*****PREPRINT VERSION*****

Final paper to be published in *TechTrends*

and available at:

http://dx.doi.org/10.1007/s11528-018-0313-6

Tweet, and We Shall Find:
Using Digital Methods to Locate Participants in Educational Hashtags

Spencer P. Greenhalgh; School of Information Science, University of Kentucky; spencer.greenhalgh@uky.edu; 320 Lucille Little Fine Arts Library, Lexington, KY 40506-0224

K. Bret Staudt Willet; Department of Counseling, Educational Psychology and Special Education, Michigan State University

Joshua M. Rosenberg; Department of Theory and Practice in Teacher Education, University of Tennessee, Knoxville

Matthew J. Koehler; Department of Counseling, Educational Psychology and Special Education, Michigan State University

This research was conducted while Spencer P. Greenhalgh and Joshua M. Rosenberg were graduate students at Michigan State University.

We would like to thank Ben Rimes, Mary Wever and everyone else who helped us reach out to the #michED community.

Keywords: digital methods, geolocation, educational hashtags, hashtag, social media, Twitter

Conflict of Interest: Spencer P. Greenhalgh declares that he has no conflict of interest. K. Bret Staudt Willet declares that he has no conflict of interest. Joshua M. Rosenberg declares that he has no conflict of interest. Matthew J. Koehler declares that he has no conflict of interest.

Abstract

Although researchers have discovered a great deal about *who* uses Twitter for educational purposes, *what* they post about, *when* they post and *why* they participate, there has so far been little work to explore *where* participants in educational Twitter contexts are located. In this paper, we establish a methodological foundation that can support the exploration of geographical issues in educational Twitter research. We surveyed 46 participants in one educational Twitter hashtag, #michED, to determine where they lived; we then compared these responses to results from three digital methods for geolocating Twitter users (human coding, machine coding and GPS coding) to explore these methods' affordances and constraints. Human coding of Twitter profiles allowed us to analyze more participants with higher levels of accuracy but also has disadvantages compared to other digital—and traditional—methods. We discuss the additional insights obtained through geolocating #michED participants as well as considerations for using geolocation and other digital methods in educational research.

Tweet, and We Shall Find: Using Digital Methods to Locate Participants in Educational Hashtags

In recent years, researchers have discovered a great deal about learning in digital spaces by examining educational uses of the microblogging service Twitter. For example, by studying *who* is active in educational Twitter contexts, scholars have found that participants in digital learning spaces are not always those who are expected (Rosenberg, Greenhalgh, Koehler, Akcaoglu, & Hamilton, 2016; Veletsianos, 2017a). Studying *what* is being tweeted in these contexts has allowed for an examination of the practices and processes that characterize learning with new technologies (Gao & Li, 2017; Gleason, 2013; Luo & Clifton, 2017; Veletsianos, 2017b). Studying *when* tweets are composed has allowed scholars to discover that even within the same educational context, participants engage in distinct learning practices at different times (Carpenter, Tani, Morrison, & Keane, 2018; Rosenberg, Akcaoglu, Staudt Willet, Greenhalgh, & Koehler, 2017). Finally, studying *why* people participate in educational Twitter contexts has lent insight into what informal learning spaces may offer that formal learning opportunities do not (Carpenter & Krutka, 2014, 2015).

Asking *where* participants in educational Twitter contexts are located also has the potential to advance our understanding, but this question has to date largely been ignored in the literature. Unless researchers investigate the locations that participants identify with, they will never be able to answer questions such as:

- How geographically-diverse are learners' Twitter networks?
- Do geographic "outsiders" participate in locally-focused educational communities on Twitter?
- Are participants in educational Twitter contexts disproportionately concentrated in wealthier areas?

However, few studies have explicitly analyzed geographical data associated with educational Twitter contexts, and those that have (e.g., Greenhalgh & Koehler, 2017) have used untested methods whose accuracy is not fully known.

