Inequalities in Social Media Use and their Implications for Digital Methods Research

Christoph Lutz
Nordic Centre for Internet and Society
BI Norwegian Business School, 0484 Oslo, Norway
christoph.lutz@bi.no

Abstract

Inequalities in social media have been investigated under the umbrella of a digital divide. Research has shown how inequalities based on social categories are perpetuated or even reinforced with digital technologies. Different levels of inequality have been differentiated, including the first-level, second-level, and third-level digital divide. When it comes to social media, all these divides have received attention, but they have not been systematically connected to digital methods research, which relies on trace data, often from social media. This chapter discusses inequalities in social media access, use, and outcomes and connects them to digital methods research from an ethical perspective. It looks at two key issues for data subjects when data about them is analysed through digital methods: representation and privacy. Both issues are tied to questions of power and inequalities that merit careful attention among social media researchers. Beyond unequal representation and privacy among data subjects, the chapter also discusses inequalities within the research community, as they pertain to unequal access to social media data, unequal opportunities for digital methods skills development, and unequal opportunities to leverage digital methods analyses for career development. Overall, the chapter argues for a stronger connection of digital inequalities and digital methods.

Keywords: digital divide, digital inequality, social media, digital methods, social media research, representativeness, Internet ethics

Citation

Lutz, C. (2022). Inequalities in social media use and their implications for digital methods research. In A. Quan-Haase & L. Sloan (Eds.), *The SAGE Handbook of Social Media Research Methods* (Second Edition) (pp. 679-690). SAGE.

Inequalities in Social Media Use and their Implications for Digital Methods Research

Introduction

Social media have become a part of the everyday lives of billions of people around the world (Statista, 2020). However, not all people use social media equally and considerable differences in uptake exist between countries and within countries. Digital inequalities research has investigated social media, finding pronounced differences in terms of gender, age, socio-economic status, and attitudinal variables. These differences exist both for social media use more broadly as well as for specific platforms. Instagram, for example, is used by 43% of female US adults but only 31% of male US adults. YouTube, on the other hand, is more widely used by US-based male adults (78% vs. 68%; Perrin & Anderson, 2019).

Social media use reflects existing power imbalances and is embedded within larger systems of structural inequality, including those that pertain to Internet access, Internet uses, digital skills, and outcomes from Internet use (Lutz, 2019). Therefore, a more holistic understanding of social media use is reached when digital inequalities are considered more broadly. The goal of this chapter is to offer a contextualized overview of digital inequalities in social media use and to reflect on the consequences such inequalities have for social media research, especially digital methods research that relies on user-generated social media data. In that sense, the chapter will approach ethical implications for social media research that come with digital inequalities. Aspects of representation and voice feature prominently but I will also discuss privacy considerations and science- and community-related issues in relation to inclusion. The chapter calls for more reflexivity regarding aspects of digital inclusion in social media research.

Digital Inequalities Research and Social Media Use

For more than 20 years, digital inequalities[1] have been an important topic in the social sciences. Various conceptualizations exist (Van Dijk, 2006), and in this chapter, I follow Kvasny and Keil (2006, p. 160), who define digital divides as 'the disparities in the structure of access to and use of ICT' as well as 'the ways in which longstanding social inequalities shape beliefs and expectations regarding ICT and its impact on life chances'. This definition is broad and mentions key dimensions of digital inequalities research, including access, use and outcomes ('impact on life chances'). Moreover, the definition addresses ICTs more generally, rather than 'just' the Internet. It thus captures the broadening of digital inequalities research and its importance with emerging technological developments such as AI and mobile technology (Robinson et al., 2020).

A rich body of literature has shown that digital inequalities reflect broader social inequalities in terms of positional variables such as age, gender, race, socio-economic status, geographic location, and employment status (Robinson et al., 2015). Traditionally excluded and marginalized people, such as migrants and the homeless (Marler, 2018), are also disadvantaged when it comes to ICTs, as they have difficulties in getting access to such technology, face restricted use opportunities, and often lack important digital skills (Hargittai, 2002; Hargittai & Hinnant, 2008; Robinson, 2009;

[1] I use the plural form 'digital inequalities' rather than the singular form 'digital inequality' to acknowledge the plurality, multi-dimensionality and complexity of social stratification in the context of digital technology. 'Digital divides' and 'digital inequalities' will be used interchangeably, although I favor the latter term because it has less of a binary notion.

Sims, 2014; Zillien & Hargittai, 2009). Thus, under-privileged groups tend to benefit less from ICTs than privileged groups (Blank & Lutz, 2018; Van Deursen & Helsper, 2015) and optimistic claims that the Internet would lead to less stratified societies have been challenged by digital inequalities scholars (Norris, 2001).

Since the first studies on the topic, digital inequalities research has progressed substantially and now includes the study of at least three connected divides: the first-level divide, the second-level divide, and the third-level divide. Table 1 provides an overview of each divide, connecting the three levels to inequality issues for data subjects and researchers that emerge from digital methods research.

[Insert Table 1 near here]

The *first-level digital divide* refers to unequal access to ICTs or 'the gap between those who do and those who do not have access to new forms of information technology' (Van Dijk, 2006, pp. 221-222). Research on the first-level digital divide was particularly prominent in the 1990s and early 2000s, when Internet penetration was still relatively low in many Western countries. Van Dijk (2006) summarized this research, showing that large differences in Internet access exist between different population segments. Increasing Internet uptake in the last 25 years, including the rapid adoption of the mobile Internet, meant that certain access gaps (e.g., gender and race gaps) have largely closed in countries such as the United States (Hargittai & Hinnant, 2008), where around 90% of the population have access to the Internet (Pew, 2019a). Nevertheless, the global divide (Norris, 2001) is still extremely large, reflecting global economic and social inequalities (ITU, 2019a). In Burundi, for example, only about 3% of the population have access to the Internet and in India about 34% (ITU, 2019b). By contrast, countries such as Qatar (99.5%), Norway (96.5%) and the United Kingdom

(95%) have almost complete saturation (ITU, 2019b). In addition, within affluent countries some groups struggle to access the Internet, especially those aged 65+, those with less than a high school degree, and those living in rural areas (Pew, 2019a). Given these persistent inequalities in access, scholars have warned against abandoning the study of Internet access (Newlands & Lutz, 2020; Robinson, 2009; Van Deursen & Van Dijk, 2019).

