METHODENZENTRUM BREMEN

Universität Bremen

Digital Traces Workshop
November 8th - 10th, 2018

University of Bremen

Abstract Volume

Social Science Methods Centre of the University of Bremen

Address:
Prof. Dr. Uwe Engel
University of Bremen, Unicom 1 - Haus Salzburg
Mary-Somerville Strasse 9
D - 28359 Bremen
E-Mail: uengel@uni-bremen.de

Office:
Silke Himmel
E-Mail: silke.himmel@uni-bremen.de
Phone number: 0421-218 67322

Website:
https://www.methodenzentrum-bremen.de/index.php?lang=en

Acknowledgment: We gratefully acknowledge the financial support received by the German Research Foundation (DFG), the Freie Hansestadt Bremen (Senator for Health, Science, and Consumer Protection), the Bremen International Graduate School of Social Sciences (BIGSSS), and the University of Bremen.

Content

Thursday, 8th November 2018

Friday, 9th November 2018

Saturday, 10th November 2018

Thursday, 8th November 2018

-Computational Social Science (CSS) – Opening session-

-Networks-

- Theory, Modelling & Simulation-

-Computational Social Science (CSS) – Opening session-

13:15 – 13:45|14:00

Claudio Cioffi-Revilla

- George Mason University, United States

Theories and Methods in Computational Social Science

Computational social science (CSS) is the interdisciplinary field that investigates social systems on multiple organizational and temporal scales and domains, from small groups to the global system of civilizations, through the medium of formal methods and applied computing to advance our scientific understanding of human dynamics and social complexity. As result of such a scope, CSS draws on and develops on numerous concepts, theories, and methods, given the nature of our subject matter. In this talk I will discuss a selection of important theories and methods in CSS with special emphasis on the specific subject of the Digital Traces Workshop and connections with big data, social media, closely related topics, such as robotics, and potential ideas for collaborative research proposals envisioned by the organizers. This includes the significance of so-called hybrid functions, which are formalism containing a mix of continuous and discrete variables – a common occurrence in numerous domains of social complexity. Although hybrid functions are commonly "hidden in plain sight," their exact calculus has been elusive but remains critical for advancing CSS through empirical and formal methods, including all algorithmic methods such as information extraction, network analysis, and computational agent-based models.

14:00 – 14:20|14:30

Peter Hedström

- Linköping University, Sweden

Marc Keuschnigg

- Linköping University, Sweden

Analytical Sociology and Computational Social Science

Analytical sociology focuses on social interactions among individuals and the hard-to-predict aggregate outcomes they bring about. It seeks to identify generalizable mechanisms giving rise to emergent properties of social systems which, in turn, feed back on individual decision-making. This research program benefits from computational tools such as agent-based simulations, machine learning, and large-scale web experiments, and has considerable overlap with the nascent field of computational social science. By providing relevant analytical tools to rigorously address sociology's core questions, computational social science has the potential to advance sociology in a similar way that the introduction of econometrics advanced economics during the last half century. Computational

social scientists from computer science and physics often see as their main task to establish empirical regularities which they view as “social laws.” From the perspective of the social sciences, references to social laws appear unfounded and misplaced, however, and in this talk we outline how analytical sociology, with its theory-grounded approach to computational social science, can help to move the field forward from mere descriptions and predictions to the explanation of social phenomena.

14:30 – 14:50|15:00

Ulrik Brandes

- ETH Zurich, Switzerland

Social Network Science and the Notion of Position

From the perspective of network science, social networks constitute a particular application domain. This, understandably, causes uneasiness among social scientists because proponents of network science often dress it as a collection of concepts and methods associated with universal theoretical underpinnings. These, however, do not lend themselves to the incorporation of domain-specific theory. I will advocate a more methodological approach that is based on a formal notion of position and intended to narrow the gap between substantive theory and mathematical analysis of social networks. As a byproduct, a wealth of computational problems are suggested.

-Networks-

15:30 – 15:50|16:00

László Lőrincz,

- Hungarian Academy of Sciences, Hungary

Anna Győr,

- Hungarian Academy of Sciences, Hungary

Júlia Koltai,

- Hungarian Academy of Sciences, Hungary

Károly Takács

- Hungarian Academy of Sciences Hungary

Collapse of an Online Social Network: Burning Social Capital to Create It

Sometimes even the largest online social networks (OSNs) collapse. Significant cascading mechanisms have been identified in the pattern of abandoning the Hungarian OSN iWiW at its peak of popularity (approx. 3.5 million active users) and after. We examine who were the key actors first to leave by contrasting explanations based on different aspects of social capital. Using heterogeneous choice models, we find that a higher number of connections as well as less clustered ego-networks

hindered early abandonment, while early adoption played only a secondary role for leaving the site early.

