Defining and Predicting the Localness of Volunteered Geographic Information using Ground Truth Data

Ankit Kariryaa[+], Isaac Johnson[*], Johannes Schöning[+], Brent Hecht[*]
[+]HCI Group, University of Bremen, Germany,
[*]PSA Research Group, Northwestern University, United States
kariryaa@uni-bremen.de, isaacj@u.northwestern.edu, schoening@uni-bremen.de, bhecht@northwestern.edu

ABSTRACT
Many applications of geotagged content are predicated on the concept of localness (e.g., local restaurant recommendation, mining social media for local perspectives on an issue). However, definitions of who is a "local" in a given area are typically informal and ad-hoc and, as a result, approaches for localness assessment that have been used in the past have not been formally validated. In this paper, we begin the process of addressing these gaps in the literature. Specifically, we (1) formalize definitions of "local" using themes identified in a 30-paper literature review, (2) develop the first ground truth localness dataset consisting of 132 Twitter users and 58,945 place-tagged tweets, and (3) use this dataset to evaluate existing localness assessment approaches. Our results provide important methodological guidance to the large body of research and practice that depends on the concept of localness and suggest means by which localness assessment can be improved.

Author Keywords
Localness, placetag, Twitter, geographic HCI

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

INTRODUCTION

Georeferenced tweets, geotagged Instagram photos, and other volunteered geographic information (VGI) are critical to research and practice across a wide swath of computing. For many applications of VGI, it is important to determine the "localness" of the VGI contributor (e.g., the content poster) to a specific region. This is true for applications ranging from recommender systems that surface venues that are "local favorites" (e.g., [14,45,60]) to research that seeks to understand the local perspective on certain issues (e.g., [25,53,57]). In fact, the concept of localness is so central to VGI that, when defining the term volunteered geographic information, prominent geographer Michael Goodchild wrote [13]:

> *"The most important value of VGI may lie in what it can tell us about local activities... that go unnoticed by the world's media, and about life at a local level. It is in that area that VGI may offer the most interesting, lasting and compelling value…"*

However, considering the importance of the concept of localness to VGI, we know surprisingly little about the concept. First and foremost, there are no broadly accepted definitions of "local" in computing, with most projects adopting definitions that are ad-hoc and often unstated. Second, while there are techniques that have been widely used to assess the localness of VGI, they have not been validated (let alone validated against a concrete definition of "local"). In other words, we have little understanding of how well these techniques work, and for which conceptions of localness.

The goal of this paper is to begin the process of addressing these two important gaps in the literature. We first asked the following question: What do we mean in computing when we describe users or information as "local"? To address this question, we conducted a review of the literature that has engaged with the concept of localness. Examining 30 papers, we identified that by "local", researchers and practitioners typically mean one of three things: where someone currently lives (which we term the "*LivesIn*" definition of local), where someone currently votes ("*VotesIn*"), and the places with which someone is very familiar ("*Familiarity*").

After identifying these three definitions, we then posed our second question: "How well do existing localness assessment techniques work, and for which definitions?" We focus on four common localness assessment techniques in particular: *LocationField* (e.g., [16,33,36]), *nDays* (e.g., [22,28,40]), *Plurality* (e.g., [31,35,43]), and *GeometricMedian* (e.g., [5,22,24]). To enable our understanding of these techniques, we collected the first ground truth localness dataset from VGI contributors. The dataset consists of information from 132 Twitter users and 58,945 total place-tagged tweets, and it was collected via a survey deployed on Twitter.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ACM 978-1-4503-5620-6/18/04…$15.00
https://doi.org/10.1145/3173574.3173839

The results of this analysis provide important methodological guidance for the many practitioners and researchers engaging with the concept of localness. In particular, our results suggest a straightforward set of best practices: If considering a population that frequently geotags (or "placetags") content, researchers and practitioners should use either *Plurality* or *GeometricMedian*. However, if researchers are considering a population that does not frequently use geotagging (or placetagging) functionality, the *LocationField* approach is an excellent second option.

Our results, however, also point to some important challenges for localness assessment. While *Plurality*, *GeometricMedian*, and *LocationField* all perform reasonably well for "single location" definitions of localness (*LivesIn* and *VotesIn*), even at the city scale, all existing localness assessment techniques perform worse for *Familiarity*. Our results additionally problematize the use of the *nDays* technique, with the three others being better alternatives in most cases.

Finally, our work additionally highlights several exciting opportunities for future work in this research area. In particular, through our evaluation of localness assessment techniques, we were able to identify a series of opportunities for the improvement of these techniques. We close the paper with a discussion of these opportunities, as well as a number of other implications for researchers and practitioners that emerge from our results.

RELATED WORK

Importance of Localness in Computing

The concept of the "local" has become important to a diverse array of research areas in computing, including in geographic HCI [19]. For instance, within the recommender systems domain, a common challenge is surfacing relevant restaurant recommendations from "locals" (e.g., [14,15,58]). Indeed, this area has become sufficiently prominent that systems that provide this service were recently featured in an article in *The New York Times* [45]. In information retrieval, it has been found that search preferences and information needs in map search differ significantly between locals and non-locals (e.g., [26,54]). For social computing, the concept of "local" has been used to reveal biases in Wikipedia [47], identify potential gaps in coverage of sharing economy and mobile crowdsourcing (e.g., [35,50]), among other applications (e.g., [59]). Data science also frequently engages with localness, e.g., for understanding geographically-variable opinions on policy (e.g., [57]) and in studies of public health (e.g., [7]). Further applications of the concept of local can be found in Table 1.

Investigations of Localness

This paper is most directly motivated by the work of Johnson et al. [22]. In this paper, Johnson and colleagues investigated the four localness assessment techniques that we consider here and found that they can give different results for the same Twitter user. However, as noted by Johnson et al., they did not assess how well each technique performed against ground truth data, in part because no such dataset existed. One of the key goals of this paper was to develop this dataset and to perform this assessment.

Another important source of motivation for this work is the small set of papers that has considered ground truth localness data, but for very specialized contexts. In particular, Sen et al. [47] sought to understand the "geographic provenance" of cited sources on Wikipedia and developed a model for assessing this provenance based on ground truth data about source locations. The known accuracy of the model enabled by this ground truth allowed the authors to make important claims about the degree to which local vs. foreign sources are used to describe different areas of the world. However, this model is limited to the assessment of the localness of the URLs of news articles and other Wikipedia sources (not people), and only works at a national scale.

