

1-1-2017

A retrospective on state of the art social media research methods: Ethical decisions, big-small data rivalries and the spectre of the 6Vs

Anabel Quan-Haase
The University of Western Ontario, aquan@uwo.ca

Luke Sloan
Cardiff University, sloanls@cardiff.ac.uk

Follow this and additional works at: <https://ir.lib.uwo.ca/fimspub>



Part of the [Communication Technology and New Media Commons](#), [Library and Information Science Commons](#), [Mass Communication Commons](#), and the [Social Media Commons](#)

Citation of this paper:

Sloan, L., & Quan-Haase, A. (2017). A retrospective on state of the art social media research methods: Ethical decisions, big-small data rivalries and the spectre of the 6Vs. In L. Sloan & A. Quan-Haase (Eds.), *The handbook of social media research methods* (pp. 662–672). Sage

A Retrospective on State of the Art Social Media Research Methods: Ethical Decisions, Big-small Data Rivalries and the Spectre of the 6Vs

Luke Sloan and Anabel Quan-Haase

Sloan, L., & Quan-Haase, A. (2017). A retrospective on state of the art social media research methods: Ethical decisions, big-small data rivalries and the spectre of the 6Vs. In L. Sloan & A. Quan-Haase (Eds.), *The handbook of social media research methods* (pp. 662–672). Sage.

Abstract

This concluding chapter offers critical reflections on some of the key themes covered in the Handbook. Ethics emerged as a concern for many scholars, both for those engaging in quantitative and qualitative approaches. Scholars agree in that there is no overarching set of rules that can be applied to all projects blindly, rather they see ethical decisions as being grounded in the specifics of the data being collected, the social group under study, and the potential repercussions for subjects. A second central theme was the value of qualitative approaches for understanding ‘anomalies’ within larger data sets. Qualitative approaches are seen as valuable and a stand-alone means of collecting, analyzing and making sense of social media data, in particular for projects where context is essential. Finally, as the contributions in this volume demonstrate that many of the challenges posed by the nature of social media data are being tackled and addressed, this chapter ends with a reorientation of the 6Vs which focuses on the primacy of the researcher in the decision-making process. We argue that the provision of technical solutions alone do not entirely address the 6V problem and clarity of thought around research design is still just as important as ever.

Key words

Social media, research methods, Twitter, Facebook, ethics, big data, qualitative methods, quantitative methods, mixed methods

Introduction

The SAGE Handbook of Social Media Research Methods brings together over 50 authors from a wide range of disciplines and scholarly traditions. This makes the Handbook truly interdisciplinary, drawing on approaches focusing on large-scale quantification to studies that stress the relevance of single cases and anomalies. It is this diversity that gives the Handbook depth and relevance and provides new perspectives and insights into the study of social media research methods. The Handbook demonstrates that social media methodology is not only about big data, but how

qualitative work is also developing quickly and leaving a mark on the field. It would also be inaccurate to say that the interdisciplinary nexus between the social and computing sciences is solely oriented around a positivist paradigm as clearly the challenges around collecting, collating and handling qualitative data on social media require researchers from all backgrounds to collaborate. Further, neither quantitatively nor qualitatively-oriented scholars can simply apply the traditional approaches developed in their disciplines to the study of social media phenomena. Rather, scholars are challenged to rethink conventional approaches and reorient themselves toward the new dimensionalities inherent in these kinds of data. This creates a real need for the development of innovative methodological approaches that are uniquely suited to social media environments.

The concluding chapter identifies several key trends that weave throughout the Handbook. We discuss ethical considerations first as one central theme that is of importance to many chapters and is considered by many scholars to still be unresolved. We briefly show what ethical issues are of most pressing relevance and ways of moving forward. We then examine the growing popularity of qualitative approaches to the study of social media. The range of approaches is astonishing, borrowed and adapted from established qualitative traditions. These approaches are singled out as not only countering big data approaches – often criticized for flattening data, losing context, and stressing large-scale trends at the expense of an individual's experiences – but also as providing unique insights into “anomalies” that would go unnoticed in large-scale scholarship (Bradley, 1993). This leads to the third central theme which focuses on the development of multi-method approaches that integrate big data analytics and small-scale studies. Regardless of whether the quantitative techniques follow the qualitative ones or vice versa, either process can be used to better illustrate current trends that are demonstrated in the initial data set and gain contextualization and reach deeper meaning. These kinds of approaches are not only time-consuming, but also require the formation of interdisciplinary teams that can bring to bear expertise on different approaches to data collection and analysis. We end the chapter with a discussion of the challenges surrounding the 6Vs first brought up in the introductory chapter. Having demonstrated throughout the book that the technical solutions to the 6V problem exist, we return to the essential agency of the researcher and the additional considerations we need to reflect on when tackling this new form of data for social scientific enquiry.

