Computational Social Science for the World Wide Web (CSSW3)

Ingmar Weber
Qatar Computing Research Institute
iweber@qf.org.qa

Claudia Wagner
GESIS – Leibniz Institute for the Social Sciences
University of Koblenz-Landau
claudia.wagner@gesis.org

Markus Strohmaier
GESIS – Leibniz Institute for the Social Sciences
University of Koblenz-Landau
markus.strohmaier@gesis.org

Luca Maria Aiello
Yahoo Labs
lajello@gmail.com

ABSTRACT

This tutorial aims at outlining fundamental methods for studying typical social science research questions with organic data (i.e., data that has not been designed for a specific research purpose but can be found on the Web). Further, social theories, statistical methods and models that help to understand the processes that generated the data will be discussed. Participants will learn (1) how to turn theoretical assumptions into models and test them, (2) how to validate measurements and (3) how to approximate causality when working with organic data.

Keywords

Computational Social Science, Social Theories, Causality, Networks, Measurements, Research Methodologies

1. INTRODUCTION

Due to the increasing availability of large-scale data on human social behavior collected on the world wide web, there is an increasing interest in applying computer science methods to address research questions in the social and behavioral sciences, and vice versa [4]. However, "Big Data" researchers and "Data Scientists" entering the interdisciplinary field of Computational Social Science (CSS) often lack the necessary background in social theories and methods of sociology, while at the same time sociologists are often not aware of cutting edge advances in computational methods and algorithmic biases in organic data (i.e., data that has not been designed for a specific research purpose) that can be found on the Web. This tutorial helps to bridge this gap by providing an introduction to statistical and computational methods that are useful for addressing typical social science research questions with organic data and to social theories and models that help to understand the processes that generated the data. The goal of this tutorial is to give participants an overview of a rich repertoire of methods that help to answer not only interesting "how" questions but also more fundamental "why" questions.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2891064 .

2. TUTORIAL OUTLINE

The tutorial will give an overview of Computational Social Science (CSS) and data- and theory-driven research methodologies in that field.

Introduction to CSS. In the introductory session, we give a high level overview of computational social science research. We start by discussing the main research trajectories that can be identified in the field. Historically, *(i) agent-based simulations and models* have long been used and developed by social scientists to explore and explain macro social behavior via micro based interactions. In addition to simulations, data-based approaches today increasingly gain interest in both computer and social science communities. Here a distinction between *(ii) deductive* and *(iii) inductive approaches* can be made. While deductive approaches focus on hypotheses testing (from theory to validation via data) thereby relying on theory and models, inductive approaches focus on hypothesis and theory generation (from data to theory via observations), thereby requiring principled methods and data-centric tools.

Social Dynamics and Social Theories In the second part of the tutorial we will discuss new methods, rooted in well-established sociological theories, to interpret the role of microscopic and mesoscopic social structures (links and groups) in social interactions through the operationalization of social science theories (see e.g. [1]). We will also discuss the potential of agent based models as a tool to simulate what theories would predict. Participants will learn (1) how to turn theoretical assumptions into models that allow to create synthetic data and (2) how to compare synthetic data with empirical data.

Measurements and Causal Inference The availability of new types of data enables computational social scientists to measure certain constructs for the first time. However, it is important to critically reflect on how to assess the validity and reliability of measurements in this context. In the third part of this tutorial, we will discuss methods and approaches for assessing the validity and reliability of measurements.

Beside describing and measuring new phenomena, computational social scientists are usually also interested in understanding the process that generated the data. Therefore, they need to be able to make causal inferences to understand the fundamental principles behind what we observe. We will discuss different methods (post-hoc control methods like matching [3] and natural experiment [2]) for making causal inference based on organic data.

3. SCHEDULE.

For each of these topics, we plan a slot of roughly 90 minutes, resulting in three sessions á 90 minutes in total. For the presentation, only standard equipment (LCD-projector and microphone) is required.