Localness Assessment Techniques

Johnson et al. [22] identified four localness assessment techniques in the computing literature: *nDays*, *Plurality*, *GeometricMedian* and *LocationField*. We evaluate each of these four techniques in our experiment, and summarize each of them briefly below:

- *nDays* is a temporal range-based technique that assigns a user as local to a region if they produced content in the place at *least n* days apart. We found the value of n varied from 2 to 30 days ([26,40]), while $n = 10$ days ([18,22,28]) was the most commonly used value.
- *Plurality*, as its name suggests, assigns a user as local to the region (or regions in case of a tie) from which she or he produced the most content.
- *LocationField* extracts the entry in the location field in a user's profile (if it exists) and turns that text location (toponym) into machine-readable coordinates using a geocoder (we use Google's Geocoder). The method then assigns the user as local to the region output by the geocoder (or, in the case of less granular scales, the containing region).
- *GeometricMedian* is the most technically complex of the approaches. It assigns a user as local to the point that minimizes the distance between all the locations from which the user has posted content, and then returns the region associated with that point. It additionally has the requirements that the user have at least five posts and that half of the user's posts be within 30km of the median point (to avoid situations in which, e.g., a user posts from Anchorage, AK and New York, NY and is assigned as local to a district in rural Canada).

Geographic Information and Twitter

The vast majority of prior work on Twitter, including most localness work, has used Twitter coordinate *geotags* to associate tweets with geographic locations. However, very recent work by Tasse et al. [49] strongly problematized the use of coordinate geotags in a research context:

> *"The geotags that are still present are getting stranger: job posting bots, weather and sports bots, deleted accounts, and other accounts are creating a growing fraction of all public geotagged tweets... It is not clear how much more research can be done with coordinate geotags."*

Instead, Tasse and colleagues suggest that researchers change their focus to Twitter's placetags, which associate tweets not with specific coordinates, but rather with named places (e.g., "Nashville, TN", "Boise River Greenbelt", "McSorley's Old Ale House"). We followed this recommendation in this paper, meaning that the methodological guidance afforded by our results applies in this new "placetag era".

DEFINITIONS OF LOCALNESS

While there exists a large literature in computing that engages with the concept of localness, there exists no formal definition(s) of who is a "local". To address this problem, we conducted a survey of 30 papers in the computing literature that engaged with the concept of localness. In doing so, we leveraged the example and straightforward methods of Johnson et al. [21], which faced a similar definitional question with respect to the vehicle routing literature (i.e. we used a set of core papers, in our case those referenced by [22], iteratively employed keyword and citation network approaches to identify further papers, and collaboratively identified themes in the found literature).

We identified three general themes in how our 30 papers had implicitly and explicitly defined what "local" means: (1) where a person *lives*, (2) where a person *votes*, (3) and areas with which a person has a great deal of *familiarity.* We also note that these definitions can be split into two categories: (1) single location definitions, which assume that a person can be local to a single region at a time, and (2) multiple location definitions, which allow people to be local to multiple regions simultaneously. We provide the details about each individual definition, which we respectively term *LivesIn*, *VotesIn*, and *Familiar,* immediately below. We also show which papers utilized which definition in Table 1.

The *LivesIn* definition of localness – a "single location" definition – is relatively straightforward. Papers that use this definition assume that a person is local *only* to the region in which they live. Often this region is defined at the city scale, but occasionally it is defined at the neighborhood scale, county scale, and even state and country scale as well. For example, Malik et al. [31] employed the *LivesIn* definition of localness in their exploration of the biases in geotagged tweets, Morais & Andrade [33] used this definition to understand the difference in annotations shared by tourists and residents, and Fiorio et al. [11] used this definition to estimate short- and long-term migration.

The *VotesIn* definition – another "single location" definition – is analogous to the *LivesIn* definition, but applies to the location in which a person votes versus that in which they live. This is an important distinction, as college students, migrants, and others often live in different constituencies than those in which they vote. Zhang and Counts [57] employed this definition of localness to predict same-sex marriage policy change in U.S. states using publicly available Twitter data.

Finally, the *Familiarity* definition of localness is different from the other definitions in that it does not restrict the assignment of localness to a single region for a given user. This makes *Familiarity* our only "multiple location" definition. *Familiarity* labels someone as a local to a given region if they have a sufficient amount of on-the-ground knowledge about the region, with that amount often being extensive. For instance, Zielstra et al. [59] used this definition to study the relationship between knowledge of a place and OpenStreetMap editing patterns and Kumar et al. [26] used this definition to characterize locations using Flickr photos.

It is important to note that the papers in our study very often did not explicitly define what they meant by "local". In these cases, determining the definition that was employed required deeply reading the paper for the underlying assumptions being made. It is our hope that our work can highlight the need to formally declare the definition of local that one is using. Our small schema of localness definitions should make it easier to do so.

Lastly, our review of localness research focused explicitly on the computing literature given the immediate need for increased structure in this literature. However, localness and related ideas like *heimat* (e.g., [4,10]), *sense of place* (e.g.,

DEFINITION	EXAMPLES
LivesIn	Hecht & Gergle, 2010 [17], Abbar et al. 2015 [1], Abdullah et al., 2015 [2], Culotta, 2014 [7], Girardin et al., 2008 [12], Li et al., 2013 [28], Malik et al., 2015 [31], Mislove et al., 2011 [32], Morais and Andrade, 2014 [33], Naaman et al., 2012 [36], Reiderer et al., 2015 [43], Tasse et al., 2017 [49], Hecht & Stephens, 2014 [18] , Popescu & Grefenstette, 2010 [40], Musthag & Ganesan, 2013 [35], Hecht et al. 2011 [16], Jurgens et al., 2015 [24], Poblete et al., 2011 [39], Johnson et al., 2016 [22], Fiorio et al., 2017 [11], Kogan et al., 2015 [25], Sen et al. 2015 [47]
VotesIn	Zhang & Counts, 2015 [57]
Familiar	Eckle & Albuquerque 2015 [9], Kumar et al., 2017 [26], Kumar et al., 2017 [27], White and Buscher, 2012 [54], Wu et al., 2011 [56], Zielstra et al., 2014 [59], Ludford et al., 2007 [30]

Table 1. The definitions for localness we identified in the literature and papers that used these definitions.

