Representation and Communication: Challenges in Interpreting Large Social Media Datasets

Mattias Rost
Mobile Life @ SICS
Stockholm, Sweden
rost@sics.se

Louise Barkhuus
Stockholm University
Stockholm, Sweden
barkhuus@dsv.su.se

Henriette Cramer
Yahoo! Research
Sunnyvale, CA, USA
henriette@yahoo-inc.com

Barry Brown
Mobile Life @ DSV
Stockholm, Sweden
barry@mobilelifecentre.org

ABSTRACT

Online services provide a range of opportunities for understanding human behaviour through the large aggregate data sets that their operation collects. Yet the data sets they collect do not unproblematically model or mirror the world events. In this paper we use data from Foursquare, a popular location check-in service, to argue for the importance of analysing social media as a communicative rather than representational system. Drawing on logs of all Foursquare check-ins over eight weeks we highlight four features of Foursquare's use: the relationship between attendance and check-ins, event check-ins, commercial incentives to check-in, and lastly humorous check-ins These points show how large data analysis is affected by the end user uses to which social networks are put.

Author Keywords

Social media, location-based services, empirical methods.

ACM Classification Keywords

H.3.5 [Information Interfaces and Presentation]: Online Information Services – Web-based services.

INTRODUCTION

The widespread adoption of online services has created a range of new human activities and behaviours. Since our online behaviour is invariably tracked, datasets of online activity can be used to make a range of inferences about behaviour, both in terms of the service itself and our broader lives. Large data sets in particular afford the analysis of broad social trends and activities. CSCW and HCI are no stranger to these sorts of analyses– for example, data on Facebook's relationship status has been used to infer differences in relationship formation and breakup [14], and research has even argued that happiness, as a broadly defined measure of general wellbeing, can be measured from social media status messages [19]. Location-based social media (from services such as Foursquare and Twitter) have been used to infer details about travel patterns and behaviours [11, 22] and with the increasing possibility of location sensing, large scale location data has be used to predict social events attendance [7] or to infer the nature of neighbourhoods [12] and human mobility [8].

These analyses rest upon the assumption that large-scale data sets are representative, in some quantifiable way, of real world behaviour. Extensive work has been done to understand the relationships between these databases and human behaviour, such as inferring location from humorous posts, or extracting 'real' from 'fake' online reviews. As we argue in this paper, however, this focus on the representativeness of such data inadvertently neglects the *communicative* features of social network data sets. Social media datasets are a record of communication – analysing it as only a representation of an underlying reality neglects the ways in which social media data is produced by social interactions between users. What is being analysed is a communicative system with its own contingencies, absences, and structure – genera of accounts produced and adapted for the purposes of the online system. This is particularly the case if we turn to location sharing systems. Systems such as Foursquare, Gowalla, and Facebook places (for example) are based not only on users' locations but also on an explicit declaration – and thus communication to others – of location. Using such data offers great opportunities [8, 11, 12], but taking advantage of these opportunities necessitates an awareness and discussion of the properties that these datasets have.

Through exploring one large social media dataset in particular – Foursquare check-ins – in this paper we examine how this dataset is better understood not as a representation or record of behaviour, but rather as communication amongst Foursquare users.

Foursquare check-ins

Foursquare is a mobile service allowing its users to share their location by 'checking in' to a venue. A check-in is a manual declaration of location at a semantically named venue. Users can check in to existing venues, or create venues themselves. Venues have a latitude longitude location, a name, an address, a category and users can add tips to venues. Users can keep their check-ins private, share them with their Foursquare friends, or push them to Facebook and/or Twitter. Venues are public, with the

CSCW '13, February 23–27, 2013, San Antonio, Texas, USA.

exception of those venues explicitly labelled as homes. Foursquare also employs ‘gamification’ elements such as ‘mayorships’ (the user checking in to a venue most becomes its mayor), points and badges for combinations of certain check-ins to increase usage [9, 20]. Commercial partnerships are part of Foursquare’s business model with for example badges being offered for check-ins related to various brands, and merchants can also offer deals and discounts to users to reward checking in to their venue.