Keywords: online social networks, social capital, early adopters, embeddedness

16:00 – 16:20|16:30

Yannis Theocharis
- University of Bremen, Germany

Maarja Luhiste
- Newcastle University, United Kingdom

Zoltan Fazekas
- University of Oslo, Norway

Sebastian Adrian Popa
- Newcastle University, United Kingdom

Pablo Barberá
- London School of Economics and Political Science, United Kingdom

When does Abuse and Harassment Marginalize Female Political Voices on Social Media?

Social media can offer female candidates opportunities for escaping the gendered coverage they often receive in traditional media. While female candidates have powerfully integrated social media into their campaigns, this comes with a price tag: an abundance of news reports claim that female authors tend to be treated differently by commenters and receive more abuse on social media sites than men. Yet, there has been no systematic empirical study documenting the quantity or quality of this abuse and its consequences for the online public discourse. We investigate whether female politicians and candidates are more likely to receive, not just more uncivil comments than men, but also comments that are qualitatively different. Our study uses comparative longitudinal Twitter data that include interactions between citizens and politicians in Greece, Spain Germany, UK and US. We rely on manual and machine-learning content classification that is aimed at capturing the dynamics of uncivil attacks (i.e. volume, candidate-level characteristics, context and consequences of uncivil attacks for politicians) and at detecting qualitative differences in the language used. We find cross-national differences both in terms of how politicians interact with citizens on Twitter, and in terms of how much incivility they receive. Contrary to expectations, female candidates do not appear to be targeted more than men, but we find evidence that uncivil language used against men differs from that used against women — who are often targeted with language alluding to their suitability for political roles. Our findings have wide-ranging implications for the health of the online public discourse, and for democracy more broadly.

16:30 - 16:50|17:00

David Garcia
- Medical University of Vienna, Austria
- Complexity Science Hub Vienna, Austria

Yonas Mitike Kassa
- Charles III University of Madrid, Spain
- IMDEA Networks Institute, Spain

Angel Cuevas
- Charles III University of Madrid, Spain

Manuel Cebrian
- Massachusetts Institute of Technology, United States
- Data61, Australia

Esteban Moro
- Massachusetts Institute of Technology, United States
- Charles III University of Madrid, Spain

Iyad Rahwan
- Massachusetts Institute of Technology, United States

Ruben Cuevas
- Charles III University of Madrid, Spain

Analyzing Gender Inequality Through Large-scale Facebook Advertising Data

Online social media are information resources that can have a transformative power in society. While the Web was envisioned as an equalizing force that allows everyone to access information, the digital divide prevents large amounts of people from being present online. Online social media, in particular, are prone to gender inequality, an important issue given the link between social media use and employment. Understanding gender inequality in social media is a challenging task due to the necessity of data sources that can provide large-scale measurements across multiple countries. Here, we show how the Facebook Gender Divide (FGD), a metric based on aggregated statistics of more than 1.4 billion users in 217 countries and regions, explains various aspects of worldwide gender inequality. Our analysis shows that the FGD encodes gender equality indices in education, health, and economic opportunity. We find gender differences in network externalities that suggest that using social media has an added value for women. Furthermore, we find that low values of the FGD are associated with increases in economic gender equality. Our results suggest that online social networks, while suffering evident gender imbalance, may lower the barriers that women have to access to informational resources and help to narrow the economic gender gap.

- Theory, Modelling & Simulation-

17:00 – 17:20|17:30

Jan Lorenz

- Jacobs University Bremen, Germany

Data-driven Agent-based Modeling as an Approach in Computational Social Science

Computational social science can serve as a unifying approach to the social sciences with potential for more cumulative scientific progress. Nevertheless, it is largely driven by three research communities with little scientific contacts and different core disciplinary backgrounds. In particular, goals differ between (i) the invention of new predictive tools about human behavior through machine learning on BigData, (ii) analysis of the societal change through digitalization, (iii) the empirically validated explanation of macroscopic social phenomena through mechanisms of social interaction. The method of agent-based modeling and simulation pursues the latter. Within computational social science, it should seek to become more data-driven without sacrificing its high aims. Data can serve three different purposes: (a) inspiration for social phenomena as captured in stylized facts, (b) definition and calibrate of rules for action and interaction on the microlevel, and (c) the validation of the model mechanisms based on macroscopic simulation output. Two examples from opinion dynamics will be presented.