Ethics in big data and small data

Ethical considerations emerged as a strong theme in discussions related to the handling of social media data. It was evident that traditional considerations and guidelines regarding ethics were not applicable to the new challenges that social media environments present. This is directly linked to the kinds of approaches developing in social media scholarship including tools for data collection that harvest information at new scales in terms of the 6Vs discussed in the introductory chapter and revisited later in this concluding chapter (i.e., *v*olume, *v*ariety, *v*elocity, *v*eracity, *v*irtue, and *v*alue).

On one hand, scholars are calling for tighter regulations and more intense debate around ethical standards (Goel, 2014). On the other hand, scholars suggest that research involving social media data may not require as rigorous an ethics and consent regime as other types of research because data are publicly available and studies will often involve “minimal risk” to participants (Grimmelmann, 2015).

What complicates the decision-making around ethics is that no pre-established set of rules or guidelines can be applied to all projects. For Beninger (Chapter 5, this volume), discussions around ethics cannot be boiled down to a checklist, but must instead take the entire research process into account including issues such as the topic under investigation, the time period of data collection, the participants to be included, and the sensitivity of the content. She contends that decisions around whether to seek consent from individuals who have posted content to public sites are closely linked with the nature of the content under study and the potential repercussions disclosure can have for research subjects.

Regardless of whether there is high or low risk to participants, it is clear that existing ethical guidelines and practices are not readily applicable because social media data blur the lines between public and private spheres. Social networking sites (SNSs) contain information intended for a specific network audience consisting of a mix of close and distant ties and thus is not *truly* public, even if users do have an understanding that a wider network of “friends” can see, and interact with, this content. Recently, boyd and Crawford (2012) have also drawn attention to the fact that even data that is *truly* public, such as data posted on a Twitter timeline from a non-private account, may not be intended for further use by those who originally created the data. In short, how do social media scholars know that users are consenting to their data being utilized and analyzed in ways they cannot predict? Where in a research design is the traditional standard of consent being addressed? The networked nature of data on social media sites also presents new challenges for researchers. Consent obtained from one participant does not automatically transfer to individuals interacting with that participant. Thus, new questions emerge around what data can be collected and included in a study. This problem of boundary specification is characteristic of all data produced in social media platforms and extends to likes, comments, and retweets of content (Quan-Haase & McCay-Peet, 2016).

Issues of anonymity also arise in deliberations about ethics. In this regard, there is considerable disagreement within the scholarly community as to what strategy is the most ethically sound. At the center of this debate lies the question of whether publicly available data is by default *public* and hence can be examined by scholars for research purposes (Stewart, Chapter 16, this volume; boyd & Crawford, 2012). From a participant's point of view, anonymizing the collected data would most likely represent the lowest risk in terms of associating content with a particular person/account. A

common practice for reviewers during the peer-review process of a social media project is to request the anonymization of all data. In some instances, this may be a reasonable expectation, but one that is also associated with data loss. If key players – for example, Google’s Twitter account, the Twitter account of the US Republican candidate, Donald Trump, or of celebrity figure Kim Kardashian West – cannot be recognized in the data set, this would preclude scholars from drawing specific interpretations based on the social status of these key players and the role they play in society. To complicate things further, for some platforms such as Twitter, anonymizing data violates the terms and conditions of use. So, decisions around anonymity create tensions between the right of users to protect their privacy and the ability of scholars to draw conclusions based on their data.

