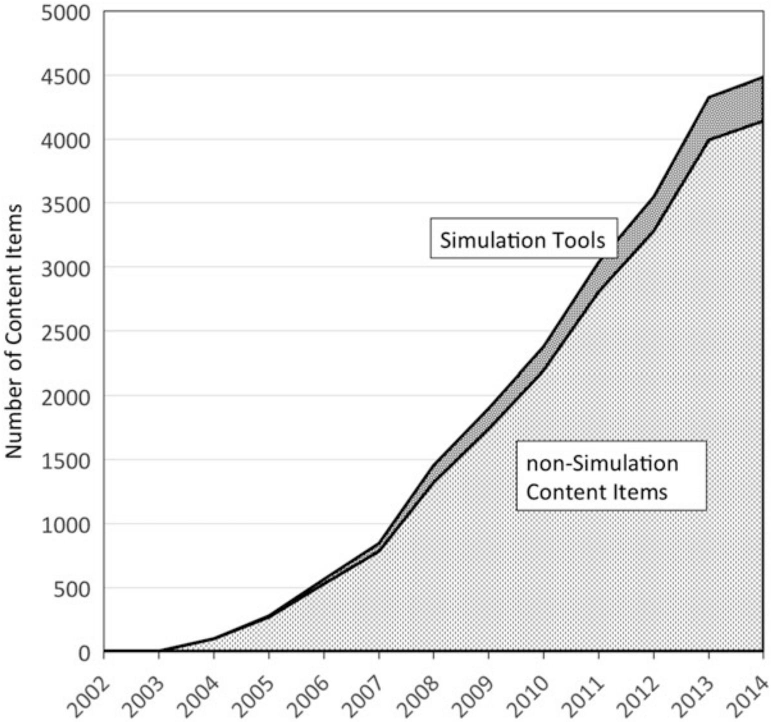


Fig. 1 This plot shows the growth in simulation tools and other resources over time on nanoHUB, culminating at 4,486 resources today. The change in slope in 2005 corresponds to the introduction of the Rappture simulation tool development kit. The corresponding increase in non-simulation content items suggests a relationship between tool creation and supporting technical materials

It is important to stress again that none of the design of nanoHUB has been from a speculative viewpoint of enabling a massive social network. Rather, each new development has been associated with increments of increased marginal utility and uniqueness.

4 Workflows Facilitated by nanoHUB

nanoHUB has been constructed on a gradual basis to fit into several workflows, all of which are centered around various aspects of the core nexus of interaction, simulation tools. Specifically, these include aspects of tool dissemination, interface construction, interface maintenance, sense-making, and publication of supporting augmentative information.

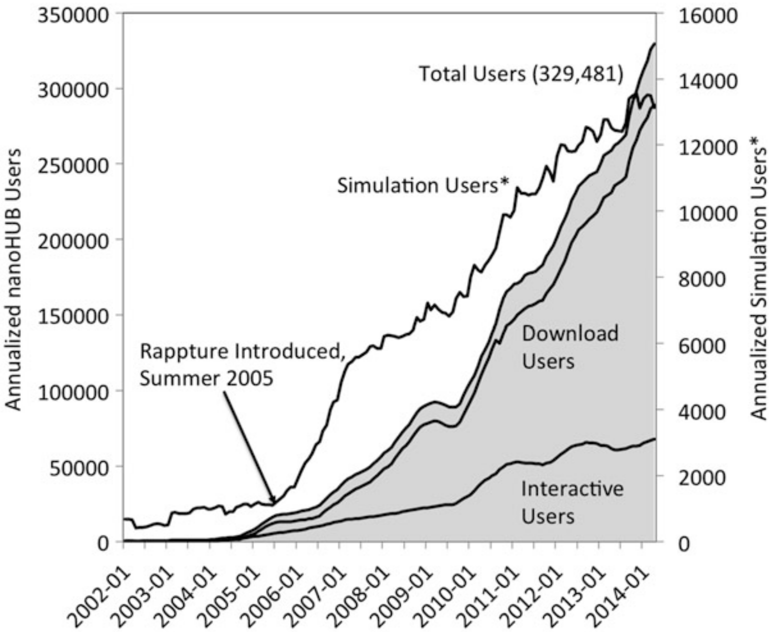


Fig. 2 Growth in total and simulation users over time. The introduction of the Rappture toolkit was instrumental in initiating the much more rapid growth rate, especially of simulation users

4.1 Tool Dissemination

Prior to nanoHUB's first publication of an online simulation tool, nanoscience researchers wishing to share their simulation tools were hindered by several difficulties. Consider this a two party relationship, where the supplying party is either an individual or a small group that has developed a simulation code and the consuming party is an individual, group, or institution that might wish to use that simulation code. The first barrier was awareness. The consumer might only become aware of the supplier's code through mention in the research literature. Such awareness has the time delay associated with publication of peer-reviewed articles, and the limitations on discoverability by common literature search methods at the time. At the same time, web search engines were becoming commonplace as a new mode of discovery. A second barrier was concerned with intellectual property. To distribute a simulation tool, the supplier would either need to distribute their source code, or create an installable binary version for the consumer. Until their research had advanced, suppliers were often uncomfortable with source code distribution and were reluctant to invest in the cost and time to create binary packages that could be easily installed on the wide variety of architectures, operating systems, and library versions owned by consumers. When a supplier was comfortable enough to distribute source code, a third barrier was the amount of effort required of the consumer to compile and install the tool locally. A fourth barrier was access to sufficient compute resources to run the tool once compiled. Not all consumers had access to enough compute power to effectively run the simulations. Finally, all of the barriers mentioned above were encountered again every time the original supplier revised the code, if the consumer desired access to the latest version. As a result, new releases reflecting the latest research results were not common.

nanoHUB interceded in this workflow by becoming a publicly available site that allowed suppliers to create and compile simulation tools, add web interfaces to them that could be accessed through a forms based web browser interface, and provide back-end compute resources (Kapadia, Fortes, & Lundstrom, 1997). As a result, suppliers could make their tools available on the web and accessible to search engines, even before publication in the research literature. They could avoid the intellectual property and installation package issues by retaining their source code and only building for one host environment. Consumers could access tools online without the significant effort of installing locally, and had access to back end compute resources that were otherwise unavailable to them. Updating to the most current version of the code became much easier. nanoHUB therefore added an active life-cycle management to the simulation tools.

4.2 Interface Construction, Interface Maintenance, and Sense-Making

With the initial workflow facilitated, a new need arose concerning the construction and maintenance of user interfaces. Nanoscience researchers are typically not web programmers, and therefore needed to engage such programmers to create interfaces for their tools. Often this meant that during the time a web interface was under construction, new research results would be incorporated into new versions of the simulation tool. By the time the web interface was finished and deployed, the underlying tool was no longer reflective of the most recent research. In addition, the interfaces were not significantly interactive, and therefore did not easily facilitate the user's examination and interpretation of results.

In response to these challenges, the Rappture (McLennan & Kennell, 2010) system was introduced. Rappture is a data and user interface management toolkit that allows a simulation tool programmer to easily construct a user interface on top of their core simulation code regardless of its language of implementation. Rappture also allows assembly and orchestration of parameters and data that feed the simulation. Further, Rappture's middleware layer allows for an interactive interface that can show tool run progress, that allows interactive inspection of results both textually and with powerful visualizations, and that enables side-by-side comparison of multiple simulation runs to help the user understand the cause and effect relationships as they investigate the simulation under various conditions. As a result, users

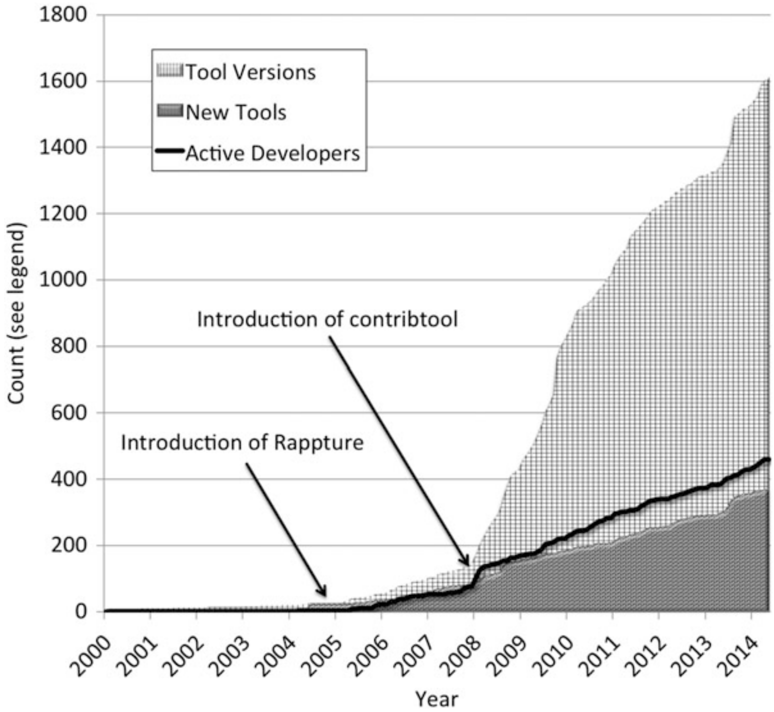


Fig. 3 This plot shows the relationship between the number of tool versions prior to and after the incorporation of Rappture and contribtool. The dramatic change in slope versus that of the rate of new tool creation indicates tool developers are much more often keeping their online simulations up to date with their most current results

were given a much richer experience. Also, tool developers no longer had long development cycles, and were able to keep their tools much more up to date as new versions were created, particularly with the introduction of a more automated update process called "contribtool" in late 2007 (Fig. 3).

4.3 Augmentative Information and Content Repurposing

From Fig. 2 there are clearly many more nanoHUB users than the subset that use simulation tools. This profile of the users evolved as a result of the community's desire to understand on a wider basis what was being offered on nanoHUB. With Rappture introduced, nanoHUB made tools with rich interfaces available to a large community. This availability for each tool was originally expected to serve users in the same or similar research areas as the domain simulated by the tool. It became clear that additional support materials might be necessary in order to provide context for the simulation tools. This support took the form of user feedback mechanisms and a variety of categories of online resources in addition to the simulation tools. Indeed, Fig. 1 shows a slope change beginning near the time of the introduction of Rappture, indicating a rise in support materials accompanying the rise in simulation tools.

Modules were created for the nanoHUB content management system that allowed forum based discussions and question and answer capability around specific simulation tools. User feedback mechanisms were put in place for support ticket management and for a quality rating system. As a result, those who are contemplating the use of a simulation tool can see how well it is ranked by the independent community of users. They can often interact directly with the tool creators through the forums to gain a better understanding of the underlying principles of the simulation.

Many additional resource types were also created and opened for the public to make contributions, including animations, courses, downloads, learning modules, online presentations, presentation materials, publications, series, teaching materials, workshops, and most recently, databases and compact models. The initial intent of these resources was to supplement the simulations provided on nanoHUB by placing them in context. For example, such resources may contain technical information about the physical phenomena underlying the simulation, or about how the simulation may be positioned in classroom based learning contexts.

Another very popular nanoHUB feature has been the hosting of unique high quality video-based courses in nanotechnology that cannot be found anywhere else. These courses were originally standard courses over a whole semester consisting of about 45 lectures of 50 min each. Recently the course format has evolved into 5-week courses that are delivered in 20-min segments with active testing. This format is embodied in nanoHUB-U.

Supplemental information has likely assisted in the repurposing of tools, particularly with respect to classroom use. Based on a user similarity calculation and clustering algorithm (to be published separately), classroom behavior can be detected as groups of users utilize simulation tools in a time-coordinated manner (Fig. 4). The classroom use has grown significantly over time, to where over 20,000 students in over 1,100 classroom-like settings have been detected over the life of nanoHUB.

5 A Platform for Sociotechnical Research

Although the primary goal of nanoHUB has been to serve nanoscience researchers with simulation tools, the evolution of the capabilities offered on nanoHUB and the growing user audience has opened an entirely different research area. Every action over a period of more than 10 years has been recorded, including every resource accessed, every simulation run, and the parameterization of those simulation runs.

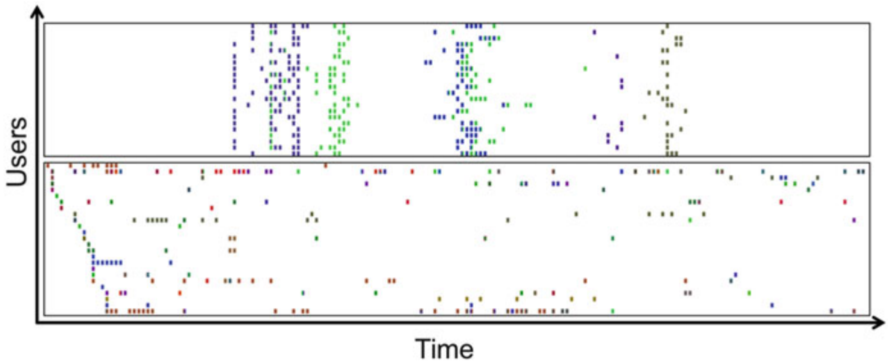


Fig. 4 This figure shows simulation activity over time. Each horizontal row corresponds to a single user. Each vertical column represents a given day. Each *dot* represents a user activating a simulation tool on a given day. Each unique shade corresponds to a different tool. The plot on *top* is an example of coordinated activity, corresponding to likely activity in a classroom scenario. The plot on *bottom* illustrates to non-coordinated activity

As such, nanoHUB has amassed a chronicle of a scientific community's online behavior over a meaningful timescale.

Figures 5 and 6 characterize the simulation data. Over the life of nanoHUB, 3.4 million simulations have been run in 697,725 sessions. The longitudinal assembly of sessions for a given user provides a trajectory of the direction, thoroughness, and pattern the user employed as they investigated the phenomena modeled by the simulation tools. Figure 6 illustrates an initial study of the novelty of the simulations during the period from early 2007 through mid 2011. One might expect simulations run in a classroom to be somewhat repetitive. Alternatively, one might expect simulations run for research to exhibit a high variety of differences. The concept of a simulation signature is used to identify the differences between simulation runs. Identical runs, such as those performed by students, will have a common signature. Different runs, like those run by researchers, will have different signatures. The growing variety of signatures indicates a collective user group that continues to innovate and explore new aspects of simulated phenomena.

For the resources that are not simulation tools, patterned use also emerges. A time and location based clustering method (to be published separately) has been developed to detect when groups of people from nearby locations exhibit time coordinated access of these resources as one might expect in a classroom setting. The analysis for the year of 2012 shows that 2,194 such coordinated clusters were found that utilized 1,319 resources, indicating that a significant number of these resources are being used as supporting information in the educational process. Figure 7 illustrates the breadth of use by such clustered users. There are 600 resources used by least 50 clustered users, and 200 resources with at least 200 clustered users. The lack of a steep decline in this relationship indicates that the classroom behaviors are served by a diverse set of resources, and not a small core.

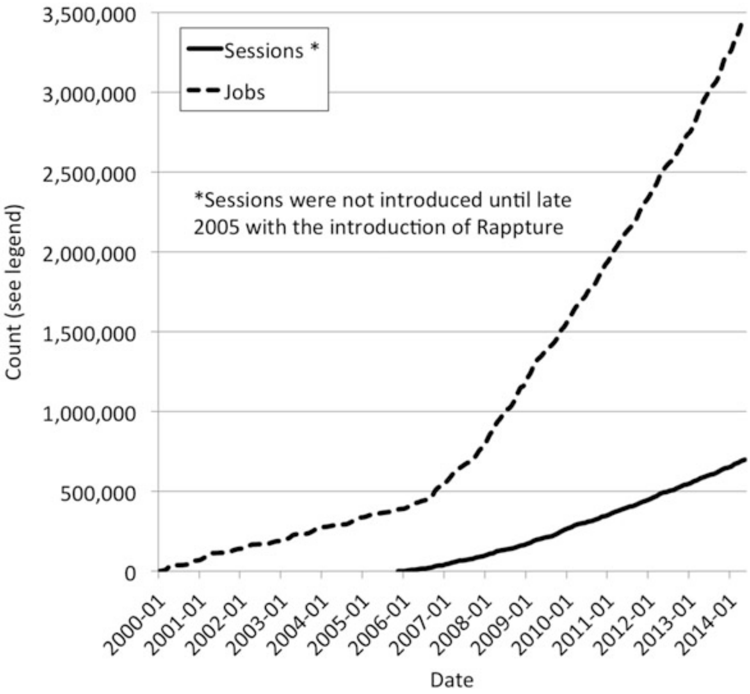


Fig. 5 This figure shows the number of simulation sessions and the number of simulation job runs over time. The significance of a session is that it represents an episode of one user's investigation, containing perhaps many individual simulation runs. The relationship between sessions and jobs indicates that simulation use is not casual: users undertake a significant amount of investigation at each sitting

These data suggest that nanoHUB has served as an enabling platform for the nanoscience community, where members can supply and consume resources, and thereby shape each other's experience. These data also suggest that nanoHUB has not played a passive role in shaping how the nanoscience community develops knowledge. To the contrary, each addition to nanoHUB to enable a new aspect of the users' workflows has changed the community's behavior while engaged in those workflows, and has also been repurposed by others for additional objectives. Simulation tool developers spontaneously increased the number of tool versions as Rappture enabled them to easily keep current. Users spontaneously began using simulation tools in classroom settings and migrated research results into the classroom with a median time of less than 6 months (Madhavan, Zentner, & Klimeck, 2013). Users spontaneously contributed supporting materials that were not simulation tools, and spontaneously began adopting those in classroom settings.

Aside from usage based metric evidence like that produced here, resource consumers have also documented their use of nanoHUB in the scientific literature by citation, as they would any other researcher. By this measure, the "persona of

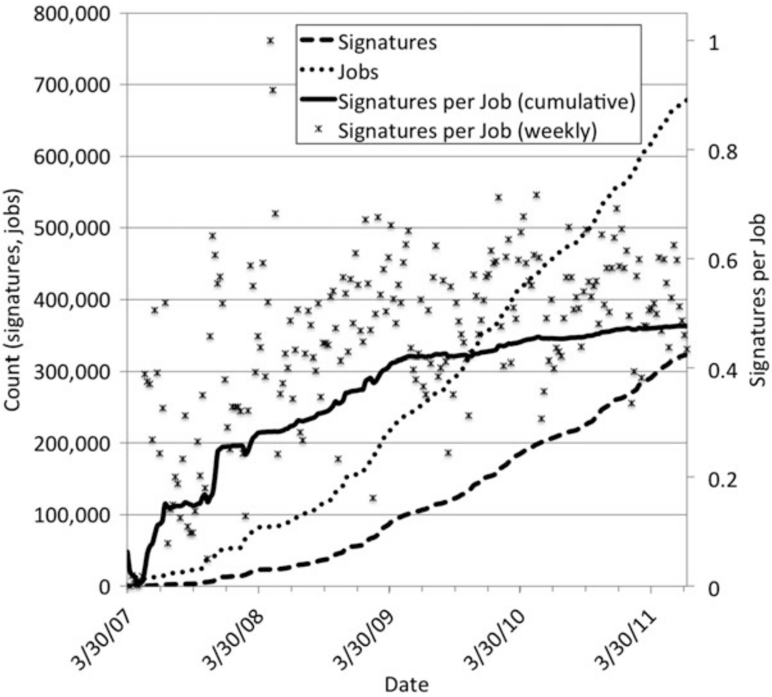


Fig. 6 This figure shows the growth in the number of signatures of tool runs from early 2007 to mid 2011. A signature is a collection of input parameters and their values. If two such collections are identical, they have the same signature. Nearly half of the jobs exhibit signature duplication (as one might expect for repetitive education use). Conversely, the other half are unique (as might be expected in a research environment). The growing number of unique signatures indicates that the community has not stagnated in its investigations, and continues to explore new regions of parameter space. This is a measure of the generative capacity of the nanoHUB user group

nanoHUB" has attained an h-index level in just over 10 years that is on par with that of career achievements of National Academy of Engineering members (Fig. 8).

Data like those described here are forming the basis of many new lines of investigation of the online behavior of this community and will be periodically released to the sociotechnical science research communities for additional studies of the online conduct of science. The purposes are for understanding the past, but more importantly for learning new ways in which the nanoHUB platform may serve and shape its community in the future. Technology changes how people interact as they engage in science (Orlikowski & Scott, 2008) and acts in shaping behavior (Bostrom, Gupta, & Thomas, 2009; Majchrzak & Markus, 2013; Orlikowski & Scott, 2008; Yoo, Boland, Lyytinen, & Majchrzak, 2012). With a large user community, nanoHUB is at the forefront of demonstrating these theories in the online conduct of science.

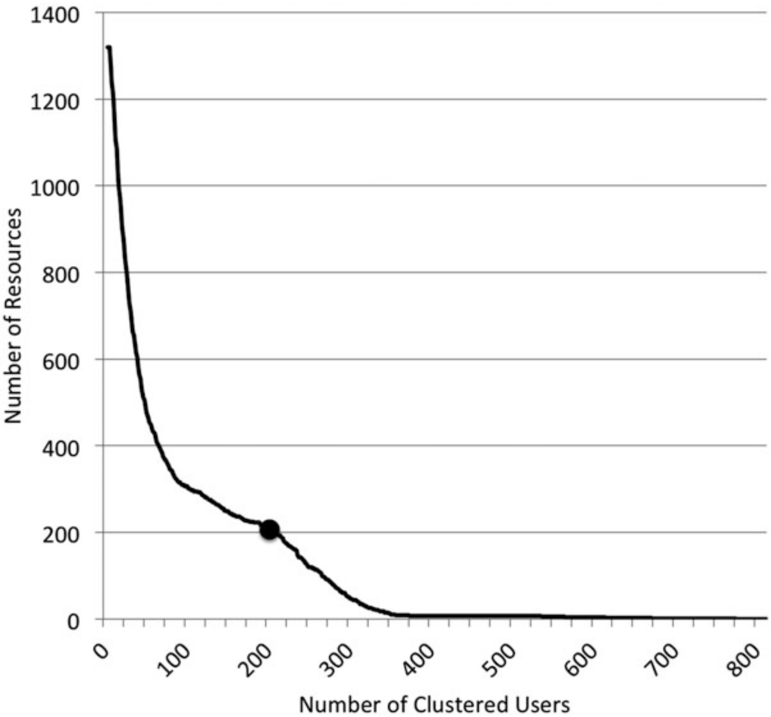


Fig. 7 This figure shows number of resources that have a set of users of at least a given size that exhibit clustered behavior. The labeled point for example indicates that slightly more than 200 resources have at least 200 distinct users exhibiting clustered behavior

6 Beyond Social Media Platforms: The Evolution of the Community Peer

In the introduction and in the previous section two provocative comments were made: (i) the medium of an online interaction environment becoming an active contributor to its own body of knowledge and (ii) that a platform like nanoHUB might have a persona. A key attribute of nanoHUB that makes it unique relative to other social media platforms resides in its nexus of interaction. Recall that for Twitter, Facebook, Wikipedia, and LinkedIn this nexus is a short text message, a status update, an article, and a personal profile, respectively. For nanoHUB, it is a simulation tool. Forums like Twitter, Facebook, and LinkedIn have passive nexuses: consuming information from them produces no new information other than that an individual consumed data. Simulation tools in nanoHUB, on the other hand, are active nexuses: consuming from them produces new information about a physical phenomenon under specified conditions. nanoHUB and science gateways like it that will emerge in the upcoming years are places for not just talking about science, but for conducting it.

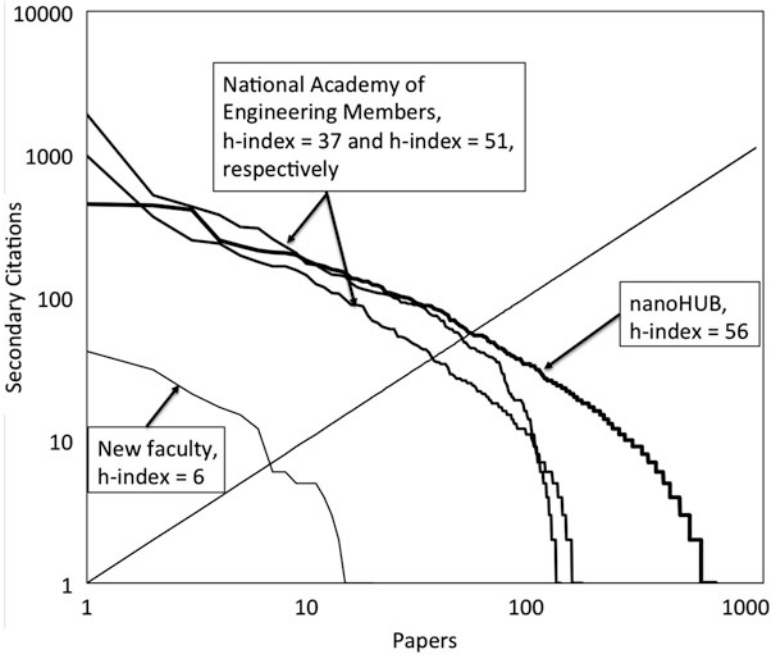


Fig. 8 Utilizing secondary citation counts, an h-index can be determined for nanoHUB-related papers and compared to both typical young researchers and high-achieving researchers, such as members of the U.S. National Academy of Engineering. This plot illustrates that nanoHUB collectively compares quite favorably to high-level researchers

A new generation of capabilities that exploit the active nexus will prompt nanoHUB more strongly into the role of a knowledge-generating participant within its community. User explorations based on simulation tools will be used as templates from which to interpolate and extrapolate, producing knowledge as nanoHUB automatically fills in gaps between parameter ranges explored by users and extends their ranges beyond the regions they tested. nanoHUB will proactively inform users of interesting discontinuities and local optima. Further, users will be able to explore the parameter spaces studied by the community as a whole, visually identifying areas of interest that have been sparsely explored, and allowing nanoHUB to optimize the detailed study of those areas of interest through uncertainty quantification approaches (Hunt et al., 2015). nanoHUB will enable new incremental publishing mechanisms where an anchor publication about a simulation tool may be automatically augmented by groups of users employing the same tool to study different regions of parameter space as they explore nanodevices under conditions not originally conceived by the anchor publication. The active nexus will allow systems like nanoHUB to do all of these things automatically; defining its place as a participant in the community.

The notion of the active nexus of interaction need not be unique to nanoHUB. Such nexuses are possible in many scientific disciplines, and therefore make it likely that the domain of science, rather than that of the everyday consumer, is where the social medium first will transform into a peer of a different sort. The social medium will become a 'shaker' rather than just a facilitator of the process of scientific discovery. It will participate in its community, fill in gaps, highlight interesting aspects of scientific phenomena, and assist with the dissemination of results within the community at rates not possible with human participation alone.

Acknowledgments Mark S. Lundstrom founded nanoHUB.org in 1998. In 2005, Michael McLennan created the Rappture Toolkit and Rick Kennell wrote the scalable middleware of HUBzero that, respectively, enable and power interactive nanoHUB simulations. The Network for Computational Nanotechnology (NCN) manages nanoHUB.org and has been funded by various NSF Awards Nos. EEC-0228390, EEC-1227110, EEC-0228390, EEC-0634750, OCI-0438246, OCI-0832623 and OCI-0721680.

References

Bostrom, R. P., Gupta, S., & Thomas, D. (2009). A meta-theory for understanding information systems within sociotechnical systems. *Journal of Management Information Systems, 26*(1), 17–48. doi:10.2753/MIS0742-1222260102.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row. ISBN 978-0-06-016253-5.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Reading, MA: Houghton Mifflin.

Hunt, Martin, et al. PUQ: A code for non-intrusive uncertainty propagation in computer simulations. *Computer Physics Communications* (2015).

Kapadia, N. H., Fortes, J. A., & Lundstrom, M. (1997, July). The semiconductor simulation hub: A network-based microelectronics simulation laboratory. In *University Government Industry Microelectronics Symposium* (pp. 72–77). Institute of Electrical & Electronics Engineers.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59–68. doi:10.1016/j.bushor.2009.09.003.

Madhavan, K., Zentner, M., & Klimeck, G. (2013). Learning and research in the cloud. *Nature Nanotechnology, 8*, 786–789. doi:10.1038/nnano.2013.231.

Majchrzak, A., & Markus, M. L. (2013). Technological affordances and constraints in management information systems (MIS). In E. Kessler (Ed.), *Encyclopedia of management theory*. Thousand Oaks, CA: Sage.

McLennan, M., & Kennell, R. (2010). HUBzero: A platform for dissemination and collaboration in computational science and engineering. *Computing in Science & Engineering, 12*(2), 48–53. doi:10.1109/MCSE.2010.41.

Miller, C. C. (2010, October 30). Why Twitter's C.E.O. demoted himself. *New York Times*. Retrieved May 29, 2014, from http://www.nytimes.com/2010/10/31/technology/31ev.html?_r=0

Orlikowski, W. J., & Scott, S. V. (2008). 10 Sociomateriality: Challenging the separation of technology, work and organization. *The Academy of Management Annals, 2*(1), 433–474. doi:10.1080/19416520802211644.

Snowden, C. (2006). Cstng a pwr4l spll: D evolshn f sms. In A. P. Kavoori & N. Arceneaux (Eds.), *The cell phone reader: Essays in social transformation* (pp. 107–124). New York: Peter Lang.

Yoo, Y., Boland, R. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science, 23*(5), 1398–1408. doi:10.1287/orsc.1120.0771.

Ostinato: The Exploration-Automation Cycle of User-Centric, Process-Automated Data-Driven Visual Network Analytics

Jukka Huhtamäki, Martha G. Russell, Neil Rubens, and Kaisa Still

1 Introduction

This chapter introduces the Ostinato Model, an exploration-automation cycle for a user-centric, process-automated, data-driven visual network analytics.

In terms of increasing the transparency of editorial processes on social media, this chapter contributes to the general theme of the book and particularly its second volume at hand in three levels. First, network analysis is a key approach in supporting explorative studies on the patterns and structures in between actors creating, curating, refining, and distributing social media content and in estimating the authority and trust these actors have, therefore allowing for increasing the transparency of the editorial structure of Wikipedia co-authors, discussion and dissemination structures on Twitter and other social media. These structures can be modeled, represented, analyzed and visualized as networks to support the investigations and exploration. Second, the presented data-driven approach allows extending these investigations beyond the boundaries of individual social media and over long periods of time. Third, actors with different sets of skills from means to crawl online sources for data to domain knowledge allowing deep sensemaking can all fully engage into the different phases of the investigative process.

J. Huhtamäki (✉)
Tampere University of Technology, Tampere, Finland
e-mail: jukka.huhtamaki@tut.fi

M.G. Russell
mediaX at Stanford University, Stanford, CA, USA

N. Rubens
University of Electro-Communications, Tokyo, Japan

K. Still
VTT Technical Research Centre of Finland, Espoo, Finland

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_11

These contributions allow the use of visual representations of the structures behind various social media phenomena to improve social interaction, estimations of trust and credibility on social media. With the data-driven approach, the investigators of social media phenomena and patterns of social interaction, trust and credibility are able to move fast in the beginning of the process. As the ways of visualizing and investigating a particular phenomena matures, the investigators may continue to follow the phenomena with the support of close to real-time dashboards adding transparency and supporting e.g. longitudinal investigations. The option for automating the process also supports developing these investigative tools toward end-user products for avid social media content authors and users.

In music, the word "ostinato" refers to both a repeating musical pattern as well as a composition that contains a repeating musical pattern. Like the repeated rhythms and melodies in Ravel's *Bolero* (Fig. 1)—small innovations are explored with each iteration, and some are incorporated into the melodic narrative—we apply the musical concept of "ostinato" to a cycle of user-centric exploration and automation that builds transparency of authorship for evidence-based decision making.

Here, data-driven means that the analysis process relies on data, is automated and conducted in a computational manner, and visual network analytics refers to taking a visual analytics (Heer & Shneiderman, 2012; Thomas & Cook, 2006) approach to network analysis. Additional data can augment the dataset selected for analysis through an automated software process. Established analytical procedures can be automated, yet new conditions for analysis-based insights can be introduced and refined incrementally with continuous computational iterations.

In this implementation of the Ostinato Model, the phenomena under investigation are modeled as a network, and highly interactive visualization tools are used to conduct the investigative process. Network analysis introduces a relationship approach to investigating the structure of many kinds of phenomena. Network analysis allows for exploratory analysis of the social roles of network actors and the phenomena of relationships, as well as for quantifying the structural properties of networks.

A key aspect of the Ostinato Model is the focal point of the user—here, the investigator of particular network-driven phenomena—in the investigative process.

Fig. 1 Ostinato patterns from Bolero's *Ravel* (Mawer, 2000)

This answers to the call for data scientists,[1] somewhat mythical multi-skilled individuals that are capable of individually running the whole investigative process from collecting data to analysis to deep sensemaking in domain of interest, by allowing both experts of the domain under investigation, developers of the technical process as well as e.g. quantitative analysis specialists to possess equal means to take a proactive role in the investigative process. Moreover, the Ostinato Model defines an overall structure for the data-driven investigative process that supports the coordination between the individual phases of the process and therefore allows all the members of the investigative team to contribute to the implementation of different phases of analysis.

Visual network analytics allows the emergence of insights on the structure and dynamics of innovation ecosystems, social media platforms and other networked phenomena. Existing research on networks shows that network analysis has a good fit for explorative analysis of (eco)systems: much is already known about structure in networks (Barabási & Bonabeau, 2003; Granovetter, 1973), the roles of individual actors in the network (Hansen, Shneiderman, & Smith, 2011), the drivers of network evolution (Giuliani & Bell, 2008) as well as the latent structures and dynamics behind the diffusion of information through networks (Leskovec, Backstrom, & Kleinberg, 2009), network control (Liu, Slotine, & Barabási, 2011) and virality (Shakarian, Eyre, & Paulo, 2013; Weng, Menczer, & Ahn, 2013). Transforming those insights into action requires communicating the insights to constituents of change (Russell et al., 2011; Still et al., 2014). Visual network analysis is a promising method for investigating social configurations and for interactively communicating their findings to others (cf. Freeman, 2009).

Data-driven visual network analytics leverages computation to analyze potentially very large datasets in order to identify the patterns driving complex phenomena. Moreno (1953), Freeman (2000, 2009), Hansen et al. (2009, 2011), Russell et al. (2011), Still et al. (2014), Basole et al. (2012), Ritala and Hallikas (2011), and Ritala and Huizingh (2014) give examples of using a network approach to investigate complex phenomena that are driven by sets of interconnected actors. The investigations of such phenomena are further complicated because data about these actors frequently come from multiple and diverse data sources, some of which are not developed for computational use. Especially in cases involving data that are heterogeneous by nature, an iterative, incremental analysis process is sometimes necessary (Telea, 2008). Analysis of complex phenomena often involves multiple pathways to actionable recommendations, and assumptions underlying decisions may change over time.

We agree with Freeman (2000) that integrated tools that can be used to collect, manage and visualize the SNA data are key in supporting network investigations. The tradeoff between usability and automation sometimes creates a barrier for new entrants into data-driven visual network analysis (Hansen et al., 2009). In order to

[1] Ideally, a data scientist is a hacker, scientist, quantitative analyst, trusted adviser and business (domain) expert, all in one person (cf. Davenport, 2014).

provide a low barrier approach to using network analysis to study complex phenomena, we prioritize usability over process automation when possible.

However, a gap exists between the vision and the practice. Manually operated processes used by individual investigators or small investigative teams rely on ready-made tools that are operated through graphical user interfaces. Using these stand-alone tools is very straightforward. The available data sources and analysis and visualization functionalities are, however, somewhat limited. The full-stack, programming-centric processes, in which massive sets of data are mined with tools that are developed and operated by experts, are generally run in complex cloud-based environments. We are aware that several process models, with different levels of abstraction, exist to structure data-driven, visualization-centric investigations; a selection of these models will be covered as part of the description of previous work in the next section.

Many of the existing models are either very general or focus on particular parts of the process. A data-driven visual network analytics approach requires drawing from a number of process models. Using parallel data sources is often not considered in the process models. Moreover, network analysis introduces specific requirements to the process, importantly including the possibility to calculate node metrics as additional data quantifying the different structural roles of the nodes.

Drawing from our experience in running multiple case studies in the context of explorative innovation ecosystem analysis, we take a design science research (Hevner et al., 2004) approach to describe a process model for data-driven visual network analytics. In this book, our chapter contributes to the body of knowledge on computational frameworks, tools and algorithms for supporting transparent authorship in social media knowledge markets by defining an interactive and iterative process model for data-driven visual network analytics to explore relationships in ecosystems. Our process model takes into account requirements stemming from a call for transparent authorship in social media knowledge markets and builds on existing models for data-driven analytics and sensemaking. It is designed to support iterative and incremental investigative processes, as well as to automatically update a visualization dashboard revealing the dynamics and evolving network structure of a phenomenon under investigation.

The rest of the chapter is organized as follows. In second section, we review previous work on which this Ostinato process model is based. The third section introduces the research methodology and a selection of cases we have used to develop the model. The fourth section describes the requirements for the process as well as the different steps that constitute the Ostinato process model (Fig. 2). In the fifth section, we discuss how this model satisfies these criteria and adapts to the exploration–automation cycle. The sixth section concludes the chapter and describes key implications and ideas for future work.

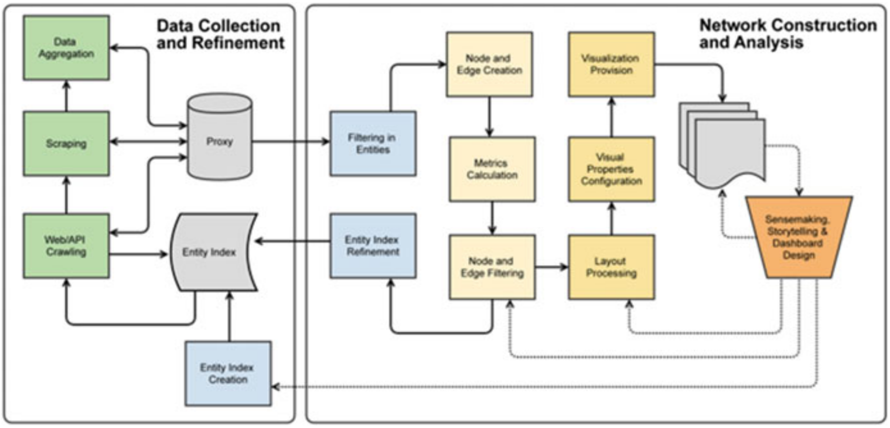


Fig. 2 Ostinato model—user-centric data-driven process model for visual network analytics

2 Previous Work

Our approach into data-driven visual network analytics builds on a number of bodies of knowledge (Fig. 3), including traditional SNA (Wasserman & Faust, 1994), information visualization (Card, Mackinlay, & Shneiderman, 1999), data-driven visualization pipelines (Nykänen et al., 2008), interactive network analysis (Hansen et al., 2009), visual analytics (Thomas & Cook, 2006), sensemaking (Pirolli & Card, 2005), interactive visualization (Heer & Shneiderman, 2012) and scientific visualization (Telea, 2008). All these fields offer models and approaches, and additionally they pose key requirements to be considered when developing next-generation analytics tools for very large networks. The objective to conduct (and publish) research in a reproducible way (Ghosh, 2013; Peng, 2009) contributes to the quality of the process and also introduces additional requirements.