17:30 – 17:50|18:00

André Grow

- KU Leuven, Belgium

Gender, Resources, and Status: An Empirically Grounded Model of Status Construction Theory

The members of different social groups (e.g., men/women, whites/non-whites) are often accorded different levels of deference, respect, and esteem–in short, status–in society. Status construction theory (SCT) offers one prominent account of processes that can create status differences between social categories. The theory focuses on goal-oriented interactions in small groups as a building block of society. Such groups often develop local status hierarchies in which some individuals appear more competent and respected than others. When such differentiation occurs consistently between members of different social categories, individuals can come to believe that the social distinction is *generally* associated with differences in competence and social worth. Once emerged, such beliefs can diffuse throughout the population, because people carry them into new interaction contexts, treat new interaction partners accordingly, and thereby create hierarchies that teach their beliefs to others. One factor that affects the formation of local status hierarchies is differences in the possession of

valuable resources (e.g., wealth), so that group members who possess relatively more of a given resource are likely to be perceived as more competent and respectable than those who possess relatively less of it. As a consequence, when the members of one social category possess on average more of a valuable resource than members of another category, the distinction is likely to become a source of status to the advantage of the more resourceful category, given that the number of local hierarchies that can bring status beliefs about will be tilted in its favor.

SCT's arguments are complex and difficult to evaluate based on mere verbal reasoning alone. Existing SCT research has therefore relied on computational simulations to study the theory's macro-level implications. One shortcoming of this earlier work is that the proposed formal models were theoretical abstractions that were not empirically grounded. As a consequence, it remains unclear how much the theory can contribute to explaining some of the actual status differences that can be observed in today's societies. In this paper, I seek to address this shortcoming. I develop an agent-based computational model of SCT that can be paired with empirical data on resource differences between social groups at the micro-level and whose macro-level outcomes can be validated against empirically observed status differences between these groups. The model is not particular to a specific social distinction and could be used to study processes of status differentiation based on any distinction. Yet, to assess the merits of the model, I focus on biological sex as one of the most pervasive sources of status differentiation across societies.

Friday, 9th November 2018

-Theory, Modelling, Simulation -

-Linkage-

-Survey plus-

-Big Data: Methodology, Statistics & Analytics-

-Theory, Modelling, Simulation

8:30 – 8:50|9:00

Annie Waldherr

- University of Münster, Germany

Advancing Social Theory with Agent-based Modeling and Simulation: Examples from Public Sphere Research

Understanding the public sphere as a complex adaptive system helps understanding dynamic macro phenomena of public communication. In terms of complexity theory, the public sphere can be conceptualized as an arena where heterogeneous, autonomous and adaptive agents interact and self-organize. Multiple feedback mechanisms lead to the emergence of non-linear macro patterns such as news waves or opinion spirals. In my talk, I show how the public sphere can be modeled as an agent-based system, and how the social simulation approach helps to develop theories of the public sphere, specifically by offering proof of concepts of existing theories, finding underlying mechanisms able to generate observed empirical patterns, and exploring possible futures. I will illustrate this approach by presenting two agent-based models. The first model simulates the typical stylized pattern of news waves by modeling the adaptive behavior of journalists and strategic issue sponsors.
The second model builds on Noelle-Neumann's (1974) theory of the spiral of silence assuming that fear of isolation causes people to only speak out their opinion if they perceive a favorable opinion climate surrounding them.

9:00 – 9:20|9:30

Stefan Bosse

- University of Bremen, Germany

Large-scale Multi-agent Simulation and Crowd Sensing with Humans in the Loop

This talk focuses on the simulation of complex and large-scale agent-based systems. There is an ongoing activity to model and study social systems using agent based modeling (ABM) and simulation. Commonly ABM is performed in a sandbox with a very limited world model. Moreover, the boundary between human beings and machines is vanishing. For example, recently exposed, automatic chat bots gain influence on society opinions and decision making processes (in politics, elections, business). Commonly ABM is performed in a closed environment only using simulated artificial agents in an artificial simulation world. There is no interaction or data exchange with real worlds. Although pure digital, real worlds include the World Wide Web, social platforms, and Clouds. The outcome of such limited scope and simplified systems is application specific.

In a large-scale agent-based simulation embedded in and connected to real world environments (so called "human- or hardware-in-the-loop" simulation) agents can represent different behaviour, goals, and individuals like chat bots or artificial humans and their interaction with virtual and real individuals, e.g., via WEB interfaces or robots (software agents meet hardware agents).

The tight coupling of simulation, technical systems (e.g., robots or WEB services), and human interaction can be established by using mobile agents and a highly portable agent processing platform that can be deployed in strong heterogeneous environments (including WEB browser and mobile devices like smart phones) and simulation simultaneously. This distributed multi-agent system is well suited to include and perform Crowd Sensing to extend the data base. Such a simulation system can be used to study a broad range of complex socio-technical systems and machine-human interactions on large-scale level. One prominent example is modeling of opinion and decision making under the influence of digital technologies. It can be expected that the simulation of large-scale agent societies with agent population beyond one Million individual agents delivers statistical strength and generality.