Zeller (Chapter 23, this volume) specifically points out that not all data sets available online are indeed *public*. For instance, the website Ashley Madison was created to romantically connect married individuals (it basically helped people cheat). The service had around 40 million users in 2015 when the site was hacked and data on user accounts were retrieved and posted online for anyone to access (Dreyfuss, 2015). Similarly, the service Snapchat was reportedly also hacked, often via third party apps (Eng, 2014). Snapchat users consider this kind of data to be ephemeral and non-retrievable (Bayer et al., 2015), but it can still be available on a company server or via a third-party app. Zeller (Chapter 23, this volume) notes that scholars have the responsibility to assess the origins of data sets and the nature of consent given by users. While data may be publicly available online, if it has been obtained illegally, it may not conform to the standards of scholarly ethical practice. Nonetheless, it is not always clear where the boundaries lie, as data sets may be of public interest, but illegally obtained, increasing researchers’ uncertainty around the usage for research purposes.

Hargittai (2015) highlights the problem of the representativeness of the big data sets available through SNSs. She points out that “if people do not select into the use of the particular site randomly, then findings cannot be generalized beyond the site’s population” (p.65) because those who are not members of the site may vary from those who are in ways that are of relevance to the research being undertaken. Indeed, Sloan et al. (2015) demonstrate that, for UK Twitter users, it appears that the distributions of tweeter age, occupation and class are not representative of the wider population and that those who enable geotagging are not demographically identical to users who do not (Sloan & Morgan, 2015). This links to yet another ethical dilemma as the absence of certain groups from social media violates ethical principles of inclusivity. Conversely, concerns can arise about representation in small data projects as individuals may be more easily identifiable and reporting such data compromises anonymity.

The most controversial discussion around ethics so far is that which surrounds collaborations between academic and corporate researchers. For Vitak (Chapter 37, this volume), the trigger for much concern emerged from the publication of large-scale studies by Facebook’s Data Science Team

in collaboration with academic collaborators (e.g., Das and Kramer, 2013; Kramer et al., 2014). Users were often not informed about the study either before it took place or after its completion and, as a result, Vitak contends, average Facebook users who served as subjects “felt uncomfortable not knowing what was going on ‘behind the scenes’ at the company” (p. ##). This led to an outcry in the media regarding the ethical practices of big data analytics and a call for increased transparency, greater communication with research subjects and more care in the design of large-scale experiments (Goel, 2014; Grimmelman, 2015; Hargittai, 2015; Tufekci, 2015).

Discussion is underway about the need for ethics standards for research involving data sets from corporate social networks. Grimmelman (2015) suggests that it might be even more important with corporate research because corporate researchers’ self-interest may be even more significant. Jeffrey Hancock, one of the academic researchers involved in the Facebook experiment manipulating users’ emotions, suggests an ‘opt-in process’ whereby users agree from the outset to participate in studies that will have a significant impact on their internet experience. He also suggests introducing a debriefing process that would provide information to users after smaller studies have been carried out, a practice that is standard today in experimental studies that involve some element of deception. May Gray, from Microsoft Research, suggests that “if you’re afraid to ask your subjects for their permission to conduct the research, there’s probably a deeper ethical issue that must be considered” (Goel, 2014). The lesson here is that, simply because it is technologically possible, does not mean that it is ethically advisable.

Social media scholars cannot turn a blind eye toward ethical considerations because academic research is based on trust. Building trust with human subjects is critical and a result of a long-standing tradition of ethical standards in academia. The ethical standards that govern research practices today are based on past experiences, such as the Stanford Prison Experiment (Zimbardo, 1971) and the Milgram (1963) experiment on obedience to authority figures. In both of these cases, researchers, in part unintentionally, breached the participants’ trust through the unexpected consequences of their study designs. If participants get the perception that scholars are unconcerned about their wellbeing and the intended and unintended consequences of their research, this long-built trust may dissipate. Salmons (Chapter 12, this volume) notes that this could jeopardize what lies at the center of much academic work, the recruitment of participants to voluntarily participate in research studies.

Big data versus small data?

Big data approaches have received considerable scholarly and media attention, being heralded for their great potential to provide new insight into human behavior and thereby transforming the nature of social science research. It is often claimed that, with large enough data sets, we will no longer need theory as powerful “knowledge discovery software tools find the patterns and tell the analyst

what—and where—they are” (Dyche, 2012). These approaches have received harsh criticism for being myopic to context and not being able to tell a full story by focusing only on large trends. Certainly their quantitative nature and the confusion around data-mining and machine learning paints a picture where theory becomes obsolete (Anderson, 2008), although this volume demonstrates that theory is seldom absent despite the hype around big data approaches.

