Tutorials

Advanced Methods for Collecting Social Science Data in the Social Media Field

Presented by Riki Conrey

Valid social and psychological data are crucial to answering the most interesting questions in social media. We need these data to parameterize agent-based models; append motivational information to social graph data; measure perceptions of risk in online decisions; and, in general, to put social media behavioral data in a human context. Collecting valid information about the unseen content of human cognition is difficult, but computational social scientists often focus their energies on the computational challenges associated with the research. Compared to calculating clustering in a massive graph, conducting an online survey seems easy. Posting an online poll is easy. Conducting an online survey that yields valid results is difficult and sometimes impossible. In this tutorial, we will discuss advanced methods for measuring human cognition, motivation, and decision-making. We will cover sampling, survey design, analysis, and alternatives to conducting a survey.

Sampling. A "probability" sample is the basis for valid social science. We will discuss the basics of sampling, challenges of sampling online, and recent innovations that can improve the quality of social media samples.

Survey design. The quality of psychological data is as much about the way you ask questions as about what you ask. We will discuss the best ways to structure questions and answer categories and to present questions on the screen to maximize response quality and response rates.

Analysis. Psychological data prevent some unique analytic challenges. We will discuss accurate variance estimation for complex samples, and inferential statistics for typical psychological distributions.

Alternatives to conducting a survey. There are some things—some of the most interesting psychological things—that people simply cannot report on a survey. We will discuss alternative ways to measure psychology using behavioral data and gamelike tasks. We will also discuss how to combine social media with existing survey data collected by the US Census, Centers for Disease Control and Prevention, and other organizations.

Information-Theoretic Tools for Social Media Analysis

Presented by Greg Ver Steeg and Aram Galstyan

Social media is a collection of moving targets. Both the platforms and the behaviors of the users of these platforms are diverse and constantly evolving. Ad hoc models based on assumptions about today's users may not hold tomorrow. Information theory provides a general framework for identifying meaningful signals without relying on assumptions about human behavior or on platform-specific implementation details. The flexibility of the information-theoretic approach allows researchers to go beyond the study of "re-tweets" to consider rich data including textual content, timing, and context.

The primary objective of this tutorial is to provide a gentle introduction to basic information-theoretic concepts and to demonstrate how those concepts can be applied in the context of social network analysis. In particular, we emphasize an interpretation of these quantities as measures of predictability. The strongest signals in social media, and the ones most amenable to research, are the ones that most predictably lead to change. We will use several case studies to illustrate how information theory can be fruitfully applied to real-world social media and to demonstrate how this analysis can be simplified with available tools.

Crisis Mapping, Citizen Sensing, and Social Media Analytics: Leveraging Citizen Roles for Crisis Response

Presented by Amit Sheth, Patrick Meier, Carlos Castillo, and Hemant Purohit

With the explosion in social media (1B+ Facebook users, 500M+ Twitter users) and ubiquitous mobile access (6B+ mobile phone subscribers) sharing their observations and opinions, we have unprecedented opportunities to extract social signals, create spatiotemporal mappings, perform analytics on social data, and support applications that vary from situational awareness during crisis response, preparedness and rebuilding phases to advanced analytics on social data, and gaining valuable insights to support improved decision making.

This tutorial weaves three themes and corresponding relevant topics — (1) citizen sensing and crisis mapping, (2) technical challenges and recent research for leveraging citizen sensing to improve crisis response coordination, and (3) experiences in building robust and scalable platforms and systems. It will couple technical insights with identification of computational techniques and algorithms along with real-world examples. We will also do live demonstrations of the Ushahidi and Twitris platforms while elaborating on the practical issues and pitfalls of the development and operation of these large-scale platforms, especially during the real-time crisis response.

Amit Sheth (PhD) is an educator, researcher and entrepreneur. He is currently the Lexis-Nexis Ohio Eminent Scholar at the Wright State University, Dayton OH and the director of Kno.e.sis - the Ohio Center of Excellence in Knowledge-Enabled Computing which works on topics in semantic, social, sensor, and services computing over the web, with the goal of advancing from the information age to meaning age. He is also an IEEE Fellow.

