

Validity, reliability and the case for participant-centred research: Reflections on a multi-platform social media study

Katy Jordan

Institute of Educational Technology, The Open University, UK

Abstract

Validity and reliability are long established as central tenets of ensuring good quality research. However, safeguarding validity and reliability can be challenging within the context of multi-platform social media research, as considerations such as access to data or self-selection place caveats upon the generalisability of any study and are exacerbated when multiple platforms are involved. In this paper, I will draw upon examples from a recent empirical study and others to illustrate how these issues can become amplified within the context of multi-platform social media studies. These include: reconsidering the concept of a representative sample across multiple platforms; issues balancing ethical considerations and terms of service in accessing data; and the effect of data repackaging strategies and extent to which similar data from different platforms can be comparable. The paper will conclude by arguing that drawing upon qualitative approaches, particularly a complimentary reframing of the unit of analysis with a focus on particular users as case studies, can ameliorate these issues and strengthen the insight of multi-platform social media studies.

Introduction

Research on social media is faced with a fundamental dilemma; whether to orient its focus upon the platforms, or the people who use them. Social media as a phenomenon is tied to a range of online technologies, but without the users the technologies would not be social. The solution is frequently to define the bounds of empirical studies by focusing on a particular platform, although this presents only a single view on a research topic, when in reality the one platform is only a single component of a social media ecosystem which individual users may have (Hall et al., 2016). While some progress has been made in tracking key influencers across multiple high-profile social media platforms (using NodeXL, for example), this functionality is yet to be extended to all platforms and levels of use.

Multi-platform social media studies acknowledge these limitations and present an alternative stance. However, while multi-platform approaches to social media have great potential for added insight and holistic understanding of a research topic, the added complexity also brings challenges for their practical execution and research integrity. The complexity of issues associated with research on a single platform can become multiplied when studies are expanded to include several platforms, although multi-platform studies may be a more accurate object of study; users are multi-platform in their use of social media. As an emergent and developing part of the field, methodological and theoretical challenges associated with multi-platform social media studies are an active area for research (Hall et al., 2016), which can be illuminated through the experiences of researchers undertaking studies of this nature.

This paper will take the experiences of undertaking a recent research project which used a mixed methods social network analysis approach to understand the structure and role of academics ego-networks on contrasting social media platforms as a starting point (Jordan, 2017). The discussion will draw out issues which were encountered during the project and have implications for safeguarding the accuracy of any multi-platform social media study.

Concepts of validity and reliability were used as a framework for thinking about accuracy in the research project. This helped to surface issues which may be emphasised in multi-platform social media studies, although also highlighting ways in which the concepts themselves may be deficient in this context. Validity addresses whether the data collected accurately reflect the phenomena under inquiry, while reliability focuses upon the reproducibility of the data produced by the research instruments involved (Hammersley, 1987). However, inconsistency is common in the use of both terms (Hammersley, 1987; Winter, 2000), and Golafshani (2003) argues that validity and reliability are concepts born of quantitative methodologies and positivist epistemology. Both terms are linked to an assumption that research should strive for generalisability, a concept which is also born out of a quantitative, post-positivistic paradigm (Maxwell, 1998). In the context of an interpretivist, qualitative-based paradigm, generalisation is not necessarily sought. Rather, the transferability of research (Guba & Lincoln, 1989) is preferable for qualitative research studies which are more concerned with the detail of the research setting in question.

By examining a multi-platform social media study in these terms, issues will be highlighted which are potentially problematic as the number of platforms involved is increased. The argument will be made for considering incorporating qualitative methods and analysis alongside platform-focused data, and a reframing to foreground individual users rather than the platform for some multi-platform studies. The empirical research context that this paper is grounded in is a recent project which has focused upon the professional use of social networking sites (SNS) by academics, which will be introduced in the next section. From this point onwards, the project will be referred to as the 'academic networking' project. The paper will be

structured around three themes for discussion drawn out from the experience of undertaking this project, and links made to similar issues in other multi-platform social media studies. The three themes include: ethical and legal frameworks; the need to carefully consider representative sampling when more than one platform is involved in a study; and the extent to which data derived from different platforms can be considered equivalent. Ethical and legal frameworks may place caveats on the extent to which similar research activities can be carried out across multiple platforms; legally, the terms of service differ, and ethically, the social norms associated with how users and communities conceptualise different platforms may mean that what would be acceptable on one site may be inappropriate for another. While ensuring use of representative samples is a longstanding mechanism to ensure validity, multi-platform studies call for a reconsideration of how representative samples are defined as an average user of one platform may be an outlier in terms of a different platform. As a result, a representative sample derived from the population of one platform cannot be assumed to be representative of a second platform. Finally, caution may be required in the extent to which data accessed from different platforms is comparable, which may have implications for the reliability of a study. The availability of APIs to access data may clarify some of the legal issues related to a platform, but it is subject to non-transparent processes of data collection and export.

Research context

Academic SNS are an interesting but under-studied type of social media platform which has developed in recent years, as they are platforms which explicitly seek to bring the benefits of online social networking to an academic community and audience (Nentwich & König, 2012). A fundamental characteristic of SNS (distinct from other forms of social media) is the ability to create connections to others, and traverse the resulting network (Ellison & boyd, 2013). At present, the two most widely used platforms aimed specifically at an academic audience are Academia.edu and ResearchGate (Van Noorden, 2014). In addition to the standard features of being able to create a profile and make connections to others, academic SNS also allow the upload and sharing of papers and other scholarly outputs, also serving as an open access publishing platform to an extent. The study sought to address whether the social network fostered by these platforms opens up new opportunities or simply replicate existing academic structures and hierarchies. While the network structure is a fundamental characteristic of the platforms, it will have implications for the types of interactions the platforms support, and few studies have examined it so far (Hoffmann, Lutz, & Meckel, 2015; Jordan, 2014).