The absence of this research is noteworthy given that answering these questions has important implications for how learners, researchers, administrators and practitioners conceive of learning in digital contexts. Stakeholders associated with degree and certificate programs may also benefit from such research. For example, an instructional technology program that encourages its graduate students to use Twitter for networking purposes may need to provide more explicit instruction if research shows that graduate students' Twitter networks tend to lack geographical diversity. A provincial ministry of education might not promote a locally-based teacher hashtag as a useful form of informal professional development if a substantial proportion of the participants are found to be from outside that province. A researcher might hesitate to describe an informal learning community on Twitter as a source of diverse perspectives on a topic if most of the participants live in more economically-privileged areas.

In this paper, we establish a methodological foundation that can support the exploration of geographical issues in educational Twitter research. In particular, we explore three methods for *geolocation*—the process of identifying or estimating the geographic location of a person or phenomenon—that could be used in future research on educational Twitter contexts. In the following sections, we explore several geolocation approaches using digital research methods, apply our efforts to a sample research problem and present the implications of our work.

Background

In this section, we introduce *digital research methods* and summarize existing *geolocation methods* used with Twitter data.

Digital Research Methods

Digital research methods are a loose collection of approaches to inquiry that implement "the use of online and digital technologies to collect and analyze research data" (Snee, Hine, Morey, Roberts, & Watson, 2016, p. 1). Although digital research methods can take many forms, many deal with the large amount of data produced by online and digital technologies—these *digital traces* (Lazer et al., 2009; Salganik, 2018; Welser, Smith, Fisher, & Gleave, 2008) serve as evidence of activity or interaction within an online or digital space. Digital traces are advantageous for researchers in that they may be collected more efficiently and more reliably than other forms of data (Munzert, Rubba, Meißner, & Nyhuis, 2015) and in that they can be gathered *unobtrusively* (Baker, 2008; Lee, 2015; Salganik, 2018)—that is, without disturbing or alerting participants in a digital space (though this also raises new ethical considerations for researchers; Fiesler & Proferes, 2018). Although digital research methods generally—and digital traces research in particular—are promising, they do warrant some caution (Lee, Fielding, & Blank, 2008; Marres, 2016) and, as is the aim of this paper, should therefore be closely investigated.

Geolocation Methods in Twitter Research

Digital trace data can be used to geolocate Twitter users in multiple ways. Twitter users have the option of identifying a location in their user profile and may also elect for their tweets to be tagged with latitude and longitude coordinates based on either the Global Positioning System (GPS) hardware in a mobile device or the geographic location of the modem a computer is connected to. These sources of data are the foundation for three common digital methods for geolocating Twitter users:

1. *Human coding*: assigning a location to a Twitter user based on human coders' interpretation of the location listed in their profile (e.g., Graham, Hale, & Gaffney, 2014; Takhteyev, Gruzd, & Wellman, 2012).
2. *Machine coding*: assigning a location to a Twitter user based on a computer program's interpretation of the location listed in their profile (e.g., Graham et al., 2014; Sloan, 2017; Takhteyev et al., 2012).
3. *GPS coding*: assigning a location to a Twitter user based on interpreting the latitude and longitude coordinates automatically associated with their tweets (e.g., Graham et al., 2014; Sloan, 2017).

Other, more complex, methods of geolocation have also been proposed—for example, researchers have attempted to geolocate Twitter users based on the content of their tweets (e.g., Cheng, Caverlee, & Lee, 2010; Sloan, 2017) or the composition of their social networks (e.g., Jurgens, Finnethy, McCorriston, Xu, & Ruths, 2015).