As with most social media systems, Foursquare collects a record of users’ activity in the system. We studied data from an unedited feed of all Foursquare check-ins worldwide, for a period of eight weeks in 2010 and 2011. In studying this data, it was apparent that the use of Foursquare is not a simple indicator of location or behaviour. Rather, Foursquare is a social medium with its own properties. Broadly, we argue that the uses to which a system is put influences the kinds of data logged [3]. In particular the data that social media systems collect is broadly data about communicative action, with the richness that entails. While this data can be mined to obtain incidental representations of other activity, this is to throw away the key features of the data that make it comprehensible. Using this data we argue for how check-in data cannot be read simply as declarations of position, but should be understood as part of a broader communication system.

PREVIOUS WORK

Many aspects of online social networks have been explored, relating to particular properties of the network in terms of friendship ties and impression management [9, 13, 20], content-based user characterisation [21], as well as large-scale network structure and behaviour [16]. These online social networks, or “social-awareness streams” [21], also provide potential for modelling activity through the data sets that are available, either through scraping websites or through companies voluntarily sharing their data. More specifically, location-based networks have provided researchers with large data sets of location data that they use for creating models of prediction in terms of issues such as friend connections and human mobility [8, 11, 24], check-ins have for example been used to identify proto-neighbourhoods [11] and city ‘rhythms’ [8]. Cranshaw et al. [11] used public data about venues that have been checked in to (rather than check-ins) to distil canonical neighbourhoods by analysing co-occurrence patterns in place categories. Scellato et al. [24] analysed the socio-spatial characteristics of the ties between users in various location-sharing services, including Foursquare. Noulas et al. [22] presented an analysis of the geo-temporal dynamics by analysing tweets that contain check-ins and Cheng et al. [8], using tweets containing check-ins, showed that users follow reproducible patterns, that socioeconomic factors are coupled with mobility, and that content analysis of check-ins can provide interesting insights in the relationships between people and locations.

While these analyses may offer interesting insights in local differences, there are a wide variety of reasons for checking in resulting in these data points (and also for *not* checking in). Actively sharing a location is not only a practical tool for coordination and serendipitous interactions; location-sharing supports self-presentation, expression of mood, sharing of life events, and a multitude of other usage motivations [2, 9, 19]. Check-ins change location from ‘a user state’ into a deliberate situated act, in which considerations of audiences of other users and social norms play an important role [9]. This includes whether to share, but also on which channel(s), such as pushing check-ins to Twitter and/or Facebook as well.

More broadly, a challenge here is that check-in data, which is inherently social in nature, is approached as data indicating behaviour, i.e. that a check-in is simply an indication that a person has been at a specific place. Although most work acknowledges that data is socially based, the models and analyses rarely take into consideration the users’ underlying motivations or social characteristics of the service itself in the modelling of the data. Analysis inevitably takes the form of seeing data as a view on behaviour beyond the social network, rather than as a way of studying the social network itself.

For example, Hecht et al. [18] describe the use of humorous ‘fake’ location descriptions in twitter users’ profiles (such as ‘Justin Bieber’s Heart’), developing an algorithm that finds users' location to overcome this deliberate obfuscation. The paper does insightful analysis so as to obtain location data from users’ tweets, broadly categorising users’ location based on their similarity to users where location is already known. Thus, according to their algorithm a user who tweets about ‘Crawfish’ is more likely to be in Louisiana, and a user who tweets about “Clegg” (the British deputy prime minister) in the U.K. While this paper acknowledges that it only looks at one part of twitter data, to our mind this analysis neglects important aspects of the twitter data feed. Twitter data is *communicative* and not *representative*, and while as a side effect it might be possible to ascertain location data this is to ignore important aspects of this data. What is absent in this analysis is an understanding of *why* location might feature and influence what is tweeted. As a communication media that supports at a distance communication, one could image that location would disappear as a feature of tweeters’ identity. Yet as this paper shows, location is so evident as a feature of actual tweets that it can be extracted with considerable accuracy from a sample of users’ tweets.