The *second-level digital divide* refers to inequalities in Internet uses and digital skills. Hargittai (2002) introduced the concept to differentiate binary divides in Internet access from more complex inequalities rooted in capabilities and practices (second-level). Research on digital inequalities has increasingly transitioned from the first- to the second-level digital divide, and substantial research on skills and use divides exists (Blank & Groselj, 2014; Gui & Argentin, 2010; Hargittai, 2010; Hargittai & Hinnant, 2008; Van Deursen & Van Dijk, 2011, 2014; Van Dijk, 2006; Zillien & Hargittai, 2009). These studies have shown differentiated inequalities along socio-economic lines, where young, male and educated users are particularly advantaged. Some second-level digital divide research has looked at online participation and social media, identifying the user characteristics of specific social media platforms or differentiating active content producers from passive consumers (Blank, 2013; Blank & Lutz, 2017; Brake, 2014; Correa, 2010; Hargittai, 2007, 2015; Hargittai & Walejko, 2008; Hoffmann, Lutz, & Meckel, 2015; Schradie, 2011). The findings are mixed, not allowing for strong associations between socio-economic status and online participation or social media use. In addition to vertical aspects such as education and income, horizontal aspects of inequalities, capturing lifestyles and life courses, strongly affect social media use and participation (Lutz, 2016). We will come back to social media below.

Finally, the *third-level digital divide* describes inequalities in outcomes from Internet use. As Van Deursen and Van Dijk (2015) note, '[t]hird-level divides […] relate to gaps in individuals' capacity to translate their internet access and use into favorable offline outcomes' (p. 30). This understanding builds on the assumption that certain individuals profit disproportionally from Internet access and use, being able to leverage these benefits to strengthen their social position (Van Dijk, 2005; Tichenor, Donohue, & Ollien, 1970 for a similar argumentation within the knowledge gap theory). The dynamic is termed a Matthew effect or rich-get-richer effect (Merton, 1968). Within the third-level digital divide, outcomes should be understood broadly to include not only benefits but also harms (Blank and Lutz, 2018). Studies on the third-level digital divide commonly investigate tangible offline outcomes from Internet use in economic, social, political, and cultural terms (Van Deursen & Helsper, 2015). Outcomes of interest include finding a job thanks to the Internet, improving one's health, and meeting a partner. However, intangible outcomes such as wellbeing are also studied (Büchi et al., 2018) and the relationship between ICT (especially smartphone use) and wellbeing has been a matter of heated debate in psychological research (Orben & Przybylski, 2019). Empirically, research on the third-level digital divide shows that skills, attitudes, and uses predict Internet outcomes better than demographic and socio-economic factors and that age has the most pronounced effect among the demographic characteristics (Blank & Lutz, 2018; Van Deursen & Helsper, 2018).

When it comes to *social media*, all three divides are important. The first-level social media divide differentiates social media access, including whether certain platforms are available or blocked in a certain country (Nabi, 2014). An example is the ban of TikTok in India, which had more than 200 million users and was particularly empowering for marginalized groups[2] such as rural youth (Economist, 2020; Shukla, 2020). Questions of accessibility also matter when it comes to the mobile Internet and mobile social media (Humphreys, 2013). Since social media are increasingly used on mobile devices and can be quite data intensive, access to smartphones, high-speed and reliable mobile Internet (4G, 5G) and Wi-Fi connections at home can make a big difference in someone's social media experience. However, compared to non-mobile Internet access, mobile Internet access has been described as inferior in terms of functionality, usability and openness, leading to fears of a 'mobile underclass' among the mobile-first and mobile-only population (Napoli & Obar, 2014; Newlands & Lutz, 2020; Tsetsi & Rains, 2020).

The second-level social media divide refers to differences in what people do once they are on social media as well as how literate and skilful they are in using social media (Blank & Lutz, 2016; Livingstone, 2014; Van Deursen et al., 2016). Blank and Lutz (2016), for example, analysed the social structuration of ten social media activities in Great Britain and found that different activities have quite distinct profiles. Posting writing and liking a company page are most strongly associated with vertical indicators of inequality (education, income), while updating personal information, liking others' content and unfollowing/unfriending do not depend on socio-economic status. Pearce and Rice (2017), relying on representative data of Armenia-based citizens (16+), showed small effects of demographic characteristics on ten social media activities. Two of the activities (taking quizzes; meeting new people and be entertained) were gendered,

[2] In a widely shared tweet, Ana (@tweetsofamuggle) commented: *'Tiktok succeeded in breaking the monopoly on content creation to a large extent. It gave a space for expression to those who don't lead instagrammable lives. Elites brush off its content as cringe ignoring how their idea of what's cringeworthy is determined by their social class.'* https://twitter.com/tweetsofamuggle/status/1277666237682561024

with men performing these activities more often than women, and only one activity was significantly - and negatively - associated with education (playing games).