Purpose

The purpose of this paper is to better understand the affordances, constraints and other considerations for educational research that are associated with digital geolocation methods for Twitter data. To accomplish this purpose, we compared Twitter users' true geographic location to the locations estimated by three different geolocation methods—*human coding*, *machine coding* and *GPS coding.* Our analysis of these three approaches is guided by the following research question: *How accurately can Twitter profiles be geolocated using each method?*

Methods

Research Context

For this study, we examine data related to the #michED Twitter hashtag. A hashtag is a key word or phrase preceded by a hash symbol (i.e., "#") that indexes conversations on Twitter; "michED" stands for "Michigan education," and this hashtag therefore serves as a space for teachers and other stakeholders to engage in conversation about education in this U.S. state. #michED is an example of a Regional Educational Twitter Hashtag, whose participants are—in theory—concentrated in a particular geographic location (see Rosenberg et al., 2016). Although this study is focused primarily on methodological issues, using #michED data as an example allows us to explore the benefits of applying geolocation methods to a particular phenomenon.

Data Collection

Our primary source of data for this study is an online survey we distributed between December 2017 and January 2018—we posted the survey on Twitter using the #michED hashtag, asking anyone who had ever used the hashtag to identify their Twitter username and their place of residence. We also coordinated with active #michED participants in our personal networks and with facilitators of the #michED chat (a weekly Twitter conversation moderated through this hashtag) to further distribute the survey. After removing invalid survey responses (i.e., incomplete responses and those associated with Twitter usernames for which we had no evidence of #michED participation), we had 46 responses.

We then collected additional data directly from Twitter to supplement the survey information. First, we used the *twitteR* package (Gentry, 2015) for the R programming language to gather profile information—including the user's location—for the 46 usernames associated with our survey. Second, we retrieved all of the tweets associated with these 46 usernames from a dataset of #michED tweets collected between September 1st, 2015 and September 30th, 2017 using a series of Twitter Archiving Google Sheets (Hawksey, 2014).

Procedures and Measures

In the following sections, we describe each of the geolocation methods we use in this paper.

Human coding. The human coding method involves human interpretation of the location listed in each user's Twitter profile. We found that 39 of the 46 survey respondents (84.8%) listed locations in their Twitter profiles. We compared each profile location to the true location for that participant (as identified in our survey).

Machine coding. The machine coding method involves inputting the text from each Twitter profile to a computer program that estimates the corresponding location. Our program was based off of the web application *Nominatim* (https://nominatim.openstreetmap.org; Rudis, 2016); *Nominatim* associates text input (e.g., "Clinton Township MI" or "West Michigan") with the place name from its database that it judges to best correspond with that text (e.g., "Clinton Township, Macomb County, Michigan, United States of America" or "West Michigan, Lenora, Norton County, Kansas, United States of America"). If *Nominatim* was unable to associate a place name with the search terms (e.g., "Michigan's Thumb"), it returned nothing (i.e., "NA"). *Nominatim* returned a place name for 34 of the 46 profiles in our dataset (73.9%); we then compared each of the *Nominatim*-derived place names to the true location for that participant.

GPS coding. The GPS coding method is based on latitude and longitude coordinates provided by Twitter for certain tweets based on the GPS hardware in a mobile device or the geographic location associated with a computer network. For each of the three usernames associated with at least one GPS-coded tweet in our secondary dataset (6.5% of total usernames), we randomly selected one tweet, retrieved the corresponding latitude and longitude coordinates and used the *Nominatim* application to turn these coordinates into human-readable place names. We then compared each place name to the true location for that participant.

Data Analysis

We compared the locations derived from each of the three methods (i.e., human coding, machine coding and GPS coding) with the true location of each of the 46 Twitter users. This involved a total of 76 estimated locations (39 related to human coding, 34 to machine coding and 3 to GPS coding). Two coders (the first two authors of this paper) categorized each estimated location as either:

- *accurate*—within the specific municipality identified in the survey (e.g., if a user identifying with East Lansing, Michigan listed "East Lansing, MI" in their profile)
- *approximate*—outside the municipality but within the same sub-national region (i.e., a state or a province) listed in the survey (e.g., if a user identifying with East Lansing, Michigan listed "Lansing, MI" or "Michigan" in their Twitter profile).
- *inaccurate*—outside the sub-national region identified in the survey (e.g., if a tweet from a user identifying with East Lansing, Michigan was tagged with GPS coordinates placing them in Indianapolis, Indiana—presumably because of travel).
- *unknown*—a returned location that is ambiguous or does not appear to correspond to an actual location (e.g., if a user identifying with East Lansing, Michigan listed "Spartan Territory" or "Firefly-class spaceship, Milky Way Galaxy" in their Twitter profile).