4. TARGET AUDIENCE & PREREQUISITE

This tutorial is aimed at participants with a basic level of data mining and data processing. The content covered in the three tutorials and hands-on sessions is designed to introduce basic and advanced computational social science methods, which require some basic level of data analysis skills. We will be seeking both PhD students and researchers interested in learning or advancing their current knowledge of computational social sciences theories and methods. Participants should have the basic skills of data harvesting, processing, and analysis. It would be advantageous if participants were familiar with Python, R, or some similar statistical packages.

5. PRESENTERS

Ingmar Weber is a senior scientist in the Social Computing group at the Qatar Computing Research Institute. His main research interest lies in using large amounts of online data to study offline phenomena including political unrest, obesity, gender inequalities, international migration, religion or relationship breakups. He has published over 80 peer-reviewed articles and has worked with medical doctors, demographic researchers, sociologists, political scientists and socio-linguists. He was PC co-chair of SocInfo'14, lead organizer of the PLEAD (Politics, Elections and Data) workshop '12, '13, co-organizer of a tutorial on "Twitter and the Real World" and serves on the PC of major conferences such as WWW, WSDM, ICWSM, CIKM, and KDD, as well as the editorial board for the Web Science journal. With Yelena Mejova and Michael Macy he has edited a CUP book on "Twitter: A Digital Socioscope".

Claudia Wagner is a post-doctoral researcher at the Computational Social Science department at GESIS – Leibniz Institute for the Social Sciences and an adjunct lecturer at the University of Koblenz-Landau. She received her PhD in Computer Science at Graz University of Technology. Claudia's work is interdisciplinary and focuses on exploring to what extent digital traces can be used to learn about the offline world (e.g., interests, knowledge, habits and biases of individuals and societies) and how the online world and new technologies may impact the offline world. She has served on the PC of major computer science conferences, including WWW, ICWSM and ESWC and has conducted internships in industry labs such as at HP Labs, Xerox PARC's and Yahoo Labs.

Luca Maria Aiello received his Ph.D. in Computer Science from the University of Torino, Italy in 2012. He is currently a Research Scientist at Yahoo Labs in London. He conducts interdisciplinary work connecting computer science, physics of complex systems, and computational social science through quantitative big data analysis. Recently his research has been devoted to study social phenomena such as homophily, influence, conversational norms, status and social attention, with applications to personalization, ranking, recommendation and link prediction. He was General Chair for the SocInfo'14 conference and organizer of the SNOW'14 Workshop. More recently, he has been appointed workshop and tutorial chair for ICWSM. He has been member of the PC of major computer science conferences, including WWW, WSDM, ICWSM, CIKM, ACM MM. He has led the work of Yahoo in the SocialSensor EU project consortium.

Markus Strohmaier is a Full Professor of Web-Science at the Faculty of Computer Science at University of Koblenz-Landau (Germany) and Scientific Director of the Computational Social Science department at GESIS - the Leibniz Institute for the Social Sciences (Germany). Markus' research focuses on the intersection between computer and social sciences. He was a visiting scientist/professor at Stanford University (USA) during the 2011/12 academic year, at XEROX Parc (2009, 2010-2011) and at RWTH Aachen (2009). He has published in top-tier conferences and high impact journals including full papers at the International World Wide Web (WWW) conference, the International Semantic Web (ISWC) conference, the ACM Conference on Human Factors in Computing Systems (CHI) and the ACM Conference on Information and Knowledge Management (CIKM). He is a co-author of the recent IEEE Intelligent Systems article "Computational Social Science for the World Wide Web" (2014) and a co-recipient of the WWW'2015 Best Paper Award.

6. REFERENCES

[1] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81(2):591–646, 2009.

[2] T. Dunning. *Natural Experiments in the Social Sciences*. Cambridge University Press, 2012.

[3] J. S. Sekhon. Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science*, 12:487–508, 2012.

[4] **M. Strohmaier** and C. Wagner. Computational social science for the world wide web. *IEEE Intelligent Systems*, 29(5), 2014.