Academic SNS sit within a broader range of both specialist and generic online tools and social media which form the ecology of tools within which digital scholarly

practices may be enacted (Cann, Dimitriou & Hooley, 2011; Weller, 2011). Existing studies have frequently used survey approaches to focus upon academics' use of single-platforms or encompassing all social media tools as a whole (Jordan, 2017). The academic networking project sought to examine in detail the structure of academics' ego-networks on two contrasting social media platforms which they use in relation to their professional work, and to understand the significance of those structures through interviews with participants. Academic SNS and Twitter were selected as contrasting platforms because while academic SNS are designed for a specifically academic audience, academics often view them as an online CV, while more actively discussing their work on Twitter (Bukvova, 2012; Van Noorden, 2014).

The academic networking study used a mixed methods social network analysis approach (Dominguez & Hollstein, 2014). Initially, an online survey was used, to gain a baseline of academics' perceptions and levels of use of SNS; 528 responses to the survey were achieved. A sample of 55 participants for network analysis were selected from this pool using a purposive sampling approach (Teddlie & Yu, 2007), stratified to include four job positions and three disciplinary areas. For each participant, two ego-networks were collected; either Academia.edu or ResearchGate (as an academic SNS, depending on which site they primarily use), and Twitter. The pair of networks sampled and visualized for one participant is shown in Figure 1 as an example. The networks were visualized and analysed using Gephi (Bastian, Heymann, & Jacomy, 2009); interactive versions of the networks were created with the sigma.js plugin (Hale, 2012) and shared with participants. Co-interpretive interviews were then held via Skype with a sub-sample of 18 participants, to gain insight into the meaning of the network structures and how they were created from the participants' point-of-view (Molina, Maya-Jariego, & McCarty, 2014).

--Figure 1 here--

The network analysis phase revealed that academics ego-networks on academic SNS networks were smaller and more highly clustered; Twitter networks were larger and more diffuse. Communities within networks are more frequently defined by institutions and research interests on academic SNS, compared to research topics and personal interests on Twitter. From analysis of the co-interpretive interviews, emerging themes link network structure to differences in how academics conceptualise and use the sites. Academic SNS are regarded as a more formal academic identity, akin to a business card, or used as a personal repository. Twitter is viewed as a space where personal and professional are mixed, similar to a conference coffee break. Academic SNS replicate existing professional connections, Twitter reinforces existing professional relationships and fosters novel connections. Several strategies underpinning academics' use of the sites were identified,

including: circumventing institutional constraints; extending academic space; finding a niche; promotion and impact; and academic freedom. The themes also provide a bridge between academic identity development online and formal academic identity and institutional roles.

Converging and contrasting ethical and legal frameworks

At the stage of planning a multi-platform social media study, legal factors can have a significant influence upon decisions about the platforms that can or cannot be included, and this can substantially restrict or skew the object of study and threaten the validity and reliability of the research project from the outset. In the context of the academic networking study, I shall describe how by involving participants directly in the research process, consent provides an ethically sound way of conducting research across platforms, as participants have the legal right to access and use their personal data.

Ethically, the 'public sphere' argument was prevalent during the early years of social media research. That is, information that can be viewed by anyone online can be considered freely available for researchers to use. However, as boyd & Crawford note, "just because it is accessible doesn't make it ethical" (boyd & Crawford, 2011, p.10); for example, Facebook users probably do not consider researchers as an audience when they are creating profiles and posts likely intended to stay between friends (van Gilder Cooke, 2011). In practice, the distinction between public and private in online settings is messy (boyd, 2010), and it is important to consider potential harm not only to the individual but the online communities they are embedded within (Krotoski, 2010).

Paying particular attention to each context and what is appropriate in that particular setting frequently emerges as the solution to such dilemmas (Eynon, Fry & Schroeder, 2008; Markham & Buchanan, 2012). However, this solution is also potentially problematic for multi-platform studies, as each platform involved may have different cultures and norms which would challenge what is considered to be appropriate in each context. Sites which may appear similar in many ways (e.g. being defined as SNS) may be perceived by participants in very different ways, bringing with them different sets of ethical issues and social norms.

This was evident in the academic networking project in two ways. Initially the intention had been to include Facebook in the study as an additional platform. However during the pilot phase, it became evident that participants did not consider it to be part of their professional practice, seeing it as private, and not wishing for data to be collected from the site. Second, the co-interpretive interviews revealed the contrasts between how the academics conceptualise each site. The metaphor of a 'coffee break' encapsulated the view of Twitter as an informal social space, whereas

academic SNS were viewed as reflecting formal academic identities, and conceived of as a type of CV or 'personal repository'. From the interviews, LinkedIn was frequently mentioned as fulfilling a similar role albeit not a specifically academic one. With hindsight, it may have been an interesting contrast to include in the network analysis phase, although the platform does not make access to network data easy following the retirement of their InMaps function (Lunden, 2014). In multi-platform studies being aware of context and social norms therefore presents a challenge as these may differ for each of the platforms included in the study. In their study of social media used during Super Bowl XLIX (including Instagram, Tumblr and Twitter), Mukherjee and Jansen (2016) noted differences in formality across the three platforms and also a temporal element, so this is an issue which can also vary over time.

To an extent, the platform is another entity concerned with preventing harm to the community; website terms of service (ToS) are a legal framework which may include stipulations to prevent harm to users, by way of protecting users from spam or commercial exploitation. However, ToS are not constructed primarily from the viewpoint of the community but rather the company behind the platform, derived from legal standards which are often unrelated to community norms and expectations (Fiesler, Lampe & Bruckman, 2016). Through restrictions placed upon data collection, ToS can introduce bias into sampling and the data collected (Fisher, McDonald, Brooks & Churchill, 2010). As a researcher, navigating the ToS for a single social media platform can be challenging enough, but it is an issue which is multiplied as further platforms are involved. Whether the ToS should be viewed as an absolute barrier to research is a key topic for debate at present (Sandvig & Karahalios, 2016). In their work on academic freedom, the American Association of University Professors (AAUP) emphasise the key principles of "maximizing access to information and protecting user privacy" and note that social media users are vulnerable to changes in ToS (AAUP, 2013). As Weller and Kinder-Kurlanda (2016) note, whether it is acceptable to share data collected from social media platforms with other researchers is a further minefield, despite having great potential for increasing validity of studies through independent analysis and verification.