[3,42,48,51]), *homeness* [46], *place attachment* [29], *place dependence* [55], *place identity* [41,52], *dwelling identity, community identity and regional identity* [6] have been studied in the humanities and social sciences for decades (e.g. in geography, sociology, economics). Additionally, further operationalizations of the term "local" appear in various legal and other contexts (e.g. in the food industry [61]). An exciting direction of future work is to engage deeply with these literatures to introduce more sophisticated systematic definitions of localness that can be adopted by the computing literature. In this study, however, our contribution lies in formalizing existing definitions in the computing literature and evaluating how well we can operationalize them with localness assessment techniques.

METHODS

Survey Design

We designed a survey to collect ground truth information such that we could compare the accuracy of each of the four localness assessment approaches with respect to each of the three definitions of localness. Specifically, we asked participants for where they live (*LivedIn*), where they vote (*VotesIn*), and locations with which they were familiar. Given recent concerns in the United States about the privacy of voter information [20], we made all *VotesIn* information optional.

Most of the 30 papers in our literature view considered localness at the city or county scales, but a few used less granular scales. As such, we focused our analyses at three scales: city, U.S. county, and U.S. state. Similarly, we selected the United States as our study area as it is the region in which much of the localness literature has been conducted (e.g., [1–3,15,16]). As is discussed in more detail below, a compelling direction of future work involves extending our study to other countries.

To gather *LivedIn* information, we asked the following question: "In which city and state do you live?" Participants were then asked about *VotesIn* with the optional question "In which city are you registered to vote?". For *Familiarity*, we allowed participants to enter up to five cities with which they were familiar. We also asked them to indicate how familiar they were with each entered location on a five-point scale ranging from 1 ("Slightly familiar") to 5 ("Very familiar"). For each location for which they indicated they were familiar, participants were asked to list their relationship with the location ("I have visited it", "I have lived in it", or "Other", with "Other" including an open text box to describe the relationship). In the below analyses, we consider any *Familiarity* rating of four or above to be "familiar", otherwise we treat the corresponding location as not familiar.

The survey, which was implemented in Qualtrics, closed with two final open-ended questions: "Do you have any additional thoughts to share about the areas to which you consider yourself local?" and "Do you have any additional comments about this survey?"

All our survey procedures followed the guidance provided by the IRB and similar organizations at our various institutions. The full text of our survey is available in the Supplementary Materials included with our submission.

Survey Sample

Since we focused on Twitter users who use placetags, we created a potential participant list by gathering a set of users for whom their most-recent placetagged tweet was in the United States from the Twitter streaming API for one week during the summer of 2017. In total, we developed a potential survey population of approximately 830,000 users in this fashion.

Our next challenge was finding a way to deploy our survey to this population, and this challenge was a serious one. A well-known approach for collecting ground truth information from social media at scale is the technique outlined by Nichols and Kang in their *TSATracker* work [37]. At a high level, this approach involves creating a Twitter bot that pings users with a request to tweet at the bot with a desired piece of information. However, this approach was not feasible for us as our research questions necessitated that users to fill out a survey (as opposed to *TSATracker*, which, e.g., asked a single question about the length of airport security lines). Unfortunately, taking a similar approach to Nichols and Kang with tweets that include a link to a survey (or any link, for that matter) is considered spam by Twitter's Terms of Service and is banned [62].

This highlights an important issue, not just for this paper, but also for work that engages with social media more generally: if a research question requires data outside of what can be gathered using the standard public behavioral trace information, how can one gather this information at scale?

To partially address this issue, we turned to a version of the *TSATracker* approach, but one that is formally sanctioned by Twitter: we used Twitter's ad platform. Specifically, instead of tweeting at users in our target population, we simply uploaded our list of users to Twitter's ad system and targeted these users via paid ads. It is interesting to note that the exact same content we would have tweeted at users using the *TSATracker* approach was no longer considered spam as soon as it became a paid ad. We used two ads: one with a monetary incentive (offering a chance to win one of four $25 gift cards) and one with an altruistic incentive. Our study ran for one week in Summer 2017 and from the two ads, we received 22,600 impressions and 222 clicks (1.0% click-through rate (CTR)), and 29,434 impressions and 237 clicks (0.8% CTR), respectively. Overall, we received 136 complete responses and 25 partial responses. Partial responses are those in which the participant did not reach the end of the survey, but did provide us with some information. As long as these partial responses contained *LivesIn* information, we considered them for the final analysis where relevant.

As we will show below, the scale afforded by the Twitter advertising platform allowed us to get a broad sense of the relative performance of each localness definition. However, the ad platform is sufficiently expensive and low-throughput that gathering information for a project that requires more ground truth data – e.g., training more complex localness models for each localness definition – would not be tractable using this approach. We return to this issue in the Discussion section.

Supplementary Data Collection and Data Cleaning

After filtering out survey responses in which the input Twitter handle was invalid or the *LivesIn* city was outside the United States (or non-existent), we were left with 132 responses. The accidental input of a Twitter display name instead of a Twitter handle was a common reason for invalid responses. Since display names are non-unique, we had to filter these users out. On inspection of the raw data, we found that some people had filled in the *LivesIn* city also as a *Familiar* city, while many others did not. We assumed that people were familiar with cities where they lived and included the *LivesIn* city in the list of *Familiar* cities when it was not explicitly included.

Next, we downloaded the most-recent tweets for each of survey participants using the Twitter API, up to 3,200 tweets per user (3,200 is the maximum allowed by the API). We then deleted all tweets that did not have placetags within the United States. On examination of our placetags, we found that approximately 80% of tags were at the city scale, less than 2% of the total placetags were at a scale more local than the city and the rest of the tags were at the state scale or less granular.

We used the Google Geocoding API to determine the city, county and state from the place names in each placetag. In our evaluation of the localness assessment techniques, we eliminated from consideration any tweets whose placetags were at a scale more general than the given scale of analysis. In other words, when analyzing tweets at the city scale, we eliminated any tweets that were tagged at a scale more general than a city, and did the same for county- and state-scale analyses. Some of our participants exclusively geotagged at a state or higher scale, or they had provided only state-scale information in the survey. When performing the county-scale analysis, we excluded 14 such participants and were left with 118 participants. Additionally, two participants had specified only county-level information in the survey and they were removed from city-level analysis, leaving us with 116 valid responses at the city level.

Localness Assessment Techniques

Johnson et al. [22] provided an open-source implementation for all the four localness assessment techniques we consider here. However, since we were dealing with placetags and not geotags as in the case of Johnson et al., we had to re-implement some aspects of the four assessment techniques. We describe these adaptations below:

nDays: For every user, we took the available placetagged tweets and aggregated them into enumeration units at each analysis scale (i.e. we grouped them into cities, counties, and states). As per the definition of *nDays*, a user was considered local to all of the cities, counties, and states in which they posted at least one pair of tweets more than *n* days apart (n = 10).