This Handbook demonstrates that there is more to big data than nomothetic, quantitative work – indeed there is an expanding body of work around innovative qualitative approaches that demonstrate completely different insights into the value of social media data. As Rasmussen Pennington (Chapter 15, this volume) points out: “The exponentially-growing presence of non-text documents on popular social media outlets such as Facebook, Twitter, Instagram, Flickr, Pinterest, Snapchat, YouTube, and Vine has created an opportunity for social science researchers to understand the products of digital society through analyzing this data in many formats” (p XXX). Qualitative approaches being developed in social media scholarship do not only consist of embedding traditional techniques into new research designs (as argued by Latzko-Toth, Bonneau and Millette, Chapter 13, this volume), rather they consist of also using small datasets to reassess their capabilities and complementarity with quantitative approaches. For example, Georgakopoulou (Chapter 17, this volume) proposes a new kind of narrative analysis based on small stories research to analyze social media data. While she borrows from the principles of narrative analysis, her approach is uniquely suited to the parameters created by social media environments. This is particularly relevant for narrative analysis, as narratives unfold differently on social media than in any other medium. In addition, the value of qualitative approaches goes beyond the type of method being employed and also expands to the populations being investigated. Salmons (Chapter 12, this volume) identifies that social media can be an entry point for more traditional studies through offering access to hard to reach individual or groups and enabling us to further understand their lived experiences.

The use of big versus small data does not have to be an either or debate. Rather, mixed methods can provide an alternative that takes the advantages of one approach to compensate for the disadvantages of the other – hence they can complement each other. Consider the use of Big Data, which uses large and complex data sets. Some argue that the massive data speaks for itself, that quantity equates to quality (Zeller, 2015). However, critics argue that such data lacks contextualization and deeper meaning. A solution to this problem would therefore be to employ qualitative strategies in order to gain more in-depth knowledge regarding one’s research topic, as well as its meaning to participants.

The value of a mixed methods approach is demonstrated by Mayr and Weller (Chapter 8, this volume) through the combination of surveys, social media and interviews. Indeed, the way in which qualitative and quantitative data complement each other is particularly visible when utilized for

social media research. Social media sites produce vast amounts of diverse content at a rapid pace, creating a dilemma for researchers who must balance keeping the size of the data manageable while gathering adequate information to develop knowledge (Latzko-Toth, Bonneau and Millette, Chapter 13, this volume). For this reason, a mixed-methods strategy can be instrumental, as quantitative data collection allows for sufficient breadth, while qualitative data collection provides the required depth. One can also combine the two methods through conversion, in which the data is either “quantitized” or “qualitized” (Zeller, 2015). In other words, one need not collect both qualitative and quantitative data, but can transform one into the other to meet the research needs of a project. Ultimately, the goals, research questions formulated, and theoretical underpinnings of the study will guide these decisions.

In his study of the relation between physical places and their social media hyper-local representations through the application Instagram, Nadav Hochman (Chapter 22, this volume) demonstrates the value of a mixed-method approach to social media research. Using Instagram’s API to gather more than 28,000 images pertaining to the elusive-yet-renowned street artist Banksy, Hochman manipulated the sample in a variety of ways to cluster such images in order to compare and contrast the ways in which various users disseminated Banksy’s art in New York. While his collection method is largely quantitative, his examination of the images has a qualitative element. Hochman informally examined each cluster of images to reveal differences that were both unintentional, as well as intentionally provided by users. Since he sought to determine what particular characteristics of hyper-locality are experienced through social media, statistical analysis simply would not suffice. Once his quantitative methods became inadequate, he transitioned to a qualitative analysis in order to draw significance and meaning from the collected images.

Zeller (Chapter 23, this volume), Hochman (Chapter 22, this volume), and Latzko-Toth, Bonneau and Millette (Chapter 13, this volume) demonstrate that in order to effectively conduct research in a field as vast and diverse as social media, one has to draw on a varied and flexible methodological toolkit. In this case, numbers do not speak for themselves, as each post (whether it be an image, a tweet, or a share) encompasses a variety of motivations, interactions and subjectivities. Employing a form of qualitative analysis is thus essential to fully understand such online activities. On the flip side, the massive amount of users flocking to each site means that the smaller samples typically required for qualitative analysis risk producing “distinct” results, distinct in that they do not speak for the majority of other users. Thus, researchers must develop a flexible approach to the study of social media data, and be prepared to develop strategies that best suit the topic at hand.