Traditional SNA (Wasserman & Faust, 1994) introduces a set of node and network level metrics that can be used to describe the structural properties of networks and to quantify the various social roles of network actors. To support the use of network analysis for novices, Hansen et al. (2009) introduce the Network Analysis and Visualization (NAV) process model that builds on top of the general sensemaking model. The NAV process starts with defining the goals for the analysis and continues through data collection and structuring, after which data are interpreted through multiple loops of network visualization and SNA metrics calculation. Finally, the insights and conclusions are formatted and summarized, then disseminated through a report. Seeking low-barrier entry, the authors introduce NodeXL, an Excel-based toolset for SNA, to conduct the analysis. Among others, Hansen et al. (2011) define ways to apply these metrics in investigating phenomena taking place in social media.

The information visualization reference model (Card et al., 1999) presents a four-step process that can be used as a blueprint for implementing data-driven visualization processes. Raw data is (1) first collected and then (2) refined into data tables to

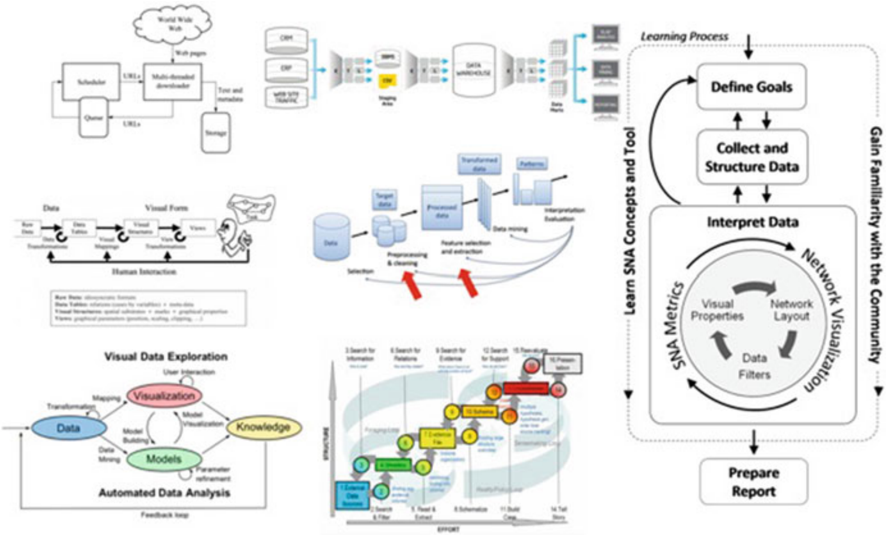


Fig. 3 Process models related to data-driven visual network analytics. The six small diagrams, from *top-left*: Web Crawling (Wikipedia, 2014), extract-transform-load (Intel, 2013), information visualization reference model (Card et al., 1999), knowledge extraction from databases (Indarto, 2013), visual analytics (Keim, Kohlhammer, & Ellis, 2010), sensemaking (Pirolli & Card, 2005). On the *right*: Network Analysis and Visualization (NAV) model (Hansen et al., 2009)

allow straightforward processing. Data tables are then (3) transformed into a portfolio of visual representations from which various concrete views are (4) served to the visualization user for sensemaking. Importantly, the reference model suggests that best practice is when the user can interact with all steps of the process (Fig. 3).

Component-based data-processing pipelines, a technical application of the information visualization reference model, introduce a viable approach for developing reusable pieces of software to support the automation of processes related to social network analysis across application domains (Huhtamäki et al., 2010; Nykänen et al., 2008). To support investigating the social structure among wiki co-creators, Huhtamäki et al. (2010) present a set of components and a process model to orchestrate the use of the components. A key benefit of the component-based approach presented by Nykänen et al. (2008) is the possibility to integrate existing software tools implemented in different technologies into the data-processing pipeline, given that they can be operated from the command line. The main restriction of the approach is the need to implement the automation through scripting, i.e. writing program code that describes rules for a particular functionality.

The general sensemaking model by Pirolli and Card (2005) divides the sensemaking process into two loops, the foraging loop and the sensemaking loop. To simplify, data is first collected and refined and then transformed into various visualizations and other representations that support the sensemaking. The process is iterated as many times as required. Similarly, the process of visual analytics "typically progresses in an iterative process of view creation, exploration, and refinement" (Heer & Shneiderman, 2012).

The sensemaking step can be applied in different ways—from purely manual processes in which humans interact with various user interfaces to conduct the analysis to automated information systems in which data are collected and processed in runtime. Sensemaking also includes the process of visual analytics (Thomas & Cook, 2006) that, by default, relies on the availability of software and tools supporting the users. Heer and Shneiderman (2012) give an insightful overview to the specific functionalities that users should be able to operate: (1) specify data and views; (2) manipulate views; and (3) process and provenance their findings.

Peng (2009) builds his definition of reproducible research on three categories: a piece of research is fully reproducible when both the data and code used to are available and, moreover, the code is executable by anyone. As Ghosh (2013) shows, reproducibility can be approached at many different levels, from policy to detailed technological solutions.

3 Methodology

In this section, we briefly describe the context in which the data-driven network analytics takes place. This illustrates the explanatory power and novelty introduced by the Ostinato Model to network analysis workflows. Further, we discuss the use of Design Science Research (Vaishnavi & Kuechler, 2007) as a method that we apply in our venture to develop the Ostinato Model in a way that is both credible in terms of scientific theory as well as practical utility. We also refer to a selection of case studies we have used to develop and validate the process model presented in this chapter. This shows that the Ostinato Model is a general approach to conducting data-driven network analysis investigations that has already been applied extensively in a series of real life experiments and investigations.

Addressing innovation ecosystems as networks allows scholars and practitioners to study their complexity, providing a means for mapping, monitoring and managing the ecosystem components. To do this, we have taken a data-driven network analysis approach to study innovation ecosystems in regional, metropolitan, national and international level as well as e.g. in the context of programmatic activities supporting innovation and growth. We have followed a design science research approach that is based on iteration through construction of network visualizations as artifacts. We have used a number of different datasets in these studies, including social media, socially constructed data available online, and proprietary sets of data represented as spreadsheets and other formats.

3.1 Context

The research that led to the development of this process model for data-driven network analysis began in the context of studying complex networks of relationships in innovation ecosystems. Russell et al. (2011) use the concept of innovation ecosystem to refer to the inter-organizational, political, economic, environmental, and technological systems through which a milieu conducive to business growth is catalyzed, sustained, and supported. A dynamic innovation ecosystem is characterized by a continual realignment of synergistic relationships that promote growth of the system (Russell et al., 2015).

Ecosystems are a complex phenomenon, with multiple entities connected through multiple level relationships, as well as multiple stakeholder perspectives into those relationships. Ecosystems that promote innovation have become a quest for companies, cities, regions and countries. It is agreed that "relationships shape the behavior and outcome of all stakeholders as well as the system-level effects" (Hwang & Horowitt, 2012), and that it is through the relationships of individuals within and across organizations in an ecosystem that knowledge transfer, technology dissemination and organizational change are accomplished (Russell et al., 2015). Program managers and policy analysts in charge of transforming innovation ecosystems seek to define and describe innovation ecosystems in order to set goals, determine interventions and evaluate change, and visualizing the innovation ecosystem has proven instrumental to strategy setting and decision-making (Still et al., 2014). By making the roles and relationships explicit, both numbers and visualizations can be used to support the creation and management of innovation ecosystems. By tracking the provenance of data and authorship of analytical refinements, the collaborative exploration gains transparency.

To manage as well as to create innovation ecosystems, network orchestration has been encouraged (Paquin & Howard-Grenville, 2013; Ritala, Armila, & Blomqvist, 2009; Still et al., 2014). A data-driven process for understanding roles allows for interactive discovery of the innovation ecosystem. Multiple perspectives can be invited and exchanged in the process of developing and orchestrating transformation programs. With subsequent automation of data updates and tracking analyses, the assumptions and contingencies underlying decisions can be monitored for changes that would impact policy and program directions.

3.2 Design Science Research

In this research, we take a Design Science Research (DSR) approach to describe a process model for data-driven visual network analytics applicable in replicable investigations of innovation ecosystems as well as other domains in which network structures over time are of interest. DSR is a research method that allows "learning and investigation through artifact construction" (Vaishnavi & Kuechler, 2007, p. 187). "Whereas natural sciences and social sciences try to understand reality, design science attempts to create things that serve human purposes" (Simon, 1969). The rationale for DSR hails from the importance of the practical utility of research (Peffers et al., 2007). Design science research aims to build a bridge between information system (IS) research and its practical application by producing results that have real-life relevance. "Design science [. . .] creates and evaluates IT artifacts intended to solve identified organizational problems" (Hevner et al., 2004).

The General Design Cycle (GDC) is a key part of any DSR process (Vaishnavi & Kuechler, 2007). The process begins from the awareness of the problem and continues to one or more suggestions for solution. Next, an implementation of the plan is developed and evaluated, and finally, the process is concluded and the results shared; in the case of a scientific process, they are published. In each of these steps, new knowledge is both created and fed back to previous phases. The phases are repeated in an iterative fashion until a satisfactory end result (one that has practical utility) is achieved.

Readers with experience in software development will notice a straightforward connection between general design cycle and agile software development (see e.g. Schwaber & Beedle, 2001). Apart from the intent to publish the results, both design and development processes move forward in an iterative and incremental fashion and are guided by feedback collected from the users and other stakeholders of the developed software or other artifact, here the process model.

To develop the Ostinato Model for data-driven visual analytics presented in this chapter, we effectively applied and repeated the General Design Cycle. To evaluate the process model for added credibility of the presented results, we applied the Experimentation Pattern defined by Vaishnavi and Kuechler (2007), more specifically the case-based prototype development pattern on which the prototype is developed in an incremental, iterative manner over a number of cases, leading to deep knowledge of the problem and the proposed solution.

3.3 Experimental Cases

The Ostinato Model has been developed over a number of cases in which a variety of innovation ecosystems have been investigated in collaboration with their stakeholders using various sets of data sources (Basole et al., 2012; Jussila et al., 2014; Rubens et al., 2011; Russell et al., 2015). Table 1 describes core cases in which the automation-exploration cycle was implemented, using structured and semi-structured data sources, involving stakeholders in the exploration process as well as the sensemaking of key visualizations and other outputs for each case.

4 Ostinato Model

This section presents summary of the results of our research. First, we describe the requirements for the data-driven visual network analytics process; these requirements stem from existing process models and are augmented through results that emerged in case studies on which we applied the method. Second, as the core contribution of this chapter, we describe the process model for the exploration–automation cycle of data-driven visual analytics, the Ostinato Model.

Table 1 Illustrative cases for developing the exploration-automation cycle and the process model for data-driven visual network analytics

Case	Data	Co-creators/case shareholders	Visualizations/outputs
Demola (Huhtamäki et al., 2013)	Proprietary data on Demola projects, the companies that initiated the project and university affiliations of project members (university students)	Demola leaders and operators and the investigative team	The animation of the evolution of Demola project sphere including projects, the affiliations of project team members and companies. Multimode networks on (1) projects and affiliated actors and (2) projects and their key competences
Tekes Young Innovative Companies (Huhtamäki et al., 2012)	Innovation Ecosystem Network Dataset on growth companies, Twitter data on Tekes Young Innovative Companies (YIC) and their followers	Policy makers at Tekes—the Finnish Funding Agency for Innovation and the investigative team	One and two-step networks of the companies part of Tekes YIC program and their affiliations to investors and key individuals
Finnish Innovation Ecosystem (Still et al., 2013)	Three separate datasets: (1) Thomson Reuters SDC for deals and alliances and IEN Dataset for (2) Executives and Finance and (3) Startups and Angels	Finnish national-level policy makers and the investigative team	Network visualizations and metrics about companies having their main office in Finland and their first-step connections to other companies, investors and key individuals
Network orchestration for EIT ICT Labs (Still et al., 2014)	IEN Dataset for Executives and Finance	EIT ICT Labs representatives and the investigative team	Network visualizations of companies having their main office in one of the EIT ICT Labs co-location centers and their first-step connections to investors and individuals as well as to other companies through investments and acquisitions

4.1 Process Requirements

Developed through several rounds of iterations following the General Design Cycle, the core guidelines and requirements for the data-driven visual network analytics process model include the following: continuous data collection; exploration; transparency; loose coupling; reproducibility; automation; enabling manual steps; low entry barrier; and interoperability. Each is described.

Continuous data collection. When collecting data from social media, persistent processes are often needed, particularly when the investigators want to capture both the structure and dynamics of a phenomenon. Twitter, for example, currently provides only limited access to its historical data, and even then data on followers and friend connections between users do not include timestamps. At times, collecting the data takes days or weeks or "forever" to complete, due to throttling or other technical limitation or the sheer size or the dynamic nature of source data.

Exploration. A visual analytics approach is key to enable users with varied technical skills to collaboratively explore and make sense of a phenomenon. Being able to follow the visual analytics approach requires process flexibility. That is, all the stakeholders of the analysis process should be able to conduct any of the individual steps by themselves even though development of the overall process requires technical development skills.

Transparency. Developers with technical skills may select to manage the network analysis data, in its different phases, with a database. To accomplish transparency and flexibility in the process, other members of the investigative team may, however, need less technical means to access the data. The use of intermediary results is key in facilitating the transparency and flexibility of the process. Intermediary results refer to data in between the individual steps of the analysis. These data should be available as files in widely used formats, such as CSV and GEXF. In addition to the enhanced transparency, these intermediary results allow for speeding up the analysis process by using cached versions of source data and intermediary results when they have not changed.

Loose coupling. At best, data-processing pipelines can be built with a range of tools and components that have been implemented with different technologies. This kind of flexibility allows the introduction and use of new expressive tools from individual software components to full-featured applications as they become available to the investigative team. Many of them introduce new opportunities for advancing the analysis process but generally it is not possible to integrate these tools to a data-processing framework in program code (API) level.

Reproducibility. In the data-driven visual network analytics approach, reproducibility is first and foremost a technical quality of the process: the investigative team should be able to repeat the study or one or more steps of the analysis process and reproduce the results. Reasons for the need to rerun the process include, among others, updates on the source data, development steps of the analysis process, and the introduction of completely new processing steps and tools that insist on the use of a particular data format or extending the existing data. Moreover, dynamic sensemaking for complex phenomena mandates being able to refresh the data and derive new results with updated data. Reproducibility at this technical level also allows the investigative team to release the process, data and results to other researchers interested in the phenomena under investigation.

Enabling manual steps. While reproducibility is important, at the same time it is important to realize that automating some of the steps may not be feasible when an analysis is conducted the first time or requires intensive tailoring. Therefore, the process should support implementing any of the process steps manually. The use of file-based intermediary results is a practical approach in enabling manual analysis steps.

Automation. Allowing the development of automatically updating dashboards as needed gives the investigative team the opportunity to continue observing particular phenomena over time. It is expected that production-ready analysis processes for dashboards will operate without supervision; however, in the context of exploratory research, some requirements may be relaxed.

Low entry barrier. Analysis of innovation ecosystems and other network-based investigations of complex phenomena require extensive domain knowledge, and hence insist on active participation from domain experts (often without extensive technical expertise) throughout the analysis process. This requirement further underlines the need for transparency of the analysis process and the individual analysis steps.

Interoperability. The investigative team should be able to use a number of existing analytics tools with high usability and rich interactivity such as Gephi, NodeXL, KNIME and Tableau for conducting the analysis. Moreover, provisioning the visualized networks and other outputs of the analysis should be possible through dashboard built with Web technologies such as D3.js, DC.js, GEXF.js and the like.

In terms of the General Development Cycle, these requirements can be used to describe the Definition of the problem that serves as the starting point of artifact development (cf. Vaishnavi & Kuechler, 2007). These requirements form a design rationale for the Ostinato exploration–automation cycles of the process model for data-driven network analysis.

4.2 Process Model

The Ostinato process model that is presented in this section is developed over multiple case studies with a design research approach. It is built on existing models and previous work, and it takes into account the process requirements presented in Sect. 4.1. Each step is described. Figure 2 shows a diagram of the process model.

Phase 1: Data Collection and Refinement

1. Entity Index Creation
2. Web/API Crawling
3. Scraping
4. Data Aggregation

Phase 2: Network construction and visualization

5. Filtering in Entities
6. Node and Edge creation
7. Metrics Calculation
8. Node and Edge Filtering
9. Entity Index refinement
10. Layout Processing
11. Visual Properties Configuration
12. Visualization Provision
13. Sensemaking, Storytelling & Dashboard Design

4.2.1 Phase 1: Data Collection and Refinement

The general rules of data-driven analytics apply here: collecting and cleaning the data will in most cases consume most of the time and resources available for the investigation.

Entity Index Creation

In some cases, the source data can be collected in full; whereas, in other cases only data on entities that are relevant for the analysis need to be collected. In one use case, we were interested in the Twitter discussions taking place in relation to a conference, #cmadfi. We collected all the Tweets sent by conference participants before, during and after the event in order to create a network representing the social structure of the conversation. For this, we created an entity index including the Twitter handles of conference participants, as well as those mentioned in the discussion (Jussila et al., 2014).

In the context of innovation ecosystem studies, the entities for which we collected data were defined by boundary specification (Basole et al., 2012). For example, in investigating the connections between companies taking part in Young Innovative Companies program[2] run by the Finnish Funding Agency for Innovation Tekes, the list of companies defined the starting point of the analysis (Huhtamäki et al., 2012).

Web/API Crawling

Collecting the data is the most heterogeneous step in the data-driven visual analytics process. Possible source data potentially includes everything digital, from proprietary offline documents and document collections to spreadsheets to Web

[2] Funding for young innovative companies, http://www.tekes.fi/en/funding/companies/funding-for-young-innovative-growth-companies/

APIs (Application Programming Interface) to Web sites that are designed primarily for human interaction.

Similarly, the functionality required to collect the source data can range from relatively simple reading of individual documents to functions similar to a fully featured Web crawler. Compared to crawling random websites, Web APIs are, by default, more straightforward for data collection as they are often designed to support reuse (Vinoski, 2008). At best, source data is available as linked data (Bizer, Heath, & Berners-Lee, 2009), i.e. data that has a clear structure with individual facts that can be interconnected with the help of unique identifiers. This is key in ensuring referential integrity.

At the end of the crawling phase, a set of web resources, or rather their representations in Hypertext Markup Language (HTML) or some other format, is made available in a local database or other storage, a proxy that significantly speeds up the subsequent processing steps.

Scraping

Once the raw source data is available locally, the next step is to filter, select and distill the utility data relevant to the analysis process. Scraping refers to the process of distilling data from documents that are published to the Web for humans to use. Scraping can be seen as a form of the Extract, Transform, Load (ETL) process that is often applied in the context of data warehousing or other business intelligence processes to collect data from different sources to be refined and normalized and finally loaded into a consistent database for later use (Petschulat, 2010; Vassiliadis, 2009).

When collecting data from Wikipedia on Finnish Young Innovative Companies (YIC), for example, we were particularly interested in the facts presented in the Infobox section[3] of the page. To collect this data, we took advantage of the HTML markup on the page to specify the semantics (meaning) of the different pieces of text.[4] Each of the facts is represented as a table row including two cells, the first of which includes the label specifying the type of the fact and the second includes the actual value. Moreover, the value is also represented as a link to a separate page, a fact that we included in the crawl.

[3] Help:Infobox, http://en.wikipedia.org/wiki/Help:Infobox

[4] The Terms of Service for a Web page must also be considered. When using Wikipedia as a data source, for example, one has to take into account the Terms of Service that specifically deny crawling Wikipedia for large amount of files. Instead of crawling the live website, users of the data are advised to download a copy of Wikipedia's contents and set up a proxy for serving further processing.

Data Aggregation

Social media studies often take place within the boundaries of an individual social media service; and therefore, ways of accessing data and identifying individual entities can be straightforward when one source of data is used. The complex context of innovation ecosystem studies, however, led us to use several sets of data in parallel. This meant that in many, if not most, of the cases, linked data was not readily available; and therefore, links between individual sets of data had to be created through finding unique entity identifiers that allow referential integrity. In innovation ecosystem studies, the name of the company or another actor is sometimes the key data point that can be used to identify an entity; in other cases, more advanced entity recognition procedures can be applied.[5] This kind of data cleaning is sometimes referred to as data wrangling (Kandel et al., 2011). Applying the methods of entity recognition provides a potentially more general solution to creating unique identifiers for entities in the data.

4.2.2 Phase 2: Network Construction and Analysis

Once the data is available on a local proxy, the utility data has been extracted from the source documents and data from different sources has been aggregated into a consistent set of linked data, the construction of the network representation of the phenomena under investigation can begin.

Filtering in Entities

The network construction phase starts by selecting the entities that will be included in the network. The selection of nodes is guided by the boundary specification designed and defined by the investigative team. At least two approaches exist to implement the selection: starting from a list of entities and rule-based entity inclusion. To continue the Finnish YIC example, we started from the list of companies participating in the program. We scraped Wikipedia data on the connections between the YIC companies and key individuals running them. If data on the individuals was not available in a clean format, we followed the crawling pattern by including the individuals in the list of web resources to be crawled. We continued to complement the dataset with data from the Innovation Ecosystems Network Dataset (IEN Startups and Angels, IEN Executives and Growth) and other sources of data about investments, acquisitions and affiliations.

[5] When using names as identifiers, one can apply fuzzy string matching and semi-automated tools such as OpenRefine (http://openrefine.org/) or DataWrangler (http://vis.stanford.edu/wrangler/) to assist in the aggregation process.

A key reason to separate the selection of entities from node and edge construction is to support the transparency, reproducibility and extensibility of the process. To create a shared understanding of the analytical results, it is absolutely vital that all the investigators taking part in a particular network study are able to understand the original raw data, in addition to any constructed variables, and the various analytics and metrics that represent the network; this means that investigation participants need access to the analysis process as a whole, including access to the raw data. In our experience, we found that answering specific questions raised by anyone interested in the study, drawing conclusions, generalizing the results, developing more specific and potentially more interesting questions all depend on transparency of the data available and used for the analysis.

Node and Edge Creation

A key part of the data-driven network analysis process is, of course, the actual creation of the network. Network creation boils down to the creation of nodes representing the actors and the creation of edges representing the connections between the actors. Several options are available, however, when specifying details of the network creation process. First, the network can be either one-mode or two-mode. In one-mode networks all the nodes are of same type: startup companies, for example. Connections between the nodes are formed through relationships: investments, affiliations to individuals, acquisitions and transactions. In two-mode networks, there are two types of nodes, for example, startup companies and individuals related to them. Hypergraphs and bipartite graphs are examples of means to visualize two-mode networks (Freeman, 2009; Jesus, Schwartz, & Lehmann, 2009).

Further, the connections between network nodes can be either valued or dichotomous. With valued connections, the strength of a connection can be expressed. In either case, the connections may be undirected or directed. Finally, the temporal dimension can be included in networks if the data used to create the connections is time-stamped. With temporal data, insights about the evolution of the network can be gained.

Metrics Calculation

Network metrics enable quantifying a variety of structural properties, both in network and node level. These range from simple metrics such as node degree (indegree, outdegree) and betweenness to hub and authority values with HITS and other more sophisticated measures. Whereas in principle, every metric can be calculated for all of the networks and their nodes, in practice this is not feasible due to reasons of efficiency. Moreover, new metrics are being developed continually, and the investigative team is likely to find—or develop—new metrics that fulfill specific investigative purposes. From an implementation viewpoint, it is unlikely to find one tool that supports all the metrics the team wishes to use. Therefore, a combination of tools may be required to calculate the metrics.

As part of this step, network metrics for the network representation should be archived for later usage. For transparency, a list of exported network nodes and edges should include the various metrics used. In practice, node and network metrics must be recalculated after each change in the network structure; however, reference to previous calculations is often needed.

Nodes and Edge Filtering

A key limitation in visual network analysis is the amount of space available, both on screen and particularly on paper, to present the visualization. Depending on the level of detail required in the analysis, hundreds or thousands of nodes can be presented in one visualization view. For networks of tens of thousands of nodes and more, only more general structures and patterns can be observed from the visualization. Two means exist to address this limitation: the best option is to allow the visualization users to filter in and out nodes and edges. If the end-user tools used to present the visualizations do not allow filtering, it can be done as one part of the automated process. Often, reducing the size of the visualized network is accomplished with a combination of filtering out edges that have the least amount of weight as well as filtering out nodes that: (1) are left without edges; (2) have a value of the degree or some other a network analysis metric under a specified threshold; or (3) are (not) of particular type (even though this can already be taken into account when filtering in the entities used to construct the network in the first place).

Entity Index Refinement

At this stage, the network is constructed and the required metrics are calculated for each of the nodes. Depending on the boundary specification applied in a particular investigation, the network is either ready to be visualized or, alternatively, additional data can be collected to complement the network. Revisiting the Finnish Young Innovative Companies case, the boundary specification was designed to include all the individuals involved in one or more of the companies in YIC program as well as all the other companies the individuals are or have been affiliated with. Moreover, the data included all the investors that had invested into any of the companies as well as all the companies that had acquired any of the YIC companies.

Layout Processing

The principle of processing network layout is simple. Nodes are given a position in two-dimensional space in a way that network structure is revealed in an intuitive way. Despite the simplicity, novel layout algorithms have continued to be developed over several decades. In our research cases, various stakeholders found a specific implementation of force driven layout, Force Atlas, to be particularly suitable for laying out networks representing innovation ecosystems at different levels. Force Atlas is implemented in Gephi and can be used as a batch process with the help of Gephi Toolkit.[6] In practice, the parameters of the layout algorithm must be adjusted manually for a particular kind of a network before fully automating layout processing. Alternatively, the layout can be processed with the UI version of Gephi and the resulting network, including the XY-coordinates for each node, can be exported, e.g. in GEXF.

Storing the network layout data is particularly important for improving the efficiency of the layout process, as well as for reducing investigators' cognitive load and promoting transparency. In particular, it is important that after the data is refreshed, the investigators are able to find the pre-existing nodes in an area of the network where the nodes were previously located. This stability can be achieved by inserting the existing positions into the network data before re-running the force driven layout algorithm. In most cases, investigators will find the pre-existing nodes close to the initial area of the network.

Future work is needed to determine how features such as layout algorithms, e.g., those implemented into NodeXL, could be used as a component of data-driven visual network analysis pipelines.

Visual Properties Configuration

In networks, there are limited selection possibilities when defining the visual appearance of nodes and edges. Nodes have size, color and perhaps a border and shape as elected visual features. Edges have color and width. Allowing the user to select and change the visual properties according to node metrics and other node properties is perhaps the easiest way to allow end user interactivity in network analysis. Depending on the tools used by the investigators to conduct the analysis, the visual properties of nodes and edges can continue to be tweaked as part of the interactive analysis process.

Visualization Provision

At this stage, a network has all the required information available and therefore can be visualized. The means to finalize this step depend greatly on the tools that have been selected for use by the investigative team. In most cases, however, the created network is serialized into a file following a selected vocabulary or format for

[6] Gephi Toolkit, http://gephi.github.io/toolkit/

representing a network. These vocabularies and formats range from different CSV based applications to XML-based languages designed for representing networks.

A minimum approach to provision the network visualizations is to export network data in GEXF or other suitable format and place the resulting file into a folder from where a library such as Gexf.js can access it. More generally, viewer composition scenarios can include the following:

Scenario 1. Network viewer component with fixed functionality, i.e. following a fully descriptive approach. Visual properties such as node size and color need to be defined into the data during its processing. Gexf.js is an example of such a component that we have found useful in adding value to a fully static PDF-based approach in disseminating network visualizations.

Scenario 2. Implementing a dashboard with Web technologies, more specifically frameworks such as Highcharts, D3.js, Crossfilter.js, DC.js and others. In this case, tailored interactive features for data exploration can be provided to the user, adding options for representing network data.

Scenario 3. Using full-feature explorative analytics tools such as Gephi, NodeXL and Tableau, which can be used to further process the data and to connect source data to visual properties of the visualization. The key here is to produce visualizations rich-enough in data that the analyst can fully utilize the critical properties of the chosen analytics tool for investigation and exploration. In Gephi, for example, it is useful to include attribute data for nodes to assist network filtering in a way the investigator desires to do.

Sensemaking, Storytelling and Dashboard Design

While information visualization includes data transformation, representation, and interaction, it is ultimately about harnessing human visual perception capabilities to help identify trends, patterns, and outliers. Sensemaking has its roots in cognitive psychology and many different models have been developed. Sensemaking procedures are cyclic and interactive, involving both discovery and creation (North, 2006). During the data collection and refinement phase, an individual searches for representations. In the network generation phase these representations are instantiated, and based in these insights the representation may be shifted, to begin the process again. Sensemaking is closely linked to the insight objectives (Konno, Nonaka, & Ogilvy, 2014), and the Ostinato cycle of exploration–automation is key in achieving actionable insights that network orchestrators can utilize.

When sensemaking requirements are satisfied for investigators and users, steps of the Ostinato process can be formalized with automated procedures for iteration over time. Key actors, relationships and events of the network can be incorporated into dashboards that will track changes in critical assumptions and into stories that will share vision for actionable change.

5 Discussion

The present chapter adds a new perspective on the heuristic and application development process that may lead to new tools, applications, services, and algorithms dedicated to understanding how social media content is created, curated and disseminated and how the authority and trust of social media content creators accrues and how this matters in terms of trust and credibility. The Ostinato Model contributes to this call in two levels. First, it can be applied to support the data-driven investigations of innovation ecosystem structure and dynamics. Moreover, in the context of our investigations, social media serves first and foremost as a source of data that is fed into the investigations of innovation ecosystems and the structure between their actors. Therefore, second, for validity and reliability of these investigations, it is key to be able to increase the transparency of the processes behind these data originating from social media.

The Ostinato Model contributes to the data-driven network investigations of social media, innovation ecosystems and other network-driven phenomena in several ways. First, the network approach has great strength in supporting the explorative studies of the patterns in between actors creating, curating and disseminating social media content. Second, referring specifically to the first phase of the Ostinato Model, data-driven approach allows tracking down processes over the boundaries of individual social media platforms and services. Third, user-centricity of the data-driven process adds to the transparency of the process itself, therefore providing means to triangulate different phases of data refinement and transformation and allowing different stakeholders of investigations to take as proactive role as they wish in moving forward a particular investigative process.

Due to the continued and rising interest in social media analytics and general big data analysis, new tools are continually introduced to support investigative work. Despite the tool development, a combination of tools is likely to continue to provide more flexibility in accessing and aggregating data and in processing and analyzing it. Finding a balance between user interface-operated low barrier tools and expressive computational strategies that require technical knowledge is key in making the investigative process as productive as possible while maintaining transparency and process flexibility.

This Ostinato Model for user-centric, process-automated, data-driven visual network analytics meets many of the requirements outlined earlier in this chapter for the exploration–automation cycle recommended for developing shared understanding.

Setting up persistent data-collecting routines requires, in general, a programmatic implementation and must be designed and implemented case by case. To maintain the transparency of the process, it is important that the investigators are able to access both the raw data as well as to track down the various steps used to derive the data that is eventually used for the analysis and visualizations.

Allowing exploration boils down to the selection of the end user tools available for investigators to visualize and explore the data. If a rather static tool such as Gexf.js, for example, is used, the user is limited to browsing and searching the data.

If importing the data into an exploration platform such as Gephi or NodeXL is permitted, it is possible to provide the user with node and edge data, enabling them to continue their explorations with more technical independence. The availability of particularly expressive visual analytics tools, such as Tableau, adds to investigation options of analyzing network data, either as a network or using node and edge level data to provide new inspirations for other kinds of data analyses.

Using files rather than databases for representing intermediary results supports both loose coupling and transparency of the process. It also allows for implementing some of the steps manually, if seen feasible, and the flexibility of the process in general is increased.

Reproducibility is both a technical and a policy requirement. For an investigative team revisiting or extending an existing case, the availability of runnable code, source data and intermediary results provides a fruitful starting point. Moreover, results of reproducible studies can be published in a way that both data and runnable code are available, allowing a solid foundation for others to add their contributions as well. A reasonable proposition is that such a piece of knowledge draws attention from other researches and therefore has true potential for impact.

Automation is a key requirement for reproducibility, as well as for creating a dashboard that continues to update visualizations of the phenomena under investigation, sometimes in close to real time.[7]

Low entry barrier is enabled through making intermediary results available to all the members of the investigative team. As the process is repeatable and its individual steps are automated, new projections of the data can be implemented in an iterative and incremental manner. Implementing completely new steps of analysis becomes possible even without technical skills. Automating the steps, however, requires developers' attention. The Ostinato process model requires a multidisciplinary data science team or the somewhat mystical multi-skilled data scientist (cf. Davenport, 2014) to conduct the investigation.

Interoperability can be built into a computational approach. This requires that the technical architecture is flexible enough to permit different software components and tools—that may be implemented with different technologies—to be introduced into the process. When an analysis pipeline is built completely from scratch, it is recognizably important to minimize the number of technologies used. However, moving fast and in an agile manner is an objective we claim can be achieved when existing tools can be integrated to implement the individual steps of the analysis process and to provide the visualizations to investigators and other end users.

An implementation of the Ostinato user-centric, process-automated model for data-driven visual network analytics can serve as the core engine of an investigation. It can also be used to develop a pre-processing pipeline that collects and

[7] Using a full stack programming language such as Python gives the developers more opportunities to turn the scripts developed for analysis into processes that run in the cloud, intermittently collecting and preprocessing the data and feeding results into dashboards implemented in Web technologies.

refines the data, creates a network representation and serializes the outputs to be analyzed and processed with expressive tools that, standing alone, allow the full visual analytics cycle for users.

6 Summary

In this chapter, we have presented the Ostinato Model of the exploration—automation cycle user-centric for data-driven visual network analytics. This model has two main phases, data collection and network analysis; they iterate through a cycle of exploration and automation. The Data Collection and Refinement step is divided into Entity Index Creation, Web/API Crawling, Scraping, and Data Aggregation. The Network Creation and Analysis step is composed of Filtering in Entities, Node and Edge Creation, Metrics Calculation, Node and Edge Filtering, Entity Index Refinement, Layout Processing, and Visual Properties Configuration. As a final step, the visualizations are provisioned to investigators and other end users with interactive exploration tools and discussion, and their feedback activates an iteration of the process. This Ostinato process model allows both an exploratory approach during the early phases of the investigation as well as the automation of the data collection and analysis process. The iteration cycle is especially beneficial in working with multi-source datasets, complex phenomena, changing externalities that may impact assumptions for decisions, and establishing a dashboard for continued observation of the phenomena over time, perhaps in real time.

A key challenge of this approach concerns the number of options for investigators and other end users to interact with the data in real-time while conducting the analysis, particularly the non-technical investigators on a multi-disciplinary team. The design research approach favors an iterative approach for both data-driven explorations and evidence-based decision making. However, investigators with limited programming skills or related technical know-how are limited in their participation, even though they may possess vital domain intelligence. Through access to data, documentation of changes in the analytical approach, flexible means to produce network representations in various formats, and exposition of intermediary results, barriers to participation can be lowered. The cycle of exploratory visual analytics, confirmation of data selection rules and analytical results made accessible through high interactivity visual analytics, allows the investigative team to confirm assumptions and investigative procedures, identify aspects of the analysis that can be automated and establish a transparent, replicable process.

The Ostinato process model has several implications for investigative teams taking the data-driven visual network analytics approach.

First, facilitation and documentation of the investigative process are required. Low barrier for entry in exploration and analysis poses risks that increase without transparency. Put another way, with added transparency and through intermediate results and easy access, the risk of false conclusions is lowered. Co-ordinated discussion on raw data and its journey to the finalized visualizations and other results is imperative; documentation of assumptions and rationale for changing data selection or analytical procedures enables transparency. Facilitation also helps in creating literacy of the processes and its outputs within the investigative team. Having the intermediate results available, all the members of the investigative team are able to maintain more of the control of the process and continue to introduce new, novel ways of analyzing the data as their skills and methodological know-how allows.

Second, the cycle of exploration–automation introduces new requirements for governance. Intermediary results require transparent authorship in their provenance. The transparent authorship of new datasets, constructed variables and analytical iterations must be ensured.

Third, starting from exploration and moving toward automation is straightforward with the help the process model. The investigative team is able to move fast in the beginning of the process while, at the same time, maintaining control over the process as its complexity increases. With appropriate technology selection, the process can eventually be relegated to the background to collect, process, analyze and visualize data in an automated manner to support a longitudinal study of a particular phenomena. And, more importantly, a mature procedure—or one or more of its components—can be reused to investigate other phenomena of interest.

Fourth, increased reproducibility is an asset for future studies but requires explicit governance. Technical reproducibility of the process allows revisiting analytical results of a case even after a long time period. Refreshing (collecting new) data or, alternatively, adding new dimensions into existing data is straightforward when the process or its individual parts can be run computationally. Curational rules must be developed, and access to code and data has to be designed at both the technical and policy levels. Governance of the data from raw to intermediate results to outputs as well as the components and software process must be articulated.

Within the constraints imposed by the level of abstraction in this article, this Ostinato process model provides blueprints for designing analytical processes with technologies ranging from Python to R to Javascript. At best, the process is able to support the inclusion of several different technologies, as implemented e.g. by the Wille Visualisation System (Nykänen et al., 2008).

Future work includes, first, the refinement of this model on basis of the feedback collected from researchers and practitioners working with the exploration–automation cycle of data-driven visual network analytics and applying the model and, second, the implementation of a software framework—perhaps similar to Grunt (http://gruntjs.com/), a popular Javascript-based task runner—to support the development of processes of data-driven visual network analytics on very large datasets.