-Linkage-

9:30 – 10:00|10:15

Peter van der Heijden

- Utrecht University, The Netherlands

An Overview of Population Size Estimation Where Linking Registers Results in Incomplete Covariates

We consider the linkage of two or more registers in the situation where the registers do not cover the whole target population, and relevant categorical auxiliary variables (unique to one of the registers; although different variables could be present on each register) are available in addition to the usual matching variable(s). The linked registers therefore do not contain full information on either the observations (often individuals) or the variables. By treating this as a missing data problem it is possible to construct a linked data set, adjusted to estimate the part of the population missed by both registers, and containing completed covariate information for all the registers. This is achieved using an Expectation-Maximization (EM)-algorithm. We elucidate the properties of this approach where the model is appropriate and in situations corresponding with real applications in official statistics, and also where the model conditions are violated. The approach is applied to data on road accidents in the Netherlands, where the cause of the accident is denoted by the police and by the hospital. Here the cause of the accident denoted by the police is considered as missing information for the statistical units only registered by the hospital, and the other way around. The method needs to be widely applied to give a better impression of the range of problems where it can be beneficial.

10:45 – 11:05|11:15

Thomas Augustin
- LMU Munich, Germany

Martin Spieß
- University of Hamburg, Germany

Combining Imprecise Information for Valid Statistical Inference

In big-data contexts the same construct or idea is often addressed by different types of information (e.g. test, images) in varying granularity (e.g. in different aggregation levels). We discuss new opportunities (and challenges) the concept of so-called `imprecise probabilities' promises in this context.
By a generalized modelling of complex uncertainty, imprecise probabilities could provide a strategy to combine the different types and granularity of information in a reliable way, allowing for valid statistical inferences.

-Survey plus-

11:15 – 11:35|11:45

Florian Keusch
- University of Mannheim, Germany

Sebastian Bähr
- Institute for Employment Research, Germany

Georg-Christoph Haas
- University of Mannheim, Germany
- Institute for Employment Research, Germany

Frauke Kreuter
- University of Maryland, United States
- Institute for Employment Research, Germany
- University of Mannheim, Germany

Mark Trappmann
- University of Bamberg, Germany
- Institute for Employment Research, Germany

Nonresponse Error in Passive Mobile Measurement

Smartphone use is on the rise worldwide, and researchers are exploring novel ways to leverage the capabilities of smartphones for data. For example, researchers can now ask smartphone users to agree to tracking of geolocation and movements to create exact measures of mobility and transportation or to automatically log app use, Internet searches, and phone calling and text

messaging behavior to measure social interaction. These new form of data collection provide richer data (because it can be collected in much higher frequencies compared to self-reports) and has the potential to decrease respondent burden (because fewer survey questions need to be asked) and measurement error (because of reduction in recall errors and social desirability). However, agreeing to engage in these forms of data collection from smartphones is an additional step in the consent process, and participants might feel uncomfortable sharing specific data with researchers due to security, privacy, and confidentiality concerns.
In spring 2018, we invited 4,293 members of the Panel Study Labour Market and Social Security (PASS), who in wave 11 had reported owning an Android smartphone, to download the IAB-SMART app. The app collected data from the smartphone users over six months (1) through short surveys pushed by the app and (2) through passive mobile measurement using different sensors on the smartphone. Around 650 people downloaded the app and participated in the study. The rich information on PASS members from previous waves of the panel allow us to compare participants and nonparticipants in the IAB-SMART study and estimate nonresponse bias for substantive PASS measures on employment and social integration. We will also analyze the effect of consenting to some but not all types of sensor data collection within the app on nonresponse bias.

11:45 – 12:05|12:15

Merja Mahrt

- Heinrich Heine University Düsseldorf, Germany

CSS and Inequality. Insights from Multi-Method Research on Online Usage and Digital Fragmentation