Finally, the third-level social media divide describes outcomes from social media and thus considers social media use in the context of other activities and life trajectories. The question here is whether social media users can capitalize on their use and experience benefits such as better social inclusion, increased job prospects, higher levels of wellbeing, or more opportunities for cultural participation. The literature has shown the social capital benefits of social media, where social capital describes ‘the sum of the resources, actual or virtual, that accrue to an individual or a group by virtue of possessing a durable network of more or less institutionalized relationships of mutual acquaintance and recognition’ (Bourdieu & Wacquant, 1992, p. 14). Social media platforms allow for the curation of ties, both strong and weak, and can be leveraged for bridging social capital in particular (Ellison et al., 2007). Utz (2016) used a large sample of working Internet users in the Netherlands to compare Facebook, LinkedIn and Twitter in terms of their informational benefits. LinkedIn was seen as most beneficial, followed by Twitter and Facebook, showing how social media platforms designed for professional purposes seem to be most useful for social capital gains. However, research on the outcomes of social media use tends to approach the topic through other theories (e.g., social capital theory, participation, and learning), rather than a digital inequalities lens. Therefore, we lack knowledge on how different groups benefit and suffer differentially from social media. However, evidence suggests that existing inequalities and power dynamics are often reinforced, particularly as social media integrate AI and big data technology (Lutz, 2019; Micheli et al., 2018; Noble, 2018; Robinson et al., 2020).

Implications for Digital Methods Research

The following discussion will point to key implications of digital inequalities for social media-based digital methods research and is divided into two sub-sections. The first one, *Implications for data subjects*, points to ethical questions about data subjects (e.g., Twitter users whose data is analysed), particularly as they pertain to inequalities. The second one, *Implications for researchers*, points to ethical questions within the research community based on unequal access, skills and outcomes from digital methods analyses. I use the term digital methods research in a descriptive and non-programmatic way, deviating from the methodological paradigm of digital methods proposed by Rogers (2013). More specifically, I use digital methods to refer to research that studies digital contexts, including social media, and where user-generated trace data as well as machine-generated data is a key part of the analysis. Thus, social network analyses and descriptive analyses of Twitter data (Bruns et al., 2013; Highfield et al., 2013), virtual ethnographies of subreddits (Darwin, 2017; Foeken & Roberts, 2019), app walkthroughs (Light et al., 2018; Duguay et al., 2020), and algorithm audits (Rieder et al., 2018) fall within this understanding. I do not consider research that uses digital technology in a supporting role as digital methods. For example, qualitative interviews where the interview is recorded digitally and the recording is uploaded to a cloud server or a standard online survey done through Qualtrics do not fall within this definition.

Implications for data subjects

Data subjects are differentially affected by digital methods analyses of their data. In the following, I will focus on two main concerns: representation and privacy. Concerns about representation describe how individuals from different backgrounds are represented in digital methods analyses. Issues of data under-representation, exclusion, and discrimination based on age, gender, language, disability, socio-economic status, lifestyle, and sexual orientation – as well as their intersection – fall within this concern. While concerns about representation question whether individuals across social groups have a substantive and appropriate voice in digital methods analyses, concerns about privacy deal with exposure, beneficence, and consent. Here, we should ask how digital methods might amplify existing vulnerabilities of those ‘at the margins’ of society (Marwick & boyd, 2018), for example by exposing them or making them susceptible to dataveillance and social sorting (Gandy, 1993; Van Dijck, 2013).

Questions of Representation

A major digital inequalities concern of digital methods research is about representation and inclusion. Again, the differentiation of first-, second- and third-level divides helps as a structuring device.

In terms of the first-level divide, the unequal adoption of social media as well as the (lack of) availability of certain platforms mean that certain groups tend to be underrepresented in digital methods analyses, while others are overrepresented. Digital methods research should be acutely aware of the underlying inequalities in social media use and be able to answer the following questions: Who is represented in my data and why? Who am I not able to capture in my data? How representative is my data of the population in question? How is data cleaned and sorted, and are certain categories systematically excluded in this process (Bechmann & Bowker, 2019)? Analyses relying on Twitter data, for example, need to be aware that Twitter use is a minority phenomenon and tends to be more prevalent among younger, more affluent individuals (Blank, 2017), as well as those living in urban areas and with strong ideological preferences (Barberá & Rivero, 2015). Thus, such user groups will likely show up prominently in the data. Consequently, such analyses will not be able to capture the full breadth of societal voices on a topic. Some groups who are difficult to capture on mainstream social media platforms are older adults, children, disabled individuals, wealthy but secretive elites, introverted people, manual workers (as opposed to office/desk workers), chronically ill, and prisoners. Content analyses of Twitter data will have a hard time giving appropriate representation to these groups and will likely over-represent younger and middle-aged adults, able-bodied individuals, office/desk workers in areas such as journalism, academia, marketing, and the creative industries, as well as extroverted people. Such studies should be careful to generalize and be mindful of the underlying user dynamics of a specific platform (Blank & Lutz, 2017; Hargittai, 2015). A promising approach to address representation issues is to combine digital methods analyses with other forms of data collection where representativeness is higher, for example via triangulating Twitter data with representative surveys (Sloan et al., 2020). Specific consideration should also be given to language and the global divide. While Facebook is available in more than 100 languages (Fick & Dave, 2019), Instagram only supports 32 languages (Wikipedia, 2020), and Twitter 34 (Twitter, n. d.). Prominent languages that are not supported by Instagram and Twitter at the time of writing are Swahili, Omoro and Hausa, together spoken by over 200 million people across Africa.

