Big Data, Big Noise::The Challenge of Finding Issue Networks on the Web / Annie Waldherr, Daniel Maier, &Peter Miltner

Abstract: In this paper, we focus on noise in the sense of irrelevant information in a data set as a specific methodological challenge of web research in the era of big data. We empirically evaluate several methods for filtering hyperlink networks in order to reconstruct networks that contain only web pages that deal with a particular issue. The test corpus of web pages was collected from hyperlink networks on the issue of food safety in the United States and Germany. We applied three filtering strategies and evaluated their performance to exclude irrelevant content from the networks: keyword filtering, automated document classification with a machine-learning algorithm, and extraction of core networks with network-analytical measures. Keyword filtering and automated classification of web pages were the most effective methods for reducing noise whereas extracting a core network did not yield satisfying results for this case.

One-step, two-step, network-step? Complementary perspectives on communication flows in Twittered citizen protests / Martin Hilbert, Javier Vásquez, Daniel Halpern, Sebastián Valenzuela, & Eduardo Arriagada

Abstract: The article analyzes the nature of communication flows during social conflicts via the digital platform Twitter. We gathered over 150,000 Tweets from citizen protests for nine environmental social movements in Chile, and use a mixed-methods approach to show that longstanding paradigms for social mobilization and participation are neither replicated nor replaced, but reshaped. In digital platforms, long standing communication theories, like the 1955 two-step flow model, are still valid, while direct one-step flows and more complex network flows are also present. For example, we show that it is no contradiction that participants mainly refer to intermediating amplifiers (39 % of the mentions from participants go through this two-step flow), while at the same time traditional media outlets and official protest voices receive 80-90 % of their mentions directly through a direct one-step flow from the same participants. While non-intuitive at first sight, Bayes' theorem allows to detangle the different perspectives in the arising communication channel. We identify the strategic importance of a group of amplifying intermediaries in local positions of the networks, who coexist with specialized voices and professional media outlets at the center of the global network. We also show that direct personalized messages represent merely 20 % of the total communication. This shows that the fine-grained digital footprint from social media enable us to go beyond simplistic views of a single all-encompassing step-flow model for social communication. The resulting research agenda builds on longstanding theories with a new set of tools.

Exploring UK Millennials' Social Media Consumption Patterns and Participation in Elections, Activism and 'Slacktivism' / Rodolfo Leyva

Abstract: This study explores how UK young adults' exposure to social networking sites (SNSs) and attention to specific SNS content relate to their political practices. Data were

collected from a diverse group of undergraduates during the 2015 General Election and Labour leadership campaign via a web-survey designed with novel instruments (e.g., simulated Facebook and Twitter newsfeeds). Results indicate that frequenting varying SNSs and attention to certain political content –e.g., newsfeeds about student-fee activism- have a positive weak association with offline formal and activist participation and a considerably stronger association with online 'slacktivism'. Moreover, exposure to SNSs was found similarly high across respondents. However, those with typical demographic and psychographic markers of participation (e.g., upper socioeconomic status, early political socialisation) showed significantly greater levels of engagement with political content and offline and online participation. Together, these findings suggest that frequent social media consumption is linked to a minimal and narrow mobilising impact.

Why do young people avoid Photo Tagging? A new service avoidance scale / Amandeep Dhir
Aalto University

Abstract: Despite the growing attention to the study of various issues concerning online self-presentation, little research has investigated the different reasons underlying the avoidance of online identity and presentation building features. Using a multi-stage investigation, this study has developed and validated a 25-item instrument for investigating the various reasons why young people avoid Facebook photo-tagging. The instrument items were developed based on open-ended qualitative essays written by 141 Indian and Pakistani adolescents who use the Facebook photo-tagging service. From their qualitative responses, 38 items were first extracted and evaluated with 780 Indian adolescent photo-taggers. This resulted in the development of a 25-item instrument which was re-examined a year later with 314 Indian adolescent and 107 young adult photo-taggers. The study participants were recruited from India and Pakistan in order to address the criticism of previous Facebook research which is chiefly based on US-only study samples. The developed instrument indicated six reasons: destruction of capital, dislike, worry, parental control, embarrassment and personal appearance. The study results indicate that the developed instrument has a stable factorial structure, validity and reliability over time. This new instrument offers various theoretical and practical implications for scholars engaged in research involving computer-mediated communication, online self-presentation and young people.

Does the recruitment of offline households increase the sample representativeness of probability-based online panels? Evidence from the German Internet Panel / Annelies G. Blom, Jessica M. E. Herzing, Carina Cornesse, Joseph W. Sakshaug, Ulrich Krieger, & Dayana Bossert

Abstract: The past decade has seen a rise in the use of online panels for conducting survey research. However, the popularity of online panels, largely driven by relatively low implementation costs and high rates of internet penetration, has been met with criticisms regarding their ability to accurately represent their intended target populations. This criticism largely stems from the fact that: 1) non-internet (or offline) households, despite their relatively small size, constitute a highly selective group unaccounted for in internet panels, and 2) the preeminent use of non-probability-based recruitment methods likely contributes a self-selection bias that further compromises the representativeness of online panels. In response to these criticisms, some online panel studies have taken steps

to recruit probability-based samples of individuals and providing them with the means to participate online. Using data from one such study, the German Internet Panel (GIP), this paper investigates the impact of including offline households in the sample on the representativeness of the panel. Consistent with studies in other countries, we find that the exclusion of offline households produces significant coverage biases in online panel surveys, and the inclusion of these households in the sample improves the representativeness of the survey despite their lower propensity to respond.

Using passive data from a meter to complement survey data in order to study online behavior / Melanie Revilla, Carlos Ochoa, & Germán Loewe

Abstract: Surveys have been used as main tool of data collection in many areas of research and for many years. However, the environment is changing increasingly quickly, creating new challenges and opportunities. This paper argues that, in this new context, human memory limitations lead to inaccurate results when using surveys in order to study objective online behavior: people cannot recall everything they did. It therefore investigates the possibility of using, in addition to survey data, passive data from a tracking application (called a "meter") installed on participants' devices to register their online behavior. . After evaluating the extent of some of the main drawbacks linked to passive data collection with a case study (Netquest metered panel in Spain), the paper shows that the data from the web survey and the meter lead to very different results about the online behavior of the same sample of respondents, showing the need to combine several sources of data collection in the future.

Reports & Communications

Fast, free, and targeted: Reddit as a source for recruiting participants online / Itamar Shatz

Abstract: Recruiting participants is a necessary step in many studies. With the advent of online research techniques, scientists are looking for new places where participants can be recruited online, in order to overcome the limitations of current sources, and avoid the issues associated with sample overuse. The social media website 'Reddit' is a potential source for recruitment, as it allows for free and rapid data collection from large samples, while enabling researchers to target specific populations when needed. The ability to recruit for free is especially important, because it enables students and early career researchers, for whom even low recruitment costs can be prohibitive, to benefit from the opportunity of conducting research that they otherwise would not be able to. The current paper therefore aims to bring this prospective, untapped resource to the attention of the research community. The paper discusses current online recruitment sources and their limitations, provides an overview of Reddit, validates its use for research purposes, examines participation data from previous studies which recruited through Reddit, highlights its advantages and limitations as a participant pool, and suggests guidelines that can improve recruitment and retention rates for scientists looking to use Reddit for their research.