

Using Social Media Data to Understand Cities

Dan Tasse, Jason I. Hong

Human-Computer Interaction Institute

Carnegie Mellon University

dantasse@cmu.edu, jasonh@cs.cmu.edu

Abstract

Understanding urban dynamics is crucial for a number of domains, but it can be expensive and time consuming to gather necessary data. The rapid rise of social media has given us a new and massive source of geotagged data that can be transformative in terms of how we understand our cities. In this position paper, we describe three opportunities in using geotagged social media data: to help city planners, to help small businesses, and to help individuals adapt to their city better. We also sketch some possible research projects to help map out the design space, as well as discuss some limitations and challenges in using this kind of data.

Introduction

Over half of the world's population now lives in cities [33], and understanding the cities we live in has never been more important. Urban planners need to plan future developments, transit authorities need to optimize routes, and people need to effectively integrate into their communities.

Currently, a number of methods are used to collect data about people, but these methods tend to be slow, labor-intensive, expensive, and lead to relatively sparse data. For example, the US census cost \$13 billion in 2010 [11], and is only collected once every ten years. The American Community Survey is collected annually, and cost about \$170 million in 2012, but only samples around 1% of households in any given year [19]. While data like this can benefit planners, policy makers, researchers, and businesses in understanding changes over time and how to allocate resources, today's methods for understanding people and cities are slow, expensive, labor-intensive, and do not scale well.

Researchers have looked at using proprietary call detail records (CDRs) from telecoms to model mobility patterns [3, 18, 24] and other social patterns, such as the size of one's social network and one's relationship with others [36]. These studies leverage millions of data points; however, these approaches also have coarse location granularity (up to 1 sq. mile), are somewhat sparse (CDRs are recorded only when a call or SMS is made), have minimal context (location, date, caller, callee), and use data not generally available to others. Similarly, researchers have also looked at having participants install custom apps. However, this approach has challenges in scaling up to cities, given the large number of app users needed to get useful data. Corporations have also surreptitiously installed software on people's smartphones (such as CallerIQ [34] and Verizon's Precision Market Insights [35]), though this has led to widespread outcry due to privacy concerns.

We argue that there is an exciting opportunity for creating new ways to conceptualize and visualize the dynamics, structure, and character of a city by analyzing the social media its residents already generate. Millions of people already use Twitter, Instagram, Foursquare, and other social media services to update their friends about where they are, communicate with friends and strangers, and record their actions. The sheer quantity of data is also tantalizing: Twitter claims that its

users send over 500 million tweets daily, and Instagram claims its users share about 60 million photos per day [23]. Some of this media is geotagged with GPS data, making it possible to start inferring people's behaviors over time. In contrast to CDRs from telcos, we can get fine-grained location data, and at times beyond when people make phone calls. In contrast to having people install custom apps (which is hard to persuade people to do), we can leverage social media data that millions of people are already creating every day.

We believe that this kind of geotagged social media data, combined with new kinds of analytics tools, will let urban planners, policy analysts, social scientists, and computer scientists explore how people actually use a city, in a manner that is cheap, highly scalable, and insightful. These tools can shed light onto the factors that come together to shape the urban landscape and the social texture of city life, including municipal borders, demographics, economic development, resources, geography, and planning.

As such, our main question here is, how can we use this kind of publicly visible, geotagged social media data to help us understand cities better? In this position paper, we sketch out several opportunities for new kinds of analytics tools based on geotagged social media data. We also discuss some longer-term challenges in using this kind of data, including biases in this kind of data, issues of privacy, and fostering a sustainable ecosystem where the value of this kind of data is shared with more people.

Opportunities

In this section, we sketch out some design and research opportunities, looking at three specific application areas. Many of the ideas we discuss below are speculative. We use these ideas as a way of describing the potential of geotagged social media data, as well as offering possible new directions for the research community.

For City Planners

First, and perhaps most promisingly, we believe that geotagged social media data can offer city planners and developers better information that can be used to improve planning and quality of life in cities. This might include new kinds of metrics for understanding people's interactions in different parts of a city, new methods of pinpointing problems that people are facing, and new ways of identifying potential opportunities for improving things.

Mapping Socioeconomic Status

It is important for governments to know the socioeconomic status of different sections of their jurisdiction in order to properly allocate resources. In England, for example, the government uses the Index of Multiple Deprivation (IMD) to measure where the problems of poverty are the most severe, and to therefore mitigate those effects. The IMD is based on surveys and other statistics collected by different areas

of government. However, even in developed countries, surveys and statistics can be difficult and expensive to collect.