In the debate about “big data,” scholars have raised the issue of a potentially widening gap between “data haves” and “data have nots” (boyd & Crawford, 2012; Driscoll & Walker, 2014). This refers to the considerable amount of resources required for some types of CSS- research, with regard to data access, storage, handling, or analysis. There is thus a fear in the research community that data-driven research could further inequality among its members, or even bar new members from joining.
An issue of inequality in CSS less often discussed, however, concerns society in general. Internet access is not distributed evenly around the globe, nor within countries (International Telecommunication Union, 2017). In industrialized countries like Germany or the US, about ten percent of the respective populations currently do not use the Internet (Anderson & Perrin, 2018; Koch & Frees, 2017). And even among users, there are diverse ways of making use of online offerings. Research based on data created through such usage thus necessarily reflects these inequalities in access and use. This has been acknowledged before and some commentators accordingly suggest to complement online with offline data (e.g., Murthy, 2008; Orgad, 2009). Yet such research is far from being the norm in CSS or elsewhere. It seems that there is still enough excitement about research exclusively using data from digital platforms that these are collected, analyzed, and then reported to the research community as well as the wider public and policy makers.
However, when we actually compare big and small data from different sources, we sometimes come to very different conclusions about the social phenomena we investigate. In a study on digital

fragmentation (discussed, among others, by Pariser, 2011; Sunstein, 2007, as potentially leading to "filter bubbles" or "echo chambers," respectively), I combined different methods to analyze usage of online content. For the CSS conference, I will illustrate how vastly results on the use of YouTube videos differed when I used surveys, clickstream data, as well as YouTube's own indications of video popularity.
That every method of data collection only captures parts of social reality is common knowledge. But I fear that in CSS, we forget this limitation far too often. The study of digital fragmentation again is an example of this. If we want to know whether there is fragmentation in how people use online content, we can look at data like those mentioned above. However, these will not tell us how a possibly fragmented use of online content compares to other media. Is it more or less fragmented than, for example, television use? I want to argue that to properly assess even a CSS phenomenon like digital fragmentation we should take into account its wider social context.
Research based on computational data may struggle to capture such contexts. And there may be entire parts of society who are not reflected in CSS data because they did not leave traces that could have been picked up via a chosen method. So as part of the CSS workshop in Bremen, I would like to discuss with colleagues from the field what kind of social science we want to engage in if we decide to use CSS methods. Ultimately, this will, in my opinion, decide what contribution CSS can bring to understanding social issues – and to society in general.

-Big Data: Methodology, Statistics & Analytics-

13:15 – 13:45|14:00

Richard D. De Veaux
- Williams College, United States

Holistic Data Science and the Seven Deadly Sins of Big Data

As we are all too aware, organizations are accumulating vast amounts of data from a variety of sources nearly continuously. Big data advocates promise the moon and the stars as you harvest the potential of all these data. There is certainly a lot of hype. There's no doubt that savvy organizations are fueling their strategic decision making with insights from data mining, and scientists are discovering new insights with big data, but what are the challenges?
Much can wrong in the data mining process, even for trained professionals. In this talk I'll discuss a wide variety of case studies from a range of problems in science and industry to illustrate the potential dangers and mistakes that can frustrate problem solving and discovery -- and that can unnecessarily waste resources. My goal is that by seeing some of the mistakes I have made, you will learn how to take advantage of data mining insights without committing the "Seven Deadly Sins."

14:00 – 14:20|14:30

Andreas Jungherr
- University of Konstanz, Germany

Normalizing Digital Trace Data

Over the last ten years, social scientists have found themselves confronting a massive increase in available data sources. In the debates on how to use these new data, the research potential of "digital trace data" has featured prominently. While various commentators expect digital trace data to create a "measurement revolution", empirical work has fallen somewhat short of these grand expectations. In fact, empirical research based on digital trace data is largely limited by the prevalence of two central fallacies: First, the n=all fallacy; second, the mirror fallacy. As I will argue, these fallacies can be addressed by developing a measurement theory for the use of digital trace data. For this, researchers will have to test the consequences of variations in research designs, account for sample problems arising from digital trace data, and explicitly link signals identified in digital trace data to sophisticated conceptualizations of social phenomena. Below, I will outline the two fallacies in greater detail. Then, I will discuss their consequences with regard to three general areas in the work with digital trace data in the social sciences: digital ethnography, proxies, and hybrids. In these sections, I will present selected prominent studies predominantly from political communication research. I will close by a short assessment of the road ahead and how these fallacies might be constructively addressed by the systematic development of a measurement theory for the work with digital trace data in the social sciences.

14:30 – 15:00|15:15

Ashley Amaya
- RTI International, United States

Paul Biemer
- RTI International, United States

David Kinyon
- RTI International, United States

Total Error in a Big Data World: Adapting the TSE Framework to Big Data

While Big Data offers a potentially less expensive, less burdensome, and more timely alternative to survey data for producing a variety of statistics, it is not without error. The AAPOR Task Force on Big Data and others have called for researchers to evaluate the quality of Big Data using an approach similar to the total survey error (TSE) framework. However, differences in the construction of, access to, and overall data structure between survey data and Big Data make application of TSE difficult. In

this paper, we sought to develop the Total Error Framework (TEF), an extension of the TSE framework, to be (1) more inclusive and applicable to many types of Big Data, (2) comprehensive in that it considers "total" error, and (3) unified in that it allows researchers to compare errors in Big Data to errors in survey data. After outlining this framework, we then illustrate an application of TEF by comparing error in housing unit area (square footage) estimates collected in a survey (the 2015 Residential Energy Consumption Survey (RECS)) to those estimates found in three Big Data databases (Zillow.com, Acxiom, and CoreLogic).