In terms of uses and the second-level divide, use intensity tends to be skewed among social media users (Antelmi et al., 2019), privileging those who are most active and most connected in digital methods analyses. Since digital methods often have to rely on visible and user-generated data (e.g., comments, tweets, status updates, likes, shares/retweets), rather than less visible and machine-generated data (e.g., clickstreams, device logs, change of settings), passive and occasional users tend to fly under the radar. Virtual ethnographies, for example, face difficulties in accessing lurkers, thus potentially neglecting their importance. Research can partly address such issues of representation of social media users by in-depth immersion and collaboration with site owners (Ferguson, 2017), who might have access to more comprehensive use data. Digital methods researchers should also be mindful of use inequalities in temporal terms. This includes power asymmetries between experienced users who have been on a platform for a long time and managed to accumulate substantial capital vs. novice users who have not had the same opportunities. Inequalities between experienced and new users have been shown in the context of the gig economy, where experienced users tend to earn substantially more (Cook et al., 2018), and in general Internet use, where experienced users can accrue more informational benefits (Lupač & Sládek, 2008). However, less research on the temporal inequalities among social media users and their implications for digital methods research exists.

Aspects of form and expression should also be considered. Given that today's mainstream social media platforms are increasingly visual, focused on videos and storytelling (Highfield & Leaver, 2016; Russmann & Svenson, 2017), users who cannot express themselves well through such means or do not want to make use of the visual affordances of social media risk being under-represented and marginalized in social media analyses. At the same time, methods for analysing visual data, rather than textual data, are more demanding and less standardized, thus complicating representation issues with such data further. For more text-based social media platforms, such as Reddit, the opposite representation challenge applies. Here, users who have difficulties expressing themselves in textual form might face disadvantages or even stigma, thus preventing them from posting and leading to an over-representation of verbally gifted users.

Finally, unequal representation in digital methods research – both through unbalanced access and use – carries third-level digital divide risks. This is particularly true for applied digital methods research by the platforms themselves, for example the Facebook Core Data Science team (Facebook, n.d.). If certain groups are overlooked in the analyses social media data science teams carry out, platforms might design for the mainstream or popular user rather than already excluded individuals (Shinohara et al., 2018). Non-inclusive design then presents a barrier for potentially interested excluded groups, thus exacerbating power imbalances. Participatory design (Muller & Kuhn, 1993), where non-users or infrequent users are considered in the design process and dedicated research attention is paid to questions of non-use, abstinence, and low-activity use, could help address such divides. However, the issue extends beyond profit-oriented user research and also applies to publicly-funded research. Milan and Treré (2020), for example, describe how the Covid-19 pandemic has further exacerbated digital inequalities, with the Global South being negatively affected in two regards. First, populations at risk (e.g., refugees, sex workers, gig workers, impoverished families, developing countries with limited testing capability) are made even more invisible because they tend to be underrepresented in public monitoring data due to a lack of documentation. This is particularly the case when governments roll out digital solutions such as contact tracing apps. Second, there can be a vicious circle in epistemological terms if governments in the Global South import models and predictions that ignore local conditions and further exclude marginalized groups.

Questions of Privacy

A second implication in relation to the data subjects is about privacy. While privacy and digital inclusion are partly different issues, there is substantial overlap. In other words, privacy questions are often intertwined with aspects of social marginalization, exclusion, and power. Recent literature has shown how those at the margins of society often suffer from disproportionate surveillance (Madden et al., 2017), so that 'privacy is especially difficult for those who are marginalized in other areas of life' (Marwick and boyd, 2018, p. 1158). This includes political activists in authoritarian countries, children, religious and ethnic minorities, gender-queer, those living in poverty or having to rely on social welfare.

Digital methods research can expose such groups and risks violating their privacy expectations. This is the case not only in more closed platforms but also in more open and – seemingly – public-facing platforms such as Twitter, where most users have a public account. Fiesler and Proferes (2018), in a survey of 368 Twitter users, found that 61 percent of the respondents were not aware that tweets are used for research purposes, indicating a widespread lack of knowledge about digital methods research. Moreover, the respondents varied in their acceptance of such research. While about half of them were comfortable about the general use of Twitter data in research, comfort levels were considerably lower if researchers were to use the entire Twitter history of the respondent. Informed consent, transparency about the study purpose, and the amount of data collected emerged as important factors that change the acceptance of Twitter research. The participants mentioned several risk-mitigating strategies such as 'being careful about anonymization, never using real names, and making certain that nothing could link published data back to a Twitter account' (Fiesler & Proferes, 2018, p. 10). Exposure, in the form of being embarrassed by something published about them, emerged as a concern in the open-ended questions in the survey. This can be a particular issue because Twitter, in its Developer Agreement, specifies that tweets should be presented with the username and verbatim. Fiesler and Proferes (2018) did not find statistically significant differences between demographic groups and their acceptance of Twitter research practices but future research should study in more detail whether certain groups feel more reluctant to have their data collected and analysed. This question could be connected to the risks of exposure, which tend to be disproportionally high for certain groups with social media in the first place (Koc-Michalska et al., 2019; Southern & Harmer, 2019; Williams et al., 2017).

In a Canadian large-scale survey, Gruzd et al. (2018) assessed respondents' comfort of having their public social media data accessed by 11 different third parties. The respondents were most comfortable with academic researchers accessing their public social media data (44% uncomfortable) and least comfortable with marketers (66% uncomfortable), political parties, financial institutions and governments (all 65% uncomfortable). Communication networks and photos were seen as the most problematic data types to be accessed, whereas posting frequency and sentiments were perceived as least intrusive. The findings of Gruzd et al. (2018) align with Williams et al. (2017), who conducted a survey of UK-based Twitter users. They found that concern about data use was lower for university research than for government research and commercial research. Moreover, 90% of respondents wanted to remain anonymous in published findings and 80% expected to be asked for consent. Similar findings occurred in a qualitative focus group study by Mikal et al. (2016) on depression monitoring based on Twitter data. Across these studies, the results suggest a contextual understanding of privacy (Nissenbaum, 2009; Vayena et al., 2015).