In the case of the academic networking study, each of the platforms involved in online data collection adopted different stances in their ToS. Data collection from Twitter was the most straightforward in this respect, as the only platform in the sample to have an API. In terms of the academic SNS involved, at the time of planning the study, Academia.edu did not explicitly prohibit data collection for academic purposes, so web scraping software was used. At that point ResearchGate had very restrictive ToS in relation to data collection, which prevented any automated data collection and implied that manual data collection may contravene the terms. However, a precedent had been set of permitting use if participants' consent has been gained (Hoffman, Lutz, & Meckel, 2014), so data in this case was collected manually by the researcher.

To offset the variation in legal stances presented by the platforms, the ethical stance adopted for the academic networking project did not use the public square argument, but sought informed consent from potential participants before their networks were collected. This was a pragmatic decision based on several factors. The research questions for the project required insight and understanding of the network structures from the participants' own viewpoints, so it was necessary for the participants to be willing to take part. As a mixed methods project, different levels of consent and identity protection are expected for the execution of different methods. The interviews were the critical method in this respect, with the greatest need for identity protection and active consent and participation from interviewees. Given the differences and lack of clarity in the websites ToS, the legal frameworks did not provide a workable solution, so the most conservative ethical stance was adopted instead, of full informed consent and identity protection throughout all parts of the research. Only the researcher and participants in co-interpretive interviews were aware of the identities of the nodes in the social networks sampled from all the platforms.

"I really love this; this was totally fascinating to me. It totally reflects my experience of Twitter." – one of the academic networking study participants on seeing their Twitter ego-network visualisation.

The network data collection and co-interpretive interviews raise the question of whether connections between profiles can be considered to be personal data (see above quote). Whether or not social network data counts as personal data has implications for data protection and the legal rights of the individuals concerned (Data Protection Act, 1998) to access and use the information. Several of the interview participants, particularly with reference to their Twitter networks, remarked about how much they had enjoyed being able to view their networks as visualisations, and great emotional attachment to their online connections. However, in the case of the academic SNS particularly, it was not clear whether collecting and visualising the networks was in contravention of their ToS. The personal data argument is being tested at the moment through freedom of information requests to platforms such as Cambridge Analytica (Dehay, 2017). Cambridge Analytica and other proprietary mobile Big Data platforms are reported to be far advanced in terms of their tracking abilities in comparison to the academic research sector; however, this is ethically questionable and the opacity of their operations is promoting debate around consent, propaganda and surveillance in social media. Whether network data is personal data or otherwise remains a question for debate. The same could be said of any user-generated data hosted on social media platforms, and in this sense, gaining informed consent and involving participants in the research process is an ethically favourable and possibly legally defensible stance for a research project.

Redefining representative samples

Creating a representative sample is a key way of safeguarding validity and presenting an accurately reflection of the phenomenon. Multi-platform studies pose an immediate challenge for sampling, in that a sample that is representative of the population of one platform may not be representative of another. This does not invalidate studies but requires careful consideration of the research context and framing the results clearly so that the extent to which results are generalizable is made explicit.

This issue was evident in the academic networking project through the example of differences in the degree distributions of the participants in the study, across the two types of platform involved. Although information about the recruitment survey and URL was posted by the researcher on all three sites, the information circulated to a greater extent on Twitter. This greater level of recruitment via Twitter is reflected in the sample, in that it is likely that the Twitter users in the sample are not representative of Twitter academics as a whole, but is biased towards more active users. This is reflected in the degree distributions shown in Figures 2 and 3.

--Figures 2 & 3 here--

We would expect degree distributions to follow a power law or heavy-tailed distribution, which has been established as a fundamental characteristic of social networks (Barabasi, 2002). The degree distribution for the academic SNS networks (Figure 2) appears to show a heavy tailed degree distribution more clearly than the Twitter data (Figure 3). The Twitter data suggests that the distribution may be truncated; that is, the expected high proportion of low degree networks is under-represented. This is likely due to the way in which the initial survey (by which participants for further network analysis were recruited) had circulated through social media. Whilst the information was posted on Twitter, Academia.edu, ResearchGate and LinkedIn, it received greater attention through Twitter. The self-selection bias that is inherent in such approaches may also favour the more active Twitter users in this case. However, the pool of respondents was large enough that purposive sampling could be applied to create a stratified sample for network analyses. Whilst this sub-sample could not be regarded as absolutely representative, it did ensure that a range of different perspectives were represented.

The issue of sampling bias is also highlighted by Spiliotopoulos and Oakley, in their 2016 study of a sample of users paired accounts on Facebook and Twitter. In their case, “a disproportionately large number of Indian participants were recruited due to

the auction-like mechanics of Facebook ads” (Spiliotopoulos & Oakley, 2016, p.4). However, they point out that if handled correctly an understanding of such biases could be used advantageously to create stratified samples. Philips Honda (2015) employed a similar purposive sampling strategy in their YouTube-focused multi-platform social media study. In order to understand motivations for users who participated in the YouTube-based ‘It Gets Better’ project, a recruitment survey was first posted on social media sites, and interviewees sampled within this to include a range of perspectives.

The example here also raises the issue that it may not be possible to construct samples which are simultaneously representative of more than one platform, if behaviour on one platform is not correlated with that on the other. Despite the differences in distribution of degree across both platforms in the academic networking project (Figure 4), degree on academic SNS was significantly correlated with degree on Twitter (Spearman’s coefficient of rank correlation $r_s=.45$, $p<.05$), although the scatterplot suggests that other factors may also be at play and the correlation is not clear cut. Other metrics which showed significant correlations included in-degree, out-degree, and network density, while betweenness centrality, the number of communities in networks and reciprocity did not.