Patrick Meier (PhD) is an internationally recognized thought leader on the application of new technologies for crisis early warning, humanitarian response and resilience. He presently serves as director of social innovation

at the Qatar Foundation's Computing Research Institute (QCRI). He is an accomplished speaker, having given talks at several major venues including the White House, UN, the Skoll World Forum, Club de Madrid, Mobile World Congress, PopTech, Where 2.0, TTI/Vanguard, SXSW and several TEDx's.

Carlos Castillo (PhD) is a senior scientist in the social computing group of the Qatar Foundation's Computing Research Institute (QCRI). Prior to QCRI, Castillo worked with Yahoo Research. He has influenced research fields on several topics including information retrieval, spam detection and demotion, usage analysis and social network analysis. His current research interest is the mining of content, links, and usage data from the Web to fuel applications in the news and crisis domains.

Hemant Purohit is an interdisciplinary (computer and social sciences) researcher at the Kno.e.sis Center where he coordinates crisis informatics research under NSF SoCS project. He is pursuing a unique approach of people-content-network analysis for analyzing social signals with insights from psycholinguistic theories of coordination to answer: whom to coordinate, why to coordinate and how to coordinate. His work also involves problem spaces of community engagement and sustainability, expert detection and presentation.

Multiple Network Models for Complex Online Social Network Analysis

Presented by Matteo Magnani and Luca Rossi

Multiple network models are fundamental to provide accurate analyses of human relationships. For example, while Facebook connections can explain a lot about a user's social life, her professional network may require an analysis of LinkedIn connections and her information consumption practices might be better explained by looking at her Twitter network. In addition, interesting patterns may emerge from the analysis of the dependencies between different combinations of these so-called network layers.

This tutorial will review the main theoretical

models, data gathering methods and analytical tools to deal with multiple networks and to understand how a multi-layer network perspective may change our knowledge of user behaviors. Multiple online network analysis is a recent and growing field, with long-standing theoretical bases rooted in classical sociological analysis and multiplex social network analysis methods. As such, it presents numerous research opportunities both for experienced researchers and young academics looking for a field of specialization.

Matteo Magnani has worked as a researcher at the University of Bologna, Italy, at Aarhus University, Denmark, and has visited the Imperial College London. He is currently a researcher at C.N.R., Italy, in the KDDLab, ISTI (Institute of Computer Science and Technology), Network Data Management division. His main research interests span database and information management systems, specifically uncertain information management and multi-dimensional database queries, network science and social computing.

Luca Rossi is an assistant professor of media analysis at the Department of Communication Studies and Humanities, University of Urbino Carlo Bo, Italy. He works on social network analysis techniques applied to social media data and to the analysis of audience practices. He is professor of sociology of new media and media analysis.

Magnani and Rossi are principal investigators of national research projects on multiple network modeling and mining and are also active as organizers of international events on this topic.

Pulse of Virtual Worlds: Behavioral Mining and SNA in Massive Online Games

Presented by Muhammad Ahmad and Jaideep Srivastava

Massive online games (MOGs) refer to massive online persistent environments (*World of Warcraft, EVE Online, EverQuest* and others) shared by millions of people. In general these environments are characterized by a rich array of activities and social interactions with a wide array of behaviors e.g., cooperation, trade, quest, deceit, mentoring etc. Such environments allow one to study human behavior at a level of granularity where it was not possible to do so previously. Given the challenges associated with analyzing this type of data traditional techniques in data mining and social network analysis have to be extended with insights from the social sciences. The tutorial will cover predictive and generative models in the study of MOGs. Additionally we will cover some SNA techniques which are more appropriate for MOGs given the multidimensionality of the data (P*/ERGM models, IR-based network analysis, hypergraph based techniques, coextensive social networks and others). Based on our published work in this area we also describe the various ways in which MOGs exhibit similarities to the real world e.g., economic behaviors, clandestine behaviors, mentoring etc). Finally, we describe the scope and limitations of analysis of MOGs based on limitations in data collection, availability and ethical concerns. An overview of commercial applications is also given.