When it comes to ethical digital methods research and aspects of privacy and data protection, the AoIR Internet Research Ethics Guidelines 3.0 (Association of Internet Researchers, 2020) offer helpful guidance, calling for a bottom-up approach. In terms of inequalities and vulnerability, the guidelines state ‘the greater the vulnerability of our subjects, the greater our responsibility and obligation to protect them from likely harms’. Thus, digital methods researchers should think about their data subjects in terms of exposure risks and conduct a thorough risk assessment of downstream harms in the research process. Particular groups and themes mentioned in the guidelines are minors, groups with special physical or emotional states (e.g., illness, disability, trauma), politically exposed, and women.

Implications for researchers

While the previous sub-section looked at how data subjects are unequally affected by digital methods analyses in terms of representation and privacy, inequalities within the research community also merit attention. Such inequalities have been a lively topic of debate, particularly in the wake of social media API changes after the Cambridge Analytica scandal (Bruns, 2019; Tromble, 2021). For Facebook and Instagram, the API changes meant that research software for collecting public platform data (e.g., pages, groups) had to go through a review and verification process (Tromble, 2021), resulting in widely used apps, such as Netvizz, being discontinued (Rieder, 2018). These API changes have made digital methods research more complicated and have pointed to systemic inequalities within the research community that can be analysed from a digital divide lens. First, academic researchers have differential access to large-scale trace data and the necessary tools to analyse it (first-level). Second, many researchers lack the skills to effectively analyse the data, even if they have access to data (second-level). Third and finally, researchers might benefit differentially from digital methods analyses, even if they have access to data and can efficiently analyse it (third-level).

The first aspect, access to digital trace data, has perhaps been most widely discussed. Andrejevic (2014) introduced the concept of a ‘big data divide’ as the differences between data subjects and data controllers in the form of companies relying on data mining or university researchers with large-scale computational infrastructure to do big data analytics. The data controllers can get unique insights thanks to privileged ‘access to the machines, the databases, and the algorithms’ (p. 1676). Ordinary users, by contrast, lack the access, skills and context to make sense of such data, even when it is their own data. However, within the research community a divide exists between 1) the select few who work for social media data science teams and thus have premium access to data; 2) those who collaborate with digital platforms and have selective but free access (e.g., Social Science One for Facebook or economists who manage to access Uber data through collaboration agreements; Newlands et al., 2019; Scheiber, 2020); 3) those who have the funds to spend on premium API access; and 4) the many researchers who have to rely on increasingly restricted APIs or (often legally murky) scraping (Bruns, 2019). Premium API access, for example for the Twitter Premium APIs (149$ per month to 2499$ per month; Perez, 2017) and Enterprise APIs (2499$+), are prohibitively expensive and thus not a viable option for resource-constrained researchers. Bruns (2019) puts the main blame on the social media platforms but offers a discussion of four ways forward for social media researchers (walk away, lobby for change, accommodate and acquiesce, break the rules). The lobby for change path seems to be the one favoured by him and by many researchers, as indicated by the considerable number of signatories to the initial reaction to Facebook's API change (Bruns, 2018). So far, the lobby efforts seemed to have some success, as Facepager (Jünger, 2019; Jünger & Keyling, 2020) managed to get Facebook API access for public pages at the end of 2019. Puschmann (2019) provides a balanced response to Bruns (2019) and argues for the value of collaborations between researchers and platforms. However, how such collaborations enable participation opportunities beyond the select few, while guaranteeing maximum user privacy, remains an open question.

Less discussed than data access and corresponding inequalities is the divide in skills. Big data analytics require relatively sophisticated skills in statistics, database management, distributed computing, programming, machine learning, data visualization/visual storytelling and context-specific knowledge. While free and open source tools exist to carry out the research effectively (see some of the chapters in this volume), the inequalities arise in the quality of training, mentoring, and capacities for self-development in these areas. High-quality training in data science is either time-consuming (if the researcher is an autodidact) or expensive (if the researcher takes summer schools and specific courses). This privileges research-intensive scholars with technical backgrounds, a flair for numbers, and a high level of capacities to invest in skills development.

Finally, the third-level divide in this context describes how researchers differentially profit from digital methods analyses. Here, we have to keep in mind the publication landscape. High prestige mainstream journals in fields such as sociology, communication, and political science are often conservative and theory-heavy, thus not very favourable towards and well equipped to deal with digital methods analyses. While specialized open access outlets (e.g., *Social Media + Society*, *Big Data & Society*, *Computational Communication Research*, *Sociological Science*) offer a suitable home for such papers, these journals are still quite young and therefore do not count as much in hiring and promotion. In some cases, publication embargoes and non-disclosure agreements also prevent a timely publication of findings based on social media data, thus not allowing the researchers to capitalize on their findings in academic terms. If researchers use the 'break the rules' approach to collect social media data against the terms and conditions of a specific platform (Bruns, 2019), they might get into legal trouble, providing substantial financial and reputational risks (Warden, 2010). Those who lack the economic and social capital to deal with such potential repercussions and are not willing to break the rules might be disadvantaged.

Conclusion

In this chapter, I reviewed the literature on digital inequalities in social media use and discussed the methodological implications for digital methods research. The overview has shown that social media are unevenly used and some of these differences reflect broader power imbalances. At the same time, different social categories intersect and a clearly linear association between vertical inequality and social media use seems too limited. Social media can be a powerful tool for under-privileged citizens but also a locus of vulnerability. It is crucial to differentiate platforms as contexts, as they vary widely in their affordances and norms. This is also an important limitation of this chapter, which has a Western bias. My discussion of methodological implications has pointed to different pressure points of social media-based research and digital inequalities: representation and privacy are key concerns for data subjects, with unequal opportunities and risks, depending on someone's social standing. I then discussed inequalities among researchers when it comes to carrying out digital methods research, touching on points such as resource constraints, platform dependence, skill divides, unequal appreciation of digital methods research, publication embargoes and non-disclosure agreements.