Combining elements of qualitative and quantitative methods can be seen as creating a strategy of data collection and analysis that is unique to the study, however researchers typically have more extensive training in one branch of methods/analysis than another. Attempting to take on elements of

both could mean employing strategies that the researcher is not well-familiarized with and this in itself is a good justification for the value of collaboration in this area (Quan-Haase & McCay-Peet, Chapter 4, this volume).

Lastly, a concern regarding combining qualitative and quantitative methods may be deciding which to employ first – an issue not limited to social media research. Should a researcher interview a small sample for insight, and then attempt to survey a larger sample of similarly-minded people in order to generalize such insight? Or vice versa, where a large sample is surveyed and the interview collection follows? While the latter may seem more straightforward, the question then becomes who from the large sample to select for qualitative data collection? Certain members of the sample may provide information that, had other members been selected, would not have been discussed. In other words, the choice of which participants a researcher selects to conduct qualitative research on could take the study in a very different direction, depending on who is used. While the solution may appear to be using the same sample for both quantitative and qualitative strategies, this could prove very costly and time-consuming for the researcher, and such practical constraints are not inconsiderate when dealing with big data.

Reorienting the 6Vs

Returning to the 6Vs discussed in the introductory chapter, we have a very different take on the nature of the challenges presented to researchers wishing to work with social media data. The chapters in this volume have demonstrated frenetic activity around the development of processes and systems to deal with the characteristics of the data, but tools and approaches are only as effective as the researcher using them. In light of this, we invite readers to reconsider the 6Vs from an alternative perspective that focuses on the individual designing and conducting the research rather than the data itself.

Volume will be an issue for any study even if technology makes collection and access easy as researchers still have to sort the *sound from the noise*. For example, although it is laudable to use Twitter to try and predict an election by looking for positive sentiment towards political parties, looking for references to the Green Party using a search term such as ‘Greens’ is going to identify many false positives – and any strategy to whittle these errors out requires time in proportion to the number of cases (the author speaks from experience: Burnap et al. 2016). Tighter search terms will reduce volume and accuracy but may exclude much relevant content, so the researcher has to evaluate how much noise is acceptable and schedule an extensive period of post-collection data cleaning.

Taking into account a *variety* of data types has always been a challenge of mixed-methods research, however as researchers we typically design such studies with tight parameters (such as the use of open and closed questions on a questionnaire) that allow us to link the data we are collecting

with a careful plan of analysis – not so with social media data. Variety in social media means an unstructured mix of text, images, and videos with some users producing only one type of data whilst others using two or all three types. An apparently simple study looking at reactions on Twitter to, for example, the London 2012 Olympics may need to take into account multiple tweets from the same users, the text of the tweet, use of images, use of hashtags and even the end content of a URL. Does this require a researcher to be an expert methodological pluralist? Should a researcher choose to focus on only one mode of data? What is being excluded by such a choice? The data can be captured, but that does not aid us in dealing with its complexity and variety.

Velocity is a key concern for any researcher interested in events or time sensitive investigations. Reacting quickly to real world events by starting live data collections using some of the tools described in this volume, such as COSMOS (Morgan, Chapter 26, this volume), SocialLab (Reips & Garaizar, Chapter 27, this volume) and Netlytic (Gruzd, Mai, & Kampen, Chapter 30, this volume), allows data to be collected whilst events unfold but deciding on an analytical strategy for the data requires an understanding of temporal granularity. The metadata associated with social media activity specifies the creation of a post/tweet/check-in to the second and it is then up to the researcher to decide at what temporal level the data is aggregated. For example, does it make sense to plot sentiment around a specific event for every second or should a summary sentiment score be computed by minute, hour, day, week or month? For studies with a high n during a short burst of time a smaller aggregation may be appropriate, but for other studies where cases are limited it may be necessary to summarize data over a longer period (see Williams, et al., 2016).