Past work with cell phone call logs suggests that it is possible to find valuable demographic information using communication records. For example, Eagle et al found that network diversity in phone calls correlated with the IMD [16]. A recent project by Smith-Clarke et al [41] explored the use of call logs to map poverty. Especially in developing countries, maps of poverty are often very coarse and out of date. This is not a simple problem of improving an already-good metric; data at different granularities can tell very different stories. For example, Figure 1 shows two different maps of socioeconomic data in the UK, one very coarse-grained and one fine-grained.

However, call log data, while more complete than surveys, still presents the limitations mentioned earlier: it is proprietary, coarse-grained, and lacking context and transparency. Much social media data, on the other hand, is and publicly visible and accessible to researchers. Different forms of social media data also offer their own advantages. For example, Twitter users follow other users, and Foursquare checkins often have “likes” or comments attached.

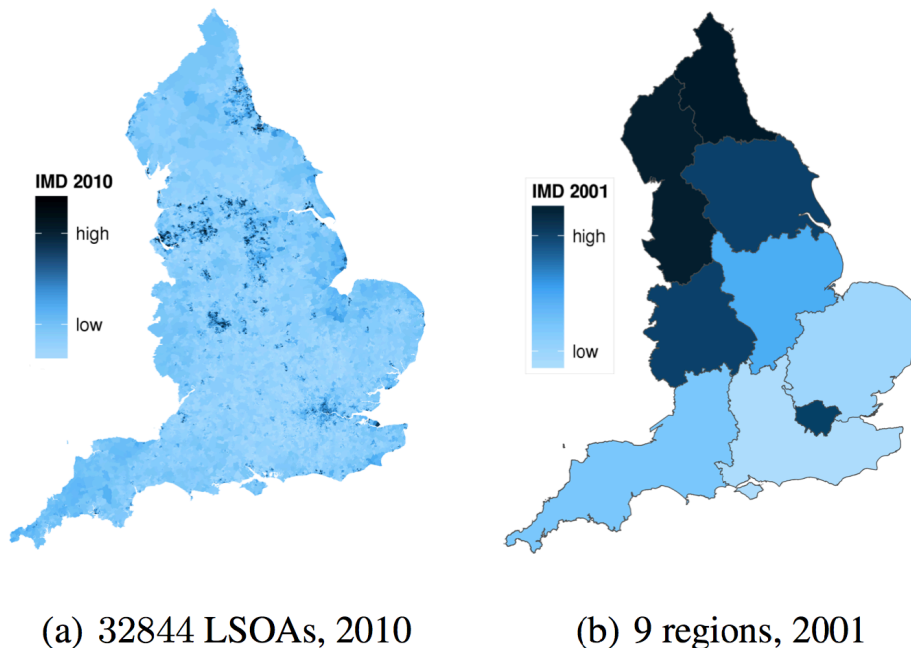


Figure 1 - Left: accurate data about socio-economic deprivation in England; darker indicates more deprivation. Right: much coarser, out-of-date information about the same index. Figure from [41].

Mapping Quality of Life

Socioeconomic status, however, is not the only metric that matters. A community can be poor but flourishing, or rich but suffering. Other metrics like violence, pollution, location efficiency, and even community coherence are important for cities to track. Some of these are even more difficult to track than socioeconomic status.

We believe some aspects of quality of life can be modeled using geotagged social media data. For example, approximating violence may be possible by analyzing the content of posts. Choudhury et al showed that psychological features associated with desensitization appeared over time in tweets by people affected by the Mexican Drug War [10]. Other work has found that sentiments in tweets are correlated with general socio-economic wellbeing [39]. Measuring and mapping posts that contain these emotional words may help us find high crime areas and measure the change over time.

As another example, location efficiency, or the total cost of transportation for someone living in a certain location, can be approximated by sites like Walkscore.com. However, Walkscore currently only relies on the *spaces*, that is, where services are on the map. It does not take into account the *places*, the ways that people use these services. A small market may be classified as a “grocery store”, but if nobody goes to it for groceries, maybe it actually fails to meet people’s grocery needs. We believe geotagged social media data can be used as a new way of understanding how people actually use places, and thereby offer a better measure of location efficiency.

Mapping Mobility

One more analysis that could be useful for city planners is in understanding the mobility patterns of people in different parts of different cities. This kind of information can help, for example, in planning transportation networks [27]. Mobility can help planners with social information as well, such as how public or private people feel a place is [42].