Overall, this chapter has shown how social media reflect existing power imbalances and are embedded within larger systems of structural inequality, including those that pertain to Internet access, Internet uses, digital skills, and outcomes from Internet use (Lutz, 2019). Digital methods researchers should be aware of these inequalities and take an active part in addressing them. Thus, digital methods research has a lot to learn from digital inequalities research.

References

Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, *8*, 1673-1689.

Antelmi, A., Malandrino, D., & Scarano, V. (2019, May). Characterizing the Behavioral Evolution of Twitter Users and The Truth Behind the 90-9-1 Rule. In *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 1035-1038).

Association of Internet Researchers (2020). Internet research: Ethical guidelines 3.0. Association of Internet Researchers. https://aoir.org/reports/ethics3.pdf

Barberá, P., & Rivero, G. (2015). Understanding the political representativeness of Twitter users. *Social Science Computer Review*, *33*(6), 712-729.

Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, *6*(1). https://doi.org/10.1177/2053951718819569

Blank, G. (2013). Who creates content? *Information, Communication & Society*, *16*(4), 590-612.

Blank, G. (2017). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review*, *35*(6), 679-697.

Blank, G. & Groselj, D. (2014). The dimensions of Internet use: Amount, variety and types. *Information, Communication & Society*, *17*(4), 417-435.

Blank, G., & Lutz, C. (2016, July). The social structuration of six major social media platforms in the United Kingdom: Facebook, LinkedIn, Twitter, Instagram, Google+ and Pinterest. In *Proceedings of the 7th 2016 International Conference on Social Media & Society* (pp. 1-10). ACM.

Blank, G., & Lutz, C. (2017). Representativeness of social media in Great Britain: Investigating Facebook, Linkedin, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, *61*(7), 741-756.

Blank, G., & Lutz, C. (2018). Benefits and harms from Internet use: A differentiated analysis of Great Britain. *New Media & Society*, *20*(2), 618-640.

Bourdieu, P., & Wacquant, L. (1992). *An Invitation to Reflexive Sociology*. University of Chicago Press.

Brake, D. (2014). Are we all online content creators now? Web 2.0 and digital divides. *Journal of Computer-Mediated Communication*, *19*(3), 591-609.

Bruns, A. (2018, April). Facebook shuts the gate after the horse has bolted, and hurts real research in the process. *Medium*, 25 April. https://medium.com/@Snurb/facebook-research-data-18662cf2cacb

Bruns, A. (2019). After the 'APIcalypse': social media platforms and their fight against critical scholarly research. *Information, Communication & Society, 22*(11), 1544-1566.

Bruns, A., Highfield, T., & Burgess, J. (2013). The Arab Spring and social media audiences: English and Arabic Twitter users and their networks. *American Behavioral Scientist*, *57*(7), 871-898.

Büchi, M., Festic, N., & Latzer, M. (2018). How social well-being is affected by digital inequalities. *International Journal of Communication, 12*, 3686-3706.

Cook, C., Diamond, R., Hall, J., List, J. A., & Oyer, P. (2018). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers (No. w24732). *National Bureau of Economic Research*. https://www.nber.org/papers/w24732

Correa, T. (2010). The participation divide among 'online experts': Experience, skills and psychological factors as predictors of college students' web content creation. *Journal of Computer-Mediated Communication*, *16*(1), 71-92.

Darwin, H. (2017). Doing gender beyond the binary: A virtual ethnography. *Symbolic Interaction*, *40*(3), 317-334.

Duguay, S., Burgess, J., & Suzor, N. (2020). Queer women's experiences of patchwork platform governance on Tinder, Instagram, and Vine. *Convergence*, *26*(2), 237-252.

Economist (2020, July). India's ban on TikTok deprives the country of a favourite pastime. *The Economist*, 4 July. https://www.economist.com/asia/2020/07/04/indias-ban-on-tiktok-deprives-the-country-of-a-favourite-pastime

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook 'friends:' Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143-1168.

Facebook (n.d.). Advancing the frontiers of data science. About Core Data Science (CDS). *Facebook*. https://research.fb.com/teams/core-data-science/

Ferguson, R. H. (2017). Offline 'stranger' and online lurker: methods for an ethnography of illicit transactions on the darknet. *Qualitative Research, 17*(6), 683-698.

Fick, M., & Dave, P. (2019). Facebook's flood of languages leave it struggling to monitor content. *Reuters*, 23 April 2019. https://www.reuters.com/article/us-facebook-languages-insight/facebooks-flood-of-languages-leave-it-struggling-to-monitor-content-idUSKCN1RZ0DW

Fiesler, C., & Proferes, N. (2018). 'Participant' perceptions of Twitter research ethics. *Social Media + Society, 4*(1). https://doi.org/10.1177/2056305118763366

Foeken, E., & Roberts, S. (2019). Reifying difference: Examining the negotiation of internal diversity on a (post-) lesbian subreddit. *Sexualities*, *22*(7-8), 1268-1287.

Gui, M., & Argentin, G. (2011). Digital skills of internet natives: Different forms of digital literacy in a random sample of northern Italian high school students. *New Media & Society*, *13*(6), 963-980.

Gruzd, A., Jacobson, J., Mai, P., & Dubois, E. (2018). Social media privacy in Canada. *Ryerson University Social Media Lab*. Available at SSRN 3195503. https://doi.org//10.2139/ssrn.3195503

Hargittai, E. (2002). Second-level digital divide: Differences in people's online skills. *First Monday, 7*(4). http://firstmonday.org/article/view/942/864

Hargittai, E. (2007). Whose space? Differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication*, *13*(1), 276-297.