Veracity is hard to establish and researchers must be reflexive around the use of demographic proxies and how users present themselves online (Sloan this volume, Yang, et al. this volume). The presentation of the self and construction of identity and group memberships is not new to the social sciences but the issues are compounded by the ‘remoteness’ of the researcher and the *virtuality* and *plurality* of social media data. Certainly respondents to a survey may answer items in light of social desirability bias, but how does this manifest in naturally occurring user-produced data? Sloan is involved in current work investigating the possibility of linking social media to survey data to test the accuracy of demographic proxy measures and the relationship between opinions expressed in survey-format and tweets made online, but in the meantime studies are (successfully) drawing on the *wisdom of crowds* using Twitter data to predict elections (Burnap et al., 2016), box office revenue (Asur & Huberman, 2010) and exchange rates (Papaioannou et al., 2013) with variable degrees of success (Lassen, La Cour, & Vatrapu, Chapter 20, this volume). Veracity may be less important to studies looking for nomothetic aggregate patterns than those interested in the intricacies of individual cases.

How do we account for *virtue*? The terms and conditions of data usage differ by platform but as long as we abide by them we are legally entitled to do things with the data that violate traditional notions of ethical research. For example, it is not possible to implement the principle of anonymity when conducting qualitative analysis on tweets because Twitter terms and conditions require the tweet content always to be reproduced alongside the Twitter handle – what are the implications of this on protecting the ‘participant’ from harm for research into sensitive topics such as the use of hate speech online? If Twitter is a broadcast medium, is it necessary to gather informed consent? Conversations with colleagues in the wider academic community demonstrate a variable approach to ethics dependent on discipline, the level of understanding that ethics committees have about the nature of social media data and whether projects using ‘scraped’ data should be classified under primary collection or secondary analysis. An advantage of differing approaches is the opportunity for researchers to share ideas and for good practice to emerge and be publicized.

The *value* of social media data for social scientific research has been demonstrated in this volume through the use of case studies and applied examples and in the wider academic literature. Social media is a different source of data to what we are used to dealing with and in turn it enables us to answer different questions, but the true value of a study is often in the reflections of the researcher on the strengths and limitations of the data and approach. Throughout this volume authors have presented transparent analytical strategies and methods to allow others to build upon their ideas – indeed this is exactly how science progresses and we should encourage open and frank discussions about what works and what does not. There is also value in academics from different disciplines working together, expanding their understanding and learning. This is reflected in the interdisciplinary nature of the volume.

In summary, whilst many of the challenges discussed in the introduction to this book appeared to be methodological and technical, following the developments outlined in this volume we can see that the challenges operate at a much more personal level. Researchers need to make good decisions informed by an understanding of the data and continue reflecting on their current practice, which may in turn involve closer collaborations with other disciplines. In many cases the technology and tools exist to enable access to the data, but just because we *can* does not mean that we *should* - there is no substitute for good research design and constant reflexive practice.

Conclusion

We started this volume by outlining the methodological mountain ahead of us, but in retrospect the climb is not so sheer. Much interest and enthusiasm has been generated around the development of this Handbook and the range of disciplines, methodological positionings and expertise demonstrated across the chapters illustrates the frenetic research activity around the use of social media data for social scientific analysis. There are still important issues to be resolved, not least around ethical

frameworks and the small data vs. big data rivalry, but it is clear that these discussions are well underway and that the thinking in this area is sophisticated, informed and grounded in a knowledge of the data, its limitations, and possibilities.

For us, it seems that ‘knowledge of the data’ is the key. The technological solutions clearly exist after a fruitful meeting of minds between the social and computer sciences and humanities, but the types of questions that can be asked, how representative our findings are and what the best plan of analysis is to answer our research questions all require a deep understanding of the purpose, functionality and idiosyncrasies of the relevant social media platform. So after 39 chapters we find that the difficult decisions around designing and conducting research using social media data are analogous to those of any traditional social scientific enquiry. At the same time, social media data has enticed scholars to develop new frameworks and approaches that are uniquely suited to the challenges and dimensions presented by social media data.

So, having established that the remaining challenges are typical of any research project, there is no reason to treat social media data with trepidation or fear. There may be a technical learning curve depending on what you want to do, but what better opportunity to learn a new skill or to partner with a colleague in a different and complementary discipline? This book is a demonstration of the ability of the social sciences and humanities to upskill and remain relevant in a fast-paced and changing world. It is also a testament to how creative, innovative and groundbreaking we can all be when we break down disciplinary silos and collaborate. We sincerely hope that this volume enables and encourages new and experienced researchers to add to the debates and that, in a few years’ time, even more colleagues will feel able contribute to the second edition!