In coding the 76 estimated locations, the coders achieved 85.5% agreement and a Cohen's kappa of .76, which can be interpreted as *substantial* agreement (Landis & Koch, 1977). They then discussed and resolved all discrepancies.

Results

In this section, we describe our findings (see also Table 1). First, because of the relative unavailability of the data necessary for GPS coding, it is difficult to determine the accuracy of this method with any great precision. However, of the three survey respondents that had tweets associated with GPS codes, one (33.3%) was coded as accurate, another (33.3%) was coded as approximate, the third (33.3%) was coded as inaccurate and none were coded as unknown.

We found the human coding and machine coding geolocation methods to be more accurate. Sixteen human-coded profiles (41.0%) were accurate, 20 (51.3%) were approximate, one (2.6%) was inaccurate and two (5.1%) were unknown. On the other hand, 11 machine-coded profiles (31.9%) were accurate, 16 (45.9%) were approximate and seven (20.6%) were *inaccurate*. A two-sided Fisher's exact test demonstrates that the classification of codes across these categories is not significantly different between the two methods (p = 0.0503).

Discussion

Although the focus of this paper is primarily methodological, we begin with a discussion of the additional insight into a sample educational Twitter context that was achieved by asking *where* its participants were found. We then discuss the methodological implications of our findings for both these specific geolocation methods and the broader use of digital methods in educational research. Throughout this section, we call attention to some of the limitations of this study as well as potentially fruitful areas for future research.

Benefits of Asking "Where?"

Despite the important differences between the three geolocation methods tested in this paper (which will be discussed in the following sections), the same broad conclusions about our sample data can be drawn from each method—that most but not all #michED participants are located in Michigan (see Table 1). This finding has immediate implications for researchers studying phenomena like #michED—implications that would not emerge without the inclusion of geolocation methods. For example, researchers can now assert that this hashtag is a largely localized phenomenon, which contrasts with the common narrative and intuitive understanding of Twitter as a tool that connects teachers across the world (e.g., Carpenter & Krutka, 2015; Gao & Li, 2017). This emphasizes the importance of recognizing that often-overlooked "social, cultural, economic, and political factors" (Veletsianos, 2017a, p. 286) shape how technologies such as Twitter are used for learning just as much as their inherent affordances and constraints.

As a second, contrasting example, researchers can now ask why some teachers would choose to participate in informal learning communities based in regions where they don't live. This is especially interesting given that in the United States (like other federal countries), education is chiefly under local control, which has previously been understood as the *raison d'être* for state-based hashtags (Rosenberg et al., 2016). Researchers who further pursue this question may lend additional insight as to the new opportunities afforded by digital learning contexts as well as to learners' motivations for participating in these contexts, which could have broad implications for the educational technology literature.

It is not difficult to imagine how these same methods could be used to lend useful insight into other extant questions about educational technology. Two examples from the literature on teachers' professional use of Twitter serve as useful illustrations. First, Krutka, Asino and Haselwood (2018) have highlighted the important role that the #OklaEd Twitter hashtag has played in teacher activism, including a recent teacher walkout in the American state of Oklahoma. In a cultural context where many are concerned about geographic outsiders engaging in local social media activism, educational researchers may wish to use geolocation to determine how many #OklaEd participants are from outside the state and to explore the implications of these results. Second, teachers have frequently referred to Twitter as a helpful antidote to the geographical and professional isolation associated with their jobs (Carpenter & Krutka, 2014, 2015; Wesely, 2013); there may therefore be interest in determining just how much these teachers' networks are locally-bound or geographically-diverse.