--Figure 4 here--

This issue also resonates with reflections from Spiliotopoulos and Oakley (2016). A more extensive range of potential correlations between Facebook and Twitter metrics were examined. Approximately 20% of the pairs tested were significant at the 0.5 level. This issue would be exacerbated if more platforms were included in the study. The authors also note that “none of the six “second-level” twitter metrics, that represent the activities of one’s followers and friends, were found to be significantly associated with Facebook activity” (Spiliotopoulos & Oakley, 2016, p.5). Factor analysis is suggested as a potential way forward to approach this (ibid.).

Depending upon the phenomenon being studied, there are two useful messages which emerge here for multi-platform social media studies. First, the importance of situating any sample in relation to broader trends in social media data. For example, knowing that degree distributions follow heavy-tailed distributions and observing the Twitter data in the sample here did not invalidate the results but revealed an important caveat to the limitations and generalisability of the findings. Second, that it may not be possible to construct a sample that is simultaneously representative of all the platforms involved. This problem also highlights the value of reframing studies from the perspective of particular individual users as case studies.

Different platforms, equivalent data?

The third issue for discussion here is a corollary of the first issue (ethical and legal considerations). Differing policies on access to data from different social media platforms may lead to implicit biases within seemingly comparable data.

For each of the platforms included in the academic networking study, different approaches were required to access network data. In order to construct and analyse the networks, data for each network was ultimately brought together in the same format (CSV data, as a series of 'source' and 'target' nodes). However, different approaches to data collection could affect the accuracy and comparability of data. It is a combination of social and technical factors that necessitated these different approaches and makes the issue hard to resolve with certainty. The approaches selected for data collection on each site took into account several factors, including: technical access to data; terms of service of the sites; and ethical considerations and participant consent. No networks were collected without the consent of the participant whose ego-network was involved; this was gained via the recruitment survey, which included links to exemplar networks.

The API represented the only feasible method of acquiring network data from Twitter; NodeXL was used to run queries and collect data from the API (Smith et al., 2009). However, the Twitter API places certain restrictions on the amount and frequency of data collection and is a 'black box' in a sense; only the first 2,000 followers or following can be collected. This immediately excluded a number of participants from network analysis and more complex metrics beyond simply in-degree and out-degree. For all others, there remained an element of uncertainty regarding how the API would treat highly-connected nodes within an ego-network. NodeXL gives the option to only collect connections between a specified list of IDs, and collects follower and following data for each ID, so any connections which would be valid in the context of an ego-network would have two chances to be collected, but it is impossible to accurately account for the extent of data loss through the API and its effect on the network. While the academic networking project focused upon follower/following relationships, the Twitter API places further restrictions upon collection of tweet-based data. Data collection is restricted to the most recent 18,000 tweets and approximately the past seven days (Hawksey, 2017). Larger potential samples are more prone to inaccuracy (González-Bailón, Wang, Rivero, & Borge-Holthoefer, 2014), although it is not possible at present to quantify the extent of data loss.

As no API is available for Academia.edu or ResearchGate, different approaches were used. Generic web-scraping software (Mozenda) was used to collect data from Academia.edu. Due to their ToS (discussed earlier in the section on ethical and legal frameworks), ResearchGate data were collected manually by the researcher, with

the aid of a Chrome browser extension (called 'Scraper') which enables data in the users' browser to be stored in a spreadsheet.

While the ego-networks were collected from different sites by different methods, each converged into a common format of CSV files. These files were ostensibly comparable – each only consisted of two columns, a 'source' and 'target' node – and whilst every care had been taken when collecting data, there remained a small and inquantifiable element of doubt that every edge had been collected. The third phase of the project, which involved co-interpretation of network structures via interactive visualisations, provided a way of checking the validity of the networks to an extent.

The element of doubt was more likely in the Twitter data. Whilst the scraper-based data collection from the academic SNS was still subject to human error or errors in the software, data were collected directly from the sites themselves. The availability of an API for Twitter brought its own benefits and constraints. It was convenient in that data collection could be easily configured and was automated to a greater extent, although rate limiting meant that some queries took several days to complete. In contrast to the academic SNS, were not harvested from the profiles in a direct sense, but processed and re-released through the API, in a process which is not transparent. With this hidden process, decisions which seemed innocuous or trivial to the developers may have unknown consequences for researchers using the resulting data. The limit on the number of followers returned is such an example. This is one example of a problem which Halford et al. (2016) describe as 'the social media pipeline' (Halford, Weal, Tinati, Carr & Pope, 2016):

"To investigate this we explore the 'pipeline' of social data production and circulation: from the user who creates the content, posting to a social media platform, to the client software on the phone, laptop, etc. that represents the data (sometimes in different ways, if there are multiple clients available for a given platform), the to the Application Programming Interface(s) (APIs), which enforces rules to determine what is passed through to the company's server software, and how, and the server software that organizes content into data bases that store data in particular formats and structures. [...] In turn, all this shapes if and how these data are circulated for re-use, back down the pipeline. This 'output' is not a simple reversal of the 'input' and is shaped by the methods that researchers use to access data, the economics and practicalities for the companies in sharing data, with whom and on what basis, both shaped by legal and sometimes even ethical considerations." (Halford et al., 2016, p.2).

The authors suggest three levels – the population, the sample, and the method by which data are produced - as a starting point for a framework toward investigating and standardising an understanding of the social media pipeline.

Conclusions

While the examples here illustrate some of the potential pitfalls of multi-platform social media studies, using a multi-platform approach also offered benefits in terms of greater insight. Taking a multi-platform approach in the academic network helped to throw the role played by academic SNS and Twitter into sharp contrast, highlighting how necessary multi-platform work is. While academic SNS networks were smaller and highly clustered, Twitter networks were larger and more diffuse in structure (Jordan, 2017). By discussing the networks with participants, the differences in network structure were found to be related to the very different ways in which academics conceptualise the role of the two types of platform and use them in their professional work (ibid). The basic measure of connection for all the platforms in the study – the follower-following relationship – held nuanced differences in meaning for participants on academic SNS in contrast to Twitter. This underscores the importance of multi-platform studies, as sites which may be technically very similar can be playing very different roles in practice and highlights the critical role for qualitative work to accompany social media data in order to account for context.