Previously, mobility information has been gathered from many sources, but they all lack the granularity and ease of collection of social media data. Cell tower data has been used to estimate the daily ranges of cell phone users [1]. At a larger scale, data from moving dollar bills has been used to understand the range of human travel [5]. Among other applications, this data could be used for economic purposes, such as understanding the value of centralized business districts like the Garment District in New York [43]. It seems plausible that pre-existing social media data could help us find the similar information without needing people to enter dollar bill serial numbers or phone companies to grant access to expensive and sensitive call logs. Geotagged social media data is also more fine-grained, allowing us to pinpoint specific venues that people are going to.

“Design Patterns” for Cities

Originating in the field of architecture, *design patterns* are good and reusable solutions to common design problems. Geotagged social media data offers new ways of analyzing physical spaces and understanding how the design of those spaces influences people’s behaviors.

For example, in his book *A Pattern Language* [1], Alexander and colleagues present several kinds of patterns characterizing communities and neighborhoods. These

patterns include Activity Nodes (community facilities should not be scattered individually through a city, but rather clustered together), Promenades (a center for its public life, a place to see people and to be seen), Shopping Streets (shopping centers should be located near major traffic arteries, but should be quiet and comfortable for pedestrians), and Night Life (places that are open late at night should be clustered together).

By analyzing geotagged social media data, we believe it is possible to extract known design patterns. One possible scenario is letting people search for design patterns in a given city, e.g. “where is Night Life in this city?” or “show the major Promenades”. Another possibility is to compare the relationship of different patterns in different cities, as a way of analyzing why certain designs work well and others do not. For example, one might find that areas that serve both as Shopping Streets and as Night Life are well correlated with vibrant communities and general well-being.

For Small Businesses

Understanding one’s customers is crucial for owners of small businesses, like restaurants, bars, and coffee shops. We envision two possible scenarios for how geotagged social media data can help small business owners.

Knowing Demographics of Existing Customers

Small businesses cannot easily compete with big-box stores in terms of data and analytics about existing customers. This makes it difficult for small businesses to tailor their services and advertisements effectively.

Businesses can already check their reviews on Yelp or Foursquare. We believe that geotagged social media data can offer different kinds of insights about the behaviors and demographics of customers. One example would be knowing what people do before and after visiting a given venue. For example, if a coffee shop owner finds that many people go to a sandwich shop after the coffee shop, they may want to partner with those kinds of stores or offer sandwiches themselves. This same analysis could be done with classes of venues, for example, cafés or donut shops.

As another example, an owner may want to do retail trade analysis [22], which is a kind of marketing research for understanding where a store’s customers are coming from, how many potential customers are in a given area, and where one can look for more potential customers. Some examples include quantifying and visualizing the flow and movement of customers in the area around a given store. Using this kind of analysis, a business can select potential store locations, identify likely competitors, and pinpoint ideal places for advertisements.

Currently, retail trade analysis is labor intensive, consisting of numerous observations by field workers (e.g. watching where customers come from and where they go, or shadowing customers) or surveys given to customers. Publicly visible social media data offers a way of scaling up this kind of process, and

extending the kind of analysis beyond just the immediate area. For example, one could analyze more general kinds of patterns. For example, what are the most popular stores in this area, and how does the store in question stack up? How does the store in question compare against competitors in the same city?

Knowing more about the people themselves would be useful as well. For example, in general, what kinds of venues are most popular for people who come to this store? If a business finds that all of its patrons come from neighborhoods where live music is popular, they may want to consider hosting musicians themselves. All of the information offered by a service like this would have to be rather coarse, but it could provide new kinds of insights for small businesses.

Knowing Where to Locate a New Business

New businesses often have many different potential locations, and evaluating them can be difficult. Public social media data could give these business owners more insight into advantages and disadvantages of their potential sites. For example, if they find that a certain neighborhood has many people who visit Thai restaurants in other parts of the city, they could locate a new Thai restaurant there.

For Individuals

There are also many opportunities for using geotagged social media to benefit individuals as well. Below, we sketch out a few themes.

Feeling At Home In New Cities

Moving to a new city can make it hard for people to be part of a community. The formally defined boundaries of neighborhoods may help people understand the *spaces* where they live, but not so much the socially constructed *places* [20]. Currently it is difficult for non-locals to know the social constructs of a city as well as locals do. This is particularly important when someone changes places, either as a tourist or a new resident.