As an ecosystem of tools and components develops and requirements for interoperability are articulated, we see the possibility of developing a community of people moving the field forward. They will need a package management framework, system components and a supportive community.

The Kredible.net initiative is an important step toward establishing a community like this.

Acknowledgement The research reported in this chapter was funded through resources provided by Tekes—the Finnish Funding Agency for Innovation and mediaX at Stanford University.

References

Barabási, A.-L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American, 288*(5), 50–59.

Basole, R. C., Russel, M. G., Huhtamäki, J., & Rubens, N. (2012). Understanding mobile ecosystem dynamics: A data-driven approach. In *Proceedings of the 2012 International Conference on Mobile Business* (*ICMB 2012*) (pp. 17–28). Delft, The Netherlands. Retrieved from http://aisel.aisnet.org/icmb2012/15/

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: The story so far. *International Journal on Semantic Web and Information Systems, 5*(3), 1–22.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann. Retrieved from http://www.amazon.com/Readings-Information-Visualization-Interactive-Technologies/dp/1558605339

Davenport, T. H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Boston: Harvard Business Press Books.

Freeman, L. C. (2000). Visualizing social networks. *Journal of Social Structure, 1*(1). Retrieved from http://www.cmu.edu/joss/content/articles/volume1/Freeman.html

Freeman, L. C. (2009). *Methods of social network visualization*. Berlin, Germany: Springer.

Ghosh, S. (2013). *Python tools for reproducible research in brain imaging*. Retrieved from https://speakerdeck.com/satra/pydata-2013-python-tools-for-reproducible-research-in-brain-imaging

Giuliani, E., & Bell, M. (2008). *Industrial clusters and the evolution of their knowledge networks: Revisiting a Chilean case*. Brighton, England. Retrieved from http://www.sussex.ac.uk/spru/documents/sewp171

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380. Retrieved from http://www.jstor.org/discover/10.2307/2776392?uid=3737976&uid=2&uid=4&sid=21104852601921

Hansen, D. L., Rotman, D., Bonsignore, E., Milic-Frayling, N., Rodrigues, E. M., Smith, M., & Shneiderman, B. (2009). *Do you know the way to SNA?: A process model for analyzing and visualizing social media data*. University of Maryland Technical Report: HCIL-2009-17. Retrieved from http://www.smrfoundation.org/wp-content/uploads/2010/05/2009-UMD-TechReport-Do-you-know-the-way-to-SNA.pdf

Hansen, D., Shneiderman, B., & Smith, M. A. (2011). *Analyzing social media networks with NodeXL: Insights from a connected world*. Burlington, MA: Morgan Kaufmann. Retrieved from http://www.amazon.com/dp/0123822297

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Communications of the ACM, 55*(4), 45–54. Retrieved January 31, 2013, from http://dl.acm.org/citation.cfm?id=2133821

Hevner, A. R., et al. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105.

Huhtamäki, J., Salonen, J., Marttila, J., & Nykänen, O. (2010, October 3–6). Context-driven social network visualisation: Case wiki co-creation. In D. Karabeg & J. Park (Eds.), *Proceedings of the Second International Workshop on Knowledge Federation: Self-Organizing Collective Mind*, Dubrovnik, Croatia. Dubrovnik, Croatia: CEUR-WS.org. Retrieved from http://urn.fi/URN:NBN:fi:tty-201201161008

Huhtamäki, J., Still, K., Isomursu, M., Russell, M. G., & Rubens, N. (2012, September 20–21). Networks of growth: Case young innovative companies in Finland. In *Proceedings of the 7th European Conference on Innovation and Entrepreneurship (ECIE)*, Santarém, Portugal

Huhtamäki, J., Luotonen, V., Kairamo, V., Still, K., & Russell, M. G. (2013, October 1–3). Process for measuring and visualizing an open innovation platform: Case Demola. In *17th International Academic MindTrek Conference 2013: "Making Sense of Converging Media*". Tampere, Finland: ACM. Retrieved from http://urn.fi/URN:NBN:fi:tty-201312201533.

Hwang, V. W., & Horowitt, G. (2012). *The rainforest: The secret to building the next silicon valley* 1.02 Edition. Los Altos Hills, CA: Regenwald. Retrieved from http://www.amazon.com/The-Rainforest-Secret-Building-Silicon/dp/0615586724

Indarto, E. (2013). *Data mining*. Retrieved December 13, 2014, from http://recommender-systems.readthedocs.org/en/latest/datamining.html

Intel. (2013). *Extract, transform, and load big data with Apache Hadoop*. Retrieved from https://software.intel.com/en-us/articles/extract-transform-and-load-big-data-with-apache-hadoop

Jesus, R., Schwartz, M., & Lehmann, S. (2009). *Bipartite networks of Wikipedia's articles and authors: A meso-level approach* (p. Article 5, 10 pages). Orlando, FL: ACM. Retrieved September 2, 2010, from http://portal.acm.org/citation.cfm?id=1641309.1641318

Jussila, J., Huhtamäki, J., Henttonen, K., Kärkkäinen, H., & Still, K. (2014, January 6–9). Visual network analysis of Twitter data for co-organizing conferences: Case CMAD 2013. In *Proceedings of the 47th Annual Hawaii International Conference on System Sciences* (pp. 1474–1483). Computer Society Press. Retrieved from http://urn.fi/URN:NBN:fi:tty-201401221053.

Kandel, S., Heer, J., Plaisant, C., Kennedy, J., van Ham, F., Riche, N. H., … Buono, P. (2011). Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization, 10*(4), 271–288. doi:10.1177/1473871611415994

Keim, D., Kohlhammer, J., & Ellis, G. (Eds.). (2010). *Mastering the information age—Solving problems with visual analytics*. Geneva, Switzerland: Eurographics Association. Retrieved from http://www.vismaster.eu/book/

Konno, N., Nonaka, I., & Ogilvy, J. (2014). Scenario planning: The basics. *World Futures, 70*(1), 28–43. Retrieved December 14, 2014, from http://www.tandfonline.com/doi/abs/10.1080/02604027.2014.875720

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. In *KDD '09* (pp. 497–506). New York: ACM. Retrieved September 14, 2012, from http://doi.acm.org/10.1145/1557019.1557077

Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature, 473*(7346), 167–173. Retrieved June 15, 2011, from http://dx.doi.org/10.1038/nature10011

Mawer, D. (2000). Ballet and the apotheosis of the dance. In D. Mawer (Ed.), *The Cambridge companion to ravel* (p. 157). Cambridge, England: Cambridge University Press.

Moreno, J. L. (1953). *Who shall survive?: Foundations of sociometry, group psychotherapy and sociodrama*. Beacon, NY: Beacon House Inc. Retrieved August 23, 2010, from http://www.asgpp.org/docs/WSS/WSS.html

North, C. (2006). Toward measuring visualization insight. *IEEE Computer Graphics and Applications, 26*(3), 6–9. Retrieved from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1626178&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D1626178

Nykänen, O., Salonen, J., Haapaniemi, M., Huhtamäki, J., Huhtamäki, J., & Huhtamäki, J. (2008). *A visualisation system for a peer-to-peer information space* (pp. 76–85). Tampere, Finland: Tampere University of Technology. Retrieved from http://matriisi.ee.tut.fi/hypermedia/events/opaals2008/articlelist.html#opaals2008-article14.

Paquin, R. L., & Howard-Grenville, J. (2013). Blind dates and arranged marriages: Longitudinal processes of network orchestration. *Organization Studies, 34*(11), 1623–1653. Retrieved March 11, 2014, from http://oss.sagepub.com/content/34/11/1623

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77. doi:10.2753/MIS0742-1222240302.

Peng, R. D. (2009). Reproducible research and biostatistics. *Biostatistics, 10*(3), 405–408. Retrieved December 14, 2014, from http://biostatistics.oxfordjournals.org/content/10/3/405

Petschulat, S. (2010). Other people's data. *Communications of the ACM, 53*(1), 53. Retrieved December 14, 2014, from http://cacm.acm.org/magazines/2010/1/55742-other-peoples-data/fulltext

Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*.

Ritala, P., Armila, L., & Blomqvist, K. (2009). Innovation orchestration capability—Defining the organizational and individual level determinants. *International Journal of Innovation*

Management, 13(04), 569–591. Retrieved March 6, 2014, from http://www.worldscientific.com/doi/abs/10.1142/S136391960900242X

Ritala, P., & Hallikas, J. (2011). Network position of a firm and the tendency to collaborate with competitors—A structural embeddedness perspective. *International Journal of Strategic Business Alliances, 2*(4), 307–328. Retrieved December 19, 2014, from http://dx.doi.org/10.1504/IJSBA.2011.044859

Ritala, P., & Huizingh, E. (2014). Business and network models for innovation: Strategic logic and the role of network position. *International Journal of Technology Management, 66*(2), 109–119. Retrieved December 19, 2014, from http://dx.doi.org/10.1504/IJTM.2014.064608

Rubens, N., Russell, M., Perez, R., Huhtamaki, J., Still, K., Kaplan, D., & Okamoto, T. (2011). Alumni network analysis. In *Proceedings of 2011 IEEE Global Engineering Education Conference, EDUCON* 2011 (pp. 606–611).

Russell, M. G., Still, K., Huhtamäki, J., Yu, C., & Rubens, N. (2011, July). Transforming innovation ecosystems through shared vision and network orchestration. In *Proceedings of Triple Helix IX International Conference: "Silicon Valley: Global Model or Unique Anomaly?"*, Stanford, CA.

Russell, M. G., Huhtamäki, J., Still, K., Rubens, N., & Basole, R. C. (2015). Relational capital for shared vision in innovation ecosystems. *Triple Helix: A Journal of University-Industry-Government Innovation and Entrepreneurship*.

Schwaber, K., & Beedle, M. (2001). *Agile software development with Scrum*. Upper Saddle River, NJ: Prentice Hall.

Shakarian, P., Eyre, S., & Paulo, D. (2013). *A scalable heuristic for viral marketing under the tipping model*. Retrieved September 23, 2013, from http://arxiv.org/abs/1309.2963

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Still, K., Huhtamäki, J., Russell, M. G., Basole, R. C., Salonen, J., & Rubens, N. (2013, June 16–19). Networks of innovation relationships: Multiscopic views on Finland. In *Proceedings of the XXIV ISPIM Conference—Innovating in Global Markets: Challenges for Sustainable Growth*, Helsinki, Finland (p. 15).

Still, K., et al. (2014). Insights for orchestrating innovation ecosystems: The case of EIT ICT Labs and data-driven network visualisations. *International Journal of Technology Management, 66*(2/3), 243–265.

Telea, A. C. (2008). *Data visualization: Principles and practice*. Wellesley, MA: A K Peters. Retrieved from http://www.amazon.com/Data-Visualization-Principles-Alexandru-Telea/dp/1568813066

Thomas, J. J., & Cook, K. A. (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications, 26*(1), 10–13.

Vaishnavi, V. K., & Kuechler, W., Jr. (2007). *Design science research methods and patterns: Innovating information and communication technology*. Boca Raton, FL: Auerbach.

Vassiliadis, P. (2009). A survey of extract–transform–load technology. *International Journal of Data Warehousing and Mining, 5*(3), 1–27.

Vinoski, S. (2008). Serendipitous reuse. *Internet Computing, 12*(1), 84–87. Retrieved February 27, 2009, from http://steve.vinoski.net/pdf/IEEE-Serendipitous_Reuse.pdf

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (1st ed.). New York: Cambridge University Press. Retrieved from http://www.amazon.com/dp/0521387078

Weng, L., Menczer, F., & Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports, 3*, 2522. Retrieved March 20, 2014, from http://www.nature.com/srep/2013/130828/srep02522/full/srep02522.html

Wikipedia. (2014). Web crawler. *Wikipedia*. Retrieved December 14, 2014, from http://en.wikipedia.org/w/index.php?title=Web_crawler&oldid=635502147

Visual Analytics of User Influence and Location-Based Social Networks

Jiawei Zhang, Junghoon Chae, Shehzad Afzal, Abish Malik, Dennis Thom, Yun Jang, Thomas Ertl, Sorin Adam Matei, and David S. Ebert

1 Introduction

Social media have transformed the way people express opinions, react to evolving and emergent events, and share their whereabouts. When an event occurs, information generated by users who witness or engage in it can provide first-hand accounts and updates. This information is propagated in the online social networks and triggers reactions from other users. Identifying influential users, monitoring the interaction between users, and analyzing information diffusion in social media can improve situational awareness in a crisis situation, and provide significant and reliable information for emergency management. Yet, inferring actionable information from raw social media data is not straightforward. The large volume of data and their multiple dimensions makes this process extremely difficult. The real time streaming nature of social media data introduces additional challenges. Reliance only on fully automated methods with minimum human intervention is not suitable while working with such datasets. Besides that, analysts working on such problems often need to look into the contextual evidence that could help them accept or reject certain hypothesis. Such contextual information could be provided only if application framework supports

J. Zhang (✉) • J. Chae • S. Afzal • A. Malik • D.S. Ebert
School of Electrical and Computer Engineering, Purdue University,
West Lafayette, Indiana, USA
e-mail: zhan1486@purdue.edu

D. Thom • T. Ertl
Department of Computer Science, University of Stuttgart, Stuttgart, Germany

Y. Jang
Department of Computer Engineering, Sejong University, Seoul, South Korea

S.A. Matei
Brian Lamb School of Communication, Purdue University, West Lafayette, IN, USA

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_12

interactive exploration, querying, and visual feedback. In this chapter, we introduce our suite of interlinked visual analytics tools that attempt to overcome these issues.

In the process, we also discuss some recent advances to support space-time indexed data. Such spatiotemporal data have immense value for increasing situational awareness of local events, providing insights for investigations regarding the extent of incidents, their severity and consequences, as well as public behavior during crises. However, the large volume of the data hinders effective exploration and examination, while its volume is significantly and constantly increasing. Also, a relatively small volume of critical data may easily be obscured by the large amounts of data generated every day. Thus, analysts need new methods for dealing with the data volume and its dynamic nature, as well as identifying abnormal events and topics within the data.

Typically, such challenges take advantage of automated data mining techniques, which are utilized to deal with large and/or high-dimensional data sets. Data mining is commonly defined as the process of discovering useful high-level knowledge from low-level data known as Knowledge Discovery in Database (KDD) (Fayyad, Piatetsky-Shapiro, & Smith, 1996). In the data mining process, data exploration is an important step to gain insight into data and create hypotheses. Adding a human intelligence touch to data discovery increases the relevance and validity to data streams. The human knowledge and the ability to quickly understand human behavior play an important role in the data exploration process. Visual analytics aims at integrating the human knowledge and the perceptual abilities of the human mind with automatic data mining procedures (Keim, 2002). Shneiderman (2002) notes that combining human cognitions supported by visualization and automated data mining can lead to novel discovery strategies that preserve user control and enable a more effective data exploration. In this chapter, we describe our visual analytics approaches that allow users to directly manipulate the algorithms in order to understand the workings and results of the data mining algorithms; thereby, enhancing their ability to contextualize data. This serves not only the human analysts but also improves the data mining methodologies. In other words, our interactive visualizations and visual analytic systems can help facilitate knowledge discovery from social media data by enabling analysts to generate, test, and refine their hypotheses.

2 Related Work

In what follows, we primarily consider the case of location-based information for discussing our broader strategy of data mining, discovery and visualization. Location-sensitive data harvested from social networks have become a popular and influential data source for many applications. However, the large volume of data and the unstructured nature of the information hinders exploration and examination. Thus, scalable computational analysis for improving spatiotemporal situational awareness and discovering of critical information within the data are vital research topics and application domains. The following subsections present previous works that have focused on LBSN analysis and the manner in which they have contributed to our own vision.

2.1 Visualization of Social Networks

In order to explore and examine the large number of nodes and links in social networks, some previous studies combine data mining algorithms and visualization techniques (Sun et al., 2009; Yang, Asur, Parthasarathy, & Mehta, 2008). Correa, Crnovrsanin, and Ma (2012) propose an analytical mechanism for measuring sensitivities of nodes in a network using eigenvector and Markov centralities to find important and influential nodes. Crnovrsanin, Liao, Wu, and Ma (2011) demonstrate a visualization system for supporting effective navigation of social networks. The system suggests directions based on the importance of the nodes in the network and past user interactions. In user activities in Twitter, retweeting and replying are important and distinguishable attributes associated with the links between Twitter users compared to traditional social network relationships. Also, these are the key mechanisms for information propagation and transfer in the social networks. There are some studies that focus on the mechanisms of information transfer, although those do not discuss visualization aspects. Suh, Hong, Pirolli, and Chi (2010) study the factors affecting the retweet ability of tweets. They focus on content features (e.g., URLs and hashtags) and contextual features (e.g., the number of followers and followees) to estimate the factors that are significantly associated with retweeting. Macskassy and Michelson (2011) focus on information diffusion behaviors underlying processes by which they decide to retweet. Ho, Li, and Lin (2011) study how information is propagated in micro-blog networks with respect to the number of users influenced, the speed of propagation, and the geographical distance of the propagation.

2.2 Location-Based Social Networks Analysis

As social media platforms move towards LBSN researchers have proposed various approaches to analyze spatiotemporal document collections, in general, and spatio-temporal social media data, in particular. VisGets (Dork, Carpendale, Collins, & Williamson, 2008) provides linked visual filters for the space, time and tag dimensions to allow the exploration of data sets. The user is guided by weighted brushing and linking, which denotes the co-occurrences of attributes. Further works demonstrate the value of visualizing and analyzing the spatial context information of microblogs for social network users (Field & O'Brien, 2010) or third parties like crime investigators (Roth & White, 2010) and urban planners (Wakamiya, Lee, & Sumiya, 2011). Andrienko et al. (2013) describe a visual analysis approach for exploring tweet text and spatiotemporal patterns. Krueger, Thom, and Ertl (2014) extract frequent visited places from vehicle movement data and further use semantics distilled from the social network to decode daily activities of people. MacEachren et al. (2011) demonstrates a visual analytics system to represent tweet density of actual or textually inferred locations. Their work also demonstrates that social media can be a potential source for crisis management. Bosch et al. (2011) provides a scalable system enabling analysts to work on quantitative findings within a large set of tweets with geo-location. Chae et al. (2012) propose a combination of LDA and Seasonal-Trend Decomposition for abnormal event detection. Researchers also present analysis of LBSN for disaster management and evacuation planning (Chae et al., 2014; Sakaki, Okazaki, & Matsuo, 2010; Terpstra, Stronkman, de Vries, & Paradies, 2012). Ying, Lee, Ye, Chen, and Tseng (2011) present various location-based metrics using spatial information of these LBSNs to observe popular people who receive more attention and relationships within the network.

3 User Influence-Based Dynamic Social Networks

Social media can be utilized as a publicly available data source to identify and gather information pertaining to events of interest. In such scenarios, identifying both individuals of interest (e.g., witnesses), and the potential information they disseminate through their social media networks, can be especially instrumental for decision makers and emergency managers. When a specific event occurs, four types of actors are mainly involved in the social networks: (1) users who engage in the event and post messages, (2) common users who participate in the propagation and forward messages, (3) popular users (including celebrities, opinion leaders, news broadcasters) who accelerate the propagation of messages, and (4) passive users who receive but do not forward the messages (Romero, Galuba, Asur, & Huberman, 2011). Typically, a message containing important information is posted by a witness, diffused through popular or common users, and finally ends with passive users. In this process, two types of influential users are of particular interest: witnesses and popular users. Witnesses provide first-hand updates of information that can help understand and respond to events as quickly as possible. However, in most cases, they are hidden in the massive noise of the crowd. For popular users, there is a delay in the messages to reach them. However, they stand out in the information diffusion process and can provide clues for tracing the source of the messages. Based on the above observations, we develop a visual analytics framework for user networks and information diffusion processes to identify these types of influential users.

3.1 Explicit Connections: Replies and Retweets

Explicit connections among Twitter users mainly include reply/retweet and follower/friend. Reply/retweet connections are generated when users explicitly establish connections with each other. Reply/retweet serve as the most popular way for Twitter users to share and deliver instant messages and can indicate strong

Table 1 Examples of retweet and reply

Type	User	Twitter messages
Retweet	@stoobush	1. RT @rtv6: #BREAKING: shooting reported at Purdue Electrical Engineering building, campus police confirm
Reply	@Hmother8	2. @calmoza shooting on campus!

relationships among them. In contrast, follower/friend connections are formed when users want to get real-time updates of the individuals they choose to follow. Since users can establish or disconnect a follower/friend relationship with few limitations, follower/friend based networks are often large-scale, show little change over short periods of time, and lack any semantic content. These are therefore of less importance than reply/retweet connections when collecting information pertaining to specific events. Consequently, we focus on reply/retweet connections in our work. For a typical retweet/reply message, we define the direction of information diffusion. We illustrate this directionality using two example messages shown in Table 1. Here, message (1) is a retweet message, where the original message is posted by the user @rtv6, and is then retweeted by the user @stoobush. Message (2) is a reply message, where the message is diffused by the user who initiates the conversation (@Hmother8) to the user who receives the message (@calmoza). The network generated based on reply/retweets can thus be considered as directed networks.

3.2 Visualization of Dynamic Networks

Dynamic Twitter networks are rich in valuable latent patterns dealing with multiple information entities, such as Twitter users and topics. Our visualization utilizes user connections in Twitter as the backbone of the dynamic network, and shows the evolution of retweet/reply communications and semantic correlations among the Twitter users. Twitter user relationships are depicted with an overview-based node-link visualization in our system. As shown in Fig. 1 the relationships have three main features:

Nodes: The nodes encode users and the size of each node depicts tweet volume of the user for the currently selected time frame.

Edges: Edges depict user connections. Users with retweet or reply communications are shown using arrow-styled edges to show the direction of tweet propagation. For the retweet communication, the node at the thicker end represents the user who posts the tweet, and the node at the thinner end represents the user who retransmits the provider's tweet. For the reply communication, the node at the thinner end represents the user who posts the tweet, and the user at the thicker end represents the one who replies to the tweet. The system allows users to perform filtering to show either retweet or reply relationships through a set of checkboxes in the interface.

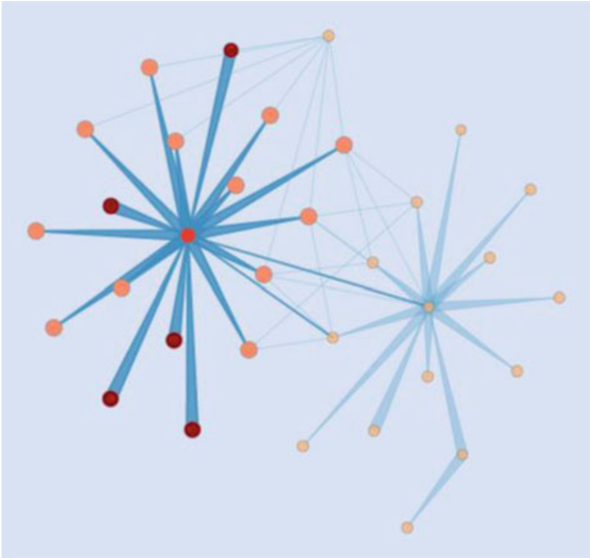


Fig. 1 Forced-directed layout with DOI based visualization

Force-directed layout (Fruchterman & Reingold, 1991): We utilize a Force-directed layout to project users on the screen based on their relationship as shown in Fig. 1. Nodes in the graphs are dynamically changing, since some users may leave, and others may join in. In order to avoid visual confusion caused by user changes, and allow analysts to perceive how users evolve over time, we provide smooth animations to visualize user evolution in different communities and topics. Aside from the above-mentioned overview-based approach, we provide a degree of interest (DOI) based visualization (Card & Nation, 2002) that shows the information propagation process related to a single user. As shown in Fig. 1, when the analyst selects a user, nodes connected to the user are highlighted to present the information diffusion pattern. Users with darker colors serve as sources of information for the users with lighter colors along the propagation path. A combination of the overview-based and DOI-based visualizations allows analysts to iteratively examine information propagation, identify influential nodes, and observe evolution of user networks.

3.3 Interaction Design

Our target end-users include decision makers for emergency and natural disasters, and public safety/law enforcement personnel. Based on close cooperation with these end-users, we base our interaction design and interactive visual tools on supporting simple, intuitive and reversible operations following Norman's Principle of Naturalness (Norman, 1993). We use an interactive details-on-demand (Shneiderman, 1996) ContentLens (Thom, Bosch, Koch, Woerner, & Ertl, 2012)

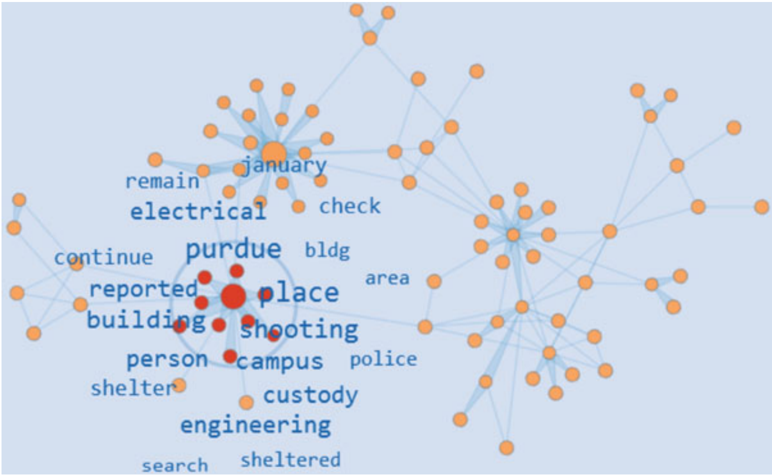


Fig. 2 ContentLens showing major keywords discussed in underlying network

that allows analysts to explore and monitor topics within the social networks. This feature provides analysts with extracted major keywords for their selected level of aggregation, from an individual user to small or large groups within the network as shown in Fig. 2. As the analysts move the ContentLens over the networks, they can focus on specific nodes within the networks. The analysts can dynamically change the diameter of the lens to either overview a large aggregation of tweets or focus on tweets of specific users. By utilizing ContentLens, the analysts are allowed to interactively investigate user activity in their identified social network.

4 Visualization of Location-Based Social Networks for Abnormal Event Detection

Social networks allow people to create a large volume of time-stamped and geo-tagged tweet messages. This requires a tool to cope with large amounts of messages in order to help analysts to explore and detect important messages. Our visual analytics system (shown in Fig. 3) allows analysts to select an initial spatiotemporal context of tweet messages to be represented in the visualization and to serve as a basis for analysis. The spatiotemporal distribution of messages can provide an initial insight to the analysts that can be relevant for their analysis tasks. In the subsequent step, the analysts start with the topic extraction on the analysis context using a data mining model, Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003), that extracts and probabilistically ranks major topics contained in textual parts of the tweets. Users can adjust the configuration parameters of LDA to interactively explore available topics by generalization and specialization.

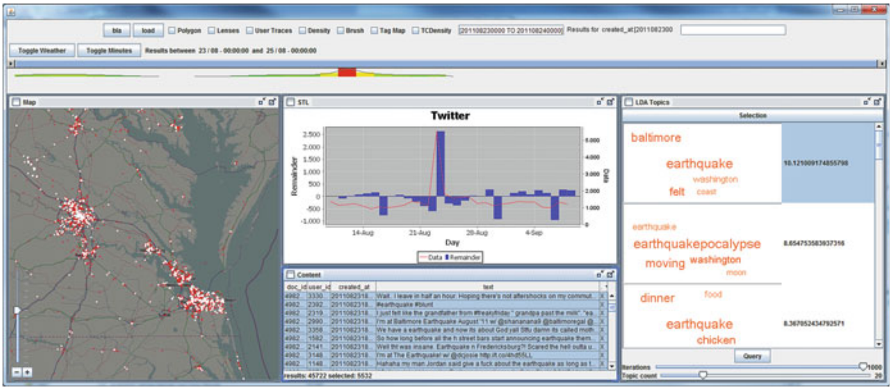


Fig. 3 Social media analysis system including message plots on a map (1), an abnormality estimation chart (2), a tweet content table (3), and a topic exploration view (4). It can be seen how the Virginia earthquake on August 23rd, 2011 is examined using the system. The system detects the earthquake event using our STL based anomaly detection model

The extracted topics can be evaluated and ordered based either on volume-based importance or abnormality estimates computed using the Seasonal Trend Decomposition procedure based on Loess smoothing (STL) (Cleveland, Cleveland, McRae, & Terpenning, 1990). In order to obtain a ranking suitable for abnormal topics analysis tasks, we discard daily chatter by employing STL. This combination of utilizing automatic data mining algorithms that are further facilitated through interactive visualizations enables analysts to discover emerging and abnormal topics from the noise. Additionally, we note that our system also provides the ability to send out email alerts to analysts when a threshold for certain keywords is met (e.g., when the number of twitter messages containing user specified keywords for a certain time step exceeds N). This further assists end-users to detect an emerging situation using social media as an information source.

In Sect. 4.1 we first describe how we utilize the LDA topic modeling to extract the inherent topics from a set of tweet messages. In Sect. 4.2, we explain how we estimate abnormalities for each given topic and re-rank the topics based on the abnormality scores to identify unusual and unexpected topics using STL.

4.1 Topic Extraction

Often when an unusual situation or an unexpected event occurs within an area and a time window, a certain number of tweet messages are generated by the community from the communication of the event. This set of messages implicitly constitutes multiple topics. In order to extract each of the topics exhibited within the collection of messages, we employ the LDA topic model which is a probabilistic topic model that can help organize, understand, and summarize vast amounts of documents.

Table 2 An example of extracted topics and their proportions (Chae et al., 2012)

Rank	Proportion	Topics
1	0.10004	Day back school today
2	0.09717	Ils bout dat wit
3	0.09443	People make hate wanna
4	0.08226	Earthquake thought house shaking
5	0.05869	Earthquake felt quake Washington

The LDA topic model extracts topics from tweets generated on August 23, 2011 around Virginia, where an earthquake occurred on this day. Topics consisting of ordinary and unspecific words have high proportion values, while the earthquake related topics have a relatively low proportion value

The LDA topic model, as presented by Blei et al. (2003), is an unsupervised machine learning technique to identify latent topics from a large document collection. Basically, it uses a "bag of words" approach and assumes that a document exhibits multiple topics distributed over words with a Dirichlet prior. In other words, the LDA assumes the following generative process for each document: First, choose a distribution over topics, choose a topic from the distribution for each word, and choose a word associated with the chosen topic. Based on this assumption, one can apply a Bayesian inference algorithm to retrieve the topic structure of the message set together with each topic's statistical proportion and a list of keywords prominent within the topic's messages. Table 2 shows an example set of extracted topics resulting from the application of LDA to tweets. The topics are ordered by the proportion ranking. The example tweets were generated on August 23, 2011 around the Virginia area. On this day, the area was struck by an earthquake with a magnitude of 5.88. As seen in the table, the last two topics indicate the earthquake event. Figure 4 shows the topic exploration view of the entire system in Fig. 3. We can see most of the topics in the view represent the earthquake event. In our system, the MALLET toolkit (McCallum, 2002) is used for the topic modeling. Prior to the topic modeling, the stemming algorithm KSTEM by Krovetz (1993) is applied to every term in the messages.

4.2 *Abnormality Estimation*

In order to prioritize the topics extracted using the LDA topic model and allow analysts to discover abnormal events from twitter messages, we employ the Seasonal-Trend Decomposition based on locally-weighted regression (Loess) known as STL (Cleveland et al., 1990) method. We define abnormal events as events that do not occur frequently and regularly. Also, abnormal events generally cover only a small fraction of the social media stream. For example, Table 2 shows the first and second ranked topics consist of ordinary and unspecific words even during an earthquake. The fourth and fifth ranked topics include words indicating the earthquake event of August 2011. From this observation in the

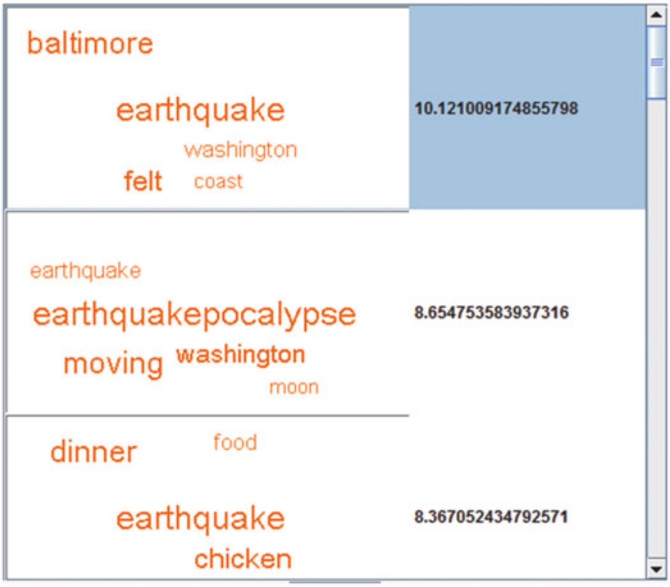


Fig. 4 Topic exploration view. The values in the right column of the view shows the z-scores of each topics. The high scores of the three topics show their strong abnormality

proportions of ordinary and unusual topics over the social media data, we need to differentiate the unusual topics from the large number of rather mundane topics. We utilize STL to identify the topics indicating unexpected and unusual situations. An abnormal event is associated with a set of tweets that provides its contents, location, and time-stamp. To detect abnormal events for a given area and a time window, users select a subset of tweets within the spatiotemporal filter. The LDA topic modeling (described in Sect. 4.1) then extracts a set of topics from the selected tweets. For each topic, we search for relevant tweets in the selected area and time period and a predefined time span of historic data preceding (e.g., 1 month). Tweets are considered relevant if they contain at least one word. A daily tweet count time series is generated from the timestamps of the tweet. The time series can be considered as the sum of three components: a trend component, a seasonal component, and a remainder:

$$Y = T + S + R \tag{1}$$

Here Y is the original time series of interest, T is the trend component, S is the seasonal component, and R is the remainder component. STL works as an iterative nonparametric regression procedure using a series of Loess smoothers (Cleveland, 1979). The iterative algorithm progressively refines and improves the estimates of the trend and the seasonal components. The resulting estimates of both components are then used to compute the remainder: $R = Y - T - S$. Under normal

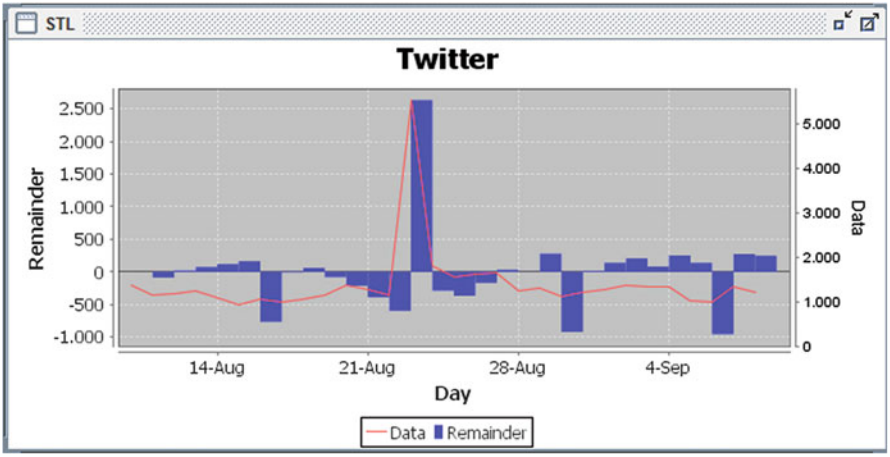


Fig. 5 Abnormality estimation chart. It shows the original time series (*red line*) and remainder values (*bar*) of the first topic in the topic view in Fig. 4. The abnormality degree is extremely high on August 23rd, 2011 (times are given in UTC)

conditions, the remainder will be identically distributed Gaussian white noise, while a large value of R indicates substantial variation in the time series. Thus, we can utilize the remainder values for control chart methods detecting anomalous outliers within the topic time series. We use a 7 day moving average of the remainder values to calculate the z-scores, $z = (R(d) - mean)/std$, where $R(d)$ is the remainder value of day d, mean is the mean remainder value for the last 7 days, and *std* is the standard deviation of the remainders, with respect to each topic. If the z-score is higher than 2, events can be considered as abnormal within a 95 % confidence interval. The calculated z-scores are thus used as abnormality rating and the retrieved topics will be ranked in the analytics environment according to this estimate. In Fig. 4, the values in the right column of the view shows the z-scores of each topic. The high values of the three topics that related to the earthquake show their strong abnormality. Figure 5 shows the original time (red line) series and remainder component values (bar) of the first topic in the topic view in Fig. 4. We can see the abnormality degree is extremely high on August 23rd, 2011.

5 Case Study

In this section, we demonstrate how our system can be used using historical data from a real-world scenario. On January 21st, 2014, a shooting event occurred around 12:10 PM at Purdue University in West Lafayette, Indiana, USA. In order to explore the evolution of the event and response from Twitter users, law enforcement analysts can utilize our system in order to explore Twitter data generated in real time. As shown in Fig. 6, our system extracts major topics from the tweets and

Fig. 6 Topics from the tweets generated within the Purdue University area during 1 h after the shooting accident on January 21st, 2014

orders the topics based on their abnormality scores. The high abnormal (more than 2) topics are highlighted in red color and the others are black. Our system identifies the shooting incident as an abnormal event and highlights it by assigning it a high abnormality score. In addition, the system generates an email alert to the intended law enforcement recipients as the threshold for law enforcement sensitive keywords is met.