Context is a more subjective research object than more easily quantifiable interactions with a single site and its specific data structure. Although multi-platform social media studies are not yet common, mixed methods approaches are emerging as the methodological choice (Hall et al., 2016; Hughes, Starbird, Leavitt, Keegan & Semaan, 2016; Spiliotopoulos & Oakley, 2016). Although social media studies enable 'big data' approaches to research, the need for careful consideration and understanding of context in multi-platform studies can be addressed by incorporating qualitative research activities into research designs. As such, a philosophical turn is also required away from big data as mixed methods designs bring with them different and contrasting research paradigms and attitudes towards validity and reliability.

In the case of the academic networking project described here, a number of safeguards to validity and reliability drawn from a qualitative perspective were used. The main mechanism for ensuring validity was a process of triangulation; that is, using different methods to verify the same phenomenon (Maxwell, 1998; Miles & Huberman, 1994). Triangulation is a key affordance of mixed methods as a methodological approach. However, by including participants in the research process, validity was further enhanced. A wide variety of strategies can be drawn upon to incorporate participatory approaches into research designs, from lighter touch approaches such as using participants as informants or verifying data through member checks, to participants acting as full research partners in approaches such as action research. Careful consideration is required to the goals of the research study and relationship with the community in selecting the appropriate approach (Maxwell, 1998). In the case of the academic networking project, co-interpretive interviews were selected as a way of gaining personal insight from participants into the network analysis phase of research. Individual participants' networks and

perspectives were then written up as case studies. The interviews served to gain insight into participants' views on their network data; while trends were identified by the researcher, the meaning was discovered through interviews. This process built in validity checks through confirmation with participants (Gray, 2009), member checks, and gaining rich data about a sub-sample of participants (Maxwell, 1998).

Data collected through participant-focused approaches to multi-platform social media studies will require different analytical approaches. There is a rich body of qualitative analysis techniques and strategies which can be drawn upon. In the case of the academic networking project, the analysis was undertaken using a grounded theory approach, so as not to place assumptions on the analytical frame but rather to allow the coding to emerge from the participants accounts (Charmaz, 2014; Glaser & Strauss, 1967). Again, a wide range of approaches and strategies for analysis may be applied to participant-centred data and depends upon the focus of the study at hand; for a practical review of strategies for drawing meaning from qualitative data, see chapter 10 in Miles and Huberman (1994).

The tension between detailed understanding and generalisability is a pervasive limitation of a case study-based approach. From undertaking the research project described here, there are two key advantages in relation to multi-platform social media studies. First, that taking a multi-platform approach adds extra depth to a project beyond the sum of its parts. For example, by looking at both academic SNS and Twitter networks with participants, insight was gained beyond simply interpretation of each individual platform. Discussing two networks with participants drew out distinctions and emphasised contrasts between the two. Furthermore, discussing two contrasting platforms also facilitated further discussion and insight about the participants' broader personal social media ecosystem and how other platforms which had not been sampled sit in relation to those under discussion.

Second, multi-platform studies can potentially locate individual cases in the broader social media landscape using metrics. For example, by comparing the degree distributions of the network samples to a broader, now well established phenomenon (heavy tailed distributions), a limitation of the dataset was surfaced and made explicit. This is one example of how trends from the existing body of single-platform, big data-style research on social media platforms could bridge a move towards multi-platform studies. This will not bring generalisability per se but can provide a way of more accurately describing the limitations of a particular case. Moving beyond metrics, typologies of social media network structures are emerging (Smith, Rainie, Shneiderman & Himelboim, 2014) which could also provide a link between individual case studies and the broader context in terms of overall characterisation of structures. Locating cases against key social media metrics and models could potentially form a meta-layer of data in addition to the case study, akin to the concept of a case record (Stenhouse, 1985; Walker, 2002).

A limitation in the approach used by the academic networking project, which compounds the tension between detailed understanding and generalisability, is the reliance upon convenience sampling. While undertaking a survey as a first stage of the research enabled a stratified sample to be constructed for network analysis based on key characteristics of interest in the study, self-selection bias was inevitably an issue to an extent. An alternative may be to identify key people of interest based on their online behaviour and to approach them for further participation, although whether this is appropriate would depend upon the research context and community involved.

Despite the limitations, adopting a person-centred approach offers the potential to ameliorate some of the difficulties of multi-platform social media studies. A combined approach would yield the greatest potential benefits and offset the limitations associated with either focusing upon solely online or experiential data. Incorporating person-centred elements into multi-platform social media research designs has the potential to enhance research in two ways. If carried out initially, the participants' views may be used to identify issues in the field and generate hypotheses to test later with larger datasets. Conversely, large-scale analyses may identify trends which are not clearly explained, in which case co-interpretation may be helpful, as was the case in the academic networking project. These approaches will be of particular value in studies where understanding the relationship between the online data traces and interactions with 'offline' contexts is critical to the research questions at hand. For example, participants' viewpoints can both illuminate online data and its relationship to organisational structures. While the examples here are drawn from a relatively small multi-platform social media study, the issues raised are of value to discuss in relation to multi-platform studies more generally. The importance of context in order to understand nuances in the social significance of technically similar platforms and data are highlighted. In the context of the study here, a case study approach was used to ameliorate these concerns to an extent, although this approach brings some limitations and the issues do require further discussion in order to be scalable to larger datasets.

Acknowledgments

This work was funded as part of a CREET doctoral studentship provided by The Open University, UK. The author would like to thank her supervisors, Professor Martin Weller and Dr. Doug Clow, for their feedback and guidance. Particular thanks to all participants who took part in the survey, network analyses and interviews.