This focus on representation can also be seen in more qualitative work. Brubaker and Hayes’ [6] insightful discussion of relationship status in Facebook and Craigslist documents the relationship between these systems and users’ lives and practices. Yet, while the paper pivots on their nature as “representational systems”, it retains a focus on these systems as descriptions of reality rather than as genera of communication.

DATABASES AND REPRESENTATION

Outside CSCW and HCI the nature of databases and their proliferation have proven to be a topic of sporadic interest for social science research. Here the nature of representations has been most directly grappled with and here the database as representation has had a prominent focus. Poster, for example, writes of databases as a 'super panopticon', resonating Foucault's writings on the prison [23]. A more subtle approach is taken by Bowker [4] who writes in detail about how the authorship of particular representations is embroiled in their use. The international classification of disease, for example, is embedded in a series of medical practices in different places. This is no less the case when we look at databases of activity. This echos Garfinkel's [15] arguments concerning the nature of representations such as medical records. Since medical records are made as part of a work process the sorts of records that are made are inherently conditioned by the needs of that work. The records that are produced are not then simply objective records of the workplace, but rather records that are used as part of that work activity, produced specifically for the purposes of treating patents.

With the recent turn to 'big data', however, understanding the conditions of the production of data has been downplayed, seen as an issue of 'data cleaning', rather than an inherent feature of records of activity. As boyd and Crawford put it: "Regardless of the size of a data set, it is subject to limitation and bias. Without those biases and limitations being understood and outlined, misinterpretation is the result." [5] We can draw an analogy with public app deployments for research purposes. Large-scale app deployments, capitalizing on easy distribution via app stores and the widespread proliferation of sensor-equipped mobile devices, offer huge opportunities for researchers, but also methodological challenges including data representation, gaining informed consent from app users, and lack of context surrounding collected data points [10].

DATA

To explore this argument, let us now turn to our data, and approach. The data set was provided to us by Foursquare and we were given access to their live stream of check-in data between September 12th 2010 and March 3rd 2011. The data is acquired through a firehose API delivering anonymised check-ins as they are performed. These data points contain venue name, venue category, venue location (long/lat), user gender, timestamp and the timezone used at the location of the venue, and whether or not the user received a badge. Due to some technical problems this data had a number of missing days – we thus selected the longest continuous period of data for closer analysis - eight weeks of continuous data at the end of 2010 and beginning of 2011 (starting Nov 8). Our approach to the data was experimental in that we took focused examinations of the data to explore what venues were being checked-into, both infrequently and frequently. We then explored the data analysing how check-ins to different venues differed over time, and over different areas.

RESULTS

We start with a broad overview of the data worldwide using roughly the first four weeks of data, from Nov 8 – Dec 5, 2010. Within this set, Foursquare users checked in to 5,499,469 different venues. The data shows a long tail in terms of the number of check-ins per venue. The mean number of check-ins to each venue was 7.6, with a median of 2 and σ=46.7. The number of check-ins to the top 20% of most checked-in venues account for 74% of all check-ins. The 2% of the most popular venues account for 32% of all check-ins. Interestingly, a rather larger set of 1,963,091 venues (37%) received only 1 check-in during this period.

When we look at all check-ins and venues, we see that 2.9M (53%) of venues are in the Americas and 24M check-ins (57%), 1.8M (33%) of the venues are in Asia, with 13M (33%) check-ins, 768,953 (14%) venues are in Europe, with 4.4M(11%) check-ins and 23,267 venues in Africa (0.4%), with 108.951 check-ins (0.26%). When we look at the top 100 venues users checked into (Table 1 shows the top 10) we see that overall, venues in the US, and in particular airports, dominate (airports: 31/100, US venues: 55/100, 39/100 in Asia, Europe: 6/100 – 3 of the latter, perhaps quite characteristic for the time of year, were ski areas in Austria).