However, we should also note that the geolocation methods explored in this paper support certain questions of *where* better than others. For example, at the beginning of this paper, we suggested that educational researchers may be interested in whether participants in educational Twitter contexts were disproportionately concentrated in wealthier areas. However, our findings suggest that *coarser-grained* questions related to geography (e.g., how many #michED participants are from Michigan?) may be better supported by these geolocation methods than *finer-grained* ones (e.g., what school districts in Michigan have the most #michED participants?). That is, although human coding was able to correctly assign a Twitter user to the region they identified with 92.3% of the time, it was only able to correctly assign a Twitter user to a specific municipality 41.0% of the time. This is due in part to our strict interpretation of the *accurate* code—in some cases, respondents identified in the survey as being from a small municipality but listed a more recognizable, nearby municipality in their Twitter profile. Nonetheless, this discrepancy highlights the issues involved with using digital methods to assign Twitter users to specific municipalities, school districts, or other small areas.

Geolocation Methods

We found that geolocation methods vary in the amount of data that is available to and accessible by them. For example, this research supports previous findings that very few Twitter users allow for their tweets to be automatically GPS coded (e.g., Graham et al., 2014; Sloan,

2017) and that not all Twitter users list interpretable locations—or any locations at all—in their profiles (e.g., Takhteyev et al., 2012; Sloan, 2017). Nonetheless, it is clear from our experience that methods based on user profiles have access to much more data than those based on Twitter's automatic GPS coding.

These methods also differ in terms of their accuracy. We found human coding of Twitter profiles to be the most accurate of the methods that we tested. The accuracy of this method stands in contrast with previous work, which has typically favored GPS coding over profile data. For example, Graham and colleagues (2014) argued that "profile locations are not a useful proxy for device locations" (p. 576). However, while the location of a Twitter user (and their device) *in the moment* may be more useful for some research questions, our analysis shows that device locations are also not a perfect proxy for the locations that Twitter users *identify with*, which will be a more important consideration for other projects—including some related to educational Twitter communities. In fact, for the one survey respondent whose GPS-derived location was coded as *inaccurate*, the human coded and machine coded locations were both found to be *approximate*.

Despite the contributions of this study, further research is needed to explore the potential of machine coding. We found that although human coding was more accurate than machine coding, this difference was not statistically significant; however, a similar test having a larger sample size and more statistical power may yield different results. Furthermore, in this paper, we have only tested one of the many available machine coding methods. Previous research has found that different APIs can interpret different amounts of data with different levels of accuracy (Graham et al., 2014); indeed, our own exploratory analysis suggests that using the Google Maps API rather than the Nominatim API for our machine coding method may have produced more accurate results (though it would have also required more time—and possibly money—to complete the analysis).

Digital Methods in Educational Research

Although this study is specifically focused on the practical benefits of employing one kind of digital method to study one particular learning context, we are writing within a broader context that is seeing an increased use of digital methods both in social science research generally (e.g., Lazer et al., 2009; Salganik, 2018; Welser et al., 2008;) and in educational research in particular (e.g., Mishra, Koehler, & Greenhow, 2016; Shaffer, 2017). Our experience and findings in this study have implications for how educational researchers should consider the relationship between traditional and digital research methods. As we elaborate in this section, digital and traditional methods can support each other in important ways.