References

- American Association of University Professors (AAUP) (2013). *Academic freedom and electronic communications*. Retrieved from <https://www.aaup.org/report/academic-freedom-and-electronic-communications-2014>
- Barabasi, A.L. (2002). *Linked*. Cambridge: Perseus.
- Bastian, M., Heymann, S. & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, San Jose, USA, 17-20 May. (pp. 361-362). Palo Alto, USA, AAAI.
- boyd, d. (2010, March). *Making sense of privacy and publicity*. Paper presented at SXSW, Austin, USA. Retrieved from <http://www.danah.org/papers/talks/2010/SXSW2010.html>
- boyd, d. & Crawford, K. (2011, September). *Six provocations for big data*. Paper presented at A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, Oxford, UK. Retrieved from <http://ssrn.com/paper=1926431>
- Bukvova, H. (2012). A holistic approach to the analysis of online profiles. *Internet Research*, 22(3), 340–360. <http://dx.doi.org/10.1108/10662241211235680>
- Cann, A.J., Dimitriou, K. & Hooley, T. (2011). *Social media: A guide for researchers*. London, Research Information Network. Retrieved from http://www.rin.ac.uk/system/files/attachments/social_media_guide_for_screen_0.pdf
- Carey, J.W. (1989). *Communication as culture: Essays on media and society*. New York: Routledge.
- Charmaz, K. (2014). *Constructing Grounded Theory: A practical guide through qualitative analysis*, 2nd ed. London: SAGE.
- Data Protection Act 1998*. (1998). London: Stationery Office.
- Dehay, P-O. (2017). *Cambridge Analytica demonstrably non-compliant with data protection law*. PersonalData.IO. Retrieved from <https://medium.com/personaldata-io/cambridge-analytica-demonstrably-non-compliant-with-data-protection-law-95ec5712b61>
- Dominguez, S. & Hollstein, B. (2014). *Mixed methods social networks research: Design and applications*. Cambridge: Cambridge University Press.
- Ellison, N.B. & boyd, d. (2013). Sociality through social network sites. In W.H.Dutton (Ed.), *The Oxford handbook of Internet Studies* (pp.151-172). Oxford: Oxford University Press.

- Eynon, R., Fry, J. & Schroeder, R. (2008). The ethics of internet research. In N. Fielding, R.M. Lee & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 23-41). London: SAGE.
- Fiesler, C., Lampe, C. & Bruckman, A.S. (2016). Reality and perception of copyright Terms of Service for online content creation. In *Proceedings of the 19th ACM Conference on Computer Supported Co-operative Work & Social Computing*, San Francisco, USA, 27 February – 2 March (pp. 1450-1461). ACM.
- Fisher, D., McDonald, D.W., Brooks, A.L. & Churchill, E.F. (2010, February). *Terms of service, ethics, and bias: Tapping the social web for CSCW research*. Panel discussion at Conference on Computer Supported Co-operative Work & Social Computing. Savannah, USA.
- Glaser, B.G. & Strauss, A.L. (1967). *The discovery of Grounded Theory: Strategies for qualitative research*. Chicago: Aldine.
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4), 597-606.
- González-Bailón, S., Wang, N., Rivero, A. & Borge-Holthoefer, J. (2014). Assessing the bias in samples of large online networks. *Social Networks*, 38, 16–27.
<http://dx.doi.org/10.1016/j.socnet.2014.01.004>
- Gray, D.E. (2009). *Doing research in the real world*, 2nd ed. London: SAGE.
- Guba, E.G. & Lincoln, Y.S. (1989). *Fourth generation research*. London: SAGE.
- Hale, S.A. (2012). *Build your own interactive network*. Interactive Visualizations blog. Retrieved from <http://blogs.oii.ox.ac.uk/vis/?p=191>
- Halford, S., Weal, M., Tinati, R., Carr, L. & Pope, C. (2016, October). *Digital data infrastructures: interrogating the social media data pipeline*. Paper presented at AoIR 2016: The 17th Annual Meeting of the Association of Internet Researchers, Berlin, Germany. Retrieved from <https://spir.aoir.org/index.php/spir/article/view/1381>
- Hall, M., Mazarakis, A., Peters, I., Chorley, M., Jens-Erik Mai, J-E., Caton, S. & Strohmaier, M. (2016, May). *Following user pathways: Using cross platform and mixed methods analysis in social media studies*. Workshop at CHI 16, San Jose, USA. Retrieved from http://jenserikmai.info/Papers/2016_following.pdf
- Hammersley, M. (1987). Some notes on the terms 'validity' and 'reliability'. *British Educational Research Journal*, 13(1), 73-81.
<http://dx.doi.org/10.1080/0141192870130107>
- Hawksey, M. (2017, March). Measuring social media impact: Google Analytics and Twitter. Presentation at the University of Oxford, UK. Retrieved from

<https://www.slideshare.net/mhawksey/measuring-social-media-impact-google-analytics-and-twitter>

- Hoffmann, C.P., Lutz, C. & Meckel, M. (2015). A relational altmetric? Network centrality on ResearchGate as an indicator of scientific impact. *Journal of the Association for Information Science and Technology*, 67(4), 765-775.
<http://dx.doi.org/10.1002/asi.23423>
- Hughes, A., Starbird, K., Leavitt, A., Keegan, B. & Semaan, B. (2016, May). *Information movement across social media platforms during crisis events*. Paper presented at 'Following user pathways: Using cross platform and mixed methods analysis in social media studies', workshop at CHI 16, San Jose, USA. Retrieved from http://www.amandaleehughes.com/MultiSMPlatformWorkshop_Final.pdf
- Jordan, K. (2014). Academics and their online networks: exploring the role of academic social networking sites. *First Monday*, 19(11).
<http://dx.doi.org/10.5210/fm.v19i11.4937>
- Jordan, K. (2017) *Understanding the structure and role of academics' ego-networks on social networking sites*. PhD thesis, The Open University, UK.
- Krotoski, A. (2010). Introduction to the special issue: Research ethics in online communities. *International Journal of Internet Research Ethics*, 3(1), 1-5.
- Lunden, I. (2014) *LinkedIn is quietly retiring network visualization tool InMaps*. TechCrunch website. Retrieved from <https://techcrunch.com/2014/09/01/linkedin-is-quietly-retiring-network-visualization-tool-inmaps/>
- Philips Honda, L. (2015). Motivations for participation in the YouTube-based "It Gets Better Project". *Proceedings of the 2015 48th Hawaii International Conference on System Sciences*, 5 – 8 January, Hawaii, USA (pp. 2033-2042). IEEE.
- Markham, A. & Buchanan, E. (2012). *Ethical decision making and internet research, Version 2.0: recommendations from the AoIR Ethics Working Committee*. Association of Internet Researchers. Retrieved from <http://www.aoir.org/reports/ethics2.pdf>
- Maxwell, J.A. (1998). Designing a qualitative study. In L.Bickman & D.J. Rog (Eds.), *Handbook of applied research methods* (pp. 69-100). London: SAGE.
- Miles, M.B. & Huberman, A.M. (1994). *Qualitative data analysis*, 2nd Ed. Thousand Oaks: SAGE.
- Molina, J.L., Maya-Jariego, I., & McCarty, C. (2014). Giving meaning to social networks: Methodology for conducting and analyzing interviews based on personal network visualisations. In S. Dominguez & B. Hollstein (Eds.), *Mixed methods social networks*