Law enforcement analysts can interactively identify popular users and witnesses for the event. They can enter several keywords that roughly describe the event (e.g., Purdue, shooting, victim, suspect, police, murder). The system then filters the tweets based on the entered keywords and generates user networks. At 12:17:35, the first relevant tweet appears: '*Shooting on campus*?' Over time, more people get involved in the conversation and post messages in their social networks related to the shooting event. Our system also allows analysts to obtain an overview of the evolving topics. The analysts can utilize the ContentLens feature of our system to examine topics over the user group. This is shown in Fig. 2, where the terms shooting, electrical, and engineering are seen as dominant terms. This indicates that the location of the event is around the Electrical Engineering department. As the conversation networks grow, users with a large number of connections with other users join the conversation. In Fig. 7 (Left), the law enforcement analysts select the user *JConline* (a local news media company) to visualize the information propagation patterns from this news source. Five users in dark colors serve as the source of information for *JConline*. We then select these five nodes to explore their information propagation patterns. From 12:30:00 to 13:00:00, these users actively post messages, most of

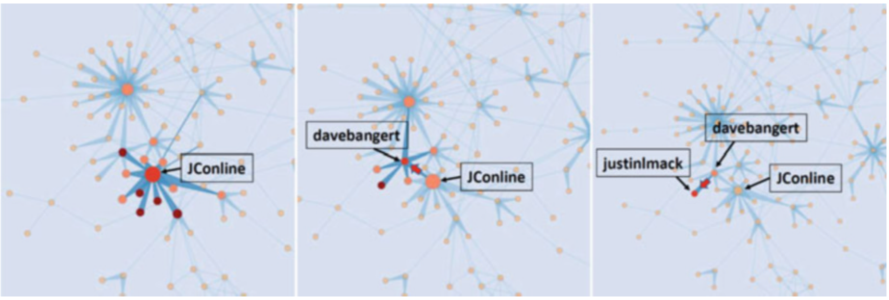


Fig. 7 Iterative approach to locate influential users. Five users serve as the source of information for *JConline* (*left*). The messages retweeted by *JConline* are propagated to other users (*center* and *right*)

which contain pictures of the event. Furthermore, most of the messages are later retweeted by *JConline* and then quickly spread to other users. Accordingly, law enforcement analysts can maintain a situational awareness of an evolving situation, and locate and track the activity of different user groups from the social networks in order to rapidly collect important information pertaining to the event.

6 Conclusion

In social media analysis, identifying influential users and analyzing user interaction and information diffusion in the networks can improve situational awareness of events and provide significant and reliable information in emergency management. Analysts working on such issues often need to look into the contextual evidence that could help them create hypothesis. To provide such contextual information, we introduced our visualizations of user influence-based dynamic social networks, that make it possible to identify user-influence from social networks and analyze information diffusion in social networks. Also, we described our visual analytics system to cope with large amounts of messages and help analysts to explore and detect important messages. Our system combines visual presentation with data mining and statistical models in order to take advantage of the synergistic impact of the multiple techniques. The system allows users to directly manipulate the algorithms, easily understand results of the algorithms and how the algorithms work in order to facilitate knowledge discovery from social media data by enabling analysts to generate, test, and refine hypotheses from data. We integrate two techniques including the LDA topic model and the time series decomposition statistical technique with our analysis environment to detect abnormal events. We demonstrated the usage and effectiveness of our system for social network analysis and anomaly analysis in abnormal situations by case studies.

References

Andrienko, G., Andrienko, N., Bosch, H., Ertl, T., Fuchs, G., Jankowski, P., et al. (2013). Thematic patterns in georeferenced tweets through space-time visual analytics. *Computing in Science and Engineering, 15*(3), 72–82.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Bosch, H., Thom, D., Worner, M., Koch, S., Puttmann, E., Jackle, D., et al. (2011). Scatterblogs: Geo-spatial document analysis. In *IEEE Symposium on Visual Analytics Science and Technology* (pp. 309–310).

Card, S. K., & Nation, D. (2002). Degree-of-interest trees: A component of an attention-reactive user interface. In *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI'02* (pp. 231–245). New York: ACM. doi:10.1145/1556262.1556300.

Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D., et al. (2012). Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *IEEE Symposium on Visual Analytics Science and Technology, October 14–19, 2012* (pp. 143–152).

Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., & Ebert, D. S. (2014). Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics, 38*, 51–60.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*(368), 829–836.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess (with discussion). *Journal of Official Statistics, 6*(1), 3–73.

Correa, C., Crnovrsanin, T., & Ma, K.-L. (2012). Visual reasoning about social networks using centrality sensitivity. *IEEE Transactions on Visualization and Computer Graphics, 18*(1), 106–120.

Crnovrsanin, T., Liao, I., Wu, Y., & Ma, K.-L. (2011). Visual recommendations for network navigation. *Computer Graphics Forum, 30*, 1081–1090.

Dork, M., Carpendale, S., Collins, C., & Williamson, C. (2008). VisGets: Coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics, 14*(6), 1205–1212.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smith, P. (1996). From data mining to knowledge discovery: An overview. In U. M. Fayyad, et al. (Eds.), *Advances in knowledge discovery and data mining* (chap. 1, pp. 1–34). Menlo Park, CA: AAAI Press/MIT Press.

Field, K., & O'Brien, J. (2010). Cartoblography: Experiments in using and organising the spatial context of micro-blogging. *Transactions in GIS, 14*, 5–23. doi:10.1111/j.1467-9671.2010.01210.x.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software Practice and Experience, 21*(11), 1129–1164. doi:10.1002/spe.4380211102.

Ho, C.-T., Li, C.-T., & Lin, S.-D. (2011). Modeling and visualizing information propagation in a micro-blogging platform. In *IEEE International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 328–335).

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics, 8*(1), 1–8.

Krovetz R. (1993). Viewing morphology as an inference process. In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'93* (pp. 191–202). New York: ACM. doi:10.1145/160688.160718.

Krueger, R., Thom, D., & Ertl, T. (2014). Visual analysis of movement behavior using web data for context enrichment. In *Proceedings of the 2014 I.E. Pacific Visualization Symposium, PACIFICVIS'14* (pp. 193–200). Washington, DC: IEEE Computer Society.

MacEachren, A., Jaiswal, A., Robinson, A., Pezanowski, S., Savelyev, A., Mitra, P., et al. (2011). SensePlace2: GeoTwitter analytics support for situational awareness. In *IEEE Symposium on Visual Analytics Science and Technology* (pp. 181–190).

Macskassy, S. A., & Michelson, M. (2011). Why do people retweet? Anti-homophily wins the day!. In *International AAAI Conference on Web and Social Media*.

McCallum, A. K. (2002). *Mallet: A machine learning for language toolkit*. Retrieved from http://mallet.cs.umass.edu

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. New York: Basic Books.

Romero, D. M., Galuba, W., Asur, S., & Huberman, B. A. (2011). Influence and passivity in social media. In *Machine learning and knowledge discovery in databases* (pp. 18–33). Berlin, Germany: Springer.

Roth, E., & White, J. (2010) Twitterhitter: Geovisual analytics for harvesting insight from volunteered geographic information. In *Proceedings of GIScience*.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: Realtime event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web, WWW'10* (pp. 851–860). New York: ACM.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages* (pp. 336–343).

Shneiderman, B. (2002). Inventing discovery tools: Combining information visualization with data mining. *Information Visualization, 1*(1), 5–12.

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *Proceedings of the 2010 I.E. Second International Conference on Social Computing (SocialCom)* (pp. 177–184).

Sun, J., Papadimitriou, S., Lin, C.-Y., Cao, N., Liu, S., & Qian, W. (2009) Multivis: Content-based social network exploration through multi-way visual analysis. In *SDM* (Vol. 9, pp. 1063–1074). SIAM.

Terpstra, T., Stronkman, R., de Vries, A., & Paradies, G. (2012). Towards a realtime twitter analysis during crises for operational crisis management. In *Proceedings of the 9th International ISCRAM Conference, Vancouver*.

Thom, D., Bosch, H., Koch, S., Woerner, M., & Ertl, T. (2012). Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *IEEE Pacific Visualization Symposium (PacificVis)* (pp. 41–48).

Wakamiya, S., Lee, R., & Sumiya, K. (2011). Crowd-based urban characterization: Extracting crowd behavioral patterns in urban areas from twitter. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, LBSN'11*. New York: ACM.

Yang, X., Asur, S., Parthasarathy, S., & Mehta, S. (2008). A visual-analytic toolkit for dynamic interaction graphs. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1016–1024). New York: ACM.

Ying, J. J.-C., Lee, W.-C., Ye, M., Chen, C.-Y., & Tseng, V. S. (2011). User association analysis of locales on location based social networks. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, LBSN'11* (pp. 69–76). New York: ACM.

Transparency, Control, and Content Generation on Wikipedia: Editorial Strategies and Technical Affordances

Sorin Adam Matei and Jeremy Foote

1 Introduction

Wikipedia is perhaps the most culturally influential example of "peer production" principles in action, and is certainly the most visible. As the sixth most popular web site on the Internet, Wikipedia has become an important source of information, not only for students, but also for academics, physicians, and many others (Hughes, Joshi, Lemonde, & Wareham, 2009).

Wikipedia is a radically inclusive way of creating an encyclopedia. With few exceptions, Wikipedia lives up to its promise as "the encyclopedia that anyone can edit". Every page has an option to "edit" the page, and edits appear immediately. In addition to being radically democratic, Wikipedia is, at least in principle, radically transparent. Again, with a few exceptions, every edit made to every page is publicly visible. The governance is also very open, with nearly all of the conversations about the policies and direction of the site held on public Wikipedia pages or public IRC channels and listservs.

Despite this prima facie inclusiveness and transparency, Wikipedia is both hierarchical and opaque in some important ways. While "anyone can edit" Wikipedia, not just anyone does. A relatively small number of contributors produce the vast majority of content, both across the entire Wikipedia project, and for most individual articles (Kittur, Chi, Pendleton, Suh, & Mytkowicz, 2007; Matei, Bruno, & Morris, 2015; Voss, 2005). While the tools for editing Wikipedia are available to everyone, the practical power of maintaining articles or categories on a certain

S.A. Matei (✉)
Brian Lamb School of Communication, Purdue University, West Lafayette, IN, USA
e-mail: smatei@purdue.edu

J. Foote
Doctoral Student, Northwestern University, Evanston, IL, USA

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_13

"line" is held by a small "adhocracy" (Matei, Tan, Zhu, Bertino, & Liu, 2014). These editors, like other active editors on Wikipedia, are much more likely to be male, young, well-educated, and from the Global North than the general population (Hill & Shaw, 2013).

In addition to this unexpected inequality on Wikipedia, there are the more obvious problems of vandalism, propaganda, and poorly researched information, which have attracted attention and reduced the credibility of the encyclopedia since its founding. Much content is simply copied and never checked (Rector, 2008). Other content, although controversial, is defended from removal by small coteries of interested individuals (Matei & Dobrescu, 2010).

Despite these concerns, the interface of Wikipedia remains simple in design and opaque with respect to authorship. It de-emphasizes everything except for the current content of a given article. While this design choice may be defended by the need to communicate the content of the article in the most direct way, it hides the social origin and potential biases of the what is written. As the value of the content depends, to a certain degree, on the nature of the collaborative process, it could be asked whether Wikipedia should reconsider its information delivery priorities. Featuring information about the nature of the collaborative process more prominently on the page could serve to make this process more transparent, and increase the perception of the content itself as accurate, credible, and unbiased.

In this chapter, we explore some of the major visualizations created to try to make Wikipedia more transparent, and theoretically more trustworthy. We also examine the conversations Wikipedians have had about whether one of these visualizations should be adopted by the site, and identify a number of possible reasons that the makers of these tools have been unsuccessful in having their visualizations accepted into the main interface. We conclude with a discussion of some possible strategies for creating and implementing visualization tools that would both increase transparency and be accepted by the Wikipedia community.

2 History of Interface Changes

Since becoming a popular site, the Wikipedia interface has changed very, very little. The Wikipedia page about the history of Wikipedia (https://en.wikipedia.org/wiki/History_of_Wikipedia#Look_and_feel) lists only nine changes to the look and feel of the encyclopedia. Three of these are changes to how the site is organized; four are changes to the look of the home page; one is a change to the logo. Only one change, made in May 2010, is a major change to the interface itself.

The way that the actual content is displayed in articles has changed very little indeed. From the beginning, content has been the focus of the page, with three tabs at the top of the page. The first is a Talk page for the article, the second opens the article for editing, and the third shows the history of changes made. However, the tabs and the information they contain are far more important than their "optional" vocabulary seems to suggest. They are entry points for understanding the social and intellectual processes that generate Wikipedia.

A number of researchers and programmers have worked to make these processes more visible. Some of them are intended as standalone visualizations, which give insight into Wikipedia, but are not intended to be part of the interface. They are mentioned here to give context to the goals and scope of visualizations that have been created. Our primary focus is on the second category of visualizations, meant to be more directly integrated in the editorial and content consumption workflow.

3 Standalone Visualizations

The first category of standalone visualizations attempt to situate Wikipedia contributions geographically. For example, Yasseri, Spoerri, Graham, and Kertész (2014) identified the most controversial articles in each language edition of Wikipedia, and then used maps to visualize where the articles with a geographic component were located (Figs. 1 and 2).

Omnipedia, a project by Bao et al. (2012), visualizes how different topics are treated differently in different language editions on Wikipedia. The topics which are linked to in a given language, but not in other languages, are highlighted. Both of these projects help to show that the way knowledge is constructed and experienced is culturally contingent (Fig. 3).

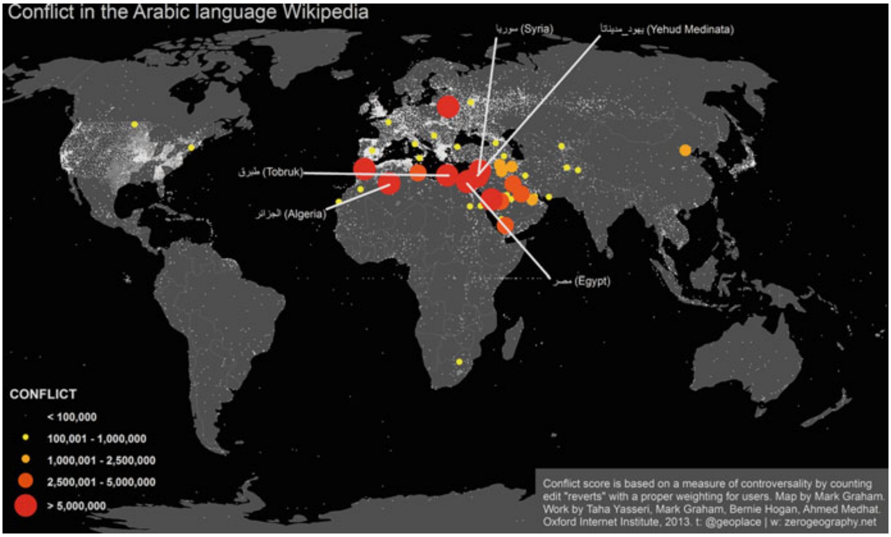


Fig. 1 Map of conflict in Czech edition of Wikipedia. Size of the *dots* is proportional to the controversy measure M

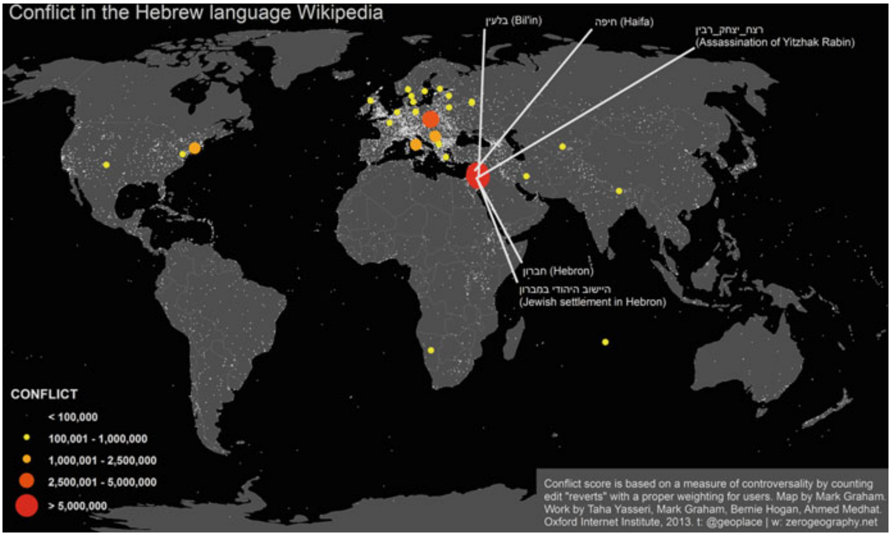


Fig. 2 Map of conflict in Hebrew edition of Wikipedia. Size of the *dots* is proportional to the controversy measure M

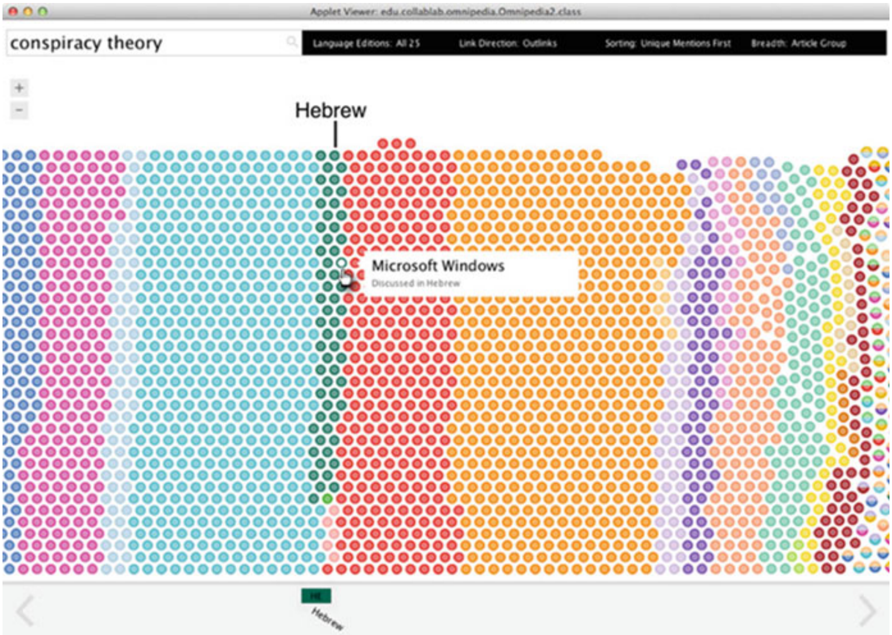


Fig. 3 This view of Omnipedia shows which articles are linked to from one language's version of an article, but not any others. In this example, "Microsoft Windows" is linked to only from the Hebrew Wikipedia's "conspiracy theory" article

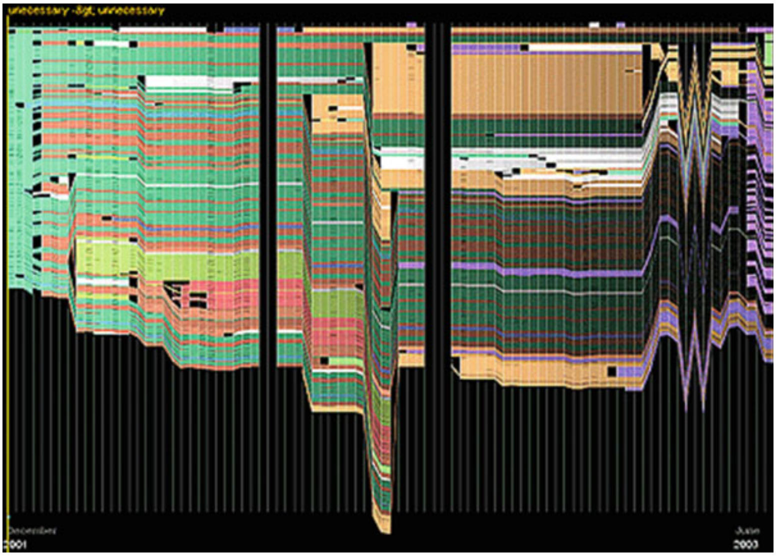


Fig. 4 history flow for “abortion” page, versions equally spaced

Other research focused on visualizing the community of editors, through summary statistics and graphs (e.g., Voss, 2005), mapping co-editing patterns by category (Biuk-Aghai, Pang, & Si, 2014), and network graphs of contributors (Keegan, Gergle, & Contractor, 2013).

Finally, Viégas and Wattenberg have worked on a number of visualizations to make the history of both articles and users more accessible. Their History Flow visualizes the way that an article has been developed over time, showing both the timing and location of revert wars, as well as giving insight into how this knowledge is produced and negotiated (Viégas, Wattenberg, & Dave, 2004). Their Chromogram visualization shows the types of edits made by users, giving a new way to identify different patterns of editing (Wattenberg, Viégas, & Hollenbach, 2007) (Figs. 4, 5, 6, and 7).

Such projects seek to provide a high-level view of Wikipedia, showing large-scale cultural differences or project-level biases or statistics. In general, they do not appear to have been created with the goal of being integrated into Wikipedia.

4 Article-Level Inequalities

While the projects so far discussed focus primarily on project-level dynamics and visualizations, much more interesting for the purposes of this inquiry are the projects that aim to visualize in a direct way the inequality of contributions to an article. This issue of paramount importance. As the bulk of most articles on Wikipedia are edited by a very small number of contributors, it could be said that while a given Wikipedia article does not have “an author” it does have a selected group of authors, who are responsible for the shape, tone, focus and often wording

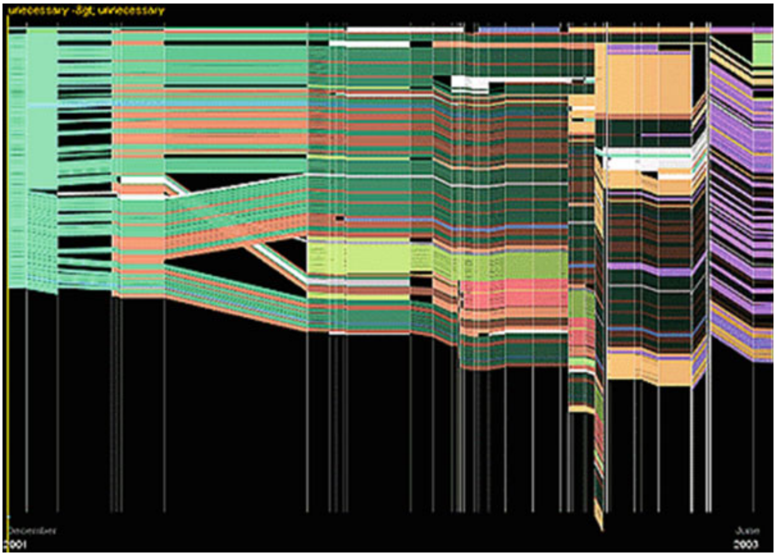


Fig. 5 history flow for “abortion” page, spaced by date

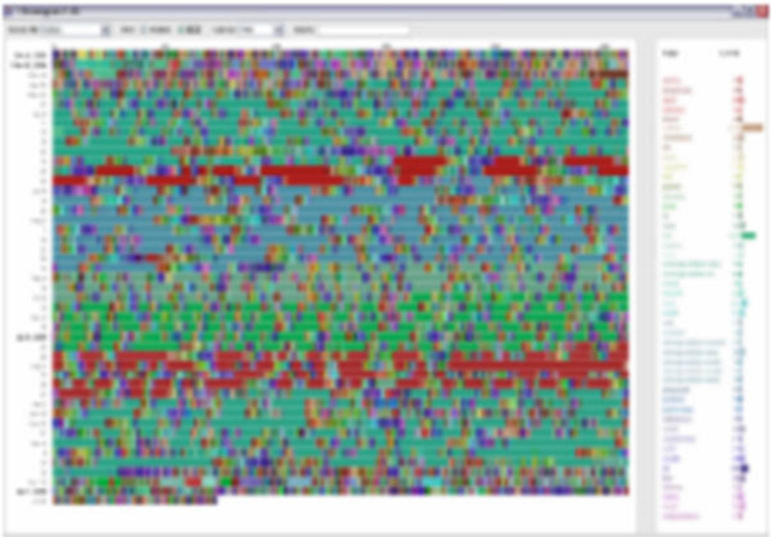


Fig. 6 The Chromogram application: block view

of the article. One would reason that the presence and identity of these selected contributors should be clearly and directly visible on each page. It is not only an issue of transparency, but also of trust. Trust in traditional encyclopedias relied on a the authority of the authors. Wikipedia is shaped in an ad hoc basis, by a group of top contributors. Trust in the content is based on trust in the social and technical structures which surround the project, by which this group emerges and works (Slattery, 2009; Swarts, 2009).

Fig. 7 The Chromogram application: time-line view, same date as Fig. 6

The issue here is not one of “unmasking” the top editors or denouncing them as frauds. While some concerns have been expressed that a system which does not rely on experts at any point in the process could not produce reliable information, research has shown that, along dimensions that are verifiable, Wikipedia’s reliability is comparable to that of the Encyclopedia Britannica for certain types of content (Giles, 2005), although not for others (Rector, 2008).

However, there are still opportunities for biases. These are functional and “perspective” biases. For example, the mere decision to create an article about a topic like a specific person’s reported alien abduction legitimizes the idea (see the article on Travis Walton’s abduction at https://en.wikipedia.org/wiki/Travis_Walton). Interested individuals or corporate authors do not shy away from repeatedly intervening to maintain some basic facts for certain articles in a certain way. For example, recent documentary evidence appeared that the Russian KGB might have reused some of the Nazi bosses for Cold War espionage, especially the head of the Gestapo, Heinrich Müller. The claim is made by Tennent Bagley, a senior retired CIA officer, who interviewed and published the biography of a major KGB leader during the Cold War, Viktor Kondrashev, the head of the American counter-espionage division of the KGB (Bagley, 2013). Attempts by one of the authors of this article (SA Matei) to include this information in the Wikipedia article about Müller were met with fierce resistance from the most prolific editor of the article, an editor with the user name Kierzek. Kierzek’s user page (https://en.wikipedia.org/wiki/User:Kierzek) reveals that he is a circuit court mediator who contributes to many World War II articles (For the debate regarding the edit proposed to the Muller page see https://en.wikipedia.org/wiki/Talk:Heinrich_M%C3%BCller_(Gestapo)#Muller_recovered_and_used_by_the_ Russians:_We_need_consensus_on_adding_this_section_to_the_article). Furthermore, the debate about the

KGB—Muller connection remains hidden from view, as does the fact that the most productive contributor to the article has become a de facto gatekeeper. In this, as in the case of many other Wikipedia articles, the nature of the authorship process remains hidden in plain sight.

Of course, the edits and the debates are still on the site (see edits on July 29, 2014 at https://en.wikipedia.org/w/index.php?title=Heinrich_M%C3%BCller_(Gestapo)&action=history), but merely looking through a list of edits makes it very difficult to discern that most pages follow an uneven distribution, or that some authors have an important role in shaping the tenor and direction of an article. This dramatic inequality of contribution and narrative direction means that for a given article, while many people may make small contributions, a few people contribute most of it, and therefore have much more control over the nature of the document. This reality is qualitatively different from the assumption that most people hold, which is that Wikipedia is fairly open and democratic.

Because the true nature of how articles are created is hidden, most readers and new contributors believe that Wikipedia's content is simply the aggregation of edits from nearly random others. This serves as a motivator of sorts. People honestly try to add new content all the time. Typically, however, only the tidbits or raw material that fit with the narrative controlled by the overall editors is preserved. Ordinary casual users never know this. Those who attempt to make more consistent contributions ultimately learn that they need to befriend the leaders and become "one of them." They can become effective editors only by recognizing that there is a community behind the content, and that Wikipedia articles are the product of a large amount of coordination, conversation, and contention (Bryant, Forte, & Bruckman, 2005).

In addition, there are a number of policy decisions, technical decisions, and administrative decisions, all of which are hidden from the typical user. Deciding, for example, which types of articles should be deleted and which should be kept, or whether a certain user should be banned, all occur in the open, but in spaces on the site that are nearly impossible for new users to find.

In brief, authorship on Wikipedia is regulated by power structures. Some are explicit while other implicit. Some users have the explicit power to ban other users, lock articles, look up the IP address of other users, etc. These are the so-called admins (a few thousand), sysops, or bureaucrats (a few dozen). In addition, there is something of an "adhocracy": a small group of editors which makes many of the edits on the site. This group has been active on the site for a long period of time, with low turnover in membership (Matei et al., 2014). Although there is a large amount of overlap with the explicit leadership, these editors are not nominated, but they also shape the nature of the content and the community. This group is composed of under .1 % of the current mass of Wikipedia editors [of which there are over 20 million, according to a study for the period 2001–2010 by Matei et al. (2014)].

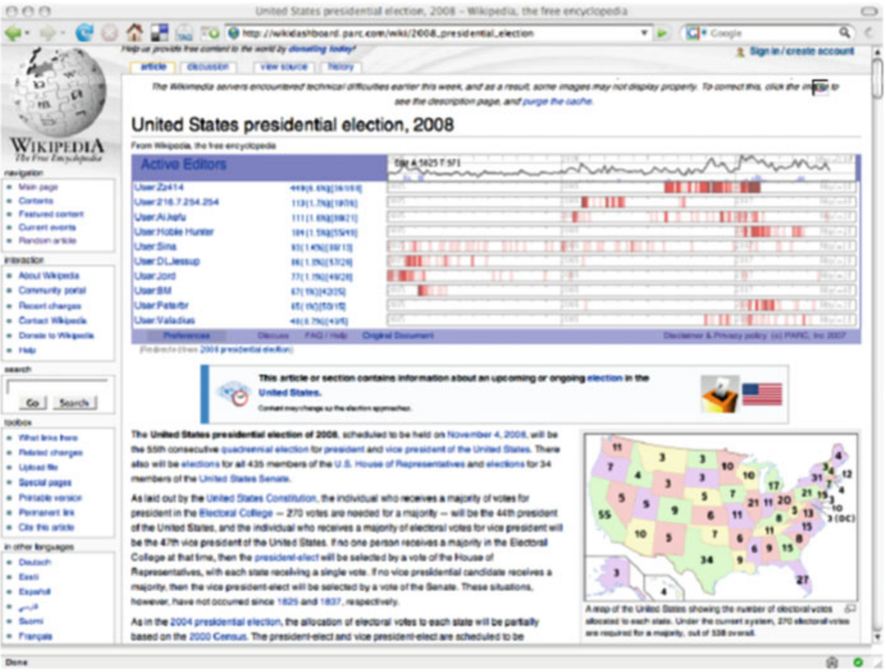
Power structures do not exercise their controls in a direct way all the time. Many times, power is inscribed in the design of the editorial tools. The edit page itself includes a number of features that are not obvious to new users. Despite the goal of transparency, the actual article page hides a lot. It doesn't show the history on the main page, doesn't show the talk page (a space for conversations about what the page should say), and doesn't show who edited each part of the article (Slattery, 2009).

The tabs the point to these features are minimalist and appear to be mere optional tools. These sorts of non-obvious features of a website are more likely to be discovered by those who already use the Internet in diverse ways (Hargittai, 2010).

5 Article-Level Visualizations

In response to these issues, a few visualizations have been proposed which are intended to actually be shown on the article page itself. These visualizations are designed to give information to readers and editors to make some of the inner workings of Wikipedia more transparent, and to help readers make more informed decisions about how credible the content is.

Suh, Chi, Kittur, and Pendleton (2008) created what they called WikiDashboard, a tool which includes a number of visualizations, one of which is active on the article page itself. It displays a list of each of the Wikipedians who have contributed to an article, together with a temporal visualization of their contributions.



Taking a different approach, WikiTrust is a project that attempts to add transparency to the actual content of Wikipedia articles (Adler et al., 2008). It changes the background color of the article text based on a trustworthiness algorithm, which takes into account how long text has been there and who authored it. New text, or text from less trusted users, is highlighted in a brighter color, while text that has been there for a long time (and has theoretically been reviewed by many others) is not highlighted at all.


the United States foreign policies

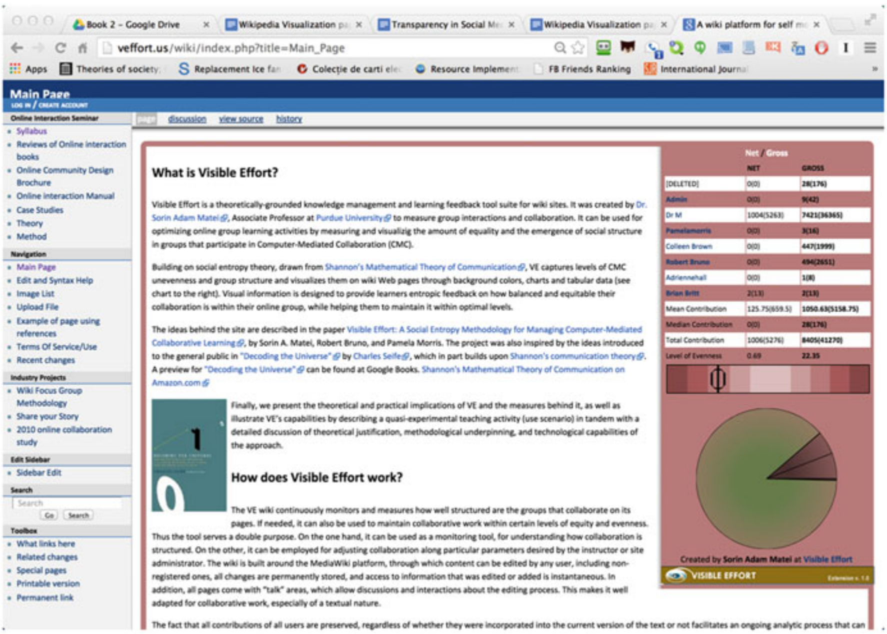
together with the leaders of his coalition partners selects the other Ministers which make up the Governments and acts as political heads of the various government departments. Cabinet members are occasionally recruited from outside the Folketing.

Since 27 November 2001, the economist Anders Fjogh Rasmussen has been Prime Minister to Denmark.

As known in other parliamentary systems of government, the executive, i.e. the Government, is answerable to the Folketing. Under the Danish constitution, no government may exist with a majority against it, as opposed to the more common rule of government needing a majority for it. It is because of this rule, Denmark often sees minority governments.

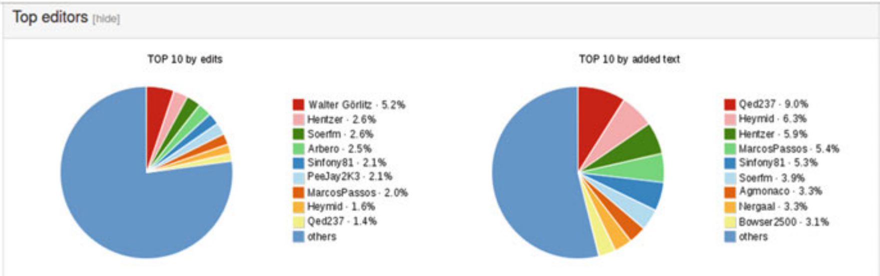
The WikiTrust interface highlights portions of Wikipedia articles which have recently been edited, indicating aspects of the article which may be less trustworthy

A related project, Visible Effort, makes the distribution of effort more visible on content pages. The project calculates the entropy for each page, lists the contribution amounts from the top editors, and changes the background color based on how unequal the contributions are. A horizontal bar indicates the level of entropy for each page, on a standardized scale between 0 and 100. This allows readers to identify pages which are primarily the work of one or a few people (Matei & Dobrescu, 2010 and chapter "Transparency, control, and content generation on Wikipedia: Editorial strategies and technical affordances" in this volume). At another level, it suggests the level of social structuration of any given article, since entropy is considered to be an index of social structuration, as explained in chapter Transparency, control, and content generation on Wikipedia: Editorial strategies and technical affordances of this volume.



6 Why Visualizations Haven't Been Accepted

As mentioned, even now, Wikipedia includes a few visualizations and statistics that are linked to from the article history page. These include the top editors, the number of views, and a chronological history of edits. These are much simpler than the tools proposed by academics, but they do still provide additional insight into the production of article content.



[1] Number of edits
[2] Average time between edits (days)

Username	Links	# [1]	Minor edits	%	First edit	Latest edit	atbe [2]	Added (Bytes)
Mjefm	ec · topedits	25	13	52%	2007-10-30, 15:41	2014-06-24, 22:06	97.2	1,766
ClueBot NG	ec · topedits	28	28	100%	2010-11-23, 01:46	2014-04-01, 21:01	43.8	79
Chanheigeorge	ec · topedits	65	9	13.8%	2006-12-15, 00:13	2014-07-14, 14:19	42.6	12,530
SQGibbon	ec · topedits	33	16	48.5%	2010-01-18, 07:29	2013-06-11, 11:44	37.6	351
Agmonaco	ec · topedits	35	0	0%	2011-10-20, 17:07	2014-06-29, 23:16	28.1	16,770
Chris1834	ec · topedits	63	27	42.9%	2010-06-11, 10:17	2014-07-10, 05:46	23.6	221
2nyte	ec · topedits	41	2	4.9%	2012-01-27, 07:46	2014-07-13, 14:22	21.9	4,008
Jlsa	ec · topedits	32	6	18.8%	2010-07-12, 06:33	2012-06-05, 11:40	21.7	1,060
PeeJay2K3	ec · topedits	143	35	24.5%	2007-12-20, 16:22	2014-06-25, 19:48	16.6	3,768
MarcosPassos	ec · topedits	141	3	2.1%	2012-06-13, 01:13	2014-06-03, 04:21	5.1	27,035
Hentzer	ec · topedits	181	0	0%	2009-11-16, 23:58	2011-12-28, 02:00	4.3	29,696
Walter Görlitz	ec · topedits	358	52	14.5%	2010-07-01, 19:57	2014-07-17, 03:38	4.1	10,473
Arbero	ec · topedits	176	16	9.1%	2012-09-28, 09:39	2014-07-15, 16:25	3.7	3,289
Rothorpe	ec · topedits	30	29	96.7%	2014-04-01, 17:59	2014-07-13, 22:03	3.4	42

However, none of the transparency visualizations created have made it onto Wikipedia article pages themselves. The pages remain as opaque as they have ever been, and indeed, they look nearly the same as they have always looked. If these tools are helpful in promoting trust and transparency, then we are led to ask why nothing has actually been incorporated into the article page, where users are likely to see it.

The discussion around WikiTrust gives some clues. In 2009, a *Wired* article reported that Wikipedia would soon be adding WikiTrust to article pages. Soon, users began discussing the proposed changes on the wikien-l mailing list. The conversation centered around a few themes. First, a few posters worried about the effect that this would have on the editors. For example, one poster said:

What's interesting about WikiTrust is that a trust score is computed for each individual. I wonder if these will be made public, and if so, how they will change the community of editors. It seems likely that they will not be made public. However, since the algorithm is published and I believe the source code as well anyone with the hardware could compute and publish how trusted each community member is.

Others questioned the validity and complexity of the algorithm for highlighting less trustworthy content. Finally, and relatedly, many of the commenters wrote about how the interface would be too confusing or too complex for readers.