There are clear advantages associated with the traditional, survey methods we employed in this study. Salganik (2018) noted that digital data, when compared to surveys and other more traditional methods, tend to be incomplete. For example, the Twitter profiles examined in this study lacked an authoritative identification with a particular location—something that was simple to obtain using a survey. Indeed, the traditional survey provided more useful data for our sample research context than any of the digital geolocation methods that we used and—self-reporting issues notwithstanding—can be assumed to provide the most accurate data. This is especially important given that our findings about the proportion of respondents who were from Michigan based on human coding (the most accurate of the digital methods) were significantly different than the actual survey results (according to a test of equal proportions; $\chi^2[1] = 4.64$, $p = .03$). Based on these findings, we argue that error is necessarily a part of the use of digital methods in educational research and is likely to be more present in digital research than in traditional educational research methods.

This study also demonstrates some of the disadvantages of traditional research methods and some of the corresponding advantages of digital approaches. For example, it took us approximately two months and a considerable amount of coordination with colleagues to collect only 46 valid responses to the survey that forms the backbone of this paper. In contrast, for one of our previous studies on hashtags like #michED (Rosenberg et al., 2016), we were able to collect data associated with 68,552 Twitter users associated with 47 hashtags (rather than just one) without expending significantly more effort. These data were “bigger” and collected less intrusively than the data collected through our survey (see Salganik, 2018). On a similar note, although the human coding of 39 profile locations did not require a substantial amount of effort, doing the same for tens of thousands of locations would be impractical, requiring researchers to sample data or consider using less accurate methods (e.g., machine coding) to save time.

Based on our experience, we argue that neither avoiding the incompleteness and error common to digital methods nor avoiding the practical difficulties that also accompany them should be sacrosanct in educational researchers’ consideration of methodological issues. Rather, when considering whether—or how—to implement digital methods in educational research, scholars should consider a continuum of options. In doing so, researchers should recognize the affordances and constraints associated with each option and select that which best supports the purpose and context of the research. Furthermore, scholars should consider the research ethics implications of incorporating digital methods into their study (e.g., Markham & Buchanan, 2012). Finally, they should be open and transparent about the weaknesses of their design when reporting their findings.

In particular, we endorse for educational research the recommendations of scholars such as Salganik (2018) and Shaffer (2017), who have argued that the decision between traditional and digital methods is not an "either-or" choice but rather a process of combining them so that each complements the strengths (and compensates for the weaknesses) of the other. The way that we have combined methods in this study differs somewhat from these recommendations in that we do not expect future research on learning in digital contexts to also carry out geolocation using both traditional and digital methods; indeed, in cases for which the goal is to assign Twitter users to specific geographic regions, we show that digital methods alone may be sufficient (if imperfect). However, we are only able to demonstrate this because we have used traditional research methods to *validate* digital approaches; without the use of our survey measure, we would not know how accurate these digital methods were. Salganik's (2018) and Shaffer's (2017) advice can therefore be applied to a sequence of research activities, not just to a single study. That is, the use of traditional methods is necessary to validate digital approaches, which can then be considered for large-scale use; conversely, the large-scale use of digital methods will likely generate questions (e.g., why are there non-Michigan residents participating in #michED?) that traditional methods are best suited to explore.

Conclusion

Asking *where* participants in educational Twitter contexts are located has the potential to lend additional insight to the understanding already gained through researchers' questions of *who*, *what*, *when* and *why*. In this paper, we have described and tested three digital geolocation methods—human coding, machine coding and GPS coding—that can help support questions related to *where*. Our application of these methods to sample data shows how answers to questions of *where* have the potential to deepen our understanding of learning in technologicallymediated contexts in important ways. Furthermore, our findings suggest that there are a number of advantages and disadvantages associated with each geolocation method, and we suggest that researchers consider these advantages and disadvantages when applying geolocation methods—or other digital research methods—to educational research.

Compliance with Ethical Standards

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

Baker, L. M. (2008). Unobtrusive research. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 905-906). Thousand Oaks, CA: SAGE.