Imagine a new person arriving in a diverse city like San Francisco, with multiple neighborhoods and sub-neighborhoods. It would be useful for that person to know the types of people who live in the city, and where each group goes: students go to this neighborhood in the evening, members of the Italian community like to spend time in this area, families often live in this neighborhood but spend time in that neighborhood on weekends.

Some work has been done in this area, but we believe it could be extended. Komninos et al collected data from Foursquare to examine patterns over times of day and days of the week [28], showing the daily variations of people's activity in a city in Greece. They showed when people are checking in, and at what kind of venue, but not who was checking in. Cheng et al, too, showed general patterns of mobility and times of checkins [9], but these statistics remain difficult for individuals to interpret.

Related, the Livehoods project [8] and Hoodsquare [45] both look at helping people understand their cities by clustering nearby places into neighborhoods. Livehoods used Foursquare checkins, clustering nearby places where the same people often checked in. Hoodsquare considered not only checkins but also other factors including time, location category, and whether tourists or locals attended the place. Both of these projects would be helpful for people to find their way in a new city, but even their output could be more informative. Instead of simply knowing that a neighborhood has certain boundaries, knowing why those boundaries are drawn or what people do inside those boundaries would be helpful. Andrienko et al [2] also describe a visual analytic approach to finding important places based on mobility data that could help newcomers understand which places were more popular or important.

Another approach to helping people get to know the city is to help them get to know the people in the city, rather than the places. We look to the work of Joseph et al [26] who used topic models to assign people to clusters such as “sports enthusiast” or “art enthusiast”. We could imagine this information being useful for individuals to find other like-minded people.

Discovering New Places

In the previous section, we described applications to help people become accustomed to a new city. However, sometimes the opposite problem may arise: people become too comfortable in their routines and they want to discover new places. Some recent projects have helped people to discover new places based on actions that can be performed there [13] or aspects that people love about the place [7]. However, we believe that this idea can be pushed further. Perhaps combining a person’s current mobility patterns with visualizations of other people’s mobility patterns would help a person to put their own actions in context. They may realize that there are entire social flows in the city that they did not even know existed.

Understanding People, Not Just Places

Tools like Yelp and Urbanspoon already exist to help people find places they would like to go or discover new places that they didn’t know about. Previous work like Livehoods [8] also worked to enable a richer understanding of the places there. The benefit of incorporating social media data, though, is that users can start to understand the people who go to places, not just the places themselves.

Potential Research in this Design Space

In this section we sketch out some potentially interesting research projects in using geotagged social media data. Our goal here is to map out different points in the overall design space, which can be useful in understanding the range of applications as well as the pros and cons of various techniques.

Who Goes Here?

The approach is simple: select a geographic region (which may be as small as an individual store) and retrieve all the tweets of the people who have ever tweeted

there. Then, compute a heat map or other geographic visualization of the other places that they tweet. This kind of visualization could help elucidate all the places associated with a given place, and could be useful to small business owners or managers of a larger organization like a university.

Groceryshed

Watershed maps show where water drains off to lakes and oceans. Researchers have extended this metaphor to map “laborsheds” and “paradesheds” [4] to describe where people who work in a certain area come from, or people who attend a certain parade. We could extend this metaphor even further to describe “sheds” of smaller categories of business, such as Thai restaurants.

More interestingly, we could map change over time in various “sheds”. This could be particularly important for grocery stores. Places that are outside any “groceryshed” could be candidate areas for a new store, and showing the change in people’s behavior after a new store goes in could help measure the impact of that store. The content of Tweets or other social data could also show how people’s behavior changed after a new grocery store was put in.

How Is This Place Relevant To Me?

We envision a system that can convey not only what a place *is*, but also what it *means*. Imagine a user looking up a particular coffee shop. Currently, they can look up the coffee shop’s web site, find basic information like store hours, and find reviews on sites like Yelp. Using geotagged social media data, however, we could surface information like:

- Your friends (or people you follow on Twitter) go here five times per week
- Friends of your friends go here much more than nearby coffee shops
- People who are music enthusiasts like you (using topic modeling as in [26]) often go to this coffee shop
- You’ve been to three other coffee shops that are very similar to this one
- People who tweet here show the same profiles of emotions as your tweets

These could help people form a deeper relationship with a place than one based on locality or business type alone. In addition, we could pre-compute measures of relevance for a particular user, giving them a map of places that they might enjoy.