One poster wrote:

The moment you give people a tool, many people will simplistically assume what it does or rely unthinkingly on it.

- WikiTrust might be described as "a way to see how long an edit endured and how much trust it seems to have"; in most users' hands it'll be "its colored red/blue so its right/wrong."
- People won't think, they'll assume and rely.

Another said:

If I understand this correctly, wouldn't trust coloring inevitably mark all new users and anonymous IPs as untrustworthy?

So, basically, wouldn't trust coloring be a way of failing to assume good faith for all anonymous IPs and new users, and institutionalising this in the software?

The overall tenor was certainly one of trepidation about making changes, and multiple posters wrote about maintaining the current experience for new and inexperienced users. While it is never written, there is a sense that these community members are concerned about pulling back the curtain, and in showing new users more than they are ready for. The implicit fear was that revealing too much would prevent new users from joining the project. In the end, the conservative viewpoint won out, and the plan to incorporate WikiTrust was abandoned.

Ideals of openness and freedom are cited as reasons that active participants edit in Wikipedia (Nov, 2007). However, there may be an unacknowledged, or even unconscious, fear of making some parts of Wikipedia more visible and transparent. Transparency might be dangerous to the project. As seen in the discussion about WikiTrust, Wikipedians are very wary about altering the experience for new users. Perhaps if readers see how uneven the levels of contribution are, or if new users know that their edits are likely to be reverted, they will be less likely to contribute. In a sense, Wikipedians may believe that the project is best served by keeping certain aspects somewhat hidden, until contributors have developed a stronger connection and dedication to the project, at which point the true nature is revealed.

Ironically, the reluctance to add greater visibility may also be driven by the hidden power structures on Wikipedia. Running these sorts of visualizations at scale on a site as large as Wikipedia requires both computing resources and programmer support. Researchers are generally not part of the programming community on Wikipedia, and may have difficulty convincing the community to take on the responsibilities of scaling and maintaining these projects. Indeed, many of the visualizations and statistics that do exist on the History page are external links to pages owned and maintained by individual programmers, supporting the idea that finding internal support for programming projects is difficult. The fact that other resource-intensive operations, such as full history dumps of the Wikipedia data, have been discontinued due to expense and difficulty, provides further evidence.

A final, related explanation for resistance to change is suggested by Shaw and Hill (2014), who looked at thousands of Wikia.com communities, and found that communities are inherently conservative, with early contributors holding much of the power. We can assume that those who are active on these sites participate because they agree with the overall goals of the site. In addition, they have spent time becoming expert in the current configuration. Therefore, suggestions of major changes to the site are more likely to be rejected by these users.

7 Possible Solutions

We offer a number of suggestions for those wishing to introduce tools to increase the transparency of Wikipedia articles, in a way that is beneficial both to contributors and to readers. Contributors and project leaders have an interest in recruiting new contributors, and maintaining current contributors, while readers have an interest in judging the trustworthiness of content, and in seeing how the encyclopedia is produced.

We suggest that tools need to be unobtrusive. The main goal of Wikipedia is the production and dissemination of knowledge, and modifications which seem to undermine or distract from this purpose are unlikely to be implemented. For example, a small warning that appears only if entropy is greater than a certain threshold, or if there are untrustworthy edits, may be more likely to be accepted. There are already manually created warnings about needed citations, articles that need to be cleaned up, etc. Automated warnings could fit this same framework, and provide increased transparency.

Academics should also be encouraged to work more closely with Wikipedia developers throughout the process of developing tools. These projects require integration into the Wikipedia socio-technical system, and researchers who work with current developers will be much more likely to overcome the technical and political barriers to successful implementation. Working together, researchers and the Wikipedia community can provide tools to make the processes of Wikipedia as open and transparent as its content.

References

Adler, B. T., Chatterjee, K., De Alfaro, L., Faella, M., Pye, I., & Raman, V. (2008). Assigning trust to Wikipedia content. In *Proceedings of the 4th International Symposium on Wikis*. New York: ACM.

Bagley, T. H. (2013). *Spymaster: Startling cold war revelations of a Soviet KGB chief*. New York: Skyhorse.

Bao, P., Hecht, B., Carton, S., Quaderi, M., Horn, M., & Gergle, D. (2012, May). Omnipedia: Bridging the Wikipedia language gap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1075–1084). New York: ACM.

Biuk-Aghai, R. P., Pang, C.-I., & Si, Y.-W. (2014). Visualizing large-scale human collaboration in Wikipedia. *Future Generation Computer Systems, 31*, 120–133. doi:10.1016/j.future.2013.04.001.

Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 1–10). New York: ACM.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature, 438*(7070), 900–901.

Hargittai, E. (2010). Digital na(t)ives? variation in internet skills and uses among members of the "net generation". *Sociological Inquiry, 80*(1), 92–113. doi:10.1111/j.1475-682X.2009.00317.x.

Hill, B. M., & Shaw, A. (2013). The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PLoS One, 8*(6), e65782.

Hughes, B., Joshi, I., Lemonde, H., & Wareham, J. (2009). Junior physician's use of Web 2.0 for information seeking and medical education: A qualitative study. *International Journal of Medical Informatics, 78*(10), 645–655.

Keegan, B., Gergle, D., & Contractor, N. (2013). Hot off the wiki: Structures and dynamics of Wikipedia's coverage of breaking news events. *American Behavioral Scientist, 57*(5), 595–622.

Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web, 1*(2), 19.

Matei, S. A., Bruno, R., & Morris, P. (2015). Visible effort: Visualizing and measuring group structuration through social entropy. In S. A. Matei, M. Russell, & E. Bertino (Eds.), *Transparency in social media—Tools, methods and algorithms for mediating online interactions*. New York: Springer.

Matei, S. A., & Dobrescu, C. (2010). Wikipedia's "neutral point of view": Settling conflict through ambiguity. *The Information Society, 27*(1), 40–51.

Matei, S. A., Tan, W., Zhu, M., Bertino, E., & Liu, C. (2014). Elite size and resilience impact on global system structuration. In *The 2014 International Conference on Collaboration Technologies and Systems (CTS 2014), Minnesota, May 19–20*.

Nov, O. (2007). What motivates Wikipedians? *Communications of the ACM, 50*(11), 60–64.

Rector, L. H. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review, 36*(1), 7–22.

Shaw, A., & Hill, B. M. (2014). Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication, 64*(2), 215–238.

Slattery, S. P. (2009). "Edit this page": The socio-technological Infrastructure of a Wikipedia article. In *Proceedings of the 27th ACM International Conference on Design of Communication* (pp. 289–296). New York: ACM. doi:10.1145/1621995.1622052.

Suh, B., Chi, E. H., Kittur, A., & Pendleton, B. A. (2008). Lifting the veil: Improving accountability and social transparency in Wikipedia with wikidashboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1037–1040). New York: ACM.

Swarts, J. (2009). The collaborative construction of "fact" on Wikipedia. In *Proceedings of the 27th ACM International Conference on Design of Communication* (pp. 281–288). New York: ACM. doi:10.1145/1621995.1622051.

Viégas, F. B., Wattenberg, M., & Dave, K. (2004, April). Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 575–582). New York: ACM.

Voss, J. (2005). Measuring Wikipedia. In *Proceedings of 10th International Conference of the International Society for Scientometrics and Informetrics, Stockholm, Sweden*.

Wattenberg, M., Viégas, F. B., & Hollenbach, K. (2007). Visualizing activity on Wikipedia with chromograms. In *Human-Computer Interaction–INTERACT 2007* (pp. 272–287). Berlin, Germany: Springer.

Yasseri, T., Spoerri, A., Graham, M., & Kertész, J. (2014). The most controversial topics in Wikipedia: A multilingual and geographical analysis. In P. Fichman & N. Hara (Eds.), *Global Wikipedia: International and cross-cultural issues in online collaboration*. Lanham, MD: Scarecrow Press.

Part IV
Transparency in Social Media: Ethical and Critical Dimensions

Truth Telling and Deception in the Internet Society

Robert B. Laughlin

1 Truth or Consequences

All human communication involves value propositions. One is accustomed to describing this idea with less inflammatory language, for example as friendship, networking, gaming or academic inquiry, but all of these things boil down in the end to strategies for getting something of value, or perceived value, for oneself. The person with whom one is communicating does the same. Both parties strive to get something for nothing, and both parties engage in psychological tactics to persuade the other party that they have succeeded. Game playing is a major component of this give-and-take, but it is not the core objective. The core objective is to get things for oneself by trade or stealth, always paying as little as possible.

This world view is obviously exaggerated, but it is a helpful starting point for clear thinking about the darker aspects of electronic communications technologies (Fish, 2011). It helps us avoid being dazzled by the Internet as the ultimate instrument of democracy (Balkin, 2004; Best & Wade, 2009; Gotlieb, 2002; Margolis & Moren-Riaño, 2009) or the idea of bandwidth as a resource, like oil or diamonds, that makes us all richer when we have more of it (Gilder, 2000; Orcutt, 2012; Reisinger, 2009; Wu, 2010). Reasoning logically through principles of self-interest leads directly to the question of value and from there to a string of extremely sobering observations: the death of privacy (Andrews, 2012; Froomkin, 2000; Garfinkle, 2001; Wicker, 2013), the death of journalism (Gerson, 2009; McChesney & Nichols, 2010; Nesbit, 2013; Shepard, 2012), the death of books (Bosman, 2012; Franek, 2010; Yi, 2011), the death of civil discourse (Brooks, 2011; Hoyt, 2007; Levmore & Nussbam, 2011), and perhaps even the death of truth itself (Damon, 2011; Gardner, 2011; Shields, 2011; Stanley, 2005), an astonishing idea captured by Stephen Colbert's wonderful neologism "truthiness" (Fig. 1).

R.B. Laughlin (✉)
Department of Physics, Stanford University, Stanford, CA 94305, USA
e-mail: rbl@large.stanford.edu; http://large.stanford.edu

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_14

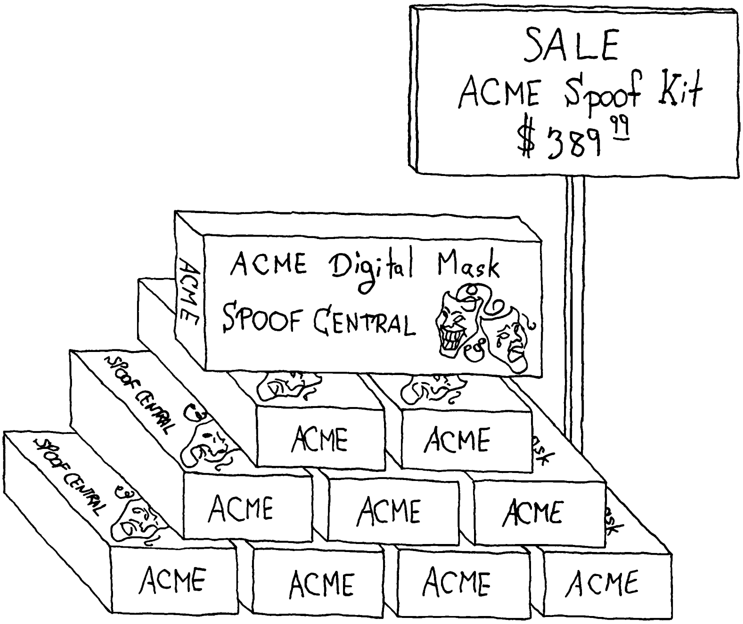


Fig. 1 The software industry creates products that enable people to distinguish themselves from the competition through mastery of the latest technology rather than through what they say.

Viewing speech as economics reveals immediately why electronic means are causing these problems: They vastly strengthen people's power to deceive without, at the same time, increasing their power to detect deception. Electronics have changed nothing qualitatively. The fundamental principles of rhetoric are the same as they always have been (Aristotle, 2004; Plato, 2008; Sprague, 2001). Electronic tools have simply given people the ability to do something they have always done a million times better. Thus, I log in anonymously and slander people and organizations I don't like (Klein, 2011; Palmer, 2013; Wilson, 2002; Wyer, 2011). I start a mass whispering campaign with an email macro (Carlson, 2011; Farhi, 2011; Parker, 2012). I generate fake ratings and opinion polls to boost sales (Charmen-Anderson, 2012; Dellarocas, 2006; Streitfeld, 2011; Tuttle, 2012). I stalk coeds under an assumed name (Abrahamson, 2013; Barbaro & Corasiniti, 2013; Reston, 2013; Zamost & Griffin, 2013). I steal artistic works and pretend that someone else did it (Bilton, 2012; Hinduja, 2008; Verrier, 2013; Yiannopoulos, 2011). I plagiarize (Brumfield, 2013; Erlangen, 2013; Gabriel, 2010; Kaufman, 2013; Nocera, 2013; Williams, 2010). I lie about my accomplishments (Grens, 2012; Irvine, 2013; Jan, 2010; Reich, 2010; Stewart, 2012; Wade & Choe, 2006). I create highly detailed and believable fictions and misrepresent them as fact (Allen, 2011; Itzkoff, 2013; Milmo, 2007; Wallace, 2013). And it isn't just individuals at fault. Organizations are empowered in the same way, and they do the same things, just on a larger scale.

The problems of electronic communication are therefore not little glitches on the way to the happy digital future but systemic failure of an institution so fundamental that nobody in the past bothered to make laws protecting it, the everyday give-and-take of speech. The argument that high-octane deception is acceptable because it's a feature of commerce, and commerce is good, cuts very little ice when the house one just bought is found to be infested with termites. It is also not correct that electronic deception is a rare deviant behavior that doesn't matter because good people don't deceive. That is like saying that blowing bank vaults doesn't matter because good people won't pick up the money. Deception is perfectly natural behavior (Kornet, 2012). It was just formerly held in check by the debunking powers of the deceived. The issue is not failure of a few people's self-control but failure of regulatory machinery designed by nature to prevent people from doing selfish things that they want to do but shouldn't because those things are hurtful to everyone else.

Absent a decision to turn the Internet off completely, a thing unlikely to happen for many reasons, fighting back against electronic deception will probably require the invention of new institutions different from any that have existed before. The reason has to do with costs. Present-day journalism is failing not because its quality has declined but because rapid electronic dissemination of news by Internet short-circuits the profit models of organizations generating the material, leading them to cut back on the more expensive parts of their operations, in particular investigative reporting. Trash information is proliferating not because people are more irresponsible but because electronic machinery has greatly lowered the cost of making false or irrelevant things look true and important. The invention that needs to be made is thus a means of distinguishing truth from falsehood that costs very little. This invention is not going to be a falsehood-detecting robot. Truth and falsehood in the sense of speech are economic concepts, not objective ones, and it is therefore fundamentally impossible to distinguish them by mechanical means. The invention will have to be something that involves people and that discourages them from generating falsehood in the first place.

2 The Principle of Witness

Fortunately, inducing people to tell the truth is actually very easy. All that's required is the power to make them comply with a rule or two. The reason is that, while everyone lies, very few of us are willing to get caught doing it. Thus, all that is usually required to obtain truthfulness is to compel the speech to take place in front of witnesses. This leads automatically to more circumspection because the person does not want to be seen in public as either dishonest or unreliable. Consumers of what is said also talk among themselves to get to the bottom of things. The more witnesses there are, the greater the pressure on the speaker to tell the truth. The pressure is entirely self-imposed. It isn't necessary to browbeat or bully. The mere threat of being whispered about behind one's back is sufficient.

A record of what was said increases the pressure because it enables third parties not present at the time to become witnesses.

However, for this simple mechanism to work effectively, several additional elements need to be in place. For one thing, the witnesses have to pay attention. They won't do so if the speech is boring, confusing or excessively long. Those things therefore have to be blocked. The speaker also has to be identifiable as the responsible party. This requires a byline that reveals his or her identity clearly, not a pseudonym or a committee of which the speaker is a member. The speech must also be recorded permanently and be displayed in public, so that responsibility for it cannot be later disavowed. This requires it to have a specific creation date and to be stored functionally forever in a place protected from post facto changes and easily accessible to anyone. Sequential updating must be banned.

These extra elements of responsibility are, of course, functionally the practices of present-day academic journals, at least in theory. Thus one might say that journals exist to enforce the Principle of Witness, and that this is why we have them. However, that is not quite right. Journals are businesses that exist to make profit. They compel the truth, if at all, as a means to this end. The profit is crucial, as it turns out, because maintaining records over the long term is expensive, and institutions that try to do it without a budget all wind up failing in the end. Unfortunately, the present-day profit needs of journals are so great that they have an even bigger cost problem than investigative newspapers and magazines have. The buyers, in this case university libraries, are tired of paying for things they can get for free on the Internet and don't want to do it any more. The profit squeeze then prevents journals from making themselves viewable to everybody because they will lose the cash flow of their remaining paying customers if they do. They are inherently secretive as a result, and find themselves forced into the business of preventing people from knowing things unless they pay. Journals also select and edit material, a practice central the journal's image and profit model but also institutional censorship in all but name. The market relationships of journals also tend to make them jargony and unreadable except by specialists, and often not many of them. Thus, journals, as they exist today, are not a viable option for pushing back against mass electronic deception, although they contain elements of the answer.

The institution that needs to be invented is thus something that achieves the Principle of Witness at least as completely as journals or investigative news instruments but at much lower cost. The specific delivery vehicle does not matter, so long as it is cheap. Cash flow and managing the value proposition for the speaker matter much more than technological details. This is highly nontrivial because deception is generally profitable while telling the truth isn't. Everybody knows that one can sell something that is not true but cannot sell something that doesn't look good. Crowdsourcing is not the solution. The large amount of wrong information on Wikipedia is proof of this.

The religious overtones implicit in the concept of witnessing are not accidental. The ideas and practices in question have roots in religious traditions and come to us indirectly through Roman law (Buckland, 2007).

3 Evasion Tactics

However, the task of compelling truth is complicated by people's endless creativity at evading the Principle of Witness while appearing to comply with it. This is why advances in communications technology never produce responsible speech all by themselves. The speakers simply use the extra power to disguise their deceptions in ever more clever ways. Analyzing their evasive tactics and blocking them is central to achieving the desired result.

One simple way people avoid responsibility for speech is to deny authorship. In the old days, one might have achieved this by pasting anonymous flyers to walls or publishing a book under a pseudonym. Today people do it most commonly by posting unattributed web pages. There are lots of variations on this theme. There is the ordinary blog, a personal Internet posting that may or may not be identifiably authored. There is the chat page with proliferating avatars. There is the official information page with no identified authors or references, but rendered very slickly and professionally so as to appear to be an authoritative source of what "they say". There is the government data site with no author and no corresponding official paper publication with file number. There is the (functionally anonymous) Wikipedia posting with reference links to documents on some server somewhere lacking accreditation, authors or dates.

Another way people avoid responsibility for speech is by rendering it obsolete. This is especially easy to accomplish with web sites because they are designed to be easily changed. Whenever the content becomes inconvenient, one just changes it, thus erasing all memory of the original statement. In a world where this is allowed, things are never wrong, just out of date. In extreme emergencies one can make a web page disappear by moving the relevant file to an inaccessible place in the computer or turning the computer off. This tactic is partly countered by web crawlers, the job of which is to acquire and store web page content for later access by search engines. But the copies they generate do not live forever, so the memory is eventually lost. For either web sites or electronic documents, one can leave off a creation date, thus allowing later re-issue with modifications. This is the same thing as claiming later that the original was only a draft and not what one really meant to say. A variation on this theme is the issuing a continual stream of formal updates and upgrades.

Another way people avoid responsibility for speech is by hiding in excessive mass. A well-known example of this tactic is the monster report so bristling with tables, appendices and unnecessary facts that an important mistake or misrepresentation then gets lost in plain sight like a needle in a haystack. The extreme modern version of excessive mass is video, the information transfer rate of which is 10,000 times greater than the human brain can process.[1] One does not worry about this

---

[1] Verbal speech rate is 100 words per minute. Assuming 5.1 letters per word and 5 bits per letter (speech effectively has punctuation marks), the baud rate of speech is 42.5 baud. Typical human reading speed is 250 w/min or 106.2 baud. A telephone audio channel is about 8,000 baud. High fidelity sound is 200,000 baud. Standard analog television is 3,500,000 baud. High-definition digital television is 19,000,000 baud. These audio and video rates are known mainly through compression technologies. The full capacity of a video channel, including improbable frames that one would never see, is much greater.

overcapacity when one is being entertained, for the video maker has gone to great lengths to make sure one perceives certain things and that one is comfortable. But in a reporting context, it is easy for even an unsophisticated person to pack a video with so much imagery that the viewer becomes overwhelmed with perceptions and cannot figure out whether there is any central content or not. This is why video never works as a vehicle for storing company records. Video and audio files are also conveniently inaccessible to computer search, so that errors or misrepresentations in them cannot be detected easily. Another notorious example of the excessive mass tactic the slick Power Point presentation so packed with detail and zoomed through so quickly that no human being could possibly follow it (Cyphert, 2004; Feith, 2009; Shaer, 2011).

Another way people avoid responsibility for speech is by employing non-standard communication formats. In writing, for example, this might be a custom font, a special layout procedure or the latest upgraded file protocol of a popular word processor. On the Internet or in a Power Point presentation, one might spice things up a bit more with twirling soccer balls, flash movies, blinking red banners, and numbers that swoop down and land in the appropriate box in a table. Speakers love this tactic because it enables them to distinguish themselves from their competition on the basis of looks rather than content. Unfortunately, it also make comparison of one work with another difficult, if not impossible, and it blocks computer search. The latter makes it especially effective when used in combination with excessive mass.

Another way people avoid responsibility for speech is by being intentionally confusing. This is somewhat sensitive because it is difficult to distinguish from actual confusion, which is not a tactic but a genuine communication difficulty. But much of the time confusion turns out to be a tactic. It nicely supplements excessive mass by requiring extra mental effort to wade through and find errors and deceptions. In writing, confusion is accomplished by means of jargon, awkward paragraphing, undefined symbols, poor organization, and deliberate non-sequiturs. Thus one might wish to implement standard object principles by linking the several subcategories of concept threads together into a portable class object, thus bringing the picture into registry with the analysis of Hendricks and Charleston. The opportunities for deliberate confusion multiply enormously when the content is technical and contains mathematics with lots of exotic symbols.

4 Robots Have Limitations

Creating a low-cost instrument of truthful speech thus comes down to the problem of automating, as much as possible, the enforcement of the Principle of Witness.

That robots can lower the base cost of submission and delivery of content is now well established. Digital delivery of conventional newspaper content is commonplace, although the revenue model associated with it is still unstable (Lee, 2013; Satell, 2013; Stynes & Launder, 2013). The most impressive example is YouTube, a company that provides free (for now) distribution of uploaded video supported by advertising revenue solely. YouTube's financial details are unfortunately closely guarded industrial secrets, but estimates for its operating budget are typically $1 billion per year (Bradshaw, 2013; Hamilton, 2013; Manjoo, 2009; Prabhu, 2012; Tsukayama, 2013). YouTube claims to be serving six billion hours of video per month at a default rate per video of 500,000 bits per second. A better documented example of robotic cost reduction is the Cornell physics bulletin board (Bachrach et al., 1998; Fischman, 2011; Ginsparg, 2011; Guinnessy, 2010; Skorton & Altschuler, 2013). Trusted authors, typically professors of physics and their graduate students, post written works by browser in a standard format. Computers then check the formatting and, if they find it satisfactory, date-stamp the work, file it, add the title to the archive's running table of contents, and make it world-viewable. For copyright and political reasons the archive represents that it is not a publication, but this is largely a fiction. Faculty routinely refer to it among themselves as "The Journal." It has 30 million downloads a year and has a 2012 operating budget of $500,000. It has processing costs per uploaded paper of approximately $7.

However, the ability of robots to enforce the Principle of Witness all by themselves is at best unproven and at worst highly questionable. Both YouTube and the Cornell bulletin board simply abrogate their enforcement responsibilities by redefining the editing task, in the first case allowing communications with an arbitrarily large number of bits and in the second by restricting the submitter pool to a set of people with an economic agenda of a very special kind. Both use robots as editing tools but rely on actual people for the final decisions about filtering uploads for copyright infringement and inappropriate content. Both minimize the (large) costs of human intervention by leveraging unpaid viewer critiques. Both transfer editing costs to the submitter by severely restricting submission formats.

Unfortunately, one cannot just define the editing problem away. Except for high transmission speed, which most ordinary servers do not have at the moment, YouTube does nothing conceptually different from an ordinary web site. It is fundamentally impermanent and anonymous, and its high bit rate is an extreme deceit opportunity (Blake, 2013; Moore, 2008; Mosher & Hernandez, 2012; Posner, 2012; Steel, 2012). The Cornell bulletin board has much more stability and credibility, but it achieves this at the cost of being inaccessible to just about everybody. One could never achieve the necessary communication discipline outside the tight confines of a small professional community. In addition, the archive suffers from impermanence and excessive mass. It is a purely electronic enterprise run on a shoestring, so it cannot be relied upon to exist on very long time scales. It allows postings to be revised, although with a permanent revision log. The absence of length limits in the archive also makes it very difficult to distinguish true and important things in it from false and unimportant ones.

Thus the core problem of enlisting robots to help enforce responsible speech is not yet solved. For better or worse, computers are very good at archiving and retrieving enormous files quickly, but they are not very good at identifying deception made deliberately by humans and taking steps to discourage it. The humans are just too smart. Thus, absent a great future breakthrough in computer intelligence, there will continue to be a need for a human editor at a key place in the loop who says, in effect, I see what you're doing and I want you to stop.

5 Using Computers Correctly

Fortunately, it is fairly easy for an editor to use computers in a way that will coax out truth rather than encourage ever more elaborate falsehood. All of the key elements are already deployed in various applications. What's required is to bring them together in one place. One must do this for the same reason one must plug all the holes in a bucket before using it to carry water. If one misses even one hole, all the water will flow out there.

First of all, one must ban anonymity. The only known way to compel truthful speech is to make the speaker personally responsible for it. Any kind of anonymous speech therefore has the potential to be deceitful. This is not to say that one bans anonymous speech generally (Flock, 2011; Gapper, 2011; Ingram, 2011; Jeffries, 2013; Kravets, 2012; Pfanner, 2011; Rooney, 2013; Wallace, 1999). The political implications of that would be very grave, and anyway it is technically impossible. One only bans anonymity in speech wishing to acquire accreditation. Not only must one disallow works without bylines, one must disallow works by committees or companies, where the responsibility is shared or diffuse, or those by fake authors or temporary symbolic ones, for example using the U.S. Secretary of Energy as the author of a report on oil supplies that is actually written by Department of Energy staff. Computers are not smart enough to see through all the subtle ways people can disguise authorship, so a human editor has to do it. Computers can, however, help with the sorting-out process and double-checking that no mistake has been made.

Secondly, the speech must have a fixed creation date that cannot be changed post-facto or disavowed by means of addenda or updates. It must also be posted in plain sight in perpetuity. These things go together. A date stamp on a file doesn't matter if nobody sees the file. A file that has no date stamp is not yet avowed and thus not worth seeing. Computers are obviously superb for posting things in plain sight, for it is hard to imagine anything more visible than world-viewability on the Internet. They are not, however, very good at maintaining records over long times. The reason has to do with the physical nature of their memories. These are deliberately engineered to be easily changed because computation is functionally the same thing as sequential modification of memory. All computer records are therefore inherently volatile. They can be made to last without corruption only through assiduous management by human institutions. Not surprisingly, many people prefer to avoid the corresponding long-term maintenance costs by defining their speech to be a product, like a bar of soap, that is impermanent. Unfortunately, this is exactly the opposite of what one must do to achieve responsibility. Not only must the computer records of the speech be fixed, computers must also be enlisted to track down and counter clever disavowals made, for example, by aggressively broadcasting a revised version of the work on the Internet, thus rendering the original invisible, or re-publishing a revision under a different title, thus rendering the original obsolete.

Thirdly, the speech must be incorruptible. The wide electronic visibility required for truthfulness necessarily implies that the work can be copied, modified and re-used by third parties. This means there is a danger that it can be digitally misrepresented. In a commercial environment, such misrepresentation is blocked as a side effect of copyright. One legally transforms the speech to property and then prosecutes the copying of it as theft. This practice has notorious logical inconsistency that perceiving the speech cannot be theft if one's objective is to compel truthful behavior through visibility. All barriers to transparency are deceit opportunities. This problem forces news outlets wishing to maintain journalistic credibility to make an artificial distinction between speech that is perceived and remembered, which is not theft, and speech that is copied using machines, which is. This distinction gets more and more nonsensical the smaller the file in question becomes. The core problem is that copyright law is designed for preventing intellectual theft, not preventing intellectual falsehood. It often fails to do the latter (Davidson, 2013; Knapp, 2011; Miller, 2004; Switak, 2012). But it is perfectly possible for an institution to guard against corruption without copyright by digitally watermarking its files, or for that matter by simply making bit-by-bit comparison of the copied file with the original one upon request.

Finally, one must stop the use of excessive mass. The obvious way to accomplish this is to require communication in plain prose. It is often claimed that writing is an outdated method of communication, but this is only when the speaker is trying to deceive. Writing is a perfectly adequate vehicle for saying anything important, and it has the immense advantage over all other vehicles in having a low data transmission rate. A speaker can get around this advantage with excessive length, so a length limit is also required. Non-standard formats must be banned for the same reasons. Making the work widely viewable then takes care of the rest. If one writes something long that is incomprehensible, people shrug and chalk it up to troubled genius. If one writes something short that is incomprehensible, people see a fool, a liar or both. Short written pieces with correct bylines and publication dates appearing in plain sight are almost never incomprehensible for this reason.

6 Institutions Are Essential

A publication institution of some kind is required to accomplish these things. The free, unfettered Internet cannot do it. The reason is that the people doing the paying, the speakers themselves, benefit from the power imbalance. People buy software for economic purposes, which is to say, to advertise themselves. Software manufacturers then satisfy this need by making products that enable a person to appear superior. In the process they secure for themselves ongoing business through never-ending upgrades. An audience may wish that a presentation should be thoroughly comprehensible, on point, and verifiable but it is not in the presenter's best interest that this should be so. It is in the presenter's best interest to be held accountable for as little as possible, to make all permanent record of the speech event disappear, and to rise above the pack using the latest technology rather than content. Plain language is a perfectly good method of communicating, but only mastery of a recently upgraded software package enables one to demonstrate that one savvier than everyone else when it comes to technology. Volatility and deceitfulness of electronic speech is therefore not pathological at all but a normal and logical consequence of market principles.

7 An Experimental Test

An appropriately structured university course is an ideal venue for testing these ideas about obtaining responsible speech at low cost. There is a pool of smart participants who need to hone and demonstrate skills of public speech. The market is already created by the parent institution. The participants are effectively being compensated for their work through receipt of a degree. Good writing practices are something that universities are supposed to inculcate. And there is a pool of technically sophisticated faculty and graduate students who can write the necessary computer programs and manage the servers at low cost. The combination of value propositions is very different from that of open-market publishing, particularly as regards tuitions, government support and teaching compensation, but this is acceptable in a proof-of-principle experiment.

For the past several years I have been using several courses I teach at Stanford to conduct such experiments. The latest and most important of these are on the subject of energy. This choice is deliberate. Energy and climate are both scientific and highly technical matters involving numbers but, at the same time, quintessentially political. Present-day public speech about them is highly polarized (Alter & Fishman, 2013; Bryce, 2009; Chazan, 2012; Gillis, 2013; Handley, 2013; Hebbert, 2004; Hendrickson, 2012; Inhofe, 2012; Pannett, 2013; Schulz & Becker, 2013; Silverman, 2013; Tertzakian, 2012; Tvergerg, 2012; Washington & Cook, 2011). Electronic rhetorical tools are routinely unleashed with fangs bared. Enormous amounts of money are flying about for the purpose of influencing public opinion (Bachman, 2011; Bell, 2013; Cryderman, 2013; Goldenberg, 2013; Harvey, 2011; Jenkins, 2013; Mayer, 2010; Mullins, 2008). All parties claim that their (mutually exclusive) positions are truth. This is thus a suitable arena in which to test whether compelling speech that conforms to rules can enforce responsible behavior automatically through the Principle of Witness (Fig. 2).

The courses have several management features essential for conducting a balanced experiment. They are voluntary and not a requirement for any degree. This is essential for guaranteeing that the speech itself is voluntary and not coerced. Individuals enroll only if they are sufficiently interested in the subject and are willing to perform public-domain writing. The courses are aggressively advertised so as to draw in participants from a wide variety of backgrounds. These wind up ranging from very applied things such as Petroleum and Mechanical Engineering through the pure sciences of Physics, Chemistry and Biology to abstract subjects such as Management, Mathematics and Economics. The writing task itself is completely unlinked from either lecture material or grading. This is essential for

PH240 - Fall 2014
Introduction to the Physics of Energy

http://large.stanford.edu/courses

Henry	Anderson	The Future of Electric Cars
		Levi's Stadium Efficiency
Alex	Blandino	Tidal Power
		Solar Thermal Energy
Natalie	Burkhard	American Garbage
		Why Invent the Hyperloop?
Carla	Co	US CO_2 Emissions from Electricity Generation
		Hydraulic Fracturing Facts
Alexandra	Crerend	Asteroid Mining
		Graphene-Based Lithium-Ion Batteries
Zheng	Cui	When Is the Future for Synthetic Fuel?
		Underground Coal Gasification
Noor	Davis	Tesla Hyperloop
		Solar Powered Pool Management
Matthew	DeGraw	The Future of the Indian Point Nuclear Energy Center
		Keystone XL Debate Update November 2014
Lucy	Dikeou	Ethanol Efficiency
		Efficiency of Incandescent Light Bulbs Vs. Florescent Light Bulbs
Colin	Epperson	European Union Sets 2030 Energy Targets
		United States as an Energy Superpower: The Role of Natural Gas
Alyssa	Fujimoto	Energy Harvesting Flooring
		The Merits of Recycling
Oscar	Galvan-Lopez	Energy After Fossil Fuels
		The Cost of Pumped Hydroelectric Storage
Dustin	Gerrard	High Altitude Tethering of Wind Turbines
		Greater Energy Efficiency via Self-Driving Cars
David	Heinz	Small Scale Generation for Electrification of Rural and Remote Areas
		Carbon Emissions From Electric Passenger Vehicles in the United States
John	Ho	Wireless Power Transfer
		Glucose Fuel Cells
Charlie	Hopkins	Ecological Damage from Hydroelectric Power
		Hydrogen Fuel Cells: Future or Fool Cells
Mason	Jiang	Large Scale Energy Storage
		Ultrafast Science and Energy
Rachel	Kalick	Off Shore Wind
		Deep Water Drilling
Gautam	Krishnamurthi	Changes in Energy Production in the US
		Three Gorges Dam: Power Generation, Economics, and Impact
Lo'eau	LaBonta	Human Energy Converted to Electricity
		Mining Distresses Water
Alexander	Liegl	Peak Oil Theory
		Deepwater Drilling
Hsiao-Hsuan	Lin	Vehicle Charging Stations
		Energy Demand Management
Xinyan	Liu	Flexible Batteries: Status and Prospects
		Lithium-Air Batteries: Liquid or Solid?
Dallas	Lloyd	Hydroelectricity in Brazil
		Wind Energy: Advantages and Disadvantages
Rachel	Marincola	The Future of Oil
		Powering Space Vehicles
Emily	Mcmilin	Powering Mobile Base Stations
		Ambient RF Energy Harvesting
Bojan	Milic	Synthetic Biology and Whole-Cell Simulations
		Bioremediation of Oil Spills
Reed	Miller	Wind Energy: Why We Don't Use It
		Molten Air Batteries
Ty	Montgomery	Tar Sand Extraction
		Hammarby Sjöstad Waste Collection
Kyle	Murphy	The Future of Solar Cars
		Wind vs. Nuclear Energy
Kyle	Olugbode	Food Energy: Effect of Meat Consumption on Energy
		Curtain Wall Efficiency
David	Parry	Coconut Oil Biofuel in the Pacific
		Corn Ethanol Use in the Midwest
Morgan	Pope	Food-Miles and the Cost of Eating
		Paperless Classrooms
Nick	Rolston	The Blue LED: Prospects for Next-Generation Lighting
		Reliability of Organic Photovoltaics
Zhi Wei	Seh	Lithium-Sulfur Batteries: Challenges and Solutions
		Titanium Dioxide Photocatalysts
David	Sell	Fiber Optic Power
		Laser Gain
Prastuti	Singh	Quantum Dot Solar Cells
		Power Grid Inefficiencies
Arul	Suresh	An Efficient Future for Aviation?
		Solar Roadways: The Roads of the Future?
Tyler	Thorne	U.S. Offshore Wind Energy
		Landfill Gas Energy
Conrad	Ukropina	Stationary Fuel Cells: Reliable Back Up Power
		Kerosene and Gasoline Flash Point
Mengyao	Yuan	Rethinking Hydrogen
		Managing Energy in Fertilizer Production and Use

Fig. 2 Screen shot of the final publication table for the most recent PH240 writing experiment, which concluded 15 Dec 14. While the experiment is running, the site is password-protected, and submissions become visible to other writers when posted but not to the world. Submissions are made through an electronic portal that forces communications standards to be primitive. When the experiment concludes, the password protection is removed, making the material world-viewable. Participants license their work for public use on the Internet in perpetuity. Submissions are edited for formatting, integrity of references, and respect of copyright only. Content is entirely self-generated, including topic choice, and is unlinked from any institutional reward or punishment (grade). The sole arbiter of quality is fear of being seen wanting on the Internet.

preventing the participants from writing to please authority. The individuals who choose these courses are extremely smart, and they are able pick up on even very tiny hints about the instructor's personal preferences and write to them. When this happens it is a disaster, for the person has then managed to evade personal responsibility for the core concept of the speech, the single most important aspect of the exercise. For this same reason the instructor also does not critique content during editing. Submissions are checked only for formatting, length, potential copyright problems, and integrity of references.

The use of computers is essential for keeping management costs under control. Some of these techniques are well known and commonly implemented, but others are not. The course syllabus resides on a password-protected web site visible to individuals enrolled in the course but no one else. The Power Point files used in the lectures are uploaded to the server and linked on this page as they are given. Real-time communication with the writer pool is handled both through this page and by email. In addition, however, there is a fleet personal cloud editors, one assigned to each author and also individually password-protected to that author. These live on the course server and are linked to the syllabus page. They immensely simplify the imposition of formatting standards and streamline the editing and resubmission, all of which take place entirely inside the server. As the course proceeds, completed works pop up one by one on the syllabus page as titles with links. When the course ends, the syllabus materials and Power Point links are removed, leaving behind an archive of written works. The password protection is then removed, making the archive world-viewable. The total cost of the process is presently about $200 per piece, nearly all of which is labor. It could be reduced to about half this value with improved software.