research: Design and applications (pp. 305-335). Cambridge: Cambridge University Press.

Mukherjee, P. & Jansen, J. (2016, May). *The changing nature of viewership: Formality of social media conversations*. Paper presented at 'Following user pathways: Using cross platform and mixed methods analysis in social media studies', workshop at CHI 16, San Jose, USA. Retrieved from https://www.researchgate.net/publication/305171094_The_Changing_Nature_of_Viewership_Formality_of_Social_Media_Conversations

Nentwich, M. & König, R. (2012). *Cyberscience 2.0: Research in the age of digital social networks*. Frankfurt: Campus Verlag.

Sandvig, C. & Karahalios, K. (2016). *Most of what you do online is illegal. Let's end the absurdity*. The Guardian. Retrieved from <https://www.theguardian.com/commentisfree/2016/jun/30/cfaa-online-law-illegal-discrimination>

Smith, M.A., Shneiderman, B., Milic-Frayling, N., Rorigues, E., Barash, V., Dunne, C., Capone, T., Perer, A. & Gleave, E. (2009). Analyzing (social media) networks with NodeXL. In *Proceedings of the fourth International Conference on Communities and Technologies*, 25-27 June, Pennsylvania, USA (pp. 255-264). ACM.

Smith, M.A., Rainie, L., Shneiderman, B. & Himelboim, I. (2014). *Mapping Twitter topic networks: From polarized crowds to community clusters*. Pew Research Centre. Retrieved from <http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/>

Spiliotopoulos, T. & Oakley, I. (2016, May). *Post or tweet: Lessons from a study of Facebook and Twitter usage*. Paper presented at 'Following user pathways: Using cross platform and mixed methods analysis in social media studies', workshop at CHI 16, San Jose, USA. Retrieved from http://tasos-spiliotopoulos.com/publications_assets/CHI2016-FUP-Spiliotopoulos-PostOrTweet.pdf

Stenhouse, L. (1985). Case study methods. In: J.P. Keeves (Ed.), *Educational research, methodology and measurement: an international handbook*, 1st Ed. (pp. 61–66). Oxford: Pergamon.

Teddlie, C. & Yu, F. (2007). Mixed methods sampling: a typology with examples. *Journal of Mixed Methods Research*, 1(1), 77-100. <https://doi.org/10.1177/2345678906292430>

van Gilder Cooke, S. (2011). *Walls have eyes: How researchers are studying you on Facebook*. Time Magazine. Retrieved from <http://content.time.com/time/business/article/0,8599,2099409,00.html>

- Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, 512, 126-129. <http://dx.doi.org/10.1038/512126a>
- Walker, R. (2002). Case study, case records and multimedia. *Cambridge Journal of Education*, 32(1), 109-127. <http://dx.doi.org/10.1080/03057640220116463>
- Weller, K. & Kinder-Kurlanda, K. (2016). A manifesto for data sharing in social media Research. In *Proceedings of the 8th ACM Conference on Web Science*, May 22 - 25, Hannover, Germany (pp. 166-172). ACM.
- Weller, M. (2011). *The Digital Scholar: How technology is transforming scholarly practice*. London: Bloomsbury.
- Winter, G. (2000). A comparative discussion of the notion of validity in qualitative and quantitative research. *The Qualitative Report*, 4(3), 1-14.

About the author

Dr Katy Jordan recently completed her doctoral studies within the Institute of Educational Technology, at the Open University, UK. Her research interests focus on the intersection of Internet Studies and Higher Education research.



Figure 1: Examples of ego-network visualisations. Left, a participants' Academia.edu ego-network; right, her Twitter ego-network.