Human Network Visualizations

We can go beyond assigning people to groups by topics, also showing where they tweet over time. This could help people understand the dynamics of neighborhoods where, for example, one group of more affluent people are pricing out a group of previous residents. One interesting work in this area is the Yelp word maps [44], which show where people write certain words, like “hipster”, in reviews of businesses. However, this still describes the places; using social media data, we could show maps that describe the people. Instead of a map of locations tagged as “hipster”, we could identify groups of people based on their check-in patterns and tag where they go during the day. Perhaps the hipsters frequent certain coffee shops

in the morning and certain bars at night, but during the day hang out in parks where they do not check in.

Cheaper, Easier, and Richer Demographics

For all of our groups and stakeholders, it is important to understand demographic information of city regions. We could improve the process in two main ways. First, we could make it cheaper and easier to infer existing demographic information. We plan to investigate whether Twitter volume, or volume of certain topics of discussion, correlates with deprivation or other measures of socioeconomic status or quality of life. If so, then we can use the social media measures as proxies for real measures, and thereby collect that information cheaply and in real time.

Second, we plan to create more descriptive demographics. In each neighborhood, we can calculate each person's average distance traveled, radius of gyration, and other measures of mobility. We could either simply output statistics or we could create interactive visualizations that show the daily movements of people in each neighborhood.

Data Sources and Limitations

For our research, we are currently using data from Twitter, due to its richness and volume. Twitter claims over 500 million tweets are posted per day [29]. Furthermore, this data is publicly available. While only a small fraction of these tweets are geotagged, even a small fraction of tweets from any given day forms a large and rich data set. Furthermore, past work suggests that the sampling bias from only selecting these tweets is limited [38].

Biases in the Data

Of course, neither Twitter nor Foursquare provides an exactly accurate view of people's mobility. Both are communicative media, not purely representative. For example, Rost et al [40] report that the Museum of Modern Art in New York has more check-ins than Atlanta's airport, even though the airport had almost three times as many visitors in the time period that was studied. In some cases this will not matter; if we are clustering people, for example, grouping people who communicate that they go to the same places will be nearly as successful as grouping people who actually go to the same places. In other cases, we hope to minimize this bias by primarily comparing similar businesses, but we must remain aware of it.

Second, these media can be performative as well. People check in not because the check-in represents their location the most accurately, but because they want to show off that they have performed the check-in [6]. Sometimes people may avoid checking in for the same reason; they do not want it to be known that they checked in at a certain venue, like a fast food restaurant [30].