Plurality: Similar to the case for *nDays*, the placetagged tweets for each user were first assigned to their corresponding cities, states, and counties. As per the definition of *Plurality* [22], a user was considered local to the city, county or state (or multiple regions in case of a tie) that contained the maximum number of tweets. This was done separately at each analysis scale (i.e. separately for cities, counties, and states).

LocationField: The *LocationField* method does not consider placetags like the other approaches; it simply involves looking at the location field in a user's profile. As such, we followed the standard practice for this method described above (relying on Google's Geocoding API). If the geocoder returned a region at the desired scale of analysis, we consider that the output of the method. If not, the method was considered to have returned no output.

GeometricMedian: We use the centroid of the bounding box of the place provided in the placetag as a representative point for the place. We then used the implementation of Johnson et al. to calculate the geometric median given this point representation.

RESULTS

The results of our city-, county-, and state-level analyses can be found in Table 2. Overall, these tables reveal extensive variation in the precision and recall of the various localness assessment techniques for each scale-definition combination.

We split the presentation of our results into two parts. We first provide a high-level overview of the most prominent trends present in Table 2, organizing our discussion by localness definition type (i.e. single location definitions and multiple location definitions). We then present a discussion of the types of failures we observed for each localness assessment technique, with an eye towards how the techniques may be improved.

Accuracy Trends

Before discussing the patterns in Table 2, we first seek to ensure all readers have the context necessary to understand the results in the table. The rows marked "C" indicate the coverage of a technique at a given scale, which is the percent of instances for which the technique could return any result. The "R" indicates the recall of the technique, which is the percent of total correct locations (as defined by the ground truth survey) returned by the technique. In other words, this counts as incorrect cases in which (1) no location was returned and (2) cases in which an explicit incorrect location was returned. The "P" indicates the precision of the

		nDays (10)			Plurality			GeometricMedian			LocationField		
		C	R	P	C	R	P	C	R	P	C	R	P
LivesIn	City Scale	94.0	81.9	22.5	100.0	74.1	73.5	93.1	71.6	76.9	85.3	47.4	78.6
	County Scale	94.9	87.3	36.9	100.0	84.7	84.0	93.2	81.4	87.3	85.6	55.1	87.8
	State Scale	96.2	94.7	50.8	100.0	97.7	97.7	96.2	90.2	93.7	86.4	72.0	92.2
VotesIn	City Scale	93.6	81.7	23.2	100.0	73.4	72.7	92.7	70.6	76.2	85.3	47.7	78.8
	County Scale	94.6	86.6	37.3	100.0	83.9	83.2	92.9	80.4	86.5	85.7	55.4	88.6
	State Scale	96.0	94.4	52.0	100.0	97.6	97.6	96.0	89.7	93.4	86.5	72.2	91.9
Familiarity	City Scale	75.1	51.8	31.4	45.1	36.6	80.3	42.0	34.2	81.5	27.2	23.7	87.1
	County Scale	74.0	57.1	44.8	53.9	48.4	89.1	50.2	45.7	90.9	33.8	32.0	94.6
	State Scale	73.0	66.0	57.7	61.4	60.9	99.2	59.1	56.3	95.3	47.9	46.0	96.1

Table 2: Coverage (C), recall (R) and precision (P) of each metric at the city, county and state scale. All values are percentages.

technique, which is calculated as the percent of correct locations *of the locations that were returned*.

Single Location Definitions

At the highest level, Table 2 shows that *LivesIn* and *VotesIn* results are very similar at all geographic scales, which is to be expected given the similarities of the two definitions. In fact, at the county and state scale, the *LivesIn* and *VotesIn* results are identical; two participants indicated that they voted in different cities than those in which they lived, but this was not true at the county scale.

The similarity of *LivesIn* and *VotesIn* means that the assessment techniques that work well for one will work well for the other, and vice versa. In this vein, we see that *Plurality* and *GeometricMedian* both have relatively good precision and recall (and coverage). Even at the city scale, both have precisions and recalls above 70%. At the county scale, both techniques have precisions and recalls above 80%, and we see continued improvement at the state scale. Interestingly, for state-scale single-location definitions of localness, our results suggest that *Plurality* exhibits *near perfect* performance.

The accuracy of the *LocationField* technique for our single location definitions is even better news for localness assessment. Table 2 shows that just by looking at the location field entry in a user's profile, one can achieve precisions at or above those of *Plurality* and *GeometricMedian*. Of course, Table 2 also shows that *LocationField* has a very poor recall, indicating that it often cannot return a local region for users. However, in the case of *LocationField*, this is much less of a concern: while only a small minority of Twitter users geographically reference their tweets – this number has been observed at 1-3% for geotags [16,34] – Table 2 shows that around 85% of users populate their location fields (and this is roughly the same percentage observed by Hecht et al. [16] as well). In other words, a recall of 47% is not a major issue if you can consider roughly 40 time the number of users in the first place; you will end up with a lot more users with local regions. That said, a number of localness projects consider only a population of users who frequently georeference their posts. In these cases, the *Plurality* and *GeometricMedian* should be preferred given their higher recalls. We discuss these dynamics in more detail below.

The accuracy of *nDays* is the worst of all the methods for single location definitions. With respect to recall, we see performance on par with *Plurality* and *GeometricMedian*. However, *nDays*' precision is terrible at all scales. Most of this low precision can be explained by a mismatch between the output of *nDays* and the nature of single location definitions of localness, a point that we discuss in our failure analysis sub-section below.

Multiple Location Definitions (Familiarity)

The most immediate pattern in the *Familiarity* results is that they are substantially worse across the board with respect to recall. A key issue here is that techniques that are designed to output one location – *GeometricMedian*, *Plurality* and *LocationField* – are not well suited to capturing the familiarity geography of users, a common need of research and applications in the localness space. However, even *nDays*, which by design regularly outputs multiple locations still has relative poor recall (and precision). We also performed a sensitivity analysis by setting the threshold of *Familiarity* to three instead of four on our survey's five-point familiarity scale. There were no meaningful changes in any of the relative patterns in Table 2 (e.g. recall expectedly dropped ~3-5% across the board, but the trends remained the same).

More generally, with regard to precision, we see roughly the same trends as we saw with the single location definitions: *Plurality*, *GeometricMedian*, and *LocationField* have quite high precisions (even greater than 80% at the city scale) and *nDays* is substantially worse.