15:15 – 15:35|15:45

Axel Mayer
- RWTH Aachen University, Germany

Florian Lemmerich
- RWTH Aachen University, Germany

Subgroup Discovery based on Structural Equation Modeling

In this talk we aim at bringing together machine learning techniques from computer sciences and social science methodology. In the machine learning literature efficient algorithms for subgroup discovery (sometimes also called pattern mining or pattern recognition) have been developed and are widely used for identifying unobserved subgroups in datasets. In the social sciences on the other hand, there are decades of research on how to model structural relations between random variables. One of the most popular and flexible methods used in this field is structural equation modeling (SEM). Some recent extensions of this method like structural equation mixture modeling (SEMM) or structural equation model trees (SEMtrees) allow for discovering potentially latent subgroups that are distinct with regard their structural relations. SEMM uses a latent class approach while SEMtrees is based on recursive partitioning. We propose an alternative to SEMM and SEMtrees that has similar aims but builds on the latest algorithms from pattern mining to more efficiently find the subgroups of interest in large datasets. The new approach is termed SubgroupSEM. To the best of our knowledge there is little to no overlap between SEM and pattern mining fields and we believe that SEM could benefit from the algorithmic knowledge developed in other fields. Within the SEM literature there are both global (e.g., χ^2-tests of model fit) and local tests (e.g., Wald tests) that can serve as interestingness measures for quantifying differences between subgroups in the algorithm. In a small scale simulation study we show that our SubgroupSEM-algorithm can adequately recover the subgroups that underly the data generating mechanism and compare the performance of our approach to SEMM and SEMtree. Limitations and potential applications in the field of computational social science are discussed.

16:25 – 16:35|16:45 (as Introductory Note to the Round Table)

Anabel Quan-Haase

- University of Western Ontario, Canada

Self-presentation Practices in Social Media: Context, Life Course, and the Attention Economy

The present talk investigates the numerous dimensions that influence the crafting of a virtual self to inform and expand data-driven approaches – both quantitative and qualitative. The talk will explore four central aspects of digital engagement that impact data analysis and interpretation, but are difficult to consider in a traditional research design. First, platform specific self-presentation practices are discussed that complicate data verification and interpretation approaches. Second, the social context of engagement is discussed as a means to shed light on self-presentation practices. Specifically the example of #metoois used to describe how context frames self-disclosure and self-presentation practices. Third, I discuss trolling—a deliberate form of misleading, provoking, and making fun of others online--. Using these three themes, I conclude that social media scholars need to take into consideration the social context in which profiles are created and interactions occur in terms of platform-specific social norms, topic-specific concerns, and domain-specific knowledge. The real meaning of data is not always readily apparent, and its decoding may require further theorizing around social behaviour and its underlying motives.

An important consideration in social media research methodology is the extent to which users' accounts, including their profile and engagement, reflect elements of the self. First, we examine online reputation management, which describes the tendency for individuals to curate a desired self-image through selective presentation of personal data. Then, we look at shifts in personality traits and the role of e-personality in influencing online self-presentation and interaction on social media.

Saturday, 10th November 2018

-Text Analysis-

-Perspectives on Artificial Intelligence and Social Robots-

-Text Analysis-

8:30 – 8:50|9:00

Raphael Heiberger
- University of Bremen, Germany

Sebastian Munoz-Najar Galvez
- Stanford University, United States

How Useful is Topic Modeling for Social Scientists? Tracing Changes in the Sociological Field with Tools from NLP

Working with texts poses important conceptual and methodological challenges. The wider the range of areas and phenomena a corpus covers, the harder it is to map. Computational linguistics provides us with the conceptual and technical tools to address those challenges in Structural Topic Modelling (STM). To link rhetorics to conditions of social structure and mechanisms of reproduction, we combine the results derived from STM in regard to changing research tastes of PhD students in sociology with the explanatory power of inferential statistics, contributing to efforts often dubbed as “Computational Social Science”. We concentrate on two points in every scholar’s career: her entry by earning a doctorate and becoming an advisor. In particular, we gather comprehensive data of over 80,000 sociology dissertations at U.S. universities and those graduates’ pursuant academic careers. The integration of well-established regression models and natural language processing tools allows us to model the rise and fall of taste expressed by young academics in sociology as a function of the resources and capital employed by their advocates. In so doing, we are “taking Bourdieu to the digital archive”revealing that popular tastes among young sociologists indeed represent morphological transformations of the field – yet, mostly at its rhetorical surface. Power positions (or faculty advisor positions) structuring sociology in the U.S. do not change at the same rate, but rather reproduce inequities of the encompassing social world while the field’s elites incorporate new tastes by employing PhD students who fit notions of valued capital (e.g. prestige, white, males).