Properly manipulating the timing and choreography of who sees what is crucial for getting a good result. At the beginning of the writing cycle, each participant emails in a proposed title to the instructor, who then posts it on the syllabus page without attribution. The topics are not critiqued but only checked for overlap with previous choices. This posting process alerts the other participants to what the competition is doing and, equally importantly, initiates a personal relationship between the author and editor. Were a robot in charge of this cross-checking, the participants would be inclined to game it. As the submission deadline approaches, the authors manipulate content inside their own editors in complete secrecy. But, at a key moment, the author declares the piece by pushing a button that date-stamps it and makes it visible to the rest of the class. This must be done before the piece receives any editorial feedback. Participants look at other people's postings as they occur, a fact verified from the server logs. This imposes tremendous psychological pressure on the authors to create strong content and to track down and eliminate errors before submitting. Editorial modification of the piece then takes place inside the cloud editor, again in complete secrecy, but with a permanent log of the requested revisions attached. This log is visible to everyone else when the author re-posts after making repairs. This also imposes pressure to do a good job because the only way to remove the stigma of the revision log is to make satisfactory repairs and complete the writing task. The instructor then removes the log, thus completing the editing process.

The effectiveness of the procedure is self-evident from the quality of the work produced, which may be seen from anywhere in the world by means of a browser. While the work is not perfect, it is, for the most part, balanced, thoughtful and reliable. It is also highly readable. In other words, it is responsible speech. What has caused this magic to happen is not any special selection of people. The authors are all smart, but they are not superhuman, and they will definitely not speak responsibly if they are managed incorrectly. They are arguably the same kind of people who are generating chaos on Internet blogs. What has caused it is simply a change in the economics of speech. The fear of looking bad, first, in front of a peer group and, later, in front of the whole world forever on the Internet powerfully enforces good behavior. The regulation pressure is not made by computers but simply brought out by them. It is entirely imaginary and comes from the speaker's own conscience.

8 Some Important Metrics

Experience with this process gained over several years has provided some insight about what might be required for making responsible speech on the Internet generally:

1. The core cost of responsible Internet speech is not editing but maintaining integrity of the records over long time scales. Permanence of records is essential for the Principle of Witness to work properly. This maintenance cost is not from power or hardware bills, which are actually negligible, but from the labor of converting file standards when the latter go obsolete. It is therefore back-loaded and does not show up at all in the first few years of operation. The only way to evade this cost and still achieve long-term stability is to enforce primitive file standards that are non-proprietary and easy to manipulate in an emergency using recursive scripts. The standard I use is plain ASCII and a highly restricted subset of html-4 commands for typesetting. I implement these standards on a linux box that I run and control myself.
2. The minimum number of participants required for the economics to work properly is approximately 30. In principle any number larger than that would do, but in practice the number cannot be larger than about 100 without a search engine and a management algorithm that obligates participants to use it. This problem is revealed clearly by the tendency of PH240 participants to become overwhelmed by the mass of past submissions in the archive, which are all inherently readable but are full of content that requires work to fully assimilate.
3. A first-time participant typically requires three rounds of editing to get their work publication quality. The reason is that most university-age people are profoundly confused about about (1) the difference between volatile and

non-volatile Internet content and (2) copyright law. They must therefore be taught about these things before they can speak responsibility in public. I have found the concepts to be so difficult for a normal intelligent person to grasp that explaining them in lecture does not work. One-on-one interaction in writing is required. For example, most people have a very hard time understanding why it perfectly acceptable (and sometimes obligatory) to show a journal figure in a Power Point seminar to prove that something is true but a violation of law to show the same image on the Internet. Most people do not realize that government web sites with data in them and electronic encyclopedias become updated (disavowed) on a regular basis and are therefore unreliable sources of facts. There is also the subtle problem of the non-publication publication, a document released in PDF form ostensibly into the public domain but with a copyright notice on it saying that it may not be reproduced or shown to third parties without permission.

4. The electronic nature of the vehicle allows a person to defend the truth of what they say by downloading a saving in perpetuity the documents on which the conclusion is based, even if the documents are large. The completed PH240 archive for 2014 is approximately 1 GB, most of which is attached references.
5. Comments must be absolutely disallowed. The reason is that they are fundamentally unwitnessed and thus an open invitation to behave irresponsibly. Nobody is allowed to make catty anonymous remarks about anybody or anything on the official PH240 site itself. Only responsible witnessed speech is allowed there.
6. Flagging of an unstable reference is usually sufficient to cause a person to rethink a hasty or prejudiced analysis. This is not always the case, particularly when the topic is highly politicized. Public conversation in such cases is hasty and prejudiced by nature. However, the right balance is achieved if the speaker takes full responsibility for what is said, and the editor intervenes only in ways that enforce communications standards.
7. Finally, at least some component of the editing process has to be human for the economics to work properly. A robot cannot do the job completely because humans are smarter than robots, and they are endlessly creative in dreaming up ways to evade the Principle of Witness. Indeed the task of controlling them would be hopeless were it not for the fact that all of us are programmed from birth to interact in certain ways with each other. This feature of the human condition is what causes social networking to be different from exchange of data. Most people respond in the most remarkable way when another person says, “I will not critique you but I am watching you and care about what you do.” The minimum amount of human intervention required depends on the situation and must be determined experimentally. In the case of PH240, it is extremely small, a handful of paragraphs of writing per submission. Such a highly efficient use of the editor’s time is possible because “who the editor is” is previously established in the mind of the writer through lectures.

9 New Institutions of Speech

The idea of using computers to circumscribe speech in ways that re-balance it toward responsibility are directly relevant to two existential problems that people talk about constantly but despair of ever solving. One is achieving responsibility in government. Practices of hiding through secrecy, confusion or excessive mass are commonplace in government in all countries, an effect easily understood as a tactic serving the interests of the people doing the speaking. Calls for transparency and accountability in government thus effectively boil down to calls for assiduous compliance with the Principle of Witness (Harper, 2011; Lathrop & Ruma, 2010; Lichtblau, 2010; Miller, 2012; Wintour, 2012). The people involved will definitely not do this without strong institutional constraints, for it is not in their best interests economically. For better or worse this includes the people making the laws. The other is creeping corporate institutional amnesia, particularly fact bases of technological means and corporate decision histories (Kransdorff, 1998; Linde, 2009; Pierson, 2013). Many crucial components of organizations living in the heads of the people who work in them never get recorded anywhere because it would not be in the person's best interest to make such a record. This works fine until the person retires or dies. A small but important piece of the organization then dies with them. Accumulation of many such failures over the course of time is insidious and probably contributes materially to the failure of companies. In principle, one could push back against this slow systemic failure using improved internal speech management machinery, although actually implementing it would be difficult for the same reason that achieving transparency in government would.

More important than either of these things, however, is the glimmer of hope these observations raise for us all if the Fourth Estate is killed off by the Internet. It is fairly clear that the truth can be had in the Internet age at significantly reduced cost, albeit imperfectly. All that is required is a proper understanding of the economics of speech plus a management system based on it that brings out the best of people instead of the worst.

That said, it is a glimmer only. The experimental numbers reveal that it would be difficult to make such systems work financially over long time scales, even with significantly reduced costs. The economic forces unleashed by electronic tools of speech are just too formidable. It is not an accident that they are at the moment steadily killing off traditional institutions of responsibility. Speaking responsibly is hard work. People will not do this work unless they are compensated for it. They will instead use the communications tools they have to benefit themselves, which is to say, to propagandize, to advertise, to expand their markets, to make entertainment, to hide their inadequacies, and to avoid giving away information that might advantage someone else. Even in the case of my experiments, where the control is great because amount of compensation is effectively very large, enormous effort is required to push back against these forces. Thus, this may be a good time to question the wisdom of letting the economic forest fire ignited by the Internet run its course.

Acknowledgment This work was supported in part by a research grant from MediaX at Stanford University, enabled by a gift from Konica Minolta.

References

Abrahamson, A. (2013, August 16). 'Carlos Danger' brand of Weiners enters the food market. *ABC News*.

Allen, J. E. (2011, April 20). Fake news internet sites make deceptive claims about acai berry weight loss, FTC alleges. *ABC News*.

Alter, B., & Fishman, E. (2013, April 27). The dark side of energy independence. *New York Times*.

Andrews, L. (2012). *I know who you are and I saw what you did: Social networks and the death of privacy*. New York: Free Press.

Aristotle. (2004). *Rhetoric*. New York: Dover.

Bachman, J. (2011, June 8). Analysis: Russia's oil lobby thwarts green ambitions. *Reuters*.

Bachrach, S., Berry, R. S., Blume, M., von Foerster, T., Fowler, A., Ginsparg, P, et al. (1998). Who should own scientific papers? *Science, 281*, 1459.

Balkin, J. M. (2004). Digital speech and democratic culture: A theory of freedom of expression for the information society. *New York University Law Review, 79*, 1.

Barbaro, M., & Corasiniti, N. (2013, July 24). Account of Weiner's online exchanges fits pattern. *New York Times*.

Bell, L. (2013, August 1). Al Gore's oil-fueled Al-Jazeera deal follows a string of green energy fiascoes. *Forbes*.

Best, M. M., & Wade, K. W. (2009). The internet and democracy: Global catalyst or democratic dud? *Bulletin of Science Technology & Society, 29*, 255.

Bilton, N. (2012, August 4) Internet pirates will always win. *New York Times*.

Blake, M. (2013, September 10). Jimmy Kimmel reveals truth behind 'twerking girl on fire' video. *Los Angeles Times*.

Bosman, J. (2012, January 28). The bookstore's last stand. *New York Times*.

Bradshaw, T. (2013, March 21). YouTube reaches billion users milestone. *Financial Times*.

Brooks, D. (2011, January 13), Tree of failure. *New York Times*.

Brumfield, B. (2013, February 6). German education minister loses Ph.D. over plagiarized thesis. *CNN*.

Bryce, R. (2009). *Gusher of lies: The dangerous delusions of 'energy independence'*. New York: Public Affairs.

Buckland, W. W. (2007). *A text-book of Roman law: From Augustus to Justinian*. Cambridge, England: Cambridge University Press.

Carlson, N. (2011, May 12). What an anti-google whisper campaign looks like. *NBC News*.

Charmen-Anderson, S. (2012, August 28). Fake reviews: Amazon's Rotten Core. *Forbes*.

Chazan, G. (2012, January 18). U.S. on path to energy self-sufficiency. *Financial Times*.

Cryderman, K. (2013, June 11). How Canada's big oil firms are trying to sway public opinion. *Globe and Mail*.

Cyphert, D. (2004). The problem of PowerPoint: Visual aid or visual rhetoric? *Business Communication Quarterly, 67*, 80.

Damon, M. (2011). Honesty. In P. Berkowitz (Ed.), *Endangered virtues*. Stanford, CA: Hoover Institute.

Davidson, J. (2013, August 7). Discovery channel provokes outrage with fake shark week documentary. *Time*.

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implication for consumers and firms. *Management Science, 52*, 1577.

Erlangen, S. (2013, April 11). France: Chief rabbi steps down over plagiarism. *New York Times*.

Farhi, P. (2011, November 17). The e-mail rumor mill is run by conservatives. *Washington Post*.
Feith, D. (2009, July 31). Speaking the truth about PowerPoint. *Wall Street Journal*.
Fischman, J. (2011, August 10). The first free research-sharing site, arXiv, turns 20 with an uncertain future. *The Chronicle of Higher Education*.
Fish, S. (2011, January 23). Anonymity and the dark side of the internet. *New York Times*.
Flock, E. (2011, June 24). Netherlands debates net neutrality and anonymity online. *Washington Post*.
Franek, M. (2010, July 22). Amazon e-book tipping point: Is the death of books upon us? *Christian Science Monitor*.
Froomkin, A. M. (2000). The death of privacy? *Stanford Law Review, 52*, 1401.
Gabriel, T. (2010, August 1). Plagiarism lines blur for students in digital age. *New York Times*.
Gapper, J. (2011, August 31). It is right to curtail web anonymity. *Financial Times*.
Gardner, H. (2011). *Truth, beauty and goodness reframed: Educating for the virtues in the age of truthiness and Twitter*. New York: Basic Books.
Garfinkle, S. (2001). *Database nation: The death of privacy in the 21st century*. New York: O'Reilly Media.
Gerson, M. (2009, November 27). Journalism's slow, sad death. *Washington Post*.
Gilder, G. (2000). *Telecosm: How infinite bandwidth will revolutionize our world*. New York: Free Press.
Gillis, J. (2013, August 19). Climate panel cites near certainty on warming. *New York Times*.
Ginsparg, P. (2011). ArXiv at 20. *Nature, 476*, 145.
Goldenberg, S. (2013, February 14). Secret funding helped build vast network of climate denial thinktanks. *The Guardian*.
Gotlieb, C. C. (2002). Does the internet promote democracy? *Human Choice and Computers, 98*, 21.
Grens, K. (2012, February 21). Many 'mistakes' found in newbie-doctors' resumes. *Reuters*.
Guinnessy, P. (2010). ArXiv survival tied to new funding. *Physics Today, 63*(3), 27.
Hamilton, A. (2013, April 22). It's getting harder to make money on YouTube. *Bloomberg Businessweek*.
Handley, M. (2013, January 16). BP projection: U.S. will be energy self-sufficient by 2030. *U.S. News*.
Harper, J. (2011, September 23). *Publication practices for transparent government* (Cato Institute Briefing Paper No. 121).
Harvey, F. (2011, April 20). Fossil fuel firms use 'biased' study in massive gas lobbying push. *The Guardian*.
Hebbert, H. J. (2004, June 8). Exxon head: Energy independence is a myth. *Fox News*.
Hendrickson, M. (2012, September 16). Climate change: 'Hoax' or crime of the century? *Forbes*.
Hinduja, S. (2008). Deindividuation and internet software piracy. *Cyberpsychology & Behavior, 11*, 391.
Hoyt, C. (2007, November 4). Civil discourse, meet the internet. *New York Times*.
Ingram, M. (2011, June 20). Online anonymity: Painful and priceless. *Bloomberg Businessweek*.
Inhofe, J. (2012). *The greatest hoax: How the global warming conspiracy threatens your future*. Washington, DC: WND Books.
Irvine, D. (2013, February 1). Fake Amazon resume proves the power of personal branding. *CNN*.
Itzkoff, D. (2013, February 26). Really cute but totally faked. *New York Times*.
Jan, T. (2010, December 28). Harvard missed sign its was being hoodwinked. *Boston Globe*.
Jeffries, S. (2013, June 12). Internet anonymity is the height of chic. *The Guardian*.
Jenkins,H. W., Jr. (2013, January 15). Green energy's baptists and bootleggers. *Wall Street Journal*.
Kaufman, L. (2013, March 26). Goodall book delayed after copying is found. *New York Times*.
Klein, K. E. (2011, October 25). How business owners counter online slander. *Bloomberg Businessweek*.

Knapp, A. (2011, September 19). An archaeologist watches the history channel and questions the part about the aliens. *Forbes*.
Kornet, A. (2012, January 3). The truth about lying. *Psychology Today*.
Kransdorff, A. (1998). *Corporate amnesia: Keeping know-how in the company*. Oxford, England: Butterworth-Heinemann.
Kravets, D. (2012, May 22). New York Legislation would ban anonymous online speech. *Wired*.
Lathrop, D., & Ruma, L. (2010). *Open government: Collaboration, transparency, and participation in practice*. Sebastopol, CA: O'Reilly Media.
Lee, E. (2013, August 1). New York Times falls as ads drop for 11th straight quarter. *Bloomberg*.
Levmore, S., & Nussbam, M. C. (Eds.). (2011). *The offensive internet*. Cambridge, MA: Harvard University Press.
Lichtblau, E. (2010, March 14). Report faults U.S.'s efforts at transparency. *New York Times*.
Linde, C. (2009). *Working the past: Narrative and institutional memory*. Oxford, England: Oxford University Press.
Manjoo, F. (2009, August 14). Do you think bandwidth grows on trees? *Slate*.
Margolis, M., & Moren-Riaño, G. (2009). *The prospect of internet democracy*. Farnham, England: Ashgate.
Mayer, J. (2010, August 30). Covert operations. *The New Yorker*.
McChesney, R. W., & Nichols, J. (2010). *The death of and life of American Journalism: The media revolution that will begin the world again*. New York: Nation Books.
Miller, L. (2004, February 22). The last word: The Da Vinci Con. *New York Times*.
Miller, H. I. (2012, October 24). A lack of government transparency: The devil in the detailing. *Forbes*.
Milmo, C. (2007, September 27). Wish you were here? Website lures tourists to imaginary beach. *The Independent*.
Moore, M. (2008, June 9). Mobile phone popcorn 'hoax' hits youtube. *The Telegraph*.
Mosher, D., & Hernandez, D. (2012, March 22). 'Bird Man' hoaxter comes clean on Dutch television. *Wired*.
Mullins, B. (2008, December 13). Green is the color of lobbying. *Wall Street Journal*.
Nesbit, J. (2013, March 8). Google, Facebook and the death of journalism. *U. S. News*.
Nocera, J. (2013, June 7). How to monetize plagiarism. *New York Times*.
Orcutt, M. (2012, March 5). Data shows a digital divide in global bandwidth. *Technology Review*.
Palmer, A. (2013, January 26). I hardly dare to say it: Libel laws could save free speech. *The Telegraph*.
Pannett, R. (2013, July 15). Australia to scrap carbon tax sooner than planned. *Wall Street Journal*.
Parker, K. (2012, April 10). Whispering campaigns take flight in new media. *Washington Post*.
Pfanner, E. (2011, September 4). Naming names on the internet. *New York Times*.
Pierson, M. E. (2013). Capturing what employees know: Characteristics of a knowledge harvesting and management system. *TechTrends, 57*, 26.
Plato. (2008). *Gorgias*. Oxford, England: Oxford University Press.
Posner, M. (2012, December 20). Behind the eagle internet hoax: Four Montreal students create video for class assignment. *The Globe and Mail*.
Prabhu, A. (2012, October 29). How much did it cost youtube to stream Gangnam style? *Forbes*.
Reich, E. S. (2010). *Plastic fantastic: How the biggest fraud in physics shook the scientific world*. London: Palgrave MacMillan.
Reisinger, D. (2009, October 14). Finland makes 1 Mb broadband access a human right. *CNET News*.
Reston, M. (2013, August 2). Ex.-L.A. County GOP official allegedly had explicit chat with girl. *Los Angeles Times*.
Rooney, B. (2013, January 16). Debate over online anonymity. *Wall Street Journal*.
Satell, G. (2013, August 1). The continued stupidity of the New York Times. *Forbes*.
Schulz, G. P., & Becker, G. S. (2013, April 7). "Why we support a revenue-neutral carbon tax. *Wall Street Journal*.

Shaer, M. (2011, July 6). PowerPoint ban: Swiss political party wants to outlaw the software. *Christian Science Monitor*.
Shepard, A. (2012, June 13). Investigative reporting now endangered. *New York Times*.
Shields, D. (2011). *Reality hunger: A manifesto*. New York: Vintage.
Silverman, R. (2013, September 18). Climate change inaction is like aids denial, says scientist. *The Telegraph*.
Skorton, D., & Altschuler, G. (2013, July 29). Open access to research: An ideal complicated by reality. *Forbe*.
Sprague, R. K. (Ed.). (2001). *The older Sophists*. Cambridge MA: Hackett.
Stanley, A. (2005, October 25). Bringing out the absurdity of the news. *New York Times*.
Steel, E. (2012, June 21). How shell was hijacked in ad hoax. *Financial Times*.
Stewart, J. B. (2012, May 18). In the undoing of a C.E.O, a puzzle. *New York Times*.
Streitfeld, D. (2011, August 19). In a race to out-rave, 5-star web reviews go for $5. *New York Times*.
Stynes, T., & Launder, W. (2013, August 1). New York Times earnings: Digital subscriptions rise. *Wall Street Journal*.
Switak, B. (2012, May 31). *Mermaids* embodies the rotting carcass of science TV. *Wired*.
Tertzakian, P. (2012, May 21). Fracking is misunderstood, it's the key to energy self-sufficiency. *Forbes*.
Tsukayama, H. (2013, May 9). YouTube paid channels officially launch. *Washington Post*.
Tuttle, B. (2012, February 3). 9 reasons why you shouldn't trust online reviews. *Time*.
Tvergerg, G. (2012, November 14). An America self-sufficient in oil? Don't bet on it. *Christian Science Monitor*.
Verrier, R. (2013, September 17). Online piracy of entertainment content keeps soaring. *Los Angeles Times*.
Wade, N., & Choe, S.-H. (2006, January 10). Researcher faked evidence of human cloning, Koreans report. *New York Times*.
Wallace, J. D. (1999, December 8). *Nameless in cyberspace: Anonymity on the internet* (Cato Institute Briefing Paper No. 54).
Wallace, L. (2013, March 5). The making of the ultimate fake UFO video (were absolutely everything is fake). *Wired*.
Washington, H., & Cook, J. (2011). *Climate change denial: Heads in the sand*. London: Routledge.
Wicker, S. B. (2013). *Cellular convergence and the death of privacy*. Oxford, England: Oxford University Press.
Williams, R. (2010, June 20). Internet plagiarism rising in schools. *The Guardian*.
Wilson, S. (2002). Corporate criticism on the internet: The fine line between anonymous speech and cybersmear. *Pepperdine Law Review, 29*, 533.
Wintour, P. (2012, June 11). Government's transparency drive stalls amid reports of ministry opposition. *The Guardian*.
Wu, T. (2010, March 11). Bandwidth is the new black gold. *Time*.
Wyer, S. (2011). *Violated online: How online slander can destroy your life and what you must do to protect yourself*. Nashville, TN: Dunham Books.
Yi, E. (2011, September 19). Is the death of borders really good for independent bookstores? *The Atlantic*.
Yiannopoulos, M. (2011, January 18). Why is online piracy considered socially acceptable? *The Telegraph*.
Zamost, S., & Griffin, D. (2013, February 22). FBI battles 'rash of sexting' among its employees. *CNN*.

Embedding Privacy and Ethical Values in Big Data Technology

Michael Steinmann, Julia Shuster, Jeff Collmann, Sorin Adam Matei, Rochelle E. Tractenberg, Kevin FitzGerald, Gregory J. Morgan, and Douglas Richardson

1 Introduction

The phenomenon now commonly referred to as "Big Data" holds great promise and opportunity as a potential source of solutions to many societal ills ranging from cancer to terrorism; but it might also end up as ". . .a troubling manifestation of Big Brother, enabling invasions of privacy, decreased civil freedoms (and) increased state and corporate control" (Boyd & Crawford, 2012, p. 664). Discussions about the use of Big Data are widespread as "(d)iverse groups argue about the potential benefits and costs of analyzing genetic sequences, social media interactions, health records, phone logs, government records, and other digital traces left by people" (Boyd & Crawford, 2012, p. 662). This chapter attempts to establish guidelines for the discussion and analysis of ethical issues related to Big Data in research, particularly with respect to privacy. In doing so, it adds new dimensions to the agenda setting goal of this volume. It is intended to help researchers in all fields, as well as policy-makers, to articulate their concerns in an organized way, and to specify relevant issues for discussion, policy-making and action with respect to the

M. Steinmann (✉) • G.J. Morgan
Stevens Institute of Technology, Hoboken, NJ, USA
e-mail: msteinma@stevens.edu; gmorgan@stevens.edu

J. Shuster • J. Collmann • R.E. Tractenberg • K. FitzGerald
Georgetown University, Washington, DC, USA
e-mail: js777@georgetown.edu; collmanj@georgetown.edu; ret7@georgetown.edu; ktf3@georgetown.edu

S.A. Matei
Brian Lamb School of Communication, Purdue University, West Lafayette, IN, USA
e-mail: smatei@purdue.edu

D. Richardson
Association of American Geographers, Washington, DC, USA
e-mail: drichardson@aag.org

© Springer International Publishing Switzerland 2015
S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_15

ethics of Big Data. On the basis of our review of scholarly literature and our own investigations with big and small data, we have come to recognize that privacy and the great potential for privacy violations constitute major concerns in the debate about Big Data. Furthermore, our approach and our recommendations are generalizable to other ethical considerations inherent in Big Data as we illustrate in the final section of the chapter.

To understand the ethical challenges that can arise from privacy concerns in Big Data, we first elucidate how privacy in Big Data can be analyzed using two dimensions: (1) different *contexts* in which privacy is relevant and (2) different *principles* that specify the ethical meaning of privacy. *Privacy contexts* refer to the various spheres of human existence and activity in which individuals might expect different forms and degrees of privacy, each of which requires a specifically targeted analysis (Nissenbaum, 2009). Simultaneously, privacy obtains normative meaning only with additional ethical *principles* that state what is permissible in each context, which means that a variety of principles pertains—and so must be considered and prioritized—in any given context. We refer to the alignment of these two dimensions as a *Privacy Matrix*. We hope that the Privacy Matrix stimulates fruitful discussions about the role of Big Data, and helps promote awareness of privacy-oriented ethical issues in Big Data, both those issues that are known and the possibility of additional issues arising in the future. Beyond the privacy of information, however, the ethical principles upon which we draw also help interpret the design, use, and evaluation of Big Data tools. Our analysis suggests as a whole that the process of building Big Data Technology (information and tools) implicitly or explicitly embeds values into its use. By highlighting these issues, we aim to help scientists, engineers, and other Big Data Technology designers, builders and users better identify and explicitly reflect upon the ethical values their work entails.

2 Defining Big Data

The rise of personal computing, the Internet, inexpensive archiving, and advanced computational infrastructures have enabled a wide range of people such as scholars, marketers, governmental agencies, educational institutions and the general public to produce, share, and analyze vast amounts of readily available information, a phenomenon known as Big Data (Boyd & Crawford, 2012). However, the notion of Big Data remains complex and multifaceted, and invites multiple definitions. Most definitions refer to, or highlight the "three V's": volume, velocity, and variety of data. For example, IBM describes Big Data as "being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big Data is arriving from multiple sources at an alarming velocity, volume and variety" (IBM, 2014). Others emphasize the tendency of Big Data to exceed the management capabilities of conventional database tools (Dumbill, 2012; Einav & Levin, 2013; Kaisler et al., 2013).

For some people, Big Data potentially offers analytic power of mythological proportions with "the widespread belief that large data sets offer a higher form of intelligence" (Boyd & Crawford, 2012, p. 663). For this chapter, we follow the National Science Foundation (2012) which defines Big Data as "large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future."

3 Ethical Issues in the Use of Big Data

The challenge that is most frequently used in discussions of the ethical issues that arise from the research on, and use of, Big Data is *privacy*. But issues of privacy are often invoked without a proper specification of the defining qualities of the concept. It is not immediately clear what privacy is, exactly. Likewise, it is often not clear whether the term is used, or should be used, in a descriptive or a normative way. While certain kinds of data simply *are* private because of their content and origin, it is not clear that they also *ought* to be treated and respected as private matters. It is also not immediately clear where the normative implications of privacy lie. Exploring the definitions of privacy is a necessary first step for any fruitful discussion of the pertaining ethical issues. In Sect. 5 it becomes clear that appreciation for the complexity of the construct of privacy can also support growth in the awareness of other issues that arise in/from Big Data.

3.1 Defining Privacy

To start with some clarifications of the terms, privacy is a quality that can be attributed to actions, things, and pieces of information. Put differently, it applies to both to tangible and intangible things. Tangible and intangible things can be qualified as private insofar as they belong or relate to individuals or groups of individuals. This means that the privacy of things and data becomes thematic only insofar as persons can be concerned. Privacy is the right of the subjects (human individuals) to determine to what extent their thoughts, sentiments, emotions, or other personal and unique information are to be released to other individuals (Solove, 2008). It is not limited to dissemination. Privacy concerns collection and processing, as well as cases of invasions of privacy via forcible interrogations (Solove, 2008). Privacy concerns can emerge in various contexts, according to the types of human activities. In fact, according to Nissenbaum (2009), it is impossible to define privacy in the abstract or in the context of the pure individual person. Nissenbaum (2009) proposed that privacy needs to be defined as control of flows of information in given socio-technical contexts. In this paper we add that for reasons of practicality and to reflect the historical evolution of privacy we need to distinguish these contexts in terms of their distance from the ego and of the social-institutional agents that might control or distort it.

Given this clarification of "privacy", we can see that privacy implies a relation among several persons or agents, both individual or corporate (organizational). Etymologically, privacy has a negative sense: the Latin *privo*, means to rob or spoil or to de-prive someone of something. Understood in this formal sense, privacy becomes thematic because one agent has an interest in something that another agent has an interest to withhold from him. Privacy can also be defined as the right of not being despoiled of something that by right belongs to the individual.

The number of the relata that are relevant for the determination of privacy is potentially unlimited. Privacy can become thematic between more than two agents. For example, Tom might not want to tell Joe anything because he thinks that Joe will tell it to Jack. Analogously, it might not be problematic if one party has access to personal data, but if that party gives them to another person, then it *might* be a problem; and so on. This character of privacy also means that privacy can be affected in indirect ways that can have indirect consequences for persons. In addition, privacy not only has straightforwardly factual implications (for example, using or not using the property of others), it also has an emotional dimension, concerning one's personal attitudes and feelings toward others (including towards ruling authorities).

From a methodological point of view privacy cannot be identified or defined in the way that other, more distinct and self-contained qualities are defined. For example, while the term "yellow" designates the same color for everyone who is able to perceive light within a specified wavelength range, the term "private" acquires meaning depending on the context in which it is used, and based on the relations it involves.

Additionally, the relational character of privacy also entails that it is practically never valued or desired for its own sake, but always because of something else. This aspect will be further explained with respect to the different normative principles that have to be used in order to specify the ethical dimension of privacy.

Finally, the relational character of privacy also suggests that it is not a static concept, in the sense that certain actions or things, or certain information, are thought to be always and essentially private, while others are not. For example, it can have a liberating effect to make certain aspects of one's private life public (e.g., "coming out," "The personal is political."). In turn, a liberating effect can occur if individuals can trust that aspects of their private life are not disclosed to, or used by, others. The concern for privacy is best understood as concern for the changing, and often fragile, demarcation between the private and the public (social, political, economic) sphere.

The term "privacy" is not unique in having a contextual meaning. Other notions that refer to social structures are used in an equally variable way. For example, privacy is similar to the notion of friendship. There is no primary and exclusive meaning of the term "friend"; individuals can be "friends" based on various roles and contexts, and with various degrees of intimacy.

The various meanings of privacy do not need to be unified. With Wittgenstein, we can assume that there are family resemblances between the various uses of the term, that is, similarities without one unifying and unchangeable core (Wittgenstein, 2001, paragraphs 65–67). In this sense, we can also assume that although there is no unified meaning, each meaning and use is definite. Therefore, as a concept, privacy is not "vague" or "difficult to define," as some might think, but rather, has meaning that is determined by the context in which it is used. For example, the users of websites can make precise and legitimate claims with regard to the use of their private data by a specific Internet application, even if it is unclear how their privacy should be handled with respect to other applications.

3.2 Four Normative Principles as Basis for the Ethical Analysis of Privacy

Privacy can acquire a normative meaning insofar as it is possible to say that privacy "ought to be respected," or is a "value" to which one should adhere. However, privacy does not have value that trumps all other values. That is, other values or principles *can* be prioritized over privacy, which can be seen in cases such as domestic violence and child abuse, which occur in the private sphere and have at times been treated as a private matter, but are now no longer considered this way. Therefore, if privacy is taken as an ethical principle, it has to remain less fundamental than, for example, the respect for the dignity and integrity of persons.

On the other hand, "respect for privacy", if taken as an ethical principle, is too vague to be meaningful in practical terms. Insofar as the meaning of privacy is contextual, one cannot "respect privacy" as such, but only in relation to specific conditions and agents. The level of privacy has to be spelled out more concretely in each context, for example as with patient-doctor confidentiality or informed consent.

This twofold limitation of privacy as an ethical principle—that it is either not fundamental enough or too general and vague—leads us to the conclusion that one cannot use privacy as ethical principle in an isolated way. If it is given a normative meaning, it has to be specified in relation to which principles or values this meaning is understood. This conclusion is also supported by the fact that privacy is desirable and obligatory not for its own sake but always because of something else, as stated above.

Four ethical principles can be used to specify the ethical meaning of privacy: Nonmaleficence; justice; autonomy; and trust. These are defined and their contributions to our specification of privacy are outlined below.

- Nonmaleficence: refers to the harm that can be experienced by an individual or a group of individuals. In the context of Big Data, "harm" has to be understood as direct harm (physical, psychological, social, economic, etc.) following the use of personal data. It might seem that all ethical concerns can be related to the principle of avoiding harm, so that only this one ethical principle would be necessary. To a certain degree this is true. However, there are cases in which the experience of harm is only an indirect consequence and the ethical concern relates more directly to practices and the values they incorporate. For example, one can be concerned about practices of democratic citizenship that are affected by the use of Big Data without having to prove that there are individuals who actually experience harm. If harm were the only ethical concern, one would overlook, for example, the role that intrinsic values play in human practice.

- Justice: refers to the distribution of opportunities, rights, and goods among the individuals or groups of individuals that are the target of data mining activities. For example, certain groups can suffer discrimination because statistical data show that they are less likely to succeed in facing specific challenges. Strong anecdotal evidence can be found that users of Big Data are looking for predictions about economically not lucrative segments of the population (Marwick, 2014).
- Autonomy: refers to the decision-making capacities of individuals or groups of individuals. While the principle of nonmaleficence conceptualizes agents as objects, or targets, of data mining practices, the principle of autonomy conceptualizes them as subjects. Autonomy can be understood in a narrow and practical sense with respect to Big Data, for example with respect to the question how much control the individual has about the use of his/her personal data. In a wider sense it concerns the question how much freedom is left to individual decision-making. The latter is relevant given the attempt to use Big Data for the prediction of individual behavior, which potentially eliminates the whole factor of individual decision-making.
- Trust: refers to the relation between the data sources and the agents who are interested in mining their data. It involves all relata, that is, agents in their commonly shared practices. Trust can be defined as collective attitude that reduces the burden of permanent mutual control, without, however, dispensing of it, as most social interactions involve a mixture of trust and control. Like "privacy" and "friend", "trust" can be instantiated in various ways and to various degrees. Although specific institutional provisions can be implemented in order to establish trust in a given setting (e.g., "checks and balances"), it is often difficult to verify empirically whether trust really exists. In the case of Big Data, trust is not a positive imperative but a category that can be used to assess critically the ways in which data are used. With respect to Big Data, agents have no obligation to trust any other agent, but they do have an obligation to be trustworthy. One can ask, for example, whether the government or companies are acting in a way to enable individuals to trust the ways in which data are used.

3.3 Remarks and Explanations

The four principles stated above are broad enough to cover other categories that are often used in order to conceptualize the ethical reach of privacy. For example, justice is a broad enough principle to cover concerns for discrimination and the access to data. It has been remarked that privacy has both ***instrumental*** and ***intrinsic*** value. Privacy can be both a means to an end and an end, desired because it is a good in itself (Moor, 1997). These two ways of conceiving of privacy are not mutually exclusive and covered by the principles mentioned here. The instrumental value of privacy can be captured by the use of non-maleficence and justice, the intrinsic value by autonomy and trust. In the case of the former, privacy is valued because it allows one to avoid harm and establish the fair treatment of all members of society, in the case of the latter, privacy is desired because being able to rely on the privacy of one's data is a necessary part of the intrinsically valued practice of individual autonomy and the trust that enables social interaction.

Privacy has also been conceptualized through the distinction between the ***restricted access theory*** and the ***control theory*** (an overview is given by Tavani, 2008). The difference between these theories results from taking data sources (that is, individuals) as objects or subjects, respectively. While the restricted access theory sees data sources passively as objects, the control theory involves them actively as subjects. It might be worth noting that there is nothing inherently wrong with taking individuals as objects in the course of an analysis. For example, a physician uses a patient's data without the latter having direct control. The patient has to have the confidence that the physician uses the data in a confidential way. Obviously, at one point at least the data sources have to be involved as subjects, for example by giving consent to the use of personal data, but the practice that ensues does not need to involve them actively. The principles stated above reflect this distinction and show again that the different approaches to privacy are not mutually exclusive: in some cases, privacy can be organized according to the idea of restricted theory while in other cases it requires active control.

It has been suggested to conceptualize privacy in terms of ***rights***. Privacy would then be violated whenever the right to "life, liberty, and property" is being violated (Volkman, 2003). The advantage of this approach is that it makes it possible to specify concerns for privacy and relate them to well-known legal principles. As rights, privacy rights are indeed "derivative" from other, more fundamental rights. However, the focus on rights seems to obscure other normative concerns. For example, if one says: "The flow of information is not the problem. It is the illegitimate use of information that is of concern" (Volkman, 2003, p. 209), then all cases in which the illegitimate use of information cannot be shown would raise no ethical concern, which is clearly not the case. A certain flow of information can, for example, erode trust even if no direct violation of privacy rights is proven. Analogously, if the commercial use of data is seen exclusively under the perspective of rights, one has to conclude that "prohibiting such capitalist acts between consenting adults is paternalistic and immoral" (Volkman, 2003, p. 210). This perspective seems to shift the burden exclusively to the side of the providers of personal data, insofar as their active consent or refusal is needed in every case and they are given a certain degree of responsibility for the use (and misuse) of data. Also, no violation of privacy could be claimed as long as some consenting consumers were to be identified, which limits drastically the range of ethical analysis.

As an additional, fifth principle one could mention ***beneficence***. It would refer to the possible goals and purposes of data mining. As a term, beneficence can be defined as active concern for the well-being of others. Like trust, it can be used as a critical category insofar as it is possible to ask whether data are used with the goal of improving the life of citizens. Critical concerns can raise the question whether the use of data is guided by genuine beneficence, and whether the principle is used in an inclusive and universal way. For example, cases where the access to Big Data is withheld from certain segments of the population because of a concern for profitability can be seen as a violation of beneficence. However, it is not clear yet whether the use of Big Data will be driven by genuine beneficence as a concern for the well-being of others, and not rather by attempts at improving managing procedures and economic outcomes. The principle of beneficence seems therefore less relevant than the other four, although this situation might change in the future.