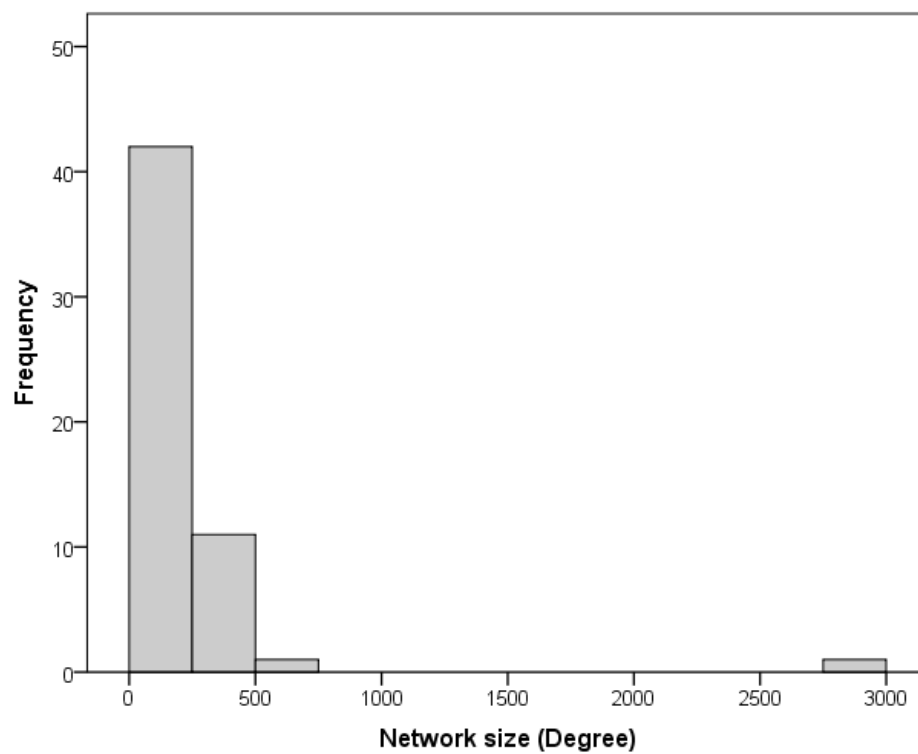


Figure 2: Histogram showing the degree distribution of the sampled academics' ego-networks on academic SNS.

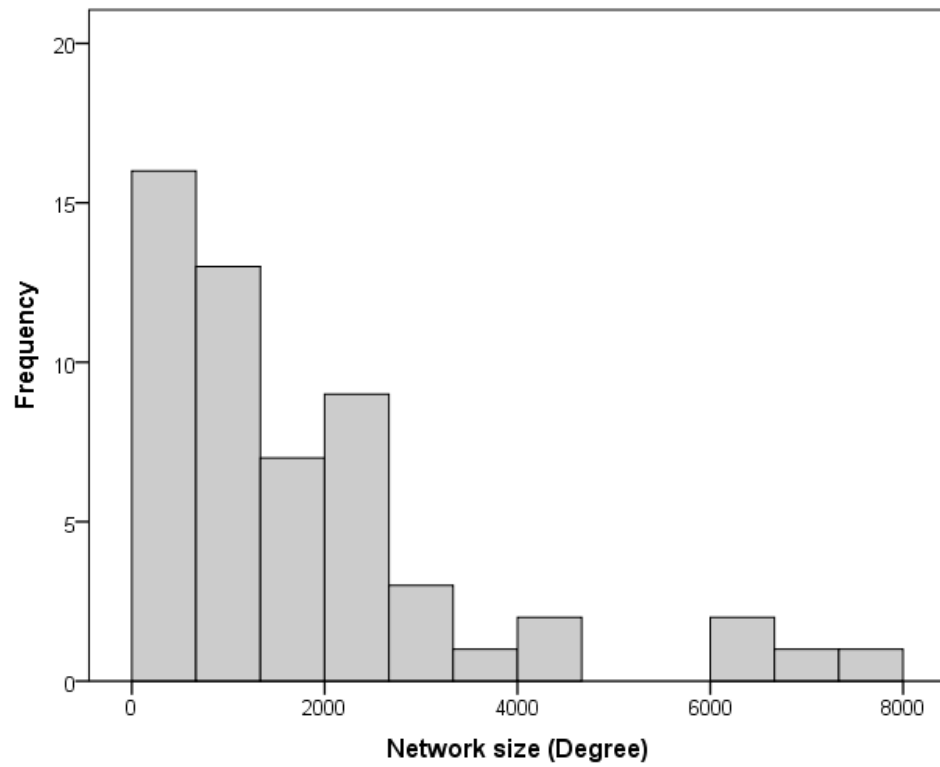


Figure 3: Histogram showing the degree distribution of the sampled academics' ego-networks on Twitter.

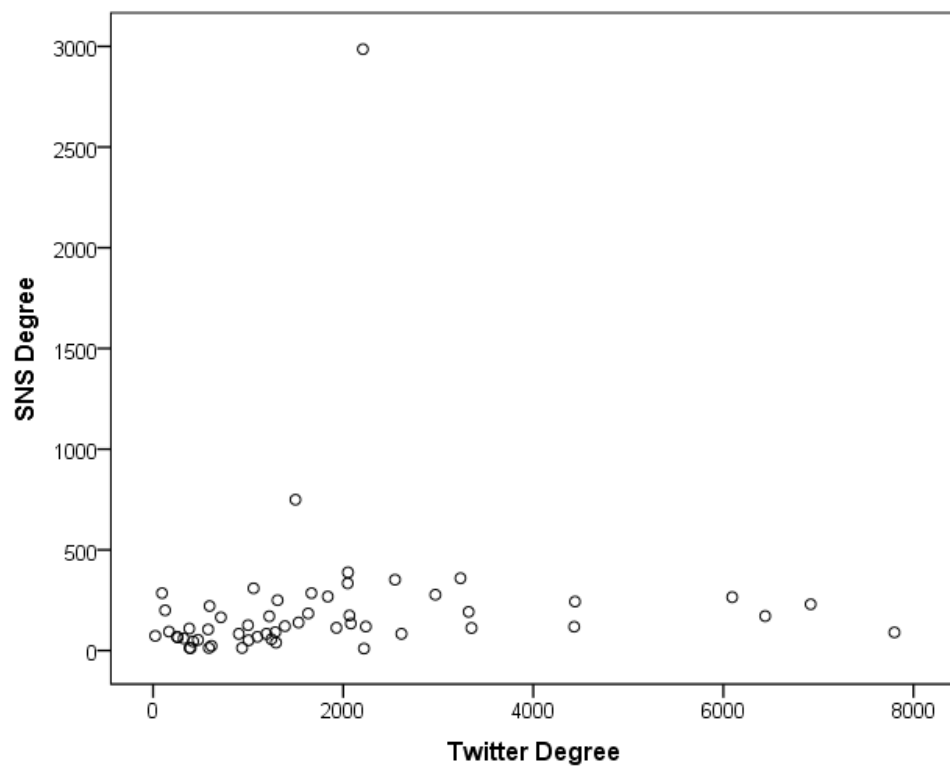


Figure 4: Scatterplot of values of degree for academic SNS and Twitter in the sampled academics' ego-networks.