Carpenter, J. P., & Krutka, D. G. (2014). How and why educators use Twitter: A survey of the field. *Journal of Research on Technology in Education*, *46*, 414-434. doi:10.1080/15391523.2014.925701

Carpenter, J. P., & Krutka, D. G. (2015). Engagement through microblogging: Educator professional development via Twitter. *Professional Development in Education*, *41*, 707-728. doi:10.1080/19415257.2014.939294

Carpenter, J. P., Tani, T., Morrison, S., & Keane, J. (2018). Exploring the education Twitter hashtag landscape. In E. Langran & J. Borup (Eds.), *Proceedings of Society for Information Technology and Teacher Education Conference 2018* (pp. 2230-3325). Waynesville, NC: Association for the Advancement of Computing in Education (AACE).

Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet: A content-based approach to geo-locating Twitter users. In X. J. Huang, G. Jones, N. Koudas, X. Wu, & K. Collins-Thompson (Eds.), *Proceedings of the 19th ACM international conference on information and knowledge management* (pp. 759-768). New York, NY: Association for Computing Machinery.

Fiesler, C., & Proferes, N. (2018). “Participant” perceptions of Twitter research ethics. *Social Media + Society*, *4*(2). doi:10.1177/2056305118763366

Gao, F., & Li, L. (2017). Examining a one-hour synchronous chat in a microblogging-based professional development community. *British Journal of Educational Technology*, *48*, 332-347. doi:10.1111/bjet.12384

Gentry, J. (2015). twitteR: R based Twitter client (Version 1.1.9) [Computer software]. Retrieved from http://CRAN.R-project.org/package=twitteR

Gleason, B. (2013). #Occupy Wall Street: Exploring informal learning about a social movement on Twitter. *American Behavioral Scientist*, *57*, 966-982. doi:10.1177/0002764213479372

Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the world are you? Geolocation and language identification in Twitter. *The Professional Geographer*, *66*, 568-578. doi:10.1080/00330124.2014.907699

Greenhalgh, S. P., & Koehler, M. J. (2017). 28 days later: Twitter hashtags as "just in time" teacher professional development. *TechTrends*, *61*, 273–281. doi:10.1007/s11528-016-0142-4

Hawksey, M. (2014). Need a better Twitter Archiving Google Sheet? TAGS v6.0 is here! [Blog post]. Retrieved from https://mashe.hawksey.info/2014/10/need-a-better-twitter-archiving-google-sheet-tags-v6-0-is-here/

Jurgens, D., Finnethy, T., McCorriston, J., Xu, Y. T., & Ruths, D. (2015). Geolocation prediction in Twitter using social networks: A critical analysis and review of current practice. In *Proceedings of the twenty-ninth AAAI conference on artificial intelligence and the twenty-seventh innovative applications of artificial intelligence conference*. Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Krutka, D. G., Asino, T. I., & Haselwood, S. (2018). Editorial: Eight lessons on networked teacher activism from #OklaEd and the #OklaEdWalkout. *Contemporary Issues in Technology and Teacher Education*, *18*(2). Retrieved from http://www.citejournal.org

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., . . . Van Alstyne, M. (2009). Computational social science. *Science*, *323*, 721-723. doi:10.1126/science.1167742

Lee, R. M. (2015). *Unobtrusive measures*. doi:10.1093/OBO/9780199846740-0048

Lee, R. M., Fielding, N., & Blank, G. (2008). The Internet as a research medium: An editorial introduction to *The Sage Handbook of Online Research Methods.* In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 2-22). Thousand Oaks, CA: SAGE.

Luo, T. & Clifton, L. (2017). Examining collaborative knowledge construction in microblogging-based learning environments. *Journal of Information Technology Education: Research*, *16*, 365-390.

Markham, A. & Buchanan, E. (2012). *Ethical decision-making and internet research: Recommendations from the AoIR ethics working committee* (Version 2.0). Chicago, IL: Association of Internet Researchers.

Marres, N. (2016). Foreword. In H. Snee, C. Hine, Y. Morey, S. Roberts, & H. Watson (Eds.), *Digital methods for social science: An interdisciplinary guide to research innovation*. (pp. viii-ix). New York, NY: Palgrave Macmillan.