The principles used in the present paper are very close to the ones used in the ***Menlo Report*** (Dittrich & Kenneally, 2012). The Menlo Report is an important document written in 2011 to provide guidelines to researchers in the field of information and communication technologies (ICT). It is modeled on the paradigm of the Belmont Report which in 1979 established principles for biomedical research. The principles used in the Menlo Report are respect for persons (equivalent to the principle of autonomy used here), justice, and beneficence. The latter is defined as avoidance of harm and concern for public welfare, which means that it is wider than the present use and covers both what is distinguished here as nonmaleficence and beneficence. As a fourth principle, the Menlo Report mentions respect for law and public interest, which covers issues such as compliance, transparency, and accountability. In the present paper, these can be subsumed under the principle of trust. In general, the Menlo Report represents the attempt at establishing ethical principles and rules from within the community of researchers in the field of ICT, and it is important to note that the Privacy Matrix suggested here is in congruence to this attempt.

Autonomy can be understood in a twofold way. The first way relates to individual decision-making as a democratic practice. It has been noted that citizens in a constitutional democracy should be given the right to opacity so that they can legitimately refuse their lives "being read" by others (Hildebrandt, 2011). Autonomy must therefore not be reduced to procedures of informed consent, but concerns the roles that agents assume in social and professional interaction. One can claim, for example, that "meaningful autonomy requires a degree of freedom from monitoring, scrutiny, and categorization by others" (Cohen, 2000). This also calls for a positive attitude toward "semantic discontinuity" which entails more "contextually specific practices of self-definition" in the use and regulation of information systems (Cohen, 2012). Instead of the assumption that the Internet and data systems are to be covered by a single global regulation that assumes all individuals follow the same ideas of agency and privacy, it seems necessary to allow for more particular, either national or group-specific regulations that reflect the respective decision-making more accurately.

While such concerns can be qualified as soft insofar as no identifiable harm to individuals has to occur (harm is a "hard" category insofar as it has to be verifiable in each case) and the concerns are related to long-term changes in mentalities and practices which can only have an indirect effect on individuals, they show that an important part in the ethical reflection on Big Data is related to the evaluative attitudes with which it is received. From an ethical point of view, there is no reason, or no possible justification, that would allow one to neglect such concerns. Even if Big Data do not necessarily have repercussions for specific individuals, they change the way in which society operates as a whole, which means that members of society can be legitimately concerned by it. Evaluative attitudes toward issues of trust and autonomy are necessary condition of individual agency and therefore need to be addressed. Otherwise, one would have to say that democratic practices are independent from the way agents experience their status vis-à-vis governing institutions, employers, and the like.

The second way in which autonomy can be understood concerns ontological conceptions of agency. Big Data can be used for the "prediction, preemption, presumption" of individual behavior (Future of Privacy Forum, 2013). Some see the risk of a reification of human cognitive processes (Hildebrandt, 2011). Pattern recognition entails a merely statistical conception of individual agency, which can have an impact on attitudes toward individual decision-making and the degree of freedom it is given in specific settings. If an inclination toward anticipatory or preemptive governance becomes an inherent part of policy-making, the autonomy of individuals or groups can be severely limited. From an ethical point of view, this means that individual decision-making has to be given an intrinsic value, especially in the light of statistical interpretations that can lead to qualifying particular decisions as arbitrary, detrimental, or defective. That is, ethically speaking individual decision-making has to be given the opportunity to define its own inherent standards, without being forced to resort to the "higher" vantage point of statistical data collection.

3.4 Contexts of Privacy

Privacy has also been described as "***contextual integrity***" (Nissenbaum, 2009). It has been remarked that approaches to privacy are often too general and do not take the "compatibility with presiding norms of information appropriateness and distribution" in given contexts into account (Nissenbaum, 2004, p. 137). This insight is particularly relevant for the fine-tuning of privacy-related policies. It follows directly from the relational meaning of privacy explained above.

Contexts of privacy cannot, however, be defined arbitrarily. A possible approach is to project privacy onto the canvass of human experience. If privacy concerns the collection, processing, and ultimate dissemination of information from the individual to others (Solove, 2008), the trajectory of private information should start with the most intimate contexts of life: the bodily, interpersonal (family, friendship), and home based ones. This context is the most strongly defended by laws and customs. The fourth Amendment of the US constitutions refers mostly to it, especially in the residential context. Privacy in this context refers to information about the self that is deeply personal, including activities consumed in one's home. It includes private diaries or other type of written records, oral communication or interactions in one's own residence, and so on. Medical records, although recorded by various institutions, refer to one's body and also benefit from one of the highest level of protection. HIPAA regulations in the US came to strengthen this point of view.

As humans due to their social obligations participate in contexts outside these realms, contexts of privacy emerge at each turn of our social journey. The contexts need to be seen, however, as layers of sociability, increasingly distant from the most intimate context of privacy (bodily, interpersonal, residential). Thus, in layers of sociability that are increasingly distant from the self, claims to privacy become increasingly weak and legal protection correspondingly thinner. Such contexts would start with the ones that are the closest to our interests, choices, and control, such as voluntary participation in various social, religious, and civic organizations. These are the communitarian contexts. Here, the expectation is that our activities, to the degree to which they are not detrimental to others (such as participation in terrorist or criminal groups) should be protected from undue scrutiny. Of course, when the social participation in these organizations is public, such as an open religious mass or open civic event, the claim to privacy cannot be called in defense. Yet, confession, some donations to charitable organizations, use of public resources (e.g., libraries), or in kind community interactions that are by definition philanthropic entail a good degree of privacy that is recognized as such. For, example, the American Library Association has staunchly and rightfully defended the right of library patrons not to have their reading records disclosed, not even in criminal cases, without a strongly determined due cause and a court order (http://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy).

The educational context of social interaction lies loosely connected and at times hard to differentiate from the community context. In an educational context some information is strongly defended, while other less, according to the social implication of the data. Personally identifiable information that might put the person at a disadvantage or reduce his or her autonomy is strongly defended (grades, courses taken, etc.). Other types of information are publically available especially if aggregated for assessment of educational policy.

Following the track of human activities, privacy contexts escape more and more the control of the individual as he or she enters in transactions with organizations and institutions that have a legitimate claim to recording, preserving, and further disseminating the activities or information pertaining to the individual. A first context is that of our interactions with legal, political, governmental, or law enforcement organizations. Here, some types of information are legitimately public, while others ought to stay private. For example, individual contributions to political campaigns are public in an attempt to keep the political process transparent. Land records are also public, as are court records for most criminal cases. While we might consider the last two types of information intimate and highly personal, the impact an individual's owning of a certain parcel of land or of their criminal activities is for the most part social and ought to be publicly accessible. On the other hand, while voter records are public, including party affiliation, voting behavior is not public, in a defense of our freedom of conscience and expression. Neither are tax records. The US Census bureau never asks questions related to religious affiliation.

More distant still from the most intimately private contexts are those pertaining to most commercial transactions, as in buying and selling or using commercial services. Such interactions demand transparency by definition, for the sake of enforcing contract laws in case of conflict. These are similar with most judicial transactions, which are to be open and subject to public scrutiny to prevent secret trials and abuse of power. At the same time, the definition of what is private and what is public is not rooted in abstract laws or principles, but in contractual obligations voluntarily accepted by the user. Most interactions and activities on social media enter in this category. What is and what is not private is subject to the contractual obligations that the users accepted when signing up and clicking the box for "accept terms of service."

Finally, privacy concerns may emerge in the context of participation as subject in scientific research. The situation is for the most part regulated by contractual terms, guaranteed by "informed consent." Yet, this is far from a clear cut situation, as some research contexts could intrude upon individual information of the most intimate kind (e.g. medical information).

In brief, our concept of concepts of privacy takes a layered approach. It orders contexts on a "distance" dimension, where some are closer, while other farther away from the most intimate and strongly defensible claims to privacy. In this respect, we follow the pragmatic approach of most legal literature, which aims to operationalize the contexts as areas of human activity with definite pragmatic implications and outcomes.

As we will explain below, the contexts become clearer and easier to comprehend when included in a Privacy Matrix that aligns different normative principles with a set of different levels and contexts of privacy.

4 The Privacy Matrix: How to Think About Privacy in Big Data

This chapter suggests that privacy should be addressed according to two dimensions, referred to here jointly as the Privacy Matrix as shown in Table 1. The first dimension comprises the possible levels, or contexts, in which privacy can become relevant. Each level or context requires specific analysis from a privacy perspective. The second dimension is based on the ethical principles that specify the normative,

Table 1 The Privacy Matrix: the columns represent possible privacy contexts while the rows represents the ethical principles of privacy

Specifying principles	Privacy contexts					
	Individual	Community	Education	Governmental	Science	Commercial
Nonmaleficence						
Justice						
Autonomy						
Trust						

or ethical meaning, that can be given to the idea of privacy. The combination of the two dimensions, finally, is based on the idea that privacy, both on the descriptive and normative level, has to be further specified if an analysis is supposed to yield meaningful results. The Matrix is based on the assumption that practical concerns regarding the different levels of individual and social life can each be combined with different ethical principles.

We suggest using the Privacy Matrix as a heuristic tool. The list of ethical principles, and the way they are understood, is not seen in opposition to existing approaches to the ethical analysis of privacy, but rather as an attempt at dealing with the necessary pluralism of principles in a more effective way. Very often, a variety of principles is suggested in a way that leaves it open which ones should be applied to the case at hand. Obviously, one would like to address all possible ethical concerns as one should not arbitrarily decide to leave some of them out if they are relevant, but not all of them can be applied to the same degree. The same can be said for the levels of privacy. With the Privacy Matrix, it is possible to start from the process of application. The question then becomes: on which level, or in which context, is privacy most relevant in the given case, and which normative concern is most relevant? It seems evident that no particular case can be limited to one combination of criteria only, but can always be conceptualized in various ways. However, one can assume, if only for heuristic purposes, that each case is relevant in one primary way, which then has to be taken as point of departure for an ethical analysis that is both specific and effective enough. Even if the search for the primary application of contextual and ethical criteria can seem arbitrary in certain cases, it supports at least one relevant issue being addressed, and it might help to identify others that have not yet been considered. The goal is to shift the focus of the analysis from the multitude of possible perspectives, which is often practically irrelevant, to the steps that are necessary to engage in a process of decision-making which is then, hopefully, practically relevant.

The privacy contexts included in Table 1 represent characteristic areas and fields for using Big Data. Electronic medical records, which the International Organization of Standardization defines as "a repository of patient data in digital form, stored and exchanged securely, and accessible by multiple authorized users" represent a good example of Big Data related specifically to **individuals** (Häyrinen, Saranto, & Nykänen, 2008). The **community** context mainly consists of social media data, which investigators in many sectors are mining for useful information. For example, a recent study published in *Preventative Medicine* revealed the attempt at tracking real-time social media like Twitter for monitoring HIV exposure and drug-related behaviors with the intention of detecting and preventing future outbreaks (Stoové & Pedrana, 2014). **Educational** services and companies now use Big Data with the aim of improving teaching and learning. For example, Knewton, an education technology company, created digital courses in which students are tracked "as they play online games, watch videos, read books, take quizzes, and run laps in physical education" (Knewton, Inc. 2014; Simon, 2014). The federal **government** employs Big Data sets from various programs for secondary purposes beyond the aim of their original collection. For example, law enforcement officials have attempted to develop predictive technologies using Big Data to anticipate, intervene, or prevent crime, including identification of terrorist networks, warning of impending attacks, and preventing the proliferation of weapons of mass destruction (Executive Office of the President, 2014). **Scientists** working on the Personal Genome Project at Harvard Medical School are investigating the utility of genetic data for enhancing health care in multiple ways such as increasing drug effectiveness, assessing predisposition to disease, and constructing microbiome profiles (Ball et al., 2014). Finally, **commercial** retailers and marketers analyze a wide range of customer activity, both on and offline, to provide, as they claim, more tailored recommendations and "optimal pricing". For example, in April 2014 Verizon Wireless notified customers that it would begin gathering data about user activities and selling them to marketers (Lazarus, 2014).

These various examples of the contexts in which Big Data arise and get used illustrate that Big Data refers to the dynamic use of data for insight rather than static archives of massive amounts of information. Research using Big Data relies on massive, continuous, real-time data streams that might predict social behavior by those both creating and analyzing the data. Thus, the phenomenon of Big Data includes the potential for relatively continuous monitoring, control, and moderation of individual and societal behavior. "(O)ur ability to modify public behavior increases as the observed individuals are more exposed to our scrutiny and tracking." (Matei, 2014). This suggests that we must acknowledge and make explicit tradeoffs between privacy and the utility of analyses based on Big Data. Our approach suggests, however, that the tradeoffs vary across the several contexts of Big Data collection, analysis and use. In the pages that follow, in which we explore the implications of this observation with respect to specific examples that serve as case studies in the contextual variation of privacy concerns, we will show the implications of privacy for the realization or protection of specific ethical values such as autonomy and social justice (5.1-4.6.). This analysis illustrates use of our Privacy Matrix as a guide for inquiry into the relationship between privacy context and ethical principles in ethical reasoning about Big Data. In the final chapter, we will see how ethical values are directly related to, or even embedded into, the tools that use and research Big Data (5.).

4.1 Profiling Individuals with Big Data

Big Data when is used to create profiles of individuals for various purposes brings risks. For example, the White House privacy report explains, “credit scores and other economic data could influence an individual’s opportunities to find housing, forecast their job security, or estimate their health outside of the protections of the Fair Credit Reporting Act. Individuals have little recourse to understand or contest the information that has been gathered about them or what that data, after analysis, suggests” (Executive Office of the President, 2014). The report further suggests that pricing and discrimination caused by Big Data could exacerbate existing socio-economic disparities in education and the workforce setting. Fung notes that the Associated Press reported a specific example of group profiling by New York City police officers who collected and plotted on a map license plate numbers of people attending services at mosques. He also comments, “The Department of Homeland Security’s more recent plan to build a national license plate database—and the outcry it provoked—suggests that minorities may be especially vulnerable to what Americans would perceive as a violation of privacy” (Fung, 2014).

4.2 Anonymity, Manipulation and User Consent in Online Communities

In 2006, researchers at Harvard began gathering anonymized data on 1,700 college age Facebook users to study how interests and friendships changed over time (Boyd & Crawford, 2012). However, “these supposedly anonymous data were released to the world, allowing other researchers to explore and analyze them. What other researchers quickly discovered was that it was possible to de-anonymize parts of the data set: compromising the privacy of students, none of whom were aware their data were being collected” (Boyd & Crawford, 2012). Among many other studies conducted using Facebook data, in 2012 Facebook completed the “emotion contagion study” in which they skewed users’ newsfeed so they would see content happier or sadder than average “and when the week was over these manipulated users were more likely to post either especially positive or negative words themselves” (Meyer, 2014). The experiment was, technically speaking, legal, according to Facebook’s Terms of Service in which users relinquish their data by joining the social media site (Sullivan, 2014). Yet, the study was conducted prior to IRB approval. Experts and casual users alike have criticized the study saying “emotional well-being is sacred” and “research is different than marketing practices” (Boyd, 2014).

In addition to Facebook, Twitter has also been accumulating user data since 2006, at a rate of five hundred million tweets worldwide everyday, and announced that it is planning to release them all. The data is promising for scientists “looking to find patterns in human behaviors, tease out risk factors for health conditions and track the spread of infectious diseases” (Moyer, 2014). However, the question arises whether researchers may collect and use such data for research without the users' consent and intention to be part of research. Other social media outlets have also used data without user consent. For example, Path was a social networking app for photo sharing and messaging. In 2012, the app was criticized for accessing and storing member phone contacts without their knowledge or permission. Path was fined $800,000 by the Federal Trade Commission (Ramirez, 2014). Similarly, a Flashlight app failed to disclose to iPhone users that it was sharing their location data with advertising networks. Finally, Snapchat is an app that allows users to send pictures to friends that "self destruct" seconds after opening. There have been several simple ways identified that allow recipients to save the pictures indefinitely. Moreover, Snapchat was fined for security failures in which attackers compiled a database of 4.6 million Snapchat usernames and phone numbers (Ramirez, 2014).

4.3 Protecting Vulnerable Populations in Educational Contexts

Educational technology firms are serving as third parties accumulating academic and behavioral data on students. The education company Knewton, as discussed above, observes students "monitoring every mouse click, every keystroke, every split-second hesitation as children work through digital textbooks, Knewton is able to find out not just what kids know, but how they think" (Simon, 2014). These companies are gathering up to 10 million unique data points on each child per day and despite extensive privacy policies and terms of service, an examination revealed "gaping holes in the protection of children's privacy" (Simon, 2014). Another case is Learnboost, a third party that allows teachers to upload their notes and student attendance, test scores, behavior and more to a digital textbook. The teachers are then eligible, for example to email these grade books with no other regulation than "as they see fit". A recent national study found that only 7 % of contracts between schools and educational technology companies agreed not to sell the data for profit. Also, "few contracts required the companies to delete sensitive data when they were done with it. And just one in four clearly explained why the company needed personal student information in the first place" (Simon, 2014). For the company InBloom, privacy concerns resulted in school districts withdrawing from contracts and ultimately shutting the company down. InBloom a non-profit corporation was financed with $100 million in seed money from the Bill and Melinda Gates Foundation as well as the Carnegie Corporation of New York to store and manage student data for public school districts across the country. However, once parents began to discover what kind of data was being collected, such as social security numbers, they began to speak out causing multiple school districts to pull out and, ultimately, InBloom to close its doors (Singer, 2014).

4.4 Unequal Access to Big Data in Scientific Research

While the previous cases relate to the undue access to Big Data, it also has to be mentioned that not all potential users have equal access to data resources. "Top-tier, well-resourced universities will be able to buy access to data, and students from the top universities are the ones most likely to be invited to work within large social media companies," resulting in a gap between researchers who have the potential to study Big Data and those who have not. Well-funded companies mostly likely will also have more access to data. This gap created by the difficulty and expenses associated with the access to Big Data results in a "restricted culture of research findings," as large data companies have no requirement or responsibility to make their data available. In addition, "Big Data researchers with access to proprietary data sets are less likely to choose questions that are contentious to a social media company if they think it may result in their access being cut. The chilling effects on the kinds of research questions that can be asked—in public or private—are something we all need to consider when assessing the future of Big Data" (Boyd & Crawford, 2012, p. 674).

4.5 Big Data and Government Surveillance

Before the Snowden affair renewed vigorous debate about government surveillance, the Terrorist Information Awareness (TIA) program generated extensive controversy about balancing privacy with national security in contemporary America (Cooper & Collmann, 2005; Department of Defense, Inspector General, & Information Technology Management, 2003). Begun in the wake of the 9/11 attacks, TIA mobilized enormous computer capability to search databases across the government in search of terrorists, an early application of Big Data before the term became popular. Critics argued that TIA posed multiple threats to the privacy of individual Americans (Crews, 2002; Safire, 2002; Simons & Spafford, 2003; Stanley & Steinhardt, 2003; Washington Post, 2002). TIA, they argue,

- Violates the Fourth Amendment of the Constitution by searching a data base containing detailed transaction information about all aspects of the lives of all Americans;
- Undermines existing privacy controls embodied in the Code of Fair Information Practices, such as improper reuse of personal data collected for a specific purpose;
- Overcomes "privacy by obscurity" including inappropriate coordination of commercial and government surveillance;
- Increases the risk of falsely identifying innocent people as terrorists;
- Increases the risk and cost of identity theft by collecting comprehensive archives of individually identifiable information in large, hard-to-protect archives;
- Accelerates development of the total surveillance society.

Critics also identified other potentially undesirable consequences in addition to invasion of privacy, including:

- Undermining the trust necessary for the successful development of the information economy and electronic commerce;
- Undesirably altering the ordinary behavior of the American population including quelling healthy civil disobedience, "normalizing" terrorist behavior, and inhibiting lawful behavior;
- Creating new, rich targets for cyberterrorism and other forms of individual malicious abuse of computerized personal information.

In addition to highlighting persistent concerns about privacy, civil liberties and government surveillance, the TIA controversy illustrates the need to reflect deeply on the ethical implications of any Big Data project during its design. Waiting until controversy erupts misses the opportunity to design a better application and sullies trust in scientific, political, educational and commercial leadership.

4.6 Sale of Big Data in Commercial Contexts

Outside of social media, any activity an individual performs online can be tracked, resulting in information for commercial purposes such as marketing and behavior studies. For example, Disconnect is a program that lets users see who is tracking their visits to websites, revealing dozens of third parties observing and following their individual "click stream" (Kroft, 2014). These third parties, known as data brokers, "are collecting, analyzing and packaging some of our most personal information and selling it as a commodity. . .to each other, to advertisers even the government, often without our direct knowledge" (Kroft, 2014). In response, the White House report on Big Data and Privacy has highlighted the need for effective consumer privacy protections for the individuals (Executive Office of the President, 2014). But the report has also received criticism for stopping short of taking effective action to protect consumers, "such as requiring that private companies disclose to consumers what they know about them" (Lazarus, 2014).

Third parties also accumulate data without online sources as well. In Boston, an automated reader attached to the front of a "spotter car" takes a picture of every license plate it passes. These images, more than 8,000 per day, are then sent to Sousa, a company in Texas, that has over 1.8 billion plates from vehicles across the country. Typically, every license plate of a stolen or defaulted vehicle results in $200–$400 for the company. In May 2014, a legislative committee was scheduled to hold a hearing on a bill that would ban most uses of license plate scanners. Jonathon Hecht, a Massachusetts representative said, "(w)e need to have a conversation about how to balance legitimate uses of this technology with protecting people's legitimate expectation of privacy" (Musgrave, 2014). Kade Crockford of the American Civil Liberties Union of Massachusetts went on to explain, "it's the wild west in terms of how companies can collect, process and sell this kind of data" (Musgrave, 2014).

5 Embedding Values in Big Data Technology

Although we may loosely refer to Big Data as if the data stand on their own free of any supporting technology, creating, analyzing and using Big Data depends on also creating complex computer infrastructures, applications, and devices. In this section, we will explore the ethical, or values-related, implications of this co-creation of Big Data and Big Data Technology. We suggest that, in the course of enabling Big Data, designers, users and analyzers embed and realize values in Big Data Tools. This process may occur in three, often interdependent ways, namely:

1. Values may guide the use of Big Data tools;
2. Big Data tools may enable realization of values, and;
3. Building tools to realize values may entail, and often requires basing their design on the target values themselves.

We will examine various chapters in the present book to elucidate each of these processes. In the course of our analysis, we will also distinguish between two types of value, technical and ethical values. For Big Data technology, computer scientists and engineers usually seek to design tools that effectively accomplish a technical purpose, such as enabling effective analysis and visualization of Twitter conversations in real time. Realizing such technical values through Big Data Technology, however, often helps realize ethical values, such as minimizing loss of life in a mass shooting or natural disaster. Performance requirements may link technical and ethical values through the effective design and functioning of Big Data Technology (Cooper & Collmann, 2005). Finally, we should recognize two types of Big Data Technology, including:

1. Technology that produces, archives, protects, and displays Big Data or its constituent components, and;
2. Technology that enables description, analysis, interpretation and understanding of Big Data.

Distinguishing between these two types of technology reflects a primary consideration of this book: much more Big Data exist than we have the tools to exploit. From the perspective of Big Data Technology design and use, we observe how values condition tools and tools help realize values in a dynamic, interdependent embedding process (Collmann & Robinson, 2010; Cooper, Collmann, & Neidermeier, 2008)

5.1 Values Guide the Use of Big Data Technology

Organizational meetings occur in a variety of formats in the twenty-first century, including face-to-face encounters, teleconferences, videoconferences and, quite commonly, mixed media meetings over the Internet. As Ahmed and Gavrilova note, meetings in all forms absorb much staff time and, thus, corporate money in the workday, including time that sometimes appears to have yielded little return. It makes good business sense to investigate the utility of employing advanced computerized technology to make best use of expensive meeting time. Ahmed and Gavrilova describe a multimodal physiological and behavioral biometric system (Microsoft Kinect v2) that records and analyses the overt participation of individuals in a meeting, including devices that record talk, gait, facial characteristics, and movement across the room. The authors emphasize the effectiveness of their system in capturing data for analysis and, thus, its potential value as a tool for increasing the efficiency and productivity of individuals during a meeting. The tool's technical efficacy in rapidly capturing traits and identifying individuals relates directly to its avowed purposes of analyzing meeting workflow, characterizing and evaluating individual contribution level and analyzing group dynamics and behavior—all with the goal of improving the ability of meetings to achieve corporate objectives.

This Big Data producing technology poses several value-related problems for reflection. First, the Kinect v2 clearly places greater value on overt meeting behaviors such as talk and note taking that monitoring devices can detect and lesser value on covert meeting behaviors such as listening or thinking that remain undetectable. Second, the Kinect v2 as a surveillance tool gives expression to the concept of the Panopticon, a means for continuously documenting all behaviors of a target population with little regard for its own desires. The Kinect v2 bears comparison with the educational technologies described above which drew comment for inserting a "third party" in the educational process with few controls by the observed population over use of the information. Adult employees in an organization usually have greater control over themselves than children in a classroom; but, without guidelines for use and protection, such minutely documented, partial information poses relatively uncontrollable risks to their well-being by becoming a yardstick for job performance. From the perspective of safeguarding personal autonomy in the workplace, Kinect v2 directly challenges employee strategies of "stage management" that establish distance between an individual's private, backstage persona and their public, onstage performance. Kinect v2 constrains meeting members to participate overtly even if covert methods of contribution match their personal working style better, they prefer "off-line" contributions to project development, or specific meeting contexts favor reticence.

5.2 Big Data Tools Enable Realization of Values

Chapters in this book specifically address how Big Data technologies enable the realization of key values such as trust in information from diverse sources on the Internet (Ignjatovic et al.) and self-organizing in task-based work groups (Matei) as well as protection against assaults on such values (Caverlee & Lee). Careful reading also gives evidence of the contributions of Big Data Tools to patterns of social injustice, for example exploitative crowdsourcing of problems from developed countries to low wage "knowledge workers" in developing countries. These examples set a precedent for making explicit reasoning about the ethical consequences of apparently value-free tools a typical dimension of their design and implementation.

In a deeply technical analysis, Ignjatovic and colleagues address an ethical value that lies at the heart of the instrumental effectiveness of Big Data technology by referring to trust in the reliability of data from disparate sources across cyberspace. We observe above that trust refers to the relation between the data sources and the agents who are interested in mining their data and involves all relata, that is, agents in their commonly shared practices. Thus, we define trust as a collective attitude that reduces the burden of permanent mutual control, without, however, dispensing of it, as most social interactions involve a mixture of trust and control. Ignjatovic and colleagues note how the failure of typical social mechanisms of information credibility and trustworthiness in cyberspace necessitate automated methods, especially when drawing vast quantities of data from vast numbers of sources. One could reasonably argue that, without some basis for trust, Big Data as currently envisioned becomes impossible. From their perspective, trust in a flow of information derives from its provence, or, the combined trustworthiness scores of sensor modes and all the nodes through which data passes including terminal, intermediate and server nodes, and the emergent trustworthiness of data elements as they pass through the network. While offering detailed technical ideas about combatting collusion attacks in online rating systems, Ignjatovic and colleagues observe that achieving effective results depends, given the current state of the art, on compromising the privacy and anonymity of participants. As one can argue, this may be entirely appropriate given the public nature of many online communities. Yet, for instances in which this condition fails because participants do not expect all information to be public, Ignjatovic and colleagues offer ideas for further technical research on Big Data tools to help realize the twin values of trust in data from an unknowable social space and the privacy of its constituent members.

Caverlee and Lee address issues that emerge from a specifically malevolent corner of cyberspace, the world of weaponized crowdsourcing. From their perspective, crowdturfing "wherein masses of cheaply paid shills can be organized to spread malicious URLs in social media, form artificial grassroots campaigns ("astroturf"), spread rumor and misinformation and manipulate search engines (p. ???)" poses clear threats to information quality and community trust of such systems. Their research focuses on developing automated means for detecting crowdturfing tasks and crowdturfing workers as well as other, off-line mechanisms such as increasing the cost of crowdturfing campaigns. In contrast to Ignjatovic and colleagues who worry about the integrity of data flowing through the network, Caverlee and Lee focus on how fraudulent use manipulates and potentially distorts the social perception of content.

The phrase "cheaply paid shills" suggests a certain view of crowdturfing workers, however, that we may want to query. Caverlee and Lee do not cite any national or international laws prohibiting crowdturfing. Their analysis suggests that crowdturfing, at least in the forms they analyze, constitutes a form of cheating or false advertising, not a form of crime. Crowdturfing workers, thus, perform no criminal acts but only knowingly or unknowingly facilitate the misrepresentation of their subject matter. The data from Bangladesh suggests a social justice issue, however, in which a crowdturfing requester takes advantage of the international division of labor and wages to exploit crowdturfing workers. Crowdturfing constitutes a form of low wage piece work with no job security, no benefits, and no form of worker organization to prevent its worse abuses. Even if they earn more money doing crowdturfing than other workers in Bangladesh, they make less than crowdturfers who work from developed countries. Hence, crowdturfers may, indeed, be cheaply paid but the noun "shills" taints the workers not the requester, the victims of the international division of labor and wealth not its perpetrators and beneficiaries. No pure technical solution exists for crowdturfing as long as poor people with access to computers can fill its labor ranks.

Matei, Bruno, Fabiola and Morris explicitly identify the values they hope to realize through use of the tool Visible Effort (VE), self-guidance and self-actualization of collaborative online workgroups. In computer-mediated collaboration (CMC), VE enables groups to (1) measure and visualize the degree of collaborative unevenness, and the emergence of social structure, and (2) actively or passively steer the collaborative processes to attain specific goals (Matei et al., see above p. 4). Matei and colleagues set their discussion of VE in the context of the debate about social hierarchy and productivity in teams. In contrast to analysts who argue that flat, decentralized teams solve problems faster and more efficiently than hierarchical teams, they state that "CMC needs division of labor, coordination and clear goals" (p. 2)" They employ Shannon's theory of social entropy to conceptualize their approach and design their application of the VE tool. Social entropy refers to varying levels of random individual participation and group structure. The greater the social entropy the more random individual participation and the less coordinated their activities. Matei and colleagues hypothesize that productive CMCs strike an effective balance between social hierarchy and social entropy and offer VE as a tool to help find the right balance for any specific project. As a tool for measuring and visualizing social entropy, VE offers a means for CMC to discover the "inflection point" between social hierarchy and social entropy that best suits their task requirements and modify their work processes to help sustain it.

5.3 Basing Big Data Tool Design on Target Values

In their chapter on bottom-up decision making in urban infrastructure projects, Bakht and El-Diraby give a striking example of how enabling realization of a specific value (effective community participation in urban planning) and its beneficial consequences (creating an innovative and socially-savvy decision-making environment) affects the requirements and design of a Big Data Tool. They describe facing a specific problem: how to track citizen discussion over social media about urban infrastructure projects in order to incorporate relevant feedback into the planning process. Specifically, they wish to discover the semantic (ideas) and social (people) characteristics of "Infrastructure Discussion Networks (IDN)" as they emerge and evolve over Twitter. IDN constitute an example of "small world phenomena" with "relatively high clustering, comparatively small diameters, and short average path lengths" that "offer a good opportunity for information diffusion and viral marketing around the project" (Bakht & El-Diraby, p. 12). In addition to discovering IDN in the Twitter flow, analysts need to follow their evolution over time with respect to the ideas under circulation and the networks of people participating in the discussions through social media. IDNs mature; that is, grow in size and density of connections among members (triadic closure) as well as display changes in sentiments about specific subjects (dynamics of opinion). In the words of Bakht and El-Diraby, "by selecting a particular context for analysis of discussions over IDNs (such as sustainability), results of (Social Network Analysis) and lexical analysis can be aggregated to form the profile of online discussions for a particular project" (Bakht & El-Dirabyp, p. 16). Their tool produces a project discussion profile represented as a series of graphs over time for specific dimensions of specific issues in a specific project. From the noise of Twitter come community-based messages from influential community members to aid planners in better meeting community needs. Without such a tool specifically designed to elucidate the dynamic social composition and meaning of ephemeral communications on Twitter, the opportunity to feed community discussion into the planning process within a realistic time horizon would not exist. From the perspective of values, the tool reflects the importance of autonomy in the twofold sense of allowing individual agency to have an impact on the planning process and integrating a plurality of viewpoints. However, the mere use of the tool realizes autonomy only in a partial and asymmetrical way, insofar as the information from the community is gathered without the possibility of interaction and dialogue, which is the reason why the authors indicate that the full realization of this value requires real-world interaction, such as meetings and public consultations. Big Data tools can stimulate, but not substitute democratic interaction in the traditional sense.

6 Conclusion

In this chapter we have developed an approach for systematically analyzing ethical values in the design, use and evaluation of both Big Data information and tools. We intend for the Privacy Matrix including the analysis of both privacy contexts and ethical principles to enhance explicit identification, discussion and reflection among scientists, engineers and other Big Data developers of the values they employ in their work. Building upon the concept of values-sensitive design (Friedman, Kahn, Borning et al., 2001), we argue that the domain of ethics in Big Data has little bearing if it remains a subspecialty discourse among professional philosophers, ethicists or social activists. We also argue, however, that, whether explicitly recognized or not, values always inform the design, development and use of Big

Data Technology. As a community of Big Data technologists, we should attempt to make our values explicit and knowingly embed the values we seek to realize in the results of our work.

References

Ball, M. P., Bobe, J. R., Chou, M. F., Clegg, T., Estep, P. W., Lunshof, J. E., et al. (2014). Harvard Personal Genome Project: Lessons from participatory public research. *Genome Medicine, 6*(2), 10.

Boyd, D. (2014, July 1) What does the Facebook Experiment teach us? *The Medium*. Online.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662–679.

Cohen, J. (2000). Examined lives: Informational privacy and the subject as object. *Stanford Law Review, 52*, 1373–1438.

Cohen, J. (2012). Configuring the networked citizen. In A. Sarat, L. Douglas, & M. M. Umphrey (Eds.), *Imagining new legalities: Privacy and its possibilities in the 21st century* (pp. 129–153). Stanford, CA: Stanford University Press.

Collmann, J., & Robinson, A. (2010). Designing ethical practice in biosurveillance: The project Argus doctrine. In D. Zeng, H. Chen, C. Castillo-Chavez, B. Lober, & M. Thurmond (Eds.), *Infectious disease informatics and biosurveillance: Research, systems, and case studies*. New York: Springer.

Cooper, T., & Collmann, J. (2005). Managing information security and privacy in health care data mining. In H. Chen, S. Fuller, C. Friedman, & W. Hersh (Eds.), *Advances in medical informatics: Knowledge management and data mining in biomedicine* (Springer's integrated series in information systems, Vol. 8). New York: Springer.

Cooper, T., Collmann, J., & Neidermeier, H. (2008). Organizational repertoires and rites in health information security. *Cambridge Quarterly of Healthcare Ethics, 17*(4), 441–452.

Crews, C. W., Jr. (2002). *The Pentagon's total information awareness project: Americans under the microscope?* Cato Institute. Retrieved 2014, November 4, from http://www.cato.org/publications/techknowledge/pentagons-total-information-awareness-project-americans-under-microscope

Department of Defense, Office of the Inspector General, Information Technology Management. (2003). *Terrorist Information Awareness Program (D-2004-033)*. Arlington, VA.

Dittrich, D., Kenneally, E. (2012). *The Menlo report: Ethical principles guiding information and communication technology research*. US Department of Homeland Security. Retrieved 2014, November 4, from http://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803.pdf

Dumbill, E. (2012, January 12). What is big data? *O'Reilly Radar*. Online.

Einav, L., & Levin, J. D. (2013). The data revolution and economic analysis. In J. Lerner & S. Stern (Eds.), *Innovation policy and the economy*. Cambridge, MA: MIT Press. doi:10.3386/w19035.

Executive Office of the President. (2014). *Big Data: Seizing opportunities preserving values*. Retrieved 2014, November 4, from http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

Friedman, B., Kahn, P. H., Jr., & Borning, A. (2001). *Value sensitive design: Theory and methods* (UW CSE Technical Report 02-12-01). Retrieved 2014, November 2, from http://www.urbansim.org/pub/Research/ResearchPapers/vsd-theory-methods-tr.pdf

Fung, B. (2014, February 27). Why civil rights groups are warning against 'big data'. *The Washington Post*. Online

Future of Privacy Forum. (2013). *Big data and privacy. Making ends meet*. Retrieved 2014, November 4, from http://www.futureofprivacy.org/big-data-privacy-workshop-paper-collection/

Häyrinen, K., Saranto, K., & Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *International Journal of Medical Informatics, 77*(5), 291–304.

Hildebrandt, M. (2011). Who needs stories if you can get the data? ISPs in the era of big data crunching. *Philosophy of Technology, 24*, 371–390.

IBM. (2014). *What is big data*? Retrieved 2014, November 4, from http://www.ibm.com/big-data/us/en/

Kaisler, S., Armour, F., Espinosa, J.A., & Money, W. (2013) Big data: Issues and challenges moving forward. In *System Sciences (HICSS). 46th Hawaii International Conference, Maui, 2013*.

Knewton Inc. (2014). *About Knewton*. Retrieved 2014, November 4, from http://www.knewton.com/about/

Kroft, S. (2014, August 24) The data brokers: Selling your personal information. *60 Minutes*. Online.

Lazarus, D. (2014, April 24) Verizon wireless sells out customers with creepy new tactic. *Los Angeles Times*. Online.

Marwick, A. (2014). How your data are being deeply mined. *New York Review of Books*. Retrieved 2014, November 4, from http://www.nybooks.com/articles/archives/2014/jan/09/how-your-data-are-being-deeply-mined/?pagination=false

Matei, S. (2014). *Email correspondence 2014*.

Meyer, R. (2014, June 28). Everything we know about Facebook's secret mood manipulation experiment. *The Atlantic*. Online.

Moor, J. (1997). Towards a theory of privacy in the information age. *Computers and Society, 27*(3), 27–32.

Moyer, M. (2014). Twitter to release all tweets to scientists. *Scientific American, 310*(6).

Musgrave, S. (2014, May 3). A vast hidden surveillance network runs across America, powered by the repo industry. *The Boston Globe*. Online.

National Science Foundation. (2012). *Directorate for computer science & information science & engineering, critical techniques and technologies for advancing big data science &engineering (BIGDATA)*. Retrieved 2014, November 4, from http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*, 101–139.

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford Law Books.