9:00 – 9:20|9:30

Gregor Wiedemann
- University of Hamburg, Germany

Tracing Utterance: Approaching Sentence-level Semantics in Computational Content Analysis

In (computational) social science, the use of topic models is becoming increasingly popular. Topic models are able to automatically detect thematic coherence in large document sets on a very distant level of discourse semantics. But, in the empirical analysis of qualitative text data, it is not only

interesting which topics occur in documents. Much more interesting is how a discourse position (or an argument, or a stance) has emerged in the first place through certain recurrence of individual utterances. Recent advancements in algorithmic models of distributional semantics from natural language processing could make a major contribution to a large-scale analysis of discourses at the level of utterances. The talk will introduce methodical thoughts and preliminary results on this issue with regard to the discourse about political “extremism” in Germany.

9:30 – 9:50|10:00

Tatjana Scheffler

- University of Potsdam, Germany

Analyzing Discourse Structure on Social Media

Social media such as Twitter, Facebook, blogs, forums, etc. are an abundant data source for texts generated by a diversity of users online. This provides unique opportunities and challenges for social scientists working with textual data. In this presentation, I address some of the specific challenges posed by social media data: For one, the large amount of data necessitates automatic methods for collecting and storing texts, as well as quantitative approaches to analyzing the resulting corpora. In addition, the language in social media contains many non-standard features which on the one hand, may prevent the use of established tools for natural language processing, and on the other hand, may themselves constitute exciting opportunities for research. In particular, the conversational nature of many kinds of social media draws attention to our lack of theoretical and practical knowledge about how to model dialog and discourse (as opposed to monogical texts).

In this talk, I will present methods from computational linguistics and computational social science that enable the collection and analysis of large corpora of social media data, with a particular focus on interactive language. Due to the spontaneous interactive nature of social media, the individual contributions are part of larger discourses that fulfill communicative intentions.

In presentation, I will address the following questions:
- What is computational linguistics/natural language processing and what research questions related to social media drive it?
- How can one collect and analyze social media corpora wrt. discourse phenomena?
- What is the linguistic structure of social media conversations like?
- How do computational linguistic models of discourse structure carry over to social media conversations?

In addition, I will discuss analyses of the nature and variability of language on social media and computational linguistic approaches to using social media data as a sensor for non-linguistic social data (e.g., health, human well-being, or politics).

10:30 – 10:50|11:00

Gertraud Koch
- University of Hamburg, Germany

Lina Franken
- University of Hamburg, Germany

Potentials of Automatizing Discourse Analysis – Lessons Learned from Studying the Phenomenon "Telemedizin"

Discourse analysis in the tradition of the sociology of knowledge is a research methodology for gaining an understanding of social orders and how they have emerged over time. Usually, the methodology of discourse analysis and discourse ethnography is hardly standardised and Grounded Theory procedures, such as theoretical sampling and saturation, guide the research process from the collection of materials for the data corpus to analyses of this data. In times of digital media, materials for discourse analyses are available more and more in digital formats, and at the same time in a growing and usually multitudinous number - often called big data. This raises questions about how the hermeneutic processes of discourse analysis can be supported by digital methods.

Within the research project "Automated modelling of hermeneutic processes – The use of annotation in social research and the humanities for analyses on health (hermA)", we scrutinize the possibilities and potentials of automatization within the hermeneutical approach, starting from annotation as a generic hermeneutic practice. Annotations are used in the process of systematic data collection when putting together the corpus to be analysed. Also, there are different sorts of annotation, such as open and axial coding, the building of categories from these codes, within the hermeneutic circle of analysis, which is mostly supported by QDA-software. Annotation is an important field of research in digital humanities too. Here, a variety of digital tools exist and support the practice of annotating texts, however with an entirely different knowledge background in computer linguistics.

Based on the example of an ongoing discourse analysis of the phenomenon "Telemedizin" in Germany, the contribution suggests that these tools can also be utilized in qualitative research when they are used as filters and are considered in terms of their particular tool quality. It will reflect how they can be integrated into the research process of discourse analysis and facilitate qualitative research in further ways. Moreover questions about the efficiency of automatizing research processes will be discussed.