#	Check-ins	Venue
1	26,159	Siam Paragon (shopping mall), Bangkok
2	18,140	Los Angeles International Airport (LAX), US
3	17,224	MoMA Museum of modern art, NY
4	16,878	John F. Kennedy International Airport (JFK)
5	16,804	NBC Studio 1A Today Show, NY, US
6	16,564	Madison Square Garden, NY, US
7	16,404	Hartsfield Jackson Atlanta Int. Airport, US
8	15,967	San Francisco International Airport, US
9	15,239	Chicago O'Hare International Airport, US
10	12,460	New York Penn Station, NY, US

Table 1: Top checked-into venues worldwide

Why this is the case cannot be established from this data alone. This could for example indicate a difference between locales in terms of places that people visit, or the type of venues people like to share. While this data does represent human activity to a certain extent, we also see that we need to be careful for which types of purposes we use this data. The data is characterized through the motivations for sharing as identified in [9], socio-economic factors and demographics of foursquare users. As examples: the MoMa (#2) for instance receives 2.5M visitors a year according to its homepage, whereas Atlanta's airport (#3) receives over 7M passengers every year, and the order of airports does not confirm to the passenger number listings of [1], nor are all airports in the world, such as the second largest airport, Beijing, present in such data.

Snowpocalypse: Reappropriation of Location

We continue by focusing on a geographical subset of the data; we extracted data from New York City, London and Boston. These are all large urban areas and looking at these enabled us to examine more detailed local phenomena. Looking at this data showed that location-sharing services are not only used to share location; their usage data also reflect local events. When we look at #5 in the top 10, we see a TV show/studio. The venues of #13 and 19 are related to the Macy Thanksgiving Parade in NY. Events can also result in spikes in the data, Foursquare for example report the 'rally to restore sanity' in Washington, DC on 30 October with 30,525 check-ins as the most checked-into event for 2010. This number of check-ins is markedly larger than the most checked-into venue in the top 10 for our Nov-Dec period (Siam Paragon with 26,159 check-ins). We also see data likely caused by incentives, rather than actual presence at a venue, such as on #39 with 7428 check-ins, 'The Conan Blimp', a promotion in which a giant blimp promoted the Conan O'Brien show (a comedy show) across America, and a check-in to this 'venue' resulted in a badge.

One characteristic of our data is the prevalence of non-location based check-ins (not unlike previous studies of prototype systems where locations were defined as hybrids between a location and status [2]). One interesting phenomena concerned weather related check-ins. Over our sampled time period there was extreme weather with a record snowfall and low temperatures across the US and Europe. On January 25th New York snowfall broke a record held since 1925. Looking at our larger time period, in our NY data set we found 85 special venues reflecting this by searching for keywords such as "…pocalypse," "snow", "freeze" and "slush". The venues were not physical venues such as a particular restaurant, but rather they captured a shared experience of extreme weather. This is reflected through the way these venues are named; of the 85 special venues identified, 17 were based on the notion of an Apocalypse, for example Snowpocalypse, Freezepocalypse, Slushpocalypse. These are all playful ways of saying that it is snowy, cold, ice is melting etc., and that this seemed at the time extreme and very much an experience that is shared by New Yorkers.

Figure 1 shows the total for check-ins to the eight most popular expressive check-ins. The most 'checked-in' venue is Snowpocalypse 2011 peaking at 4590 check-ins on January 27th. This contrasts with other top checked in physical venues the same day – for example 378 check-ins to all Starbucks and 268 to Penn Station. There are two major spikes in the data on January 24th and 27th where Freezepocalypse accounts for 2877 check-ins on the 24th, followed by 599 the day after (but notably none the days before). Snowpocalypse on the other hand exhibits a different characteristic with 3431 check-ins on the 26th and 4590 on the 27th. In the figures we also see the snowfall during the same dates. It is clear that on days of heavy snow, large numbers of check-ins to a venue reflect an event, but the online phenomenon is also affected by the actual check-in activity itself.