Failure Analysis

As noted above, a key goal of our research project was not only to gain an understanding of the accuracy of localness assessment techniques, but also to inform the design of improvements to these techniques where possible. To address this goal, we examined the users for which each technique failed at each scale and attempted to determine the cause for the failures. In this section, we outline some of the common reasons for error for each of the assessment techniques.

LocationField

Although we saw that *LocationField* performed surprisingly well, especially given the size of the population of users that input *LocationField* information, there were some clear opportunities for improvement to the technique. In particular, we identified that the scale and information quality of location field entries were two of the main reasons the *LocationField* technique would fail.

With respect to the former, the findings of Hecht et al. [16] from 2010 appear to hold with regard to the use of the location field in 2017: some people disclose location information in their location fields at scales that is less precise than many applications need. For instance, in our case, we saw a number of location field entries at the state scale (e.g., "Florida, U.S"), which made it impossible for the location field technique to perform accurately for these users at the county or city level (and explains the increase in recall at the state scale).

We also saw the same phenomenon that Hecht et al. observed with regard to information quality: people certainly are continuing to put non-geographic information in their location fields (e.g., "close enough to help" and "in the studio"). However, it appears that geocoder accuracy may have improved somewhat since 2010 when Hecht et al. ran their study, and geocoders are now more capable of identifying this information as non-geographic information. This means that rather than returning a latitude and longitude coordinate for non-geographic location field entries, the geocoder more often returns no location, thereby increasing precision (although recall is still affected). In our case, we observed that 3 of our 7 non-geographic location field entries were misinterpreted as geographic information by the Google geocoder (e.g., "here and there" was located to Florida). For Hecht et al., the equivalent number was 82% (although comparisons must be made very cautiously given our limited sample size and the different origins of the geocoders). With respect to other causes of precision errors, we found that some users indicated that they were "local" to places that were different than those in their location field, likely because of outdated location field information.

Finally, the tendency to include multiple locations in the location field that was observed by Hecht et al. [16] was also observed in our study (e.g., "Fairfax, VA & Savannah, GA"). The geocoder reacted in different ways to multiple locations: sometimes it would return coordinates for just one of the locations, other times it would return no coordinates at all. It is likely that *Familiarity* performance in particular would increase if the *LocationField* workflow included accommodation for multiple locations.

Plurality

The *Plurality* method tends to fail for single location definitions when a person spends or has spent a good deal of time in multiple regions at a given scale. For instance, one user moved to a small town in Maryland from Chicago, but *Plurality* still located him in Chicago and Cook County, IL, likely due to the backlog of placetagged tweets in Chicago. Similarly, another participant who indicated that he lived in Pueblo, Colorado had most of his placetags from Denver, causing *Plurality* to assign the wrong city (and county). Overall, for single location definitions, *Plurality* may be particularly vulnerable to commuting (especially at the city and county scales) and moves (due to backlogged tweets).

Interestingly, the very property of *Plurality* that causes problems for single location definitions could in theory enable the approach to be more effective for *Familiarity*. However, the way *Plurality* has been defined means that it only picks the region that is the mode of the geographic distribution of the user's posts. If *Plurality* were extended to return more of the head of the distribution (e.g., the top-three regions), our results suggest that it could perform better for *Familiarity*. Indeed, Chicago, IL is a place that the Maryland participant reported a *Familiarity* of "*4*".

nDays

The poor precision of *nDays* for our single location definitions of local is largely due to a mismatch between the definition and the technique: *nDays* often gives multiple locations, but these definitions are only interested in one location. In this case, the methodological guidance we can provide comes more from our effort to provide structure around definitions of localness rather than specific low-level improvements that can be made to the *nDays* technique. Put simply, if a researcher or practitioner believes that a single location definition of localness is most appropriate for their project, they likely should not use an *nDays* approach.

However, *nDays* not only struggles for single location definitions, its recall and precision is also mediocre for *Familiarity*, the definition to which it is perhaps best suited. Examining the many false negatives and false positives available to us, we identified a few clear failure modes. With respect to false positives, it appears that *nDays* is quite susceptible to confusing tourism and business travel with self-reported familiarity. For instance, one participant is an advertising manager and clearly travels regularly; *nDays* reported 13 different states with which this person is supposedly familiar, and the participant only reported four states with which he was sufficiently familiar. We noticed a similar situation for a participant who described herself as an "urban hiker" and another that was a university professor.

For false negatives, *nDays* appears to be liable to have at least two issues: (1) *nDays* cannot report places with which people are familiar in which they did not tweet and (2) the n = 10 threshold was non-optimal in some cases. With respect to the former, many participants likely have lived in places with which they became quite familiar, but before they used placetagged tweets (or Twitter at all). For instance, one user who now lies in Muncie, IN wrote in response to the survey's final questions that s/he was very familiar with Largo, FL from growing up there, but *nDays* (nor any of the other techniques) was not able to capture her/his familiarity with Largo. This is a major, likely unsolvable issue for projects utilizing a *Familiarity* definition and for which recall is important, regardless of the localness assessment technique being used.

With respect to the n = 10 threshold, we saw at least three cases in which participants had a large number of placetags but *nDays* did not assign a single local region. On manual analysis, it was discovered that these people had a single burst of placetagged tweets in a short period of time (8, 9 and 6 days respectively), missing the n = 10 threshold.

GeometricMedian

GeometricMedian, by the virtue of definition that the median absolute deviation should be less than 30km, fails when a user has highly distributed posts. For example, one participant – a software developer – had only 35 placetags but equally distributed in the states of California, Colorado and New Jersey. In this case the *GeometricMedian* implementation failed to the meet the median absolute deviation clause and did not produce a result.

DISCUSSION

Implications for Localness Methodology

The work above provides new methodological guidance for the large literature associated with localness. First and foremost, it provides a lightweight framework for deciding upon and formally stating the type of localness being considered in a study or an application. Once the definition is decided, our empirical results can help researchers and practitioners choose a localness assessment technique, as well as understand the limitations of the chosen technique.

For instance, our findings suggest that a research project that defines localness using *LivesIn* (or *VotesIn*) is best served by utilizing either the *LocationField* technique or one of the *Plurality* or *GeometricMedian* techniques. As noted above, if the project is considering a general population of users, *LocationField* is the ideal choice as its much-lower recall will be more than offset by the much larger eligible sample. If the project considers people who frequently georeference their posts, however, then *Plurality* or *GeometricMedian* is the appropriate choice as their higher recalls are desirable in this case. Researchers should also consider making the improvements to each of these three methods that are discussed in the section above when doing so is possible.