Hargittai, E. (2010). Digital na(t)ives? Variation in Internet skills and uses among members of the 'net generation'. *Sociological Inquiry*, *80*(1), 92-113.

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.

Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults' use of the Internet. *Communication Research*, *3*5(5), 602-621.

Hargittai, E. & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication & Society*, *11*(2), 239-256.

Highfield, T., Harrington, S., & Bruns, A. (2013). Twitter as a technology for audiencing and fandom: The #Eurovision phenomenon. *Information, Communication & Society*, *16*(3), 315-339.

Highfield, T., & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, *2*(1), 47-62.

Hoffmann, C. P., Lutz, C., & Meckel, M. (2015). Content creation on the Internet: A social cognitive perspective on the participation divide. *Information, Communication & Society*, *18*(6), 696-716.

Humphreys, L. (2013). Mobile social media: Future challenges and opportunities. *Mobile Media & Communication*, *1*(1), 20-25.

ITU (2019a). Measuring digital development: Facts and figures. *International Telecommunications Union (ITU)*, Geneva. https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf

ITU (2019b). Statistics: Percentage of individuals using the Internet [Excel sheet]. *International Telecommunications Union*, Geneva. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

Jünger, J. (2019, December 17). #Facepager has been granted public pages content access by FB [Tweet]. https://twitter.com/jajuenger/status/1207024036032786432

Jünger, J., & Keyling, T. (2020). Facepager. An application for generic data retrieval through APIs. Source code and releases available at https://github.com/strohne/Facepager/

Koc-Michalska, K., Schiffrin, A., Lopez, A., Boulianne, S., & Bimber, B. (2019). From online political posting to mansplaining: The gender gap and social media in political discussion. *Social Science Computer Review*, online first. https://doi.org/10.1177/0894439319870259

Kvasny, L. (2006). Cultural (re) production of digital inequality in a US community technology initiative. *Information, Communication & Society, 9*(2), 160-181.

Livingstone, S. (2014). Developing social media literacy: How children learn to interpret risky opportunities on social network sites. *Communications, 39(*3), 283-303.

Lupáč, P., & Sládek, J. (2008). The deepening of the digital divide in the Czech Republic. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 2*(1). https://cyberpsychology.eu/article/view/4210/3251

Lutz, C. (2016). A social milieu approach to the online participation divides in Germany. *Social Media + Society, 2*(1). https://journals.sagepub.com/doi/abs/10.1177/2056305115626749

Lutz, C. (2019). Digital inequalities in the age of artificial intelligence and big data. *Human Behavior and Emerging Technologies, 1*(2), 141-148.

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881-900.

Madden, M., Gilman, M., Levy, K., & Marwick, A. (2017). Privacy, poverty, and Big Data: A matrix of vulnerabilities for poor Americans. *Washington University Law Review*, *95*, 53-126.

Marler, W. (2018). Connective ambition and creative caution among homeless users of Facebook. *AoIR Selected Papers of Internet Research* (SPIR). https://www.spir.aoir.org/ojs/index.php/spir/article/view/10493

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science, 159*(3810), 56-63.

Micheli, M., Lutz, C., & Büchi, M. (2018). Digital footprints: an emerging dimension of digital inequality. *Journal of Information, Communication and Ethics in Society, 16*(3), 242-251.

Mikal, J., Hurst, S., & Conway, M. (2016). Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. *BMC Medical Ethics*, 17(1), 22.

Milan, S., & Treré, E. (2020). The rise of the data poor: The COVID-19 pandemic seen from the margins. *Social Media + Society, 6*(3), 1-5.

Muller, M. J., & Kuhn, S. (1993). Participatory design. *Communications of the ACM, 36*(6), 24-28.

Nabi, Z. (2014). ~~Resistance~~ censorship is futile. *First Monday, 19*(11). https://doi.org/10.5210/fm.v19i11.5525

Napoli, P. M., & Obar, J. A. (2014). The emerging mobile Internet underclass: A critique of mobile Internet access. *The Information Society, 30*(5), 323-334.

Newlands, G., Lutz, C., & Fieseler, C. (2019). Trading on the Unknown: Scenarios for the Future Value of Data. *Law & Ethics of Human Rights, 13*(1), 97-114.

Newlands, G., & Lutz, C. (2020). Crowdwork and the mobile underclass: Barriers to participation in India and the United States. *New Media & Society*, online first. https://doi.org/10.1177/1461444820901847

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press.

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*, 173-182.

Pearce, K. E., & Rice, R. E. (2017). Somewhat separate and unequal: Digital divides, social networking sites, and capital-enhancing activities. *Social Media + Society*, *3*(2), 1-16.

Perez, S. (2017, November). Twitter launches lower-cost subscription access to its data through new Premium APIs. *Tech Crunch*, 14 November. https://techcrunch.com/2017/11/14/twitter-launches-lower-cost-subscription-access-to-its-data-through-new-premium-apis/

Perrin, A., & Anderson, M. (2019). Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. *Pew Research Center*, April 10 2019. https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

Pew (2019a). Internet/Broadband fact sheet. *Pew Research Center: Internet & Technology*, 12 June 2019. https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Pew (2019b). Social media fact sheet. *Pew Research Center: Internet & Technology*, 12 June 2019. https://www.pewresearch.org/internet/fact-sheet/social-media/

Puschmann, C. (2019). An end to the wild west of social media research: A response to Axel Bruns. *Information, Communication & Society*, *22*(11), 1582-1589.