We contrasted this data with similar check-ins for Boston, see figure 2 (also affected by the extreme weather). Interestingly, while Snowpocalypse-styled check-ins appear sporadically here after January 2nd, at the start of the extreme weather this was not a check-in phenomenon. The phenomenon in Boston came after NYC, despite the weather starting in Boston. Foursquare check-ins can thus be read as an indicator of a large shared experience – the severe winter of 2010/2011. The social network here has been repurposed and effortlessly reappropriated to this experience, 'redefining' location and the purpose of the service itself. The check-in data does not directly correlate with the real-world occurrences, especially Boston that did not catch on with the first big snowfall.

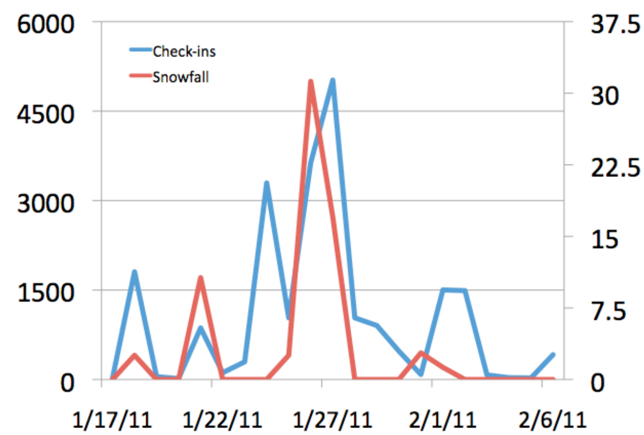

Figure 1: Snowfall and snow related check-ins for NYC

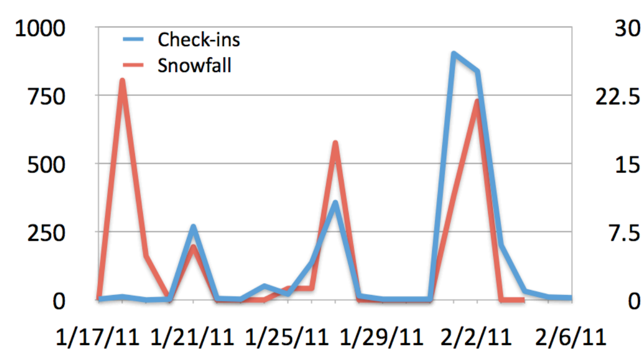

Figure 2: Snowfall and snow related check-ins for Boston

External Motivation Factors

A second feature of the check-ins we examined was the role of external motivation factors. Services such as Foursquare understandably employ various ways to increase usage, and have incentives built in to increase the number of check-ins. Game-like elements, and rewards such as discounts at retailers aim to motivate users to check in. Foursquare contains elements such as points, badges and 'mayorship' of a venue for the user who checks in the most to a venue. These game-like aspects have been highlighted in other research as well [20], and have been shown to in some cases conflict with other usage motivations to share one's location [9]. Another factor at play are third-party collaborations that are part of the Foursquare business model in which organisations can for example acquire a dedicated badge related to their brand.

The effect of these intertwined service design features can be illustrated by the introduction of a set of five new badges on November 16th 2010. In New York we found that Radioshack (an electronics chain) spiked rather quickly over a short time, going from 5 and 12 check-ins per day on 13th to 14th November, then jumping to 58, 196 and then 1323 the following days, bringing it to the top 10 venues in number of check-ins on Manhattan for three days in a row. Radioshack had introduced a special offer on November 15th where customers who checked in would receive 10% off, 15% off if the customer was a 'mayor' and 20% if the customer got a 'holiday hero badge' (obtained through checking in to a set of venues). It is perhaps not a surprise that the specific features of Foursquare, or that commercial tie-ins influence check-in data. Yet this underlines how online data can be influenced by a myriad of small features that produce anomalies and local deviations.