Luke & Anabel (Co-Editors)

References

- Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. Retrieved April 22, 2016 from http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory
- Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference* (Vol. 1, pp. 492–499). Toronto, Canada: IEEE. doi:10.1109/WI-IAT.2010.63
- Bayer, J. B., Ellison, N. B., Falk, E. B., & Schoenebeck, S. Y. (2016). Sharing the small moments: Ephemeral social interaction on snapchat. *Information, Communication & Society, 19*(7), 956–22. doi:10.1080/1369118X.2015.1084349
- boyd, d., & Crawford, K. (2011, September). Six provocations for Big Data. In *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, Symposium conducted at the Oxford Internet Institute, University of Oxford. <http://doi.org/10.2139/ssrn.1926431>

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. <http://doi.org/10.1080/1369118X.2012.678878>

Bradley, J. (1993). Methodological issues and practices in qualitative research. *The Library Quarterly: Information, Community, Policy*, 63(4), 431–449. doi:10.1086/602620

Burnap, P., Gibson, R., Sloan, L., Southern, R., & Williams, M. (2015). 140 characters to victory? Using Twitter to predict the UK 2015 general election. *Electoral Studies*, 41, 230–233. doi:10.1016/j.electstud.2015.11.017

Burnap, P., Morgan, J., Sloan, L., & Williams, M. (2015). Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLoS ONE* 10(3), e0115545. doi: 10.1109/WH-IAT.2010.63.

Burnap, P., Sloan, L., & Williams, M. (2016). Crime sensing with big data: the affordances and limitations of using open source communications to estimate crime patterns. *British Journal of Criminology*. Advanced Access published March 31, 2016. doi:10.1093/bjc/azw031

Das, S., & Kramer, A. (2013). Self-Censorship on Facebook. *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, Washington DC: AAAI, pp. 120–127. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6093/6350>

Dreyfuss, E. (2015, August 19). How to check if you or a loved one were exposed in the Ashley Madison hack. Retrieved from <http://www.wired.com/2015/08/check-loved-one-exposed-ashley-madison-hack/>

Dyche, J. (2012). Big Data ‘Eurekas!’ Don’t Just Happen. *Harvard Business Review*. Available from: <https://hbr.org/2012/11/eureka-doesnt-just-happen>.

Eng, J. (2014, January 1). Snapchat hacked, info on 4.6 million users reportedly leaked. *NBC News*. Retrieved 27 April 2016 from <http://www.nbcnews.com/business/snapchat-hacked-info-4-6-million-users-reportedly-leaked-2D11833474>

Goel, V. (2014, August 12). As data overflows online, researchers grapple with ethics. *The New York Times*. Retrieved April 16, 2016 from <http://www.nytimes.com/2014/08/13/technology/the-boon-of-online-data-puts-social-science-in-a-quandary.html>

Grimmelmann, J. (2015). The law and ethics of experiments on social media users. *Colorado Technology Law Journal*, 13(2), 221–272. Retrieved from <http://ctlj.colorado.edu/wp-content/uploads/2015/08/Grimmelman-final.pdf>

Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 63–76. doi: 10.1177/0002716215570866

- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1–12. doi: 10.1177/2053951714528481
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. doi: 10.1073/pnas.1320040111
- Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, *67*(4), 371–378. <http://dx.doi.org/10.1037/h0040525>
- Papaioannou, P., Russo, L., Papaioannou, G., & Siettos, C. I. (2013). Can social microblogging be used to forecast intraday exchange rates? *Netnomics: Economic Research and Electronic Networking*, *14*(1–2), 47–68. doi:10.1007/s11066–013-9079–3
- Quan-Haase, A. & McCay-Peet, L. (2016). Social network analysis. In *International Encyclopedia of Communication Theory and Philosophy*. Cambridge, MA: Wiley.
- Sloan, L., & Morgan, J. (2015). Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter. *PloS One*, *10*(11), e0142209. doi:10.1371/journal.pone.0142209
- Tufekci, Z. (2015). Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency. *Colorado Technology Law Journal*, *13*(2), 203–217. Retrieved from <http://ctlj.colorado.edu/wp-content/uploads/2015/08/Tufekci-final.pdf>
- Zimbardo, P. G. (1972). Stanford prison experiment: A simulation study of the psychology of imprisonment. Philip G. Zimbardo, Incorporated.