-Perspectives on Artificial Intelligence and Social Robots -

11:00 – 11:30|11:45

Sunny Xun Liu

- Stanford University, United States

Social Dynamics of Human-Social Robots Interactions

Social robots are beginning to leave factories and labs and enter consumer markets. It is imperative for us to understand what factors will impact people's evaluations of social robots. In this talk, I will discuss two projects that investigate the social dynamics of human-robot interactions. In project 1, our research team at Stanford examined ten years of published articles about social robots, summarizing features of the robots and cataloging images of each robot (n=342) from the publications, creating the first large collection of social robots. We then asked people (n=3,920) to evaluate all of the robots on personality and social attributes. We found that two evaluations of the robots, competence and warmth, were as important for the evaluation of robots as they have been for perceptions of people. We measured 21 different physical attributes of all the robots and showed that perceived age, mobility, and surface textures were the best predictors of robot competence, while attributes of the face, perceived youthfulness and the absence of mechanical features were the best predictors of robot warmth. In project 2, we replicate classic audit studies on biased social decision making and conduct an experiment to investigate the effects of social robot characteristics (warm vs. competent) and design features (appearance design, personality design, and combinations of appearance and personality design) on people's perceptions of warmth, competence, job suitability and overall evaluations of social robots. The data indicate that, in general, competent robots are preferred over warm robots and appearance is the most effective method of conveying product information. The findings contribute to our understanding of human-robot interactions, social robot design, human psychology and social behavior.

11:45 – 12:05|12:15

Sirko Straube

- Robotics Innovation Center Bremen, Germany

DFKI and the Robotics Innovation Center in Bremen

The talk will give a short overview on the German Research Centre for Artificial Intelligence (DFKI GmbH) and focus on the projects and developments at the Robotics Innovation Center in Bremen, headed by Prof. Dr. Frank Kirchner. Here, more than 100 scientists work on the development of mobile, next generation robotic systems which are able to safely cooperate with humans and to solve

complex tasks autonomously. In the framework of publicly funded joint projects or direct industrial orders, we design and implement our systems for a variety of applications. The aim is a rapid transfer of research results into application in economy and society. The talk should provide some contact points to for an interdisciplinary exchange in the field of computational social sciences.

13:30 - 13:50|14:00

Mona Abdel-Keream

- University of Bremen, Germany

Service Robots Learning from Humans

The idea of having service robots assisting people in everyday activities is coming closer to reality with recent advancements. Everyday activities seem simple to perform for a human, for a robot, however, they represent immense challenges due to their variations and complexities. Robots require extremely detailed task descriptions in order to be able to fulfill such tasks. Moreover, such activities take place in highly unstructured environments, therefore one activity may require several task descriptions to account for all variations. It is evident that hard-coded plans for such actions are impractical, so rather than manually programming each task, the robot is taught by observing and learning from human's demonstrations thus allowing the robot to adapt easily to novel situations.

We propose a game-based learning infrastructure to learn from human activities. A virtual character, referred to as Avatar, can either be steered by collected VR or Motion Capture data that is gained from participants fulfilling a specific task. Mapping this motion data onto the Avatar will result in a Task Sequence which is a photorealistic rendering of human activities in a virtual environment over time. A set of data, including the environment the human is operating in, the objects and tools in the scene, the sequences of steps and all specific motion performed during a task execution are recorded into a knowledge processing system.

Our learning infrastructure has several advantages. First, it would provide us deep insight into human task planning, which is essential for robots in order to learn about the structure of everyday activities. Furthermore, our framework can extrapolate data by mapping captured motion data to different human models, allowing, therefore, the generation of a vast amount of training data. Since our learning infrastructure is closely linked to a powerful knowledge processing system, robots can infer and reason about the executed actions and develop a better understanding of the various effects of the observed actions.

14:00 – 14:20|14:30

Suat Can

- University of Bremen, Germany

The Benefits of Computer Vision for CSS

Computer Vision is a field of computer science and artificial intelligence that works on enabling computers to see, identify and process images in the same way that human vision does, and then provide appropriate output. It is like imparting human intelligence and instincts to a computer. The progressive development of Computer Vision algorithms and methods in recent years, allows the identification of objects or facial and emotion recognition in images and videos. Particularly in the field of image content analysis, the use of the algorithms and approaches of Computer Vision is promising: Modern algorithms and methods can help automatically classifying thousands of images in a fraction of a second, where a group of humans may need weeks for it to accomplish. The hurdle to getting started with Computer Vision is usually easier than it may seem at first glance. The presentation will give a short overview on the state-of-the-art Computer Vision methods and algorithms (e.g. facial and emotion recognition, object segmentation and classification) for CSS.