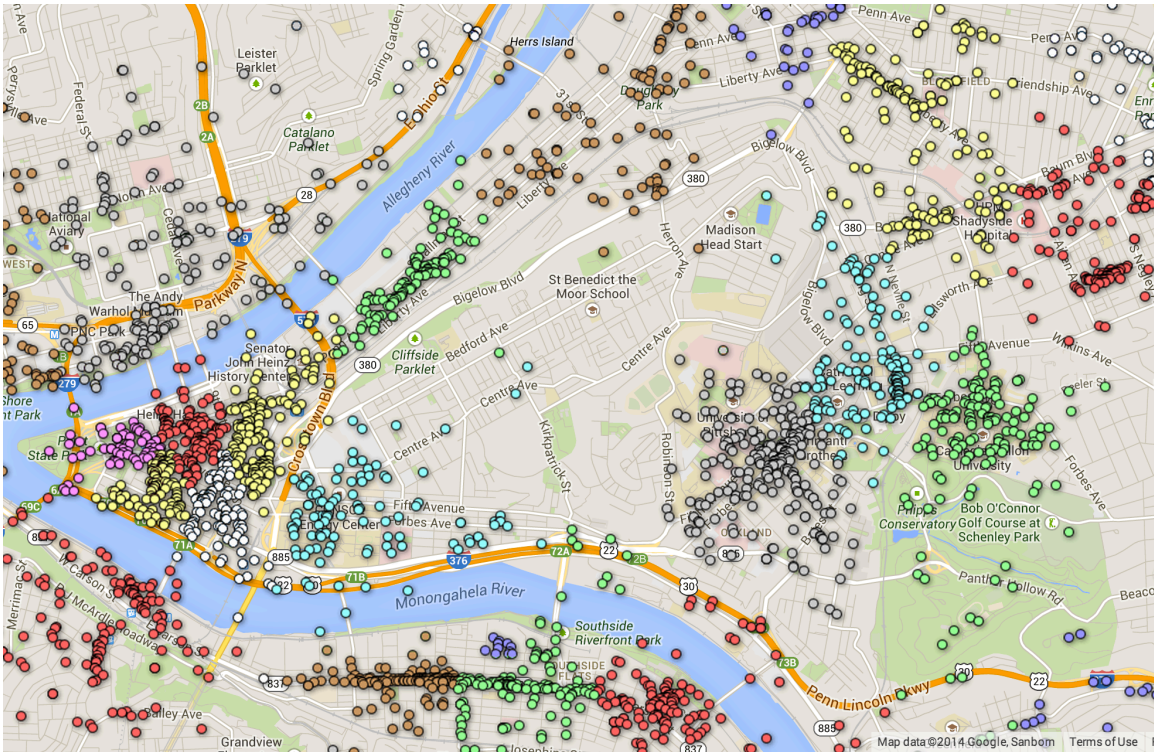


Figure 2 – Several livehoods in Pittsburgh. Each dot represents one Foursquare location. Venues with the same color have been clustered into the same livehood (the colors are arbitrary). The center of the map is a residential area where people with relatively low socioeconomic status live. There is also a notable lack of foursquare data in this area.

Third, there are currently several demographic biases in these data sets. For example, Twitter, Flickr, and Foursquare are all more active per capita in cities than outside them [21]. Furthermore, these social media sites are all used by predominantly young, male, technology-savvy people.

One effect of this bias is shown in Figure 2. This screenshot shows the results of some of our clusters in Livehoods [8], with each dot representing a venue, and different colors representing different clusters. Note the lack of data in the center of the figure. This area is Pittsburgh’s Hill District, a historic area which was the center of Pittsburgh’s jazz scene in the early 20th century. Currently, the median income for residents in the Hill District is far lower than in other parts of Pittsburgh. This neighborhood has also seen some revitalization with new senior housing, a library, a YMCA, several small office buildings, and a grocery store. However, there is still a notable lack of geotagged social media data in this area.

In short, while geotagged social media data has great potential, we also need to be careful because this data may not necessarily be representative of all people that live in a city. It is possible that this demographic bias may solve itself over time. Currently, about 58% of Americans have a smartphone [37], and the number is rapidly growing. However, it may still be many years before demographics are more representative, and there is still no guarantee that the demographics of geotagged

social media data will follow. For now, one approach is to look for ways of accounting for these kinds of biases in models. Another approach is to make clearer what the models do and do not represent.

Privacy Implications

Privacy is also a clear concern in using geotagged social media to understand cities. From an Institutional Review Board (IRB) perspective, much of social media data is considered exempt, because the researchers do not directly interact with participants, the data already exists, and the data is often publicly visible. However, as researchers, we need to go beyond IRB and offer stronger privacy protections, especially if we make our analyses available as interactive tools.

Here, there are at least two major privacy concerns. The first is making it easy to access detailed information about specific individuals. Even if a person's social media data is public data, interactive tools could make a person's history and inferences on that history more conspicuously available. Some trivial examples include algorithms for determining a user's home and work locations based on their tweets [28]. More involved examples might include other aspects of their behaviors, such as their activities, preferences, and mobility patterns. In the Livehoods project, we mitigated this aspect of user privacy by only presenting information about locations, not people. We also removed all venues labeled as private homes.

Second, we need to be more careful and more thoughtful about the kinds of inferences that algorithms can make about people, as these inferences can have far-reaching effects, regardless of whether they are accurate or not. There are numerous examples of inferences outside of geotagged social media that might be viewed as intrusive, embarrassing, or even harmful. For example, Jernighan and Mistree [25] found that, given a social network with men who did not self-report their sexuality, they could identify gay men simply by analyzing the self-reported sexuality of an individual's friends. As another example, the New York Times reported on how Target had developed algorithms that could infer if a customer was pregnant [14]. A separate New York Times article reported on how people were assessed for credit risks based on what they purchased as well as where they shopped [15].

It is important to note that these risks are not just hypothetical. At least one person had his credit card limit lowered, with the explanation that "other customers who have used their card at establishments where you recently shopped have a poor repayment history with American Express" [12].

It is not yet clear what the full range and extent of inferences is with geotagged social media. A significant concern is that inferences like the ones above and ones using social media data can become a proxy for socioeconomic status, gender, or race, inadvertently or even intentionally skirting around charged issues under the guise of an "objective" algorithm. It is also not clear if there is anything that can be done about these kinds of inferences, given that these inferences would be done on

private servers. It is unlikely that there is a technical solution to this problem. It may very well be the case that society will require new kinds of laws governing how these inferences are used, rather than trying to control the inferences themselves.