Mishra, P., Koehler, M. J., & Greenhow, C. (2016). The work of educational psychologists in a digitally networked world. In L. Corno & E. M. Anderman, *Handbook of educational psychology* (3rd ed., pp. 29-40). New York, NY: Routledge.

Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2015). *Automated data collection with R: A practical guide to web scraping and text mining*. West Sussex, England, UK: Wiley.

Rosenberg, J. M., Akcaoglu, M., Staudt Willet, K. B., Greenhalgh, S. P., & Koehler, M. J. (2017). A tale of two Twitters: Synchronous and asynchronous use of the same hashtag.

In P. Resta & S. Smith (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2017* (pp. 283-286). Waynesville, NC: Association for the Advancement of Computing in Education (AACE).

Rosenberg, J. M., Greenhalgh, S. P., Koehler, M. J., Akcaoglu, M., & Hamilton, E. (2016). An investigation of State Educational Twitter Hashtags (SETHs) as affinity spaces. *E-Learning and Digital Media*, *13*, 24-44. doi:10.1177/2042753016672351

Rudis, B. (2016). nominatim: tools for working with the "Nominatim" API. (Version 0.2.2.9000) [Computer software]. Retrieved from https://github.com/hrbrmstr/nominatim

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton, NJ: Princeton University Press.

Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.

Sloan, L. (2017). Social science 'lite'? Deriving demographic proxies from Twitter. In L. Sloan & A. Quan-Haase (Eds.), *The SAGE Handbook of Social Media Research Methods* (pp. 90-104). London, UK: SAGE.

Snee, H., Hine, C., Morey, Y., Roberts, S., & Watson, H. (2016). Digital methods as mainstream methodology: An introduction. In H. Snee, C. Hine, Y. Morey, S. Roberts, & H. Watson (Eds.), *Digital methods for social science: An interdisciplinary guide to research innovation* (pp. 1-11). New York, NY: Palgrave Macmillan.

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of Twitter networks. *Social Networks*, *34*, 73-81. doi:10.1016/j.socnet.2011.05.006

Veletsianos, G. (2017a). Three cases of hashtags used as learning and professional development environments. *TechTrends, 61*, 284-292. doi:10.1007/s11528-016-0143-3

Veletsianos, G. (2017b). Toward a generalizable understanding of Twitter and social media use across MOOCs: Who participates on MOOC hashtags and in what ways? *Journal of Computing in Higher Education*, 29, 65-80. doi:10.1007/s12528-017-9131-7

Welser, H. T., Smith, M., Fisher, D., & Gleave, E. (2008). Distilling digital traces: Computational social science approaches to studying the Internet. In N. Fielding, R. M. Lee, & G. Blank, *The SAGE handbook of online research methods* (pp. 116-141). Thousand Oaks, CA: SAGE.

Wesely, P. M. (2013). Investigating the community of practice of world language educators on Twitter. *Journal of Teacher Education*, *64*, 305-318. doi:10.1177/0022487113489032

Tables

Table 1: Availability of Data for, Accuracy and Derived Location of Three Digital Geolocation Methods

Method	Available	Accurate	Approximate	Inaccurate	Unknown	Within Michigan	Outside Michigan	Unknown
survey	n/a	n/a	n/a	n/a	n/a	93.5% (43)	6.5% (3)	0.0% (0)
human coding	84.8% (39)	41.0% (16)	51.3% (20)	2.6% (1)	5.1% (2)	84.6% (33)	10.3% (4)	5.1% (2)
machine coding	73.9% (34)	31.9% (11)	45.9% (16)	20.6% (7)	0.0% (0)	73.5% (25)	26.5% (9)	0.0% (0)
geocoding	6.5% (3)	33.3% (1)	33.3% (1)	33.3% (1)	0.0% (0)	66.6% (2)	33.3% (1)	0.0% (0)