Our results suggest that a project that requires the *Familiarity* definition now has an idea that the accuracy will likely be somewhat limited regardless of the assessment technique (in particular with regard to issues like pre-social media familiarity). Additionally, Table 2 reveals that the choice of assessment technique in this context may be complicated: using a technique that can only output one location will reduce the *Familiarity* definition problem to effectively a single location definition problem. However, Table 2 shows that these techniques may be the best we have available, with the *nDays* technique not living up to its potential as a higher-recall solution. Overall, it is clear that improved *Familiarity* approaches are needed, a point to which we return below.

As we discuss below in Limitations, we do caution that our study is small, and we strongly advocate that future work replicate our experiment in different contexts (including on other platforms of interest in the localness literature, e.g., Flickr, Yelp, Instagram). However, our results provide the best information available thus far on which to base the important definitional and assessment decisions in a localness project.

Gathering Ground Truth Data from Twitter

While the ability to observe behavioral traces on social media has proven tremendously important for research and practice ([44]), important research questions – e.g., those associated with localness and ground truth – cannot be answered without interacting with actual social media users at scale. In other words, many research questions require gathering information that is not available via APIs and public datasets.

In this light, our experience gathering ground truth location data from Twitter users is somewhat troubling. With *TSA Tracker*-like approaches not permitted for collecting information that is more complex than that which can be replied to in a tweet, the Twitter ads platform was likely our only straightforward choice. However, this platform has important limitations. First and foremost, the sampling procedure used by Twitter's ad optimization algorithms is a black box, and this not ideal for any study that interacts with these algorithms, including this one. Second, Twitter ads are sufficiently expensive that they will prevent some researchers from engaging in the exciting class of large-scale social media research that requires more than just behavior traces. Third, if alternatives to the Ads platform are not identified, this also means that certain types of projects in this space will not be tractable in general, e.g., a project to gather a large amount of training data for sophisticated localness assessment models that use machine learning (an exciting direction of future work for us).

Overall, we paid \$1.75 per participant in this study, which assuming our expected 2-3-minute completion time, is approximately \$35 per hour (but paid to Twitter, not the participant). Additionally, because ads are surfaced to users at a rate that is tied to the amount of money one bids [63], our throughput suffered: we were only able to collect about 11 responses per day, another impediment to scaling our study significantly. We perhaps could have increased our throughput by increasing our bid, but this would have raised costs significantly.

The Nature of Familiarity

We saw in Table 2 that all four localness assessment techniques do not perform as well in assessing the *Familiarity* of users when compared to the single location definitions. While part of this issue is methodological – and we have some suggestions for ways to improve the techniques below on this front – some portion of this might also be due to the complexity of the concept of familiarity. For instance, in the open response fields at the end of the survey, participants reported that they felt different levels of familiarity for different parts of cities and neighbourhoods, often at a very fine scale. Also, some people can be familiar with a very large number of places, so collecting this information correctly without overburdening the user remains a challenge.

Eventually, a goal for familiarity assessment research might be to be able to produce a fuzzy familiarity surface, rather than defining specific areas with which one is "familiar" and "not familiar".

Improving Localness Assessment Techniques

One exciting output of our empirical work is a clear roadmap for improving the performance of localness assessment techniques. A top priority on this front is improving *Familiarity* assessment, and our results point to some potential solutions. Above, we have already discussed ways to make *Plurality* and *LocationField* more amenable to *Familiarity* prediction through the use of a wider range of the region distribution and through supporting multiple locations entered in the location field, respectively. For *Plurality*, this change would be relatively trivial to implement. For *LocationField,* making this adaption would be more difficult, but likely still tractable. For instance, one could use a geo-parsing filter (e.g., [23]) prior to submitting the location field entry to the geocoder.

For *GeometricMedian*, there appears to be an opportunity to better support *Familiarity* as well. In particular, if one were to cluster placetags and apply the technique to each cluster, this would afford multiple output locations. It would also likely increase coverage and accuracy.

More generally, it is likely worthwhile to take a step back and consider entirely different paths towards *Familiarity* assessment. For instance, to capture familiarity with locations in users' pre-social media lives, one option would be to use natural language processing on the content of users' posts to detect familiarity with regions that are not present in users' placetag or location field information. For instance, affiliations with certain sports teams or the use of certain regional dialects [24] may be detectable using this approach. Of course, privacy concerns must be considered here as well.

Lastly, in our analysis of the errors in *nDays*, we noticed some opportunities to salvage its performance. For instance, *nDays* can be made more amenable to single location definitions by converting it into a "*maxDays*" approach, in which the region that is returned is the region that has the greatest temporal range in posts. Similarly, *nDays*' accuracy across the board might be improved by identifying means to adapt the *n* threshold to each users' posting behavior.

From Coordinate Geotags to Placetags

Tasse and colleagues' work on geotags represents a bit of a methodological sea change in the large literature that relies on VGI from Twitter. When their results were presented in Spring 2017, it was immediately clear that our study design needed to be converted from one that focused on geotags to one that focused on placetags. However, as noted above, all the localness assessment techniques had previously only been employed on geotagged data. Through this lens, the relatively strong performance of *Plurality* and *GeometricMedian* on the single location definitions is more important: it not only means that these techniques can reasonably replicate our ground truth data, but they can do this with placetagged data, allowing them to be used on a much more reliable type of Twitter geographic information. In other words, these techniques have a degree of "future proofing" for Twitter-based research.

Limitations and Future Work

While, as noted above, we believe our study provides an important ground truth-based lens on localness assessment, our findings must be put in the context of the size of the sample and the black box nature of sample (both of which are products of needing to use Twitter's ad platform). In particular, small differences in accuracy rates, e.g., those between *LocationField*, *Plurality* and *GeometricMedian,* should be interpreted with caution, and we do not make such comparisons above.

Beyond the black box ads placement algorithm, there are also additional sources of potential sampling bias. First, we only considered users that use placetags. Second, the altruistic and monetary incentives in our ads may have affected our sample. We did a basic pass for high-level sampling bias by comparing the *LivesIn* locations of our users to the population distribution of the United States at the state level and found that the two distributions are quite similar (e.g. California is #1 in both). However, there are many types of sampling bias that would not be detected using this type of approach. Finally, another type of bias might come from our decision to count *LivesIn* locations as *Familiar* locations. While we believe this was a valid assumption, future work may want to verify this.