Creating a Sustainable Ecosystem

We hope to find a way to co-create value both to social media users and to people in our work. As it exists now, value flows only from users to marketers and analysts. To create a more sustainable tool, and to avoid impinging on users' freedoms, it is important that the users gain some benefit from any system we create as well. Some of our projects point in this direction, especially the ones aimed at individual users. People may be more amenable to a tool that offers businesses insights based on their public tweets if they can have access to those insights as well.

A successful example of co-creation of value is Tiramisu [46]. This app aimed to provide real-time bus timing information by asking people to send messages to a server when they were on a bus. Users were allowed to get more information if they shared more information. In contrast, OneBusAway [17] provides real-time bus information using sensors that are installed on buses. Using a collaborative approach, it may not be necessary to implement a costly instrumentation project. In addition, people may feel more ownership of a system if they contribute to it.

Conclusion

Despite these challenges, social media remains a potentially transformative, yet untapped, source of geographic data. Analytics tools based on geotagged social media data can help city planners plan future developments, businesses understand the pulse of their customers, and individuals fit into new cities much more seamlessly. As our cities grow exponentially and more of the world moves into heavily urbanized areas, instead of using costly methods to understand our cities, researchers of all kinds will be able to mine existing data to understand social patterns that are already there.

References

1. Alexander, Christopher, et al. *A Pattern Language: Towns, Buildings, Construction*. (1977).
2. Andrienko, G., Andrienko, N., Hurter, C., Rinzivillo, S., and Wrobel, S. From Movement Tracks through Events to Places: Extracting and Characterizing Significant Places from Mobility Data. *IEEE VAST*, (2011), 159–168.
3. Becker, R.A., Cáceres, R., Hanson, K., and Varshavsky, A. A Tale of One City: Using Cellular Network Data for Urban Planning. *Pervasive Computing*, (2011), 2–10.
4. Becker, R., Volinsky, C., Cáceres, R., et al. Human mobility characterization from cellular network data. *Communications of the ACM* 56, 1 (2013), 74-82.
5. Brockmann, D., Hufnagel, L., and Geisel, T. The scaling laws of human travel. *Nature* 439, 7075 (2006), 462–5.

6. Cramer, H., Rost, M., and Holmquist, L.E. Performing a Check-in: Emerging Practices, Norms and 'Conflicts' in Location-Sharing Using Foursquare. *MobileHCI*, (2011).
7. Cranshaw, J., Luther, K., Gage, P., and Norman, K. Curated City: Capturing Individual City Guides Through Social Curation. CHI, (2014).
8. Cranshaw, J., Schwartz, R., Hong, J.I., and Sadeh, N. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. ICWSM, 2012.
9. Cheng, Z., Caverlee, J., Lee, K., and Sui, D. Exploring Millions of Footprints in Location Sharing Services. ICWSM, (2011).
10. Choudhury, M.D., Monroy-Hernández, A., and Mark, G. Narco emotions: affect and desensitization in social media during the mexican drug war. CHI, (2014).
11. Costing the count. *The Economist*, June 2, 2011.
12. Cuomo, C., Shaylor, J., McGuirt, M., and Francescani, C. 'GMA' Gets Answers : Some Credit Card Companies Financially Profiling Customers. ABC News, 2009.
13. Dearman, D., Sohn, T., Truong, K.N., and Road, P.M. Opportunities Exist: Continuous Discovery of Places to Perform Activities. *Proceedings of CHI*, (2011), 2429–2438.
14. Duhigg, C. How Companies Learn Your Secrets. *New York Times*, Feb. 16, 2012. Retrieved from <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.
15. Duhigg, C. What Does Your Credit - Card Company Know About You? *The New York Times*, 2009. <http://www.nytimes.com/2009/05/17/magazine/17credit-t.html>.
16. Eagle, N., Macy, M., and Claxton, R. Network Diversity and Economic Development. *Science* 328, May (2010), 1029–1031.
17. Ferris, B., Watkins, K., and Borning, A. OneBusAway: results from providing real-time arrival information for public transit. CHI, (2010), 1–10.
18. González, M.C., Hidalgo, C. a, and Barabási, A.-L. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–82.
19. Griffin, D. and Hughes, T. Projected 2013 Costs of a Voluntary American Community Survey. United States Census Bureau, 2013.
20. Harrison, S. and Dourish, P. Re-Place-ing Space : The Roles of Place and Space in Collaborative Systems. *Proceedings of CSCW 7*, (1996), 67–76.
21. Hecht, B. and Stephens, M. A Tale of Cities: Urban Biases in Volunteered Geographic Information. ICWSM, (2014).
22. Huff, D.L. "A Probabilistic Analysis of Shopping Center Trade Areas," *Land Economics*, vol. 39, no. 1, pp. 81-90, 1963
23. Instagram: Our Story. <http://www.instagram.com/press>, retrieved May 24, 2014.
24. Isaacman, S., Becker, R., Cáceres, R., et al. Human Mobility Modeling at Metropolitan Scales. *MobiSys*, (2012).
25. Jernigan, C. and Mistree, B.F.T. Gaydar: Facebook friendships expose sexual orientation. *First Monday* 14, 10 (2009).