Humour and Communication

Past researchers have remarked upon the humorous use of location systems - where locations are used as part of banter or social repartee. In particular Barkhuus et al. [2] report on how location statuses amongst a group of friends were frequently used as a part of longstanding jokes. Hecht et al. [18] describe the use of humorous location descriptions in twitter profile location fields (such as 'Justin Bieber's Heart'), developing an algorithm that finds users' location, overcoming this deliberate obfuscation. Although it is clear that the users' intention is humour in both these papers, the nature of the humour was unexplored. In here we explored in more detail categories of humorous entries. To categorise the humorous check-ins that we had identified, we sorted the forms of humorous check-ins and agreed three categories: mis-categorisation, profanity-based invented venues, and real venue/'from-a-distance' check-ins where the user's location does not match the venue checked into.

Mis-categorisations were observed when a user specified venue name did not match its category. For example on Manhattan we found 21 venues that were clearly homes but categorized as strip clubs (verified using Google Street View). Theses were only checked into on average 2.8 times, perhaps as a short-lived joke. Examples include: "Dans apartment" [sic], "Gian's rooftop" and "Tom's bachelor pad" - all categorized as strip clubs. *Invented locations* involved imaginary location names that included profanities, often expressed as a general commentary or as an opinion of a real venue. For example, searching for the one of most common profanities we found 18 such venues just on Manhattan. They were 'places' such as 'Faux-tini: This is not f****** bartini' or 'It's My F****** Birthday!'. The snowpocalypse phenomenon was in a sense a large-scale example of such miscategorisation.

Lastly, we found several types of likely *check-ins 'from-a-distance'* due to the time or situation. For example, there were around two check-ins per day to Buckingham Palace in London between midnight and six in the morning, indicating that either people were in close enough range outside the British Royal residence or that they specifically searched for it and checked in from afar. Other check-ins from-a-distance include check-ins to venues that were closed at the time, or would be difficult to reach (such as top of the empire state building, Statue of Liberty). These check-ins highlight the *communicative* role of checking in. These are not 'fake', 'incorrect' or 'invalid' check-ins - since their intention is as a means of communication, not as a literal description of location. This means that they provide a resource for understanding particular forms of humour, and communication; they are perhaps less clearly understandable as indicating a record of location or behaviour.

DISCUSSION AND CONCLUSIONS

The core argument of this paper has been that social network data should be seen and analysed as communicative data. That is: data that is produced as a side effect of communication between users, rather than as a representation of some underling activity. What these examples above give us are cases where the check-ins make little sense as representations of a 'reality', yet make more sense as communicative events between Foursquare users.

Reviewing these behavioural characteristics in terms of check-ins, it is clear that there are broader concerns for users than simply reporting their location to the record; we here specified four topics. Looking at the data overall we observed how the number of check-ins has no clear correspondence with visitor numbers. We then described how the 'Snowpocalypse' unfolded, and had a significant affect on east coast US check-ins. Third, we described the role of promotions as a motivation for check-ins. Lastly, evidence of humour on the other hand, showed itself to be fairly sophisticated and so individually tailored that it hardly seemed to affect any large scale analysis. On the other hand this might indicate another discrepancy: that it will not be discovered as an actual feature of use. Textual search for example will not uncover all such instances; this data illustrates well how humorous use is very difficult to look for in large data sets because of its subtlety and contextual base.

While an obvious solution when approaching the data as a representation of human activity would seem to 'simply filter out such noise', this is too simplistic. Data cleaning processes are inherently subjective [5] and without taking a qualitative look at big data faulty assumptions are likely to arise, not fitting local or temporal differences in the anomalies we have described. Our contribution here does not centre on the issue of deciding what these large datasets can be used for, but rather about exploring emerging genres of communication. In conclusion we would argue more broadly for the role of check-in services like Foursquare not as location-based services, but rather as particular communication genre, with particular developing forms and style. This role may turn out to be ultimately more important than its role as a record of location. Check-in based location systems, or at least Foursquare, are distinct to continuous location updating systems. The act of 'checking in' is such that it can be used to communicate in a way quite different to the background updating of ones location. The difference between a computer and an individual reporting their location makes for a service that fits quite different purposes. Rather than asking whether this data is representative of location, we should ask what the data represents.