Rieder, B. (2018). Facebook's app review and how independent research just got a lot harder. *The Politics of Systems: Thoughts on Software, Power and Digital Method*.

http://thepoliticsofsystems.net/2018/08/facebooks-app-review-and-how-independent-research-just-got-a-lot-harder/

Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algorithms to 'ranking cultures' Investigating the modulation of visibility in YouTube search results. *Convergence*, *24*(1), 50-68.

Robinson, L. (2009). A taste for the necessary: A Bourdieuian approach to digital inequality. *Information, Communication & Society*, *12*(4), 488-507.

Robinson, L., Cotten, S., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., ... & Stern, M. (2015). Digital inequalities and why they matter. *Information, Communication & Society*, *18*(5), 569-582.

Robinson, L., Schulz, J., Dunn, H., Casilli, A., Tubaro, P., Carveth, R., ... & Ball, C. (2020). Digital inequalities 3.0: Emergent inequalities in the information age. *First Monday*, *25*(7). https://journals.uic.edu/ojs/index.php/fm/article/view/10844/9562

Rogers, R. (2013). *Digital methods*. The MIT Press.

Russmann, U., & Svensson, J. (2017). Introduction to visual communication in the age of social media: Conceptual, theoretical and methodological challenges. *Media & Communication*, *5*(4), 1-5.

Scheiber, N. (2020, July). When scholars collaborate with tech companies, how reliable are the findings. *New York Times*, 14 July. https://www.nytimes.com/2020/07/12/business/economy/uber-lyft-drivers-wages.html

Schradie, J. (2011). The digital production gap: The digital divide and web 2.0 collide. *Poetics*, *39*(2), 145-168.

Shinohara, K., Bennett, C. L., Pratt, W., & Wobbrock, J. O. (2018). Tenets for social accessibility: Towards humanizing disabled people in design. *ACM Transactions on Accessible Computing (TACCESS)*, *11*(1), 1-31.

Shukla, P. (2020, July). How TikTok's 'Cringe' Empowered Indians More Than FB or Insta. *The Quint*, 1 July. https://www.thequint.com/neon/hot-take/tik-tok-ban-empowered-indians-more-than-facebook-instagram

Sims, C. (2014). From differentiated use to differentiating practices: Negotiating legitimate participation and the production of privileged identities. *Information, Communication & Society*, *17*(6), 670-682.

Sloan, L., Jessop, C., Al Baghal, T., & Williams, M. (2020). Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving. *Journal of Empirical Research on Human Research Ethics*, *15*(1-2), 63-76.

Southern, R., & Harmer, E. (2019). Twitter, incivility and 'everyday' gendered othering: An analysis of tweets sent to UK Members of Parliament. *Social Science Computer Review*, online first. https://doi.org/10.1177/0894439319865519

Statista (2020, January). Active social network penetration in selected countries as of January 2020. *Statista*. https://www.statista.com/statistics/282846/regular-social-networking-usage-penetration-worldwide-by-country/

Tichenor, P., Donohue, G., & Olien, C. (1970). Mass media flow and differential growth in knowledge. *Public Opinion Quarterly*, *34*(2), 159-170.

Tromble, R. (2021). Where have all the data gone? A critical reflection on academic digital research in the post-API age. *Social Media + Society*, *7*(1), 1-8.

Tsetsi, E., & Rains, S. A. (2017). Smartphone Internet access and use: Extending the digital divide and usage gap. *Mobile Media & Communication*, *5*(3), 239-255.

Twitter (n.d.). Supported languages and browsers. *Twitter Developer*. https://developer.twitter.com/en/docs/twitter-for-websites/twitter-for-websites-supported-languages/overview

Utz, S. (2016). Is LinkedIn making you more successful? The informational benefits derived from public social media. *New Media & Society*, *18*(11), 2685-2702.

Van Deursen, A., & Helsper, E. (2015). The third-level digital divide: Who benefits most from being online? In *Communication and Information Technologies Annual* (pp. 29-52). Emerald.

Van Deursen, A. J., Helsper, E. J., & Eynon, R. (2016). Development and validation of the Internet Skills Scale (ISS). *Information, Communication & Society, 19*(6), 804-823.

Van Deursen, A., & Van Dijk, J. (2011). Internet skills and the digital divide. *New Media & Society, 13*(6), 893-911.

Van Deursen, A., & Van Dijk, J. (2014). The digital divide shifts to differences in usage. *New Media & Society, 16*(3), 507-526.

Van Deursen, A., & Van Dijk, J. (2019). The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New Media & Society, 21*(2), 354-375.

Van Dijk, J. (2005). *The deepening divide: Inequality in the information society*. Sage.

Van Dijk, J. (2006). Digital divide research, achievements and shortcomings. *Poetics, 34*(4-5), 221-235.

Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of big data in public health. *PLoS Computational Biology, 11*(2), e1003904.

Warden, P. (2010, April). How I got sued by Facebook. *Pete Warden's Blog*, 5 April. https://petewarden.com/2010/04/05/how-i-got-sued-by-facebook/

Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology, 51*(6), 1149-1168.

Zillien, N. & Hargittai, E. (2009). Digital distinction? Status-specific types of Internet usage. *Social Science Quarterly, 90*(2), 274-291.

Author biography

Christoph Lutz is an associate professor at the Nordic Centre for Internet & Society and at the Department of Communication and Culture, BI Norwegian Business School (Oslo). With a PhD in management from the University of St.Gallen, Christoph's research interests cover a broad spectrum of topics and lie in the field of social media and Internet-mediated communication. More specifically, he investigates digital inequalities, privacy, the sharing economy, new forms of work, and social robots. Christoph has published widely in top-tier journals in this area such as *New Media & Society*, *Mobile Media & Communication*, *Information, Communication & Society*, *Social Media + Society*, and *Big Data & Society*.