Ramirez, E. (2014). *Protecting consumer privacy in the big data age*. Washington, DC: Federal Trade Commission.

Safire, W. (2002, November 14). You are a suspect. *New York Times*. Online.

Simon, S. (2014, May 15). Data mining your children. *Politico*. Online.

Simons, B., & Spafford, E. H. (2003, January 23). Co-chairs, US ACM Policy Committee, Association for Computing Machinery, Letter to Honorable John Warner, Chairman, Senate Committee on Armed Forces.

Singer, N. (2014, April 21). InBloom student data repository to close. *New York Times*. Online.

Solove, D. J. (2008). *Understanding privacy*. Boston: Harvard University Press.

Stanley, J., & Steinhardt, B. (2003). *Bigger monster, weaker chains: The growth of an American Surveillance Society*. American Civil Liberties Union, Technology and Liberty Program. Retrieved 2014, November 4, from https://www.aclu.org/technology-and-liberty/bigger-monster-weaker-chains-growth-american-surveillance-society

Stoové, M. A., & Pedrana, A. E. (2014). Making the most of a brave new world: Opportunities and considerations for using Twitter as a public health monitoring tool. *Preventive Medicine, 63*, 109–111.

Sullivan, G. (2014, July 3). Sheryl Sandberg not sorry for Facebook mood manipulation study. *Washington Post*. Online.

Tavani, H. (2008). Informational privacy: Concepts, theories, and controversies. In K. E. Himma & H. Tavani (Eds.), *The handbook of information and computer ethics* (pp. 131–164). Hoboken, NJ: Wiley.

Volkman, R. (2003). Privacy as life, liberty, property. *Ethics and Information Technology, 5*, 199–210.

Wittgenstein, L. (2001). Philosophical investigations. The German text, with a revised English translation (3rd ed., pp. 27–28). Malden, MA: Blackwell.

Washington Post (2002, November 16). Total information awareness. Washington Post, Saturday.

Critical Thinking and Socio-Technical Methods for Ascertaining Credibility Online

Howard Rheingold and Sorin Adam Matei

1 Introduction

Howard Rheingold is the author of the classic monograph "Virtual Community: Homesteading on the Virtual Frontier" (1993). The volume is the product of participant observation of the emerging "tribe" of cybercitizens (netizens), who at the beginning of the 1990s inhabited the early online world. The tribe had created their own subculture, with a jargon, values, and identity all their own. It was an utopian world, yet a world that evolved and over time gave us the social customs and expectations with which we approach online social spaces today.

The book "Virtual Community" recounted the daily lives of the WELL (Whole Earth Lectronic Link) bulletin board users over a period of several years. The WELL was an offshoot of the counterculture, founded and run for a while by the editor of the Whole Earth Catalogue, a flagship publication of the commune movement. One of Rheingold's conclusions was that given enough time and emotional investment, online spaces can become as "real" and vibrant as any other social spaces with a communal dimension. This conclusion was in counterpoint to the stereotypes of the time, which pigeonholed online users as anti-social nerds. The book Virtual Community is part of a series of meditations that Rheingold published on the socio-technical evolution of contemporary life. It builds on two previous volumes, one on computer assisted cognition and collaboration (Tools for Thought, 1987) and one on virtual reality (Virtual Reality, 1992). It was followed by a tract that aims to understand the mobile, "always on" revolution of the communication world introduced by the smart mobile phones (Smart Mobs).

H. Rheingold
Virtual Community, Smart Mobs and Net Smart, San Francisco, CA, USA
http://rheingold.com

S.A. Matei (✉)
Brian Lamb School of Communication, Purdue University, West Lafayette, IN, USA
e-mail: smatei@purdue.edu

© Springer International Publishing Switzerland 2015

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_16

A former member of the editorial team of the countercultural flagship publication Whole Earth Review, he is one of those early activists who has served as a Socratic midwife for what later was named the "cyberculture." A mix of libertarian and communitarian ideas, the cyberculture promoted or supported ideas such as net neutrality, supporting maximum freedom of expression online, or using technology to foster access to information and equality of public participation. Howard's vision encouraged his contemporaries to take the idea of online interaction seriously and to use it for building a new type of public sphere. It justified intellectually the massive investment, not only economic, but also social and intellectual, in virtual communing projects. In a way, the book made possible the social media revolution. Just like Rousseau's "Social Contract" prepared the French Revolution, Howard Rheingold prepared the new media social revolution.

Howard continues to be a keen observer of the socio-technical development of the Internet. His online projects aim to create educational experiences for young people, teaching them how to critically use online resources and social interactions (rheingold.com). The interview that follows, which he gave to the KredibleNet team leader, Sorin Adam Matei, in October 2013, looks at the implications the cybercultural revolution has on transparency, authorship, and credibility. After reviewing the promises, achieved or missed, of the cyberculture, Howard discusses his own philosophy and methodology for facilitating credibility and trust online. He emphasizes the importance of critical thinking and of developing a system of intellectual and cultural "checks and balances."

The interview is divided into two parts. The first one deals with the socio-cultural roots of our expectations (sometimes overly optimistic) for an online universe dedicated to the betterment of the human condition. The second discusses the theoretical principles and approaches for dealing with the reality that our expectations have often fallen short of the ideal.

Sorin Adam Matei

2 Online Interaction, Virtual Community: The Promises of the Internet Revolution

SM: Howard, I would like to start with the deep past of online sociability, online social interaction of the 1980s—when you were a member of the Well, the famous bulletin board that pioneered the idea of online community. We all read your book Virtual Community, Homesteading on the Electronic Frontier, and we remember vividly the justification you offered for writing the book and for theorizing about virtual communities—which was that, at the time, people thought of the idea of getting together with other people online as weird, as nerdy, as disturbing, even as a form of social deviance; that nothing good can come out of it, that it is just an adulteration of social life, a weakening of social ties. Was the fear of online interaction that intense in the 1980s?

HR: I don't know that I would use the word fear, but I would say that it was a stereotype and this is not a big part of public conversation yet. So now, if you were to ask a question like that today, digital culture and communication is a big part of our conversation. It was not a very big part of our conversation back then.

SM: But it was a part of a puzzled reaction to a certain subculture, wasn't it?

HR: Oh yes, it was a very active subculture. Well, Usenet has existed since 1980. That was worldwide. There must have been, by the time WELL came along, there must have been 100 countries or more on Usenet and probably millions of messages per year on Usenet, so [it was] a very big subculture for its time [...]. There were MUDs, there was IRC, which was big during world events—the way people go to Twitter now. People went to IRC when there were wars during the fall of the Soviet Union, and BBSes—BBS culture is what I was exploring before I got to the Well and it would be interesting to try to actually unearth actual figures, but the fellow that had the magazine called "Boardwatch" or something like that, estimated somewhere around 50,000 BBSes in the US. Who knows how true that estimate is? And that was definitely skewed toward teenage boys with modems on their PC's in their bedrooms.

SM: But the culture, the subculture, wasn't just that of teenage boys.

HR: No, but what I am saying is that you had teenage boys with BBSes, you had scientists and hackers and others on Usenet, you had people amusing themselves in MUDs. You had people tuning into the IRC to have persistent chat spaces. There was a pretty rich ecosystem when I came along to look at it. I was just one of the very first writers to come along.

SM: And you connected to this subculture within the context of the WELL, which was a subculture within a subculture in a way because it had its own countercultural roots in the San Francisco Bay area, and Stewart Brand and the Whole Earth Catalog was linked to it in a very direct way.

HR: It kind of recedes into the background of your memory, but in the book, "The Virtual Community", I talked about a virtual community in England that I visited and hung out with and people that I visited in France and MUDs; and I went to remote villages in Japan and there was a virtual community there, so there were things happening that I observed worldwide. The WELL was really what immersed me in that world. Then, when I went to write the book, I went to see what else was happening in the world and was not surprised to find that there was a lot happening and certainly it was a lot more than I uncovered.

SM: Now, there was an assumption at the time that this new subculture, this new technosubculture, was in large measure the product of the technology itself. And there are still people who think that what made this subculture was the technology—once computers became mini and micro and modems became affordable, all of the sudden the subculture formed by the spontaneous associations of individuals with interest in technology. There's also this other opinion, and I am more likely to subscribe to it than to the other one, that actually the individuals that latched upon these technologies and have selected themselves as self-declared members of a subculture had more to share than a love for technology, they also embraced some cultural values. Especially countercultural values of individual freedom, of activism, of non-conformism, and so on.

HR: Self-reliance…

SM: Do you recognize self-reliance and the other concepts and values mentioned above as the running thread through the subculture that invented the virtual community concept and practice?

HR: Self-reliance has been a running thread through American culture. It goes back at least around Ralph Waldo Emerson's essay of that name, do-it-yourself hobbyists in the 1950s, the Wright brothers, the automobile industry, practically most of what is seen as American industry came from individuals who started playing with things themselves. And for a lot of reasons that's associated with America. And of course, I thought Fred Turner got a lot of things right with "From Cyberculture to Counterculture" and one of the things he got right was the WELL being a network forum—a place where different networks met. So the hackers, as they called themselves then, the computer enthusiasts, they were only one of the different networks that met […] If you were to read that, you would think that Kevin Kelly and Stewart Brand and John Perry Barlow and Louis Rossetto and I would often get together to discuss our many political agreements. And, and that's like saying, OK, American politicians, you know, Newt Gingrich and Barack Obama. You sort of need a higher magnification to really see what's going on.

And what was really going on in the WELL was that there were people who were schoolteachers who happened to be curious and this was a way of connecting with other people. There were artists who were doing some of the earliest online experiments. There were the hackers, the tool builders, and then there was the Whole Earth crowd which, itself, seems monolithic now. But actually, that was a network forum as well. The people who were interested in self-reliance to the point where they're going to go out and create communes and, and build windmills and try to subsist on their own, those were quite different from the community people. So there were different categories among the Whole Earth Catalog. There was community and land use, people interested in home birthing. There were the people who you would call environmentalists today. The word didn't exist back then, but they were interested in Rachel Carson's Silent Spring and concerns about the environment.

These were not people who regularly communicated who were considered part of the same culture. The WELL was a place […] it was really the Whole Earth Catalog and the Whole Earth Reviews. Those were communities that were interested in their own heterogeneity. So when I was editor of the Whole Earth Review some years later, in the 1990s, I had people from the intelligence community, active CIA analysts; I had people in the Pentagon, all kinds of people you would not expect.

SM: Right, but the heterogenization came at a time when the movement had become mainstream, in a way? You know, what used to be a subculture in the 1980s had become a part of the culture, during the 1990s? [...]

HR: If you look at the beginning, the model of the Whole Earth Catalog, thinking for yourself is a value there. And I see that as something that drew people together I wouldn't say that you had a lot of people from what we would consider the right. There are a certain amount of people from what you would consider to be right and libertarian, and then you had leftie activists. And what they had in common was not accepting what the Democratic Party or the Republican Party or the Communist Party told you, but thinking through things for yourself. And I think that there was a strong connection there with the ethos. If you go to Steve Jobs' Stanford Commencement Speech, he talks about the influence of the Whole Earth Catalog. And the part of the ethos of personal computer makers was to give individuals the ability to have access to these tools. And that was Doug Engelbart's vision of augmenting human intellect. This is really hard to disentangle, and in fact [it's] impossible to disentangle all of the different influences. Without that, if it had just been the military [as some say] probably never would have happened.

SM: But at the same time, the way you tell the story now, it makes it sound as if there was a direct continuity between the mainstream strand of American individualism and self-reliance and the techno-culture of the 1980s and the 1990s. I'm wondering if, actually, there wasn't something extra mixed into the virtual community movement. Wasn't it a way also stimulated by a streak of—how should I put it?—non-conformism writ large. I'm using this word again, as countercultural opposition to the mainstream. So it wasn't quite in the mainstream.

HR: I think that it's fair but myopic to conflate thinking for yourself with what's called the counter-culture. [The counter-culture is a reification]. The anti-Vietnam War movement, Civil Rights, the budding technological movement, the let's go out and create our own utopian communes....

SM: The Whole Earth Catalog itself.

HR: ... the psychedelics, those, that was not like some movement. These were people defined as not the American conformity stereotype. So, there are lots of ways; to call someone a non-conformist is similar to saying non-American. OK, are you talking about an Australian Aborigine or are you talking about someone who lives in Stockholm? Again, I keep using this metaphor of what magnification are you looking at it. In the reification of history, there was a counter-culture. But there were lots of different people who didn't think different aspects of American culture were right and they wanted to either change it or adopt something else.

SM: Well, when I used the word counter-culture, I was trying to suggest that there was an activist aspect to embracing the techno-culture.

HR: Yes.

SM: It wasn't just done for the purposes of passing time.

HR: I would say that at the beginning. . ..

SM: There was a personal investment in, in it.

HR: . . . people who were activists, political activists . . . were a part of the mix. And they sort of infected the [movement]. It was very interesting to have people who had no real technical knowledge, like me, in contact with people who were making the tools. And there was a definite dialogue about—"these are the kind of tools we'd like". "Oh, yeah, we can maybe build something like that". That sort of becomes blurred. Remember, also, there was not a commercial aspect to this, initially. In 1994, if you said "I'm going to make money on the internet", you would have been laughed at. 1994, this is quite a bit later than what we're talking about. In fact, [the early Internet, pre 1994] you couldn't use it for commercial enterprise. It was after the Netscape IPO. That was definitely the event that signaled that, oh, yeah, you can make money here [. . .].

SM: I want to steer the discussion in a slightly different direction that is connected to the conversation that we've had so far about the cultural assumptions of the techno culture in the early 1990s. The techno-culture wasn't just about self reliance. It was very much about democracy and equality and access, right?

HR: Yes.

SM: There was a great expectation that adopting these tools we would be able to communicate more freely, more equally, while people who are marginalized would be able to talk with each other better and more.

HR: Yeah.

SM: This expectation did impact the world and it changed many things, but at the same time, in many instances, and especially those instances where communication was free and completely voluntary such as Wikipedia, we do have as a basic fact of life a tremendous inequality of participation. Pareto's Law still applies: 80 % of the content, or more, is produced by 20 %, fewer, of the individuals. This leads to inequality of power because the ones who spend the most time in these fora become old timers and those old timers get to know the system a little bit better and at times they even game it. Or, let us be delicate about it, if they don't game it, but they steer it in a specific direction. This actually, was the case for the WELL, right? You know, some of the people who were more present on the WELL, they became conference moderators and being conference moderator is—

HR: The [WELL] was a postocracy.

SM: Yeah.

HR: You wanted to influence the WELL then you posted.

SM: Yes.

HR: You wanted to have more influence, you posted more or you posted more persuasively.

SM: Right. But not just the WELL, but more recent phenomena such as Wikipedia and all the other sites where people share knowledge which have a tremendous influence on the world. We have some people who are much more equal than others and this has created a specific pecking order that's even more than a division of labor, a pecking order with leaders and followers. Some of the leaders have real power and they can imprint their own visions, needs and desires upon the sites and upon the knowledge they produce. Is this a problem? And if it is, should we do anything about it? And if it's not a problem, why isn't it a problem?

HR: So—oh, okay. Let me say several things about that. First of all, I encourage you to think to some degree "both/and". It's not very "either/or". Secondly, I'm not the first person to say this, but as this medium that was used by a relatively small portion of the population came to encompass the majority of the population, the cultures that emerged from the medium began to more closely reflect the way things went in society. So there are a lot of aspects of the broader society that were not very magnified in the early days that become more visible now. Secondly, to the degree that Wikipedia wants to meet its own ideals then they have these issues of inequality. I'm sure you know about recently this last week at Brown University they had this edit-a-thon in which a number of people edited and created Wikipedia articles because they felt that women scientists were not adequately represented. So there's not some kind of official prohibition from doing that. They are going to come up against the same issues that anybody who edits an article on Wikipedia does, which is they're going to deal with the entrenched elite who are going to want to delete their article if it doesn't meet all of the many, many, many norms that have accreted over the years, and you have to fight that in order to fight your way in. That reminds me of Nancy Fraser's article on the public sphere; the theory of the public sphere from Habermas was about bourgeois white men, and Nancy Fraser said people have fought their way into the public sphere by creating this subaltern publics. Women, the Civil Rights Movement, people who were excluded from the conversation, the political conversation, they became part of it despite the exclusionary forces by forcing their way in, by creating their own conversation and creating a conversation that had enough power of public opinion. So, Wikipedia. There are people like my friend who'd very much would like to have more women editing. They would like to have more people editing. They're trying to think of ways to do that and, in a sense, the people at Wikipedia who are interested in that, they come up against this community and the community like—I mean, any online community that's successful [has] bonds, and so it becomes harder for newcomers to break in. And on Wikipedia they have developed, I think, to their benefit, a culture where there's a tussle about something and there's a resolution about that, then there's a norm that is applied. So the next time that tussle comes up, they say, oh, that's the neutral point of view norm. This is what we have decided. And that's often used to suppress change […].

SM: Okay. But going back to my question: is there—should we do anything about the uneven distribution of power on social media that produces knowledge? As far as I understand it, your answer to this question is: let's look at how the uneven distribution of power can do good or bad for us. On the one hand, it can do good for us in the sense that it creates a division of labor and creates investment in the project. When some people are more involved, they care more about it. This is how I understand your reference to postocracy, on the Well and this applies, I suppose, to Wikipedia. The more you post, the more you care and the more you care, the more you post. And, the better for the community.

HR: WELL, when you say "we" should do something about it, are you talking about white male academics in the USA?

SM: No, I'm talking about us, the people who use the Internet, who use online media for getting information.

HR: But the people who use the Internet—people come together around interests.

SM: Of course.

HR: So there are feminists at Brown University who are taking it upon themselves to address an inequality that they perceive. Who are [the] others out there who want to fight their way in? So my friend who works there, he's trying to find ways to get—the number of people who use Wikipedia versus the number of people who edit it at all is, as you know, huge. And so he's coming up with things, ways to involve people so that they're just adding a kind of a "like" feature to it so that people can comment on edits. Just say that they like an edit, or they don't like an edit, or an article, or they don't like an article. So they're very much thinking about how do they draw people in, and the people at Wikipedia I know who are concerned with that, they come up against these kind of cranky elite.

SM: Which it is?

HR: Which it is!

HR: But I say that because Nancy Fraser—when you're talking about the public sphere, you have to define "we" because different "we-s" fight their way in and that's how it happens.

SM: Let me be a bit more specific. Since I am talking to an early explorer of online sociability, I was using that "we" designating those who imagined online interaction as being naturally open, democratic, and egalitarian in the 1980s. That was the "we" I was referring to. But I could call it "they," too, if you do not feel a part of that. I don't care much. Those "we" or "they," are faced by a reality that has distanced itself from the ideal. The Internet is not a promised land, it is not an egalitarian utopia. And the question that we have on the table now is which vision came to pass? Who won in the end? Did the principle of practical realism win? Did the common sense idea that humans are as humans do, win? Should we accept the fact that humans engage in collective action with various degree of commitment, allowing some to lead and some to follow? Is egalitarianism possible? Should we push harder for egalitarianism on the Net? What should the people who dreamed of egalitarianism do today when confronted by the the rife inequality of cyberspace?

HR: Definitely. I definitely have answers [. . .]. I've spent the last 40 years involved with the trade book publishing industry. Trade book publishing industry funnels a very large number of ideas through a tiny elite of agents, who funnel it to a tiny elite of editors, who go into meetings in which a tiny proportion of what they decide to publish is promoted. Is Wikipedia so inegalitarian; Is the Internet so undemocratic and enclosed compared to what there was until very recently? I mean, we're seeing, of course, it's either great contention: the pay-walling of knowledge by commercial enterprises versus open publishing in the sciences. The whole Aaron Swartz JSTOR issue. So who gets to publish and put whatever their product is out for public inspection, public consumption? That's still very much in contention in a relative sense; compared to before the Internet the whole thing is broken wide open. Okay. You have to have the five dollars to have access to the Internet so that you can publish on Tumblr but, you know, anybody in the world with any idea in the world can publish it and if somehow or another they can convince people to link to it, it comes to other people's attention. Are well funded, already powerful individuals, and factions, and enterprises going to have a heavier thumb on the scales? Well, yes, hasn't that ever been the case? On the other hand, boy, we could spend all day talking about stories of people that you would never have heard of otherwise. I think YouTube is a whole other site for talking about this.

SM: Oh, definitely. It's a major learning center.

HR: So remember the Numa Numa Boy?

SM: Yes.

HR: Okay. So here was like a folk song in Slovenia.

SM: No, actually it's a pop song from Romania.

HR: A pop song. From Romania.

SM: Yes.

HR: There were something like 50 or 60 thousand tribute or parody things done [...]. But what kind of phenomenon is that? It's certainly not the music publishing industry as we know it, certainly not the entertainment industry as we know it. And it's certainly not the, by far, the only example of something coming from nowhere to influence a lot of people. So we've got—so capital, power, entrenched elites, whether they're commercial or noncommercial. That's not going to go away. That's they way humans operate. There's the, what is it? The Pareto principle, you know? 80 %, 20 %.

SM: Right, and that's what we need to grapple with.

HR: We have a much larger marketplace of ideas. The doors that admit people and ideas to the marketplace of ideas, I would contend, have been radically enlarged. What happens after that is a number of mechanisms that come into play. The old fashioned who's got money, who's got power. Definitely ABC is going to have more clout than my news operation. On the other hand, you get your WikiLeaks and your Glen Greenwalds, people who are definitely not part of the establishment, are breaking big news. I think [...]

SM: I agree that we have a much larger elite and a much more diverse elite because of this, which I completely agree with. Isn't it still an issue that we need to recognize and discuss?

HR: Oh, okay.

3 Trust, Credibility, Equality? Understanding the Social Processes That Define Content Production and Consumption Online

SM: The idea is not to reject the idea of elites, but to recognize that there's an emerging elite, the digital elite, the digerati elite, that dominates public discourse. In the more cultivated circles we have the digerati that we recognize from the books, Weinberger and the other people, right? Or maybe yourself. But in the trenches we have the handful of people who edit Wikipedia at a phenomenal rate. Or the people who spend their days answering questions on AskYahoo.com or Quora, or the people who are on YouTube showing us how to fix the sink and all that, who are becoming quite famous, right? Most of the time we have no idea who they are, although they influence our lives to a tremendous degree. The most prolific editor on Wikipedia could be someone who calls itself—we do not know if it is a he or a she—wonderful_midget_22. Shouldn't we know who these people are so that we could keep them accountable for their words? How can we do that? Trust and credibility are still an issue online, even if the net social effect of these people being online and doing things for us is positive? Even if their work is socially beneficial, we still have to hold them accountable, especially since under the guise of being one of the good new elites, some people can do really terrible things? They could actually use their position of post-o-crat or Youtube-crat to spread racism, or fundamentalism of any kind and shape. How do we deal with this?

HR: You know, I detect an assumption behind your saying "hold them accountable". You're still operating on the assumption that comes from the tradition within which you, as a researcher and an academic and I as a writer come from. We can still remember when it existed, which was the authority of the text. There was a series of gatekeepers that held accountable the truth claims of an author. So, you get a book out of the library and you might disagree with the book but for the most part, you don't assume that they say George Washington was born in 1732, that you have to go check that. But, for better and worse, anybody can publish anything, and a search engine, maybe they're getting better, maybe they won't, but search engines are not going to tell you that this is authoritative and that is bogus. The holding-accountable is not "let's hold the experts accountable", it's "let's hold ourselves accountable" for whether we know how to separate the wheat from the chaff.

So, I see an analogy to the Lippman–Dewey debate of a century ago. So, Lippman said Americans are notoriously ignorant and they're willing to believe anything and they're easily misled. How many years ago that was? I wonder if that's still true? We can't afford to have these people making decisions about an increasingly complex, increasingly technological society. We need an elite. We need a technocratic elite. John Dewey's response was, if Americans are ill-informed, we need better journalism. And, if Americans are easily misled, we need better education so that they can make their own decisions. They can think for themselves. This "think for yourself" business is really important, because there are enormous forces other than your Wikipedia elite, who have a lot of power, knowledge and capital devoted to manipulating what other people think. And that was, of course, Habermas's fear of the public sphere and he was really speaking before the science of public relations really developed. So, my answer is Dewey's answer. It's a literacy issue.

More people need to understand what you're saying. Wikipedia reflects a tiny elite and that you can't believe everything you read on Wikipedia or everything you read in the Britannica. According to the Nature study, they're approximately equally inaccurate. You have to, somehow, use the tools that are available to you to check for yourself. And then, when you're talking about people Googling for symptoms, you're talking about health care, you can actually kill yourself if you don't know what to believe online. On the other hand, ask practically any cancer patient or caregiver for someone with Alzheimer's, whether their online connections are useful to them.

Now, part of that is that you educate yourself about what's legit or not. Part of it is that I think that there is a definite role for the National Institute of Health, your experts, the authority of the text, to produce some kind of add-on to a web browser, so that they can certify some sites that make medical claims as not being bogus. On the other hand, a lot of people who are patients of diseases—cancer is just one that I have experience with, one that a lot of people have experience with, but you know, there are thousands of diseases—people develop communities that argue about things among themselves.

So, you know, there's the Patients Like Me and lithium study. Patients Like Me is a commercial enterprise, and their idea is that they can put patients with particular diseases together with other patients who have that disease. That's a service they offer and they're hoping to monetize that by gathering data from those people, anonymizing that data and selling it to pharmaceutical companies, for example. People there who are concerned with ALS, Lou Gehrig's Disease, there was a rumor that lithium would ameliorate the symptoms of that. So, they, a group of patients on Patients Like Me, got together, studied what are your protocols for doing a clinical trial. Let's do our own clinical trial, and they concluded that lithium was not useful, 2 years before the peer-reviewed the medical journals did. Are citizens going to be doing legitimate peer-research in the future? I'm not sure. But, clearly, that was a group of patients who had educated themselves beyond believing whatever it was that they could Google up about their disease.

SM: I want to go back to the idea I started with, holding people accountable. For me, to "hold them accountable" does not mean that there should be judges who should license authors and creators of content, and if you don't do it the way you are licensed then they will revoke your license and you should not talk. It was not this censorious understanding of the word ‘accountable’ that I was using. The word meant to say: “shouldn’t we know all that is needed to be known in order to make an informed decision?” And then, in this respect, I’m all with you, in it is ultimately up to the reader, to the user of online resources, be it Wikipedia or whatever, to decide if what they hear or read or see is truthful or not. But, sometimes you cannot make this decision until you know everything that is to be known, not only about what you read but of the context in which it was written or created, and the motivations of the person who has written or read, and, the ultimate advantages, benefits that might accrue to this person from what they have said or written. And, until we know who that person is, why she did it, how she did it, we cannot really judge the acts of that person and we cannot hold them accountable, so to speak, if you are just a screen name.

HR: In education circles, there’s a name for what you’re talking about. It’s called “critical thinking”. And, the idea of teaching critical thinking, teaching critical thinking about media in schools, predates the internet and is widely recognized by the right wing as a communist plot. So this ‘thinking for yourself’ business, there are a lot of people in the world, and a lot of people in America, and a lot of people in the U.S. Congress, who absolutely do not want people thinking for themselves. The truth is written, it is written in the Bible, written in a number of other sources, and deviation from that is subversive. That’s also a very long American tradition. There was something called “The Know-Nothing Party” at one point. It’s not just an ideological resistance. If you are a parent, or you are a teacher, it’s a lot of work and a pain in the ass to encourage your kids or your students to question authority. At some point, you say “Because Dad said so!”.

SM: What you’re saying is that you should apply the touchstone of critical thinking to anything and the truth will be revealed. Yet, in certain situations just critical thinking is not enough. Again, you know, if all you know about a person is a screen name and if you know nothing about that person’s intentions, interests, ultimate benefits; if all you have to go by is the accurateness or the logical consistency of the argument, if you just look at what they said formally, well, the devil himself would make a very good spokesperson or writer, right? And, whatever he or she has to tell us might make sense if we don’t know it was the devil, right? [. . .] Shouldn’t there be a much more vigorous debate about knowing the identities and the reasons of the people who create our content today? When somebody signs a piece of legislation, or an article in a newspaper, they’s a signature or a byline, right? And there’s a title that is verifiable and we know who that person is and that this is a liberal, this is a conservative, or this is a smut peddler, or this is a great writer. We kind of know something about them, right? It’s all behind a veil of ignorance. More often, actually, it is the very veil of ignorance that somehow reassures us. The current assumption is that if we do not know much about them is OK. These are good people, trust me. I’m not saying they’re not good people. I’m just saying trust, but verify. I mean, shouldn’t we have a discussion about verification?

HR: Well, if you have the time and patience to do so you could go and look at the pages of the Wikipedia editors who are influential in topics that you’re interested in and see what they've edited, and what edit wars they've gotten into, and what talk wars they've gotten into and you can infer their biases by doing that. Wikipedia makes all that information available. So I talk to Stanford students, educated elite college students, and I ask: who uses Wikipedia? Everybody raises their hand. Who looks at the discussion pages? Who looks at the talk pages of Wikipedia. Nobody. Nobody raises their hand.

SM: Isn't that a problem? Shouldn't we have a discussion about that? That's what I'm trying to say.

HR: That's why I'm teaching them.

SM: Now we're coming around.

HR: So when you talk—we're have a discussion, that's why I wrote Net Smart, because I think that what people know is the critical uncertainty. You know, I think there's a large percentage of people—as I said, I don't want to repeat myself too much. They don't want to know. They don't want their kids to know. They don't want to investigate. But then there are, I think, are a lot of people who simply don't know. They haven't been taught. It's not taught in school. I keep saying that you need to—what is it? It's about 8, or 9, or 10 is the age in which most kids in affluent countries get smartphones. And, increasingly, it's not just affluent countries. There's a billion smartphones in the world. Shouldn't they be taught how do you get the answer to any question? Something that was not available ever before, unless maybe you had access to a university. Maybe they wouldn't have the answer to every question. And then how do you test those answers to find out which ones are legit or not?
[. . .]

SM: I know that for the past probably several years you've been teaching and you've been very concerned about developing this particular type of critical thinking. It's the critical thinking that helps you not only understand the message and its face value credibility but also to understand its ramifications in terms of consequences and also of its origins. So I know that you are very concerned about understanding digital products in all their complexity. And you've been talking about "bullshit detection," as you name it. Now I understand that this in another vocabulary should be called a special type of critical thinking, maybe heuristic critical thinking. Now I imagine this to be like a tool kit, made of several procedures, several types of skills, things that we need to do. Can you talk about these things? Can you detail them? What are the things that we need to know and do?

HR: Well, first of all, there are technologies that could help. So I'm sure if you're investigating Wikipedia you've come across people who are trying to come up with overlays. So there was one—I can't remember the name of it. You know, I think I mention it in my book in which you can see the edits—if you use this filter you can see edits that are made by editors whose edits have withstood challenges.

SM: WikiTrust. And that was created by my colleague, Luca De Alfaro, who supports some of our own work on mining Wikipedia for leadership positions.

HR: Yeah.

HR: So, okay, so that's an example of a lens or a technology that could help you make those decisions. Heuristics. Well, the simplistic thing is who's the author? And if there's not an author use "who is" to find out who's responsible for the website. You've got a name, search on the name. I mean, that's not difficult but it's really the first step. And I don't even think that people need to use their real names as long as they use a consistent name. And one of the things—when people want to—back in the days when I consulted with people about creating virtual communities I said if you want to foster a culture of trust then give everybody a homepage, a personal profile, and have links to all of their posts on that page. If you want to see: is this guy just trolling me? Or is this person always taking a vegetarian point of view, or whatever their bias might be? You can then go look and, whether they have a identifiable name that you can find in a phonebook, or whether it's a pseudonym, if it's the same pseudonym that has made all of these posts, then you can go look and maybe make up your mind for yourself. So being able to identify people's—to aggregate what people do under a particular name—of course people can have multiple names. They can create sock puppets. They can create a whole army of people agreeing with each other and that's, you know, done. That's another combination of tool and heuristic. You can't do the heuristic without having the tool.

SM: Let's broaden the conversation. Wikipedia is just a small, small aspect of our information of our lives. What are the kinds of things do we need to do in our everyday digital lives to be more reassured that we're not fooled or deluded led astray, so to speak? To become more self-reliant digitally.

HR: Well, I encourage people, when their grandma or cousin sends them the article about the little kid in England who wants to get postcards, to send them the link to the Snopes that says the little kid in England hasn't wanted postcards in 20 years. Please stop besieging the postmaster in this little town in England with postcards. Again, that's sort of cultural. They're the people who put together things like Snopes. And, if you were to go into the business of seeking it out and aggregating them, there are a large number of well-known urban legends that keep circulating. At one extreme, there's the "Barack Obama was born in Africa" business, but I think there are a lot of things that are less politicized than that [which] people keep spreading. It doesn't seem to have done a lot of good. I mean, I've been doing this for a long time, and it doesn't seem to be killing those things. People sort of want to believe these things that they're sent. I think that's a very large problem. It's about public education.

I think, even before you could talk about what is the heuristic, you have to have some kind of mindset that empowers you to, it's not just a kind of political issue of freedom to think for yourself or are you going to believe whatever your orthodoxy is. I think that there's the practice of it, and you know, we're going back to Ivan Illich and Paulo Friere and John Taylor Gatto about schooling is about compliance. You stand in a row and hold your hand over your heart and say the Pledge of Allegiance and you march in an orderly fashion into your classroom and you sit in your designated row and column and shut up while the teacher talks and then when the bell rings, you move to the next room. It's difficult to think of how else you're going to do it, but of course, if you can afford to send your kid to a Montessori school, there are other ways of early education, but you know, certainly this critique is not original with me. We are having a conversation about twenty-first century problems and twenty-first century skills, but the people you're talking about are educated in a system that was created to turn farmers into factory workers. It's about the nineteenth century assembly line broadcast model, industrial-era mindset. And so, you know, you begin pulling on this thread and it leads you to public education every time, I think. I mean, that's the issue. If you were to, to answer your question, are we going to hold people accountable, I think we need to start teaching people what that means fairly early.

SM: Recently, you published Net Smart. It's your fourth major book, after Tools for Thought, Virtual Community, and Smart Mobs. Some put them in that school of thought that could be called 'technoenthusiasm,' which claims that technology is for the most part a positive force in our daily life with great potential for social and cultural change. Correct?

HR: It is put in that category, but has nobody read the last chapter of "The Virtual Community"? Do they all give up before they get to the chapter called "Dis-Informocracy"? Didn't the first page of "Smart Mobs" say that terrorists can use these things to do bad things? Collective action. The Red Cross is collective action. Al Qaida is collective action. Humans. Again, what I said a long time ago is, as the medium expands from a small number of enthusiasts to more closely resemble the human population, it takes on human characteristics. So, gee whiz. There are people in history who've had Utopian ideas, like 'humans shouldn't be slaves. We should abolish slavery' that eventually become the mainstream. Or, "women should vote". Utopian idea. So, I don't know, I think it's, there are, I think, people who have a lot more uncritical enthusiasm, but I've been the main target for so long that, I think, people will put me into that category. I just don't see that it holds up. I'm aware of that criticism. I've been aware of that criticism for 20 years. I try to think critically about what I'm talking about.

I do think there are tremendous advantages to these tools, and I also think that, like any tool that's used by a large portion of the population, some people are going to figure out how to use it to increase their power over others. So, now we're in difficult territory. Power, counter-power. Knowledge, power. This is an arms race. It's evolutionary. You know, I've warned about surveillance, and I can send you columns I wrote in 1995 about, I think it's a great idea that I don't have to stop to pay my toll on the bridge. On the other hand, I am aware that I am now creating a digital signal that, together with my credit card swipes and my phone calls, are 1 day going to create a digital dossier and the technology is not there for putting all these pieces together yet, but when it does, we will see a surveillance state far beyond what Orwell envisioned. Now is the time to think about what kind of regulation we might want to make. Because, after all, people used to listen at doors, they were able to tap phones, and we don't, it's impossible to stop them from doing that, but if they're going to arrest you and take you to court and try to take your freedom and your life away, presumably they have to have a warrant that a judge issues on your evidence of probable cause that a crime has been committed. Otherwise, the evidence is inadmissible. So, they can listen at your door all they want, but they can't put you in jail unless they did that according to the regulations that our democracy has come up with. And, you know what? Nobody cared. Nobody really cared.

SM: Final question, since you mentioned the surveillance society that the Information Revolution brought about. There are two schools of thought here. There's the school of thought that says we need to regulate technologies more and that we need to regulate the regulators, to monitor them better. We e need to create laws and controls that forbid or prevent individuals, corporations and governmental organizations from snooping on us. And then, there is the school of thought that says, no, actually, the best way to handle the crisis of privacy it is to open up all the protocols and all the software and let everybody, basically, stand naked in front of everybody else. In the manner proposed by David Brin in The Transparent Society.

HR: You can't, you can't regulate with secret code. One thing the [recent Snowden] NSA revelations [. . .] are reinforcing is that you can't ever really trust that code, unless your security code is open so that people can see if there are [some] backdoors in it. So, I think there are two issues here. The technical issue is that security measures like encryption are more secure. That is Bruce Schneier; he would say that. He's the security guy [. . .]. He would say, if they're open, then all kinds of people are gonna pry into them and see whether there are backdoors into it. Who is gonna buy a chip from China? Who's gonna buy a chip from the USA? I mean, more people are gonna have to make their own chips because they now know that there are backdoors built into them. So, that's a technical issue. But in Germany, they care a lot more. Americans don't really care.

The Patriot Act, nobody really cares. In Germany, they have [fought] Google from doing Street View because they feel it's a privacy intrusion. They have, you know, they have an acute knowledge of their recent history and there's a lot of faith in the ability of state regulation to somehow ameliorate the intrusions of technology. So that's a public sphere issue. The public sphere in Germany has a very, very different opinion about regulation and technology and free enterprise than Americans do. I sort of gave up on writing jeremiads about "we really need to do something about this while there's still time." There's not still time. It's all over. It's too bad, but that's the reality. I mean, Admiral Poindexter proposed total information awareness right after 9/11 and there was an uproar about it and it didn't happen. Oh, except it did. They just didn't talk about it. And it'll be interesting to see what the result is of the [next revelations will be]. There was a comment attributed to Edward Snowden yesterday [October 18, 2013] saying the revelations to come are going to dwarf what we've seen so far. I can't imagine what that possibly could be. And also, apparently, not that many Americans are that [excited] over the NSA. For one thing, people don't really understand metadata and social network analysis very well, or at all.