26. Joseph, K., Tan, C.H., and Carley, K.M. Beyond “Local”, “Categories” and “Friends”: Clustering foursquare Users with Latent “Topics.”(2012).
27. Kitamura, R., Chen, C., Pendyala, R.A.M.M., and Narayanan, R. Micro-simulation of daily activity-travel patterns for travel. *Transportation*, (2000), 25–51.
28. Komninos, A., Stefanis, V., Plessas, A., and Besharat, J. Capturing Urban Dynamics with Scarce Check-In Data. *Pervasive Computing*, (2013), 20–28.
29. Krikorian, R. New Tweets Per Second Record, and How! Twitter Engineering Blog, Aug. 16, 2013. Retrieved from <https://blog.twitter.com/2013/new-tweets-per-second-record-and-how>
30. Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., and Zimmerman, J. I’m the Mayor of My House: Examining Why People Use foursquare - a Social-Driven Location Sharing Application. CHI, (2011).
31. Lowenthal, T.A. American Community Survey: Evaluating Accuracy. 2006.
32. Mahmud, J., Nichols, J., and Drews, C. Home Location Identification of Twitter Users. *ACM Transactions on Intelligent Systems and Technology*, 2013.
33. Martine, G., et al. The State of World Population 2007: Unleashing the Potential of Urban Growth. United Nations Population Fund, 2007. Retrieved from <https://www.unfpa.org/public/home/publications/pid/408>.
34. McCullagh, D. Carrier IQ: More privacy alarms, more confusion | Privacy Inc. - CNET News. C|Net, 2011. http://news.cnet.com/8301-31921_3-57335110-281/carrier-iq-more-privacy-alarms-more-confusion/.
35. McCullagh, D. Verizon draws fire for monitoring app usage, browsing habits | Politics and Law - CNET News. C|Net, 2012. http://news.cnet.com/8301-13578_3-57533001-38/verizon-draws-fire-for-monitoring-app-usage-browsing-habits/.
36. Palchykov, V., Kaski, K., Kertész, J., Barabási, A.-L., and Dunbar, R.I.M. Sex differences in intimate relationships. *Scientific reports* 2, (2012), 370.
37. Pew Internet. Mobile Technology Fact Sheet. <http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/>, January 2014.
38. Friedhorsky, R., Culotta, A., and Valle, S.Y. Del. Inferring the Origin Locations of Tweets with Quantitative Confidence. *CSCW*, (2014), 1523–1536.
39. Quercia, D., Ellis, J., Capra, L., et al. Tracking “Gross Community Happiness” from Tweets. *CSCW*, (2011).
40. Rost, M., Barkhuus, L., Cramer, H., and Brown, B. Representation and Communication : Challenges in Interpreting Large Social Media Datasets. (2013).
41. Smith-Clarke, C., Mashhadi, A., and Capra, L. Poverty on the Cheap: Estimating Poverty Maps Using Aggregated Mobile Communication Networks. CHI, 2014.
42. Toch, E., Cranshaw, J., Drielsma, P.H., et al. Empirical Models of Privacy in Location Sharing. *Ubicomp*, (2010), 129–138.
43. Williams, S. and Currid-Halkett, E. Industry in Motion: Using Smart Phones to Explore the Spatial Network of the Garment Industry in New York City. *PLoS ONE* 9, 2 (2014), 1–11.
44. Yelp Word Maps. <http://www.yelp.com/wordmap/sf/>

45. Zhang, A.X., Noulas, A., Scellato, S., and Mascolo, C. Hoodsquare: Modeling and Recommending Neighborhoods in Location-based Social Networks. *SocialCom*, (2013), 1–15.
46. Zimmerman, John, Anthony Tomasic, Charles Garrod, Daisy Yoo, Chaya Hiruncharoenvate, Rafee Aziz, Nikhil Ravi Thiruvengadam, Yun Huang, and Aaron Steinfeld. 2011. Field Trial of Tiramisu: Crowd-Sourcing Bus Arrival Times to Spur Co-Design. In *Proc. CHI 2011*