Another interesting direction of future work involves expanding this study to other VGI domains to see if the findings can be replicated. Tasse et al. [49] noted that much geotagging activity now occurs on platforms like Instagram and Facebook, but studying these platforms is challenging because it is difficult to gather public data at scale. However, since it is likely that one would need to use the ad-based recruiting technique employed here, scale is already a limitation, meaning that these platforms may present roughly equivalent data collection challenges when it comes to replicating this study in particular.

Beyond studying other platforms, future work should seek to expand our study area beyond the United States. It would be interesting to see if there would be substantial differences in Table 2 if a country with a substantially different human geography (e.g., India) were studied.

Another promising direction of future work involves doing a systematic literature review [8,38] that considers definitions of localness beyond the computing literature. As we discuss above, our approach was less rigorous than a formal systematic literature review and only considered definitions of localness used in the computing literature.

CONCLUSION

In this paper, we worked to make more concrete the idea of "localness" as it is used in computing. We formalize three definitions of localness and assess the accuracy of localness assessment techniques that have been employed in the literature according to these definitions. We find that while some techniques are relatively accurate, others are very noisy, especially for certain definitions. Researchers and practitioners can utilize our results to provide methodological guidance when building applications or doing studies for which the concept of localness is important.

ACKNOWLEDGMENTS

We would like to thank Christine Lohmeier for her input to and feedback on this research. This research was supported in part by the Volkswagen Foundation through a Lichtenberg Professorship and the U.S. National Science Foundation (NSF IIS-CAREER, 1707296, NSF IIS-1707319, and IIS-1702440).

REFERENCES

1. Sofiane Abbar, Yelena Mejova, and Ingmar Weber. 2015. You tweet what you eat: Studying food consumption through twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3197–3206.
2. Saeed Abdullah, Elizabeth L. Murnane, Jean MR Costa, and Tanzeem Choudhury. 2015. Collective smile: Measuring societal happiness from geolocated images. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 361–374.
3. Mark Bjelland, Daniel Montello, Jerome Fellmann, Arthur Getis, and Judith Getis. 2013. *Human Geography: Landscapes of Human Activities*. McGraw-Hill Higher Education.
4. Peter Blickle. 2004. *Heimat: A Critical Theory of the German Idea of Homeland*. Camden House.
5. Ryan Compton, Craig Lee, Jiejun Xu, Luis Artieda-Moncada, Tsai-Ching Lu, Lalindra De Silva, and Michael Macy. 2014. Using publicly visible social media to build detailed forecasts of civil unrest. *Security informatics* 3, 1: 4.
6. Lee Cuba and David M. Hummon. 1993. A place to call home: Identification with dwelling, community, and region. *The sociological quarterly* 34, 1: 111–131.
7. Aron Culotta. 2014. Estimating county health statistics with twitter. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 1335–1344.
8. Tawanna R. Dillahunt, Xinyi Wang, Earnest Wheeler, Hao Fei Cheng, Brent Hecht, and Haiyi Zhu. 2017. The Sharing Economy in Computing: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW: 38:1–38:26. https://doi.org/10.1145/3134673
9. Melanie Eckle and João Porto de Albuquerque. 2015. Quality Assessment of Remote Mapping in OpenStreetMap for Disaster Management Purposes. In *ISCRAM*.
10. Friederike Eigler and Jens Kugele. 2012. *"Heimat": At the Intersection of Memory and Space*. Walter de Gruyter.
11. Lee Fiorio, Guy Abel, Jixuan Cai, Emilio Zagheni, Ingmar Weber, and Guillermo Vinué. 2017. Using Twitter Data to Estimate the Relationship between Short-term Mobility and Long-term Migration. In *Proceedings of the 2017 ACM on Web Science Conference*, 103–110.
12. Fabien Girardin, Francesco Calabrese, Filippo Dal Fiore, Carlo Ratti, and Josep Blat. 2008. Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive computing* 7, 4.
13. Michael F. Goodchild. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal* 69, 4: 211–221.
14. Jungkyu Han and Hayato Yamana. 2015. Why People Go to Unfamiliar Areas?: Analysis of Mobility Pattern Based on Users' Familiarity. In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services* (iiWAS '15), 28:1–28:10. https://doi.org/10.1145/2837185.2837190
15. Jungkyu Han and Hayato Yamana. 2017. Familiarity-aware POI Recommendation in Urban Neighborhoods. *Journal of Information Processing* 25: 386–396.
16. Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H. Chi. 2011. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '11)*, 237–246. https://doi.org/10.1145/1978942.1978976
17. Brent J. Hecht and Darren Gergle. 2010. On the localness of user-generated content. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 229–232.
18. Brent J. Hecht and Monica Stephens. 2014. A Tale of Cities: Urban Biases in Volunteered Geographic Information. *ICWSM* 14: 197–205.
19. Brent Hecht, Johannes Schöning, Thomas Erickson, and Reid Priedhorsky. 2011. Geographic Human-computer Interaction. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (CHI EA

