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

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.