REFERENCES

1. Airports Council Int., Monthly Traffic ratings, Dec 2010.
2. Barkhuus, L., Brown, B., Bell, M., Sherwood, S., Hall, M., Chalmers, M. (2008) From Awareness to Repartee: Sharing Location Within Social Groups. In Proc. of CHI '08.
3. Bittner, E. and Garfinkel, H. (1967) Good organisational reasons for bad clinical records. In Studies in ethnomethodology: Prentice Hall.
4. Bowker, G.C. (2006) Memory Practices in the Sciences, MIT Press.
5. boyd, d. and Crawford, K. Six provocations for big data, A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, Sept 2011.
6. Brubaker, J. R. and Hayes, G. R. (2011) SELECT * FROM USER: Infrastructure And Socio-Technical Representation. In Proc. of CSCW '11.
7. Calabrese, F., Pereira, F.C., Lorenzo, G.D., Liu, L. and Ratti, C. (2010) The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events. In Proc. of Pervasive 2010.
8. Cheng, Z., Caverlee, J., Lee, K., Sui, D. (2011) Exploring Millions of Footprints in Location Sharing Services. In Proc. ICWSM '11.
9. Cramer, H., Rost, M. and Holmquist, L. E. (2011) Performing a check-in: emerging practices, norms and 'conflicts' in location-sharing using Foursquare. In Proc. MobileHCI '11.
10. Cramer, H. Rost, M., Bentley, F: (2011) An introduction to Research in the Large, Special Issue on 'Research in the Large', International Journal of Mobile Human-Computer Interaction.
11. Cranshaw, J. and Yano, T. (2010) Seeing a home away from the home: Distilling proto-neighborhoods from incidental data with Latent Topic Modeling. Proc. NIPS'10.
12. Cranshaw, J. Schwartz, R. Hong, J. and Sadeh, N. (2012) The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. Proc. ICWSM'12,.
13. Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), article 1.
14. Facebook Data Team: The Right Time for Love: Tracking the Seasonality of Relationship Formation. Accessed at https://www.facebook.com/notes/facebook-data-team/the-right-time-for-love-tracking-the-seasonality-of-relationship-formation/10150643989093859
15. Garfinkel, H. (1967) Good organizational reasons for "bad" clinic records, Studies in Ethnomethodology. Englewood Cliffs.
16. Golder, S. A., Wilkinson, D., & Huberman, B. A. (2007, June). Rhythms of social interaction: Messaging within a massive online network. In Proc. of Communities and Technologies (pp. 41–66). London: Springer.
17. Hacking, I. (1995) Rewriting the Soul: Multiple Personality and the Sciences of Memory, Princeton University Press.
18. Hecht, B. Hong, L., Suh, B., Chi, E.H. (2011) Tweets from Justin Bieber's Heart: The Dynamics of the "Location" Field in User Profiles. In Proc. of CHI '11.
19. Kramer, A.D.I. (2010) An unobtrusive behavioural model of "gross national happiness". In Proc. CHI '10.
20. Lindqvist, J., Cranshaw, J., Wiese, J., Hong, H. and Zimmerman, J. (2011) I'm the Mayor of My House: Examining Why People Use Foursquare. In Proc. of CHI '11.
21. Naaman, M., Boase, J. Lai, C. (2010) Is it Really About Me? Message Content in Social Awareness Streams. In Proc. of CSCW '10.
22. Noulas, A Mascolo, C., Scellato, S., Pontil, M. (2011) An Empirical Study of Geographic User Activity Patterns in Foursquare. In Proc. ICWSM '11.
23. Poster, M. (1990) The mode of information: post structuralism and social context, Polity press.
24. Scellato, S., Noulas, A., Lambiotte, R., Mascolo, C. (2011) Socio-spatial Properties of Online Location-based Social Networks. In Proc. of ICWSM '11.
25. Vasalou, A., Oostveen, A. M. and Joinson, A. M. (2012) A case study of non-adoption: the values of location tracking in the family. In Proc. of CSCW '12.