'11), 447–450. https://doi.org/10.1145/1979742.1979532
20. Christopher Ingraham. How Trump's nationwide voter data request could lead to voter suppression. *The Washington Post*. Retrieved September 16, 2017 from https://www.washingtonpost.com/news/wonk/wp/2017/06/30/how-trumps-nationwide-voter-data-request-could-lead-to-voter-suppression
21. Isaac Johnson, Jessica Henderson, Caitlin Perry, Johannes Schöning, and Brent Hecht. 2017. Beautiful… but at What Cost?: An Examination of Externalities in Geographic Vehicle Routing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2: 15. https://doi.org/10.1145/3090080
22. Isaac L. Johnson, Subhasree Sengupta, Johannes Schöning, and Brent Hecht. 2016. The geography and importance of localness in geotagged social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, 515–526. https://doi.org/10.1145/2858036.2858122
23. Christopher B. Jones and Ross S. Purves. 2008. Geographical information retrieval. *International Journal of Geographical Information Science* 22, 3: 219–228.
24. David Jurgens, Tyler Finethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. *ICWSM* 15: 188–197.
25. Marina Kogan, Leysia Palen, and Kenneth M. Anderson. 2015. Think local, retweet global: Retweeting by the geographically-vulnerable during Hurricane Sandy. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 981–993.
26. Vikas Kumar, Saeideh Bakhshi, Lyndon Kennedy, and David A. Shamma. 2017. Modeling Characteristics of Location from User Photos. In *Proceedings of the 2017 ACM Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces*, 1–6.
27. Vikas Kumar, Saeideh Bakhshi, Lyndon Kennedy, and David A. Shamma. 2017. Adaptive City Characteristics: How Location Familiarity Changes What Is Regionally Descriptive. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 131–139.
28. Linna Li, Michael F. Goodchild, and Bo Xu. 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *cartography and geographic information science* 40, 2: 61–77.
29. Setha M. Low and Irwin Altman. 1992. Place Attachment. In *Place Attachment*. Springer, Boston, MA, 1–12. https://doi.org/10.1007/978-1-4684-8753-4_1
30. Pamela J. Ludford, Reid Priedhorsky, Ken Reily, and Loren Terveen. 2007. Capturing, sharing, and using local place information. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07)*, 1235–1244. https://doi.org/10.1145/1240624.1240811
31. Momin M. Malik, Hemank Lamba, Constantine Nakos, and Jurgen Pfeffer. 2015. Population bias in geotagged tweets. *People* 1, 3,759.710: 3–759.
32. Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist. 2011. Understanding the Demographics of Twitter Users. *ICWSM* 11: 5th.
33. Aline Morais and Nazareno Andrade. 2014. The Relevance of Annotations Shared by Tourists and Residents on a Geo-Social Network During a Large-Scale Touristic Event: The Case of São João. In *COOP 2014-Proceedings of the 11th International Conference on the Design of Cooperative Systems, 27-30 May 2014, Nice (France)*, 393–408.
34. Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M. Carley. 2013. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In *ICWSM*.
35. Mohamed Musthag and Deepak Ganesan. 2013. Labor dynamics in a mobile micro-task market. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, 641–650. https://doi.org/10.1145/2470654.2470745
36. Mor Naaman, Amy Xian Zhang, Samuel Brody, and Gilad Lotan. 2012. On the Study of Diurnal Urban Routines on Twitter. In *ICWSM*.
37. Jeffrey Nichols and Jeon-Hyung Kang. 2012. Asking Questions of Targeted Strangers on Social Networks. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (CSCW '12), 999–1002. https://doi.org/10.1145/2145204.2145352
38. Luke Pittaway. 2011. *Systematic Literature Reviews. In The SAGE Dictionary of Qualitative Management Research, Richard Thorpe and Robin Holt (Eds.)*. SAGE Publications Ltd., 217–218.
39. Barbara Poblete, Ruth Garcia, Marcelo Mendoza, and Alejandro Jaimes. 2011. Do all birds tweet the same?: characterizing twitter around the world. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, 1025–1030.
40. Adrian Popescu and Gregory Grefenstette. 2010. Mining User Home Location and Gender from Flickr Tags. In *ICWSM*.
41. Harold M. Proshansky, Abbe K. Fabian, and Robert Kaminoff. 1983. Place-identity: Physical world socialization of the self. *Journal of environmental psychology* 3, 1: 57–83.
42. Edward Relph. 1997. Sense of place. *Ten geographic ideas that changed the world*: 205–226.
43. Christopher J. Riederer, Sebastian Zimmeck, Coralie Phanord, Augustin Chaintreau, and Steven M. Bellovin. 2015. "I don't have a photograph, but you can have my footprints.": Revealing the Demographics

of Location Data. In *Proceedings of the 2015 ACM on Conference on Online Social Networks* (COSN '15), 185–195. https://doi.org/10.1145/2817946.2817968

44. Derek Ruths and Jürgen Pfeffer. 2014. Social media for large studies of behavior. *Science* 346, 6213: 1063–1064.
45. Justin Sablich. How to Explore a City Like a Local Using Your Smartphone. *The New York Times*. Retrieved September 16, 2017 from https://www.nytimes.com/2017/09/08/travel/local-tourist-apps.html
46. David Seamon. 1979. *A Geography of the Lifeworld*.
47. Shilad W. Sen, Heather Ford, David R. Musicant, Mark Graham, Oliver SB Keyes, and Brent Hecht. 2015. Barriers to the localness of volunteered geographic information. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 197–206.
48. Fritz Steele. 1981. *The Sense of Place*. CBI Pub Co.
49. Dan Tasse, Zichen Liu, Alex Sciuto, and Jason I. Hong. 2017. State of the Geotags: Motivations and Recent Changes. In *ICWSM*, 250–259.
50. Jacob Thebault-Spieker, Loren G. Terveen, and Brent Hecht. 2015. Avoiding the south side and the suburbs: The geography of mobile crowdsourcing markets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 265–275.
51. Yi-Fu Tuan. 2013. *Landscapes of fear*. Pantheon.
52. Clare L. Twigger-Ross and David L. Uzzell. 1996. Place and identity processes. *Journal of environmental psychology* 16, 3: 205–220.
53. Alessandro Venerandi, Giovanni Quattrone, Licia Capra, Daniele Quercia, and Diego Saez-Trumper. 2015. Measuring urban deprivation from user generated content. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 254–264.
54. Ryen White and Georg Buscher. 2012. Characterizing local interests and local knowledge. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1607–1610.
55. Daniel R. Williams, Michael E. Patterson, Joseph W. Roggenbuck, and Alan E. Watson. 1992. Beyond the commodity metaphor: Examining emotional and symbolic attachment to place. *Leisure sciences* 14, 1: 29–46.
56. Shaomei Wu, Shenwei Liu, Dan Cosley, and Michael Macy. 2011. Mining collective local knowledge from google mymaps. In *Proceedings of the 20th international conference companion on World wide web*, 151–152.
57. Amy X. Zhang and Scott Counts. 2015. Modeling ideology and predicting policy change with social media: Case of same-sex marriage. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2603–2612.
58. Kaiqi Zhao, Gao Cong, and Aixin Sun. 2016. Annotating Points of Interest with Geo-tagged Tweets. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 417–426.
59. Dennis Zielstra, Hartwig H. Hochmair, Pascal Neis, and Francesco Tonini. 2014. Areal delineation of home regions from contribution and editing patterns in OpenStreetMap. *ISPRS International Journal of Geo-Information* 3, 4: 1211–1233.
60. Like a Local. Retrieved September 17, 2017 from https://www.likealocalguide.com/
61. Local Foods. Retrieved September 17, 2017 from https://www.nal.usda.gov/aglaw/local-foods
62. The Twitter Rules. Retrieved September 17, 2017 from https://support.twitter.com/articles/18311
63. Twitter Ads. Retrieved September 17, 2017 from https://dev.twitter.com/ads/campaigns/target-bidding