LOCATION BASED SOCIAL NETWORKS – TRACKING ACTIVITY IN AN URBANENVIRONMENT USING TWITTER DATA

F. Neuhaus

Centre for advanced Spatial Analysis, UCL, 1-19 Torrington Place WC1E 7HB London – fabian.neuhaus@ucl.ac.uk

KEY WORDS: Social network, Location based information, Tracking, Geographical visualisation, Urban studies, Spatio-temporal

ABSTRACT:
Increasingly people use digital or online networks to communicate and interact. This changes the socialscape of the urban area and with it the interactive hot spots change and fluctuate throughout the city as individuals follow the narrative path of their everyday routines. People leave messages, distribute news and respond to conversations not only in traditional locations anymore but potentially anywhere in the city.
This paper discusses the emerging potential of social media data used for urban area research and city planning. Working with crowd sourced data in a web 2.0 manner as described for example by Hudson-Smith et All. (2009). Specifically we look at the connections between the emerging social network, as for example described by Boccaletti et All (2006), and the local physical surrounding and conditions. Also aspects of visualisation as well as privacy and ethical implications are discussed.
The information gathered from social media networks, is gathered directly of the platform used by the network participants as for example already employed by Eagle et all (2009) in their study of social networks using mobile phones. The twitter data however, usually can be associated with a physical location for example via the GPS of the smart phone. Research using this location based technology together with a temporal structure has been demonstrated for example by Reads et All (2009). For this virtual social network and infrastructure-mapping project, the data is derived from the Twitter micro blogging service directly via the API and aims to merge the previously listed approaches into a combined location based temporal network.
These local activity are analysed and visualised based on networks of interaction. Who knows whom and get in touch with whom? However the social networks in the sense of specific interest are these datasets in relation to place and how this location based network enable the individual to shape a distinct sense of place.

1. INTRODUCTION

Where is the city active and does it physically change over time? Urban areas are no static artefacts as they are preferably described in texts and theories. Urban areas are buzzing hot spots of human activity that, to some extend, manifest themselves as or utilise built structure, but are largely temporal and ephemeral. Meaning that no constant being of this 'artefact' is present, but merely a past aggregate is telling tales of memories and rumours.

In an attempt to listen to these stories and narrative as they unfold through the streets, alleyways, in courts, buses, on roof terraces or in swimming pools the social networking platform twitter was employed to reconstruct the cities activity hotspot as a time-frozen 'New City Landscape' drawing out the ever changing locations of people's presence and power of spatial creation through narratives and activity.

From the collected data a new landscape based on density is generated. The features of this landscape of digital activity correspond directly with the physical location of their origin but at the same time represent with hills the peaks of locations from where the activity tales are submitted. The flanks and valleys stand for areas with lesser activity and vast plains and deserts of no twitter tales stretch across the townscapes that lay dormant. These New City Landscape (See for example Figure 1) maps don't represent any physical features, but the interaction with physical features on a temporal basis. The digital realm has become as much part of the urban environment as the physical features and with these tweetography maps they are made visible for the first time. The maps allow us to make a direct comparison between real word activity, physical location and digital message. In a globalised world this local reference develops an increased importance as a sense of place, a source of identity and memory. The digital social media data allows us to investigate into this realm of peer groups' social location interaction, combining the global scale with its local source.

Some of the physical features of the city that are shining through are the major infrastructure installations. The airports on one hand are the examples of quite intense activity and the parks on the other hand manifest themselves through the absence of activity, virtual social networking activity. Where as at airports users might be bored waiting or excited to just have landed, people in the park are engaged in physical activities other than tweeting and these locations are left virtually empty. A great example is the Central Park in New York, a virtual twitter activity desert, where around it and Manhattan as such is a very high tweet area.

2. METHODS AND TECHNOLOGY

The technology to collection twitter data is based on the technology developed for the Tweet-O-Meter (tom). This service was developed at CASA by Steven Gray. Similar to the tom service the data is collected using the twitter API. Twitter offers two different services through the API. One is the Streaming API and the other one is the Search API. For NCL we are using the Search API because of the built in spatial search function.

2.1 Process

With this spatial search we can filter the incoming messages as for a specific urban area. For the NCL maps we have defined the urban areas consistently as an areas with a 30km radius around an urban centre.

The search query will pass down from the twitter feed all messages to fit this criteria. The software will store all these results in a database continuously.

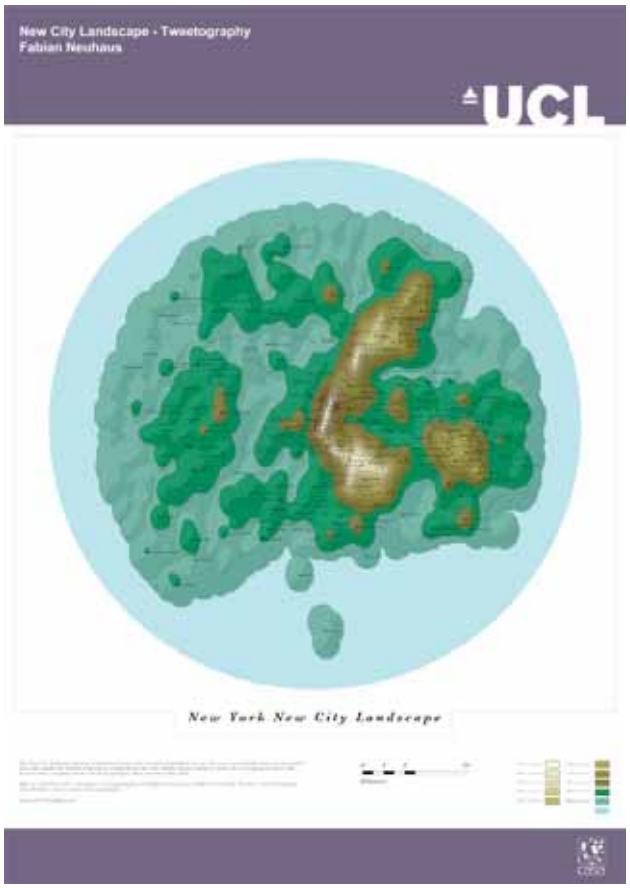

Figure 1. - New York New City Landscape Map showing the twitter activities over the period of one week as a density surface within a radius of 30 km of the NY urban area.

Due to IP limitations imposed by twitter and infrastructure limitations, we are only able to run four parallel search and collect queries at the time. Depending on the search location the resulting amount of data can be quite large, putting quite some pressure on the infrastructure. In order not to miss out on messages, the responding times of the system cannot be compromised.

The data collection per location as been limited to one week, seven days, of consecutive logging of messages sent using the twitter service. One weeks provides good comparison data over a number of days but also shows the different patterns between weekdays and weekends as well as within 24 hours.

The data collected as such therefore is already a spatially defined subset of the total number of tweets sent. However, the collected material needs to be reprocessed because the location information quality is not the same in all messages. Some messages are reverse geocoded from profile information, which generates generic place information.

The resulting data set holds all messages containing real GPS information as Latitude and Longitude coordinates. With this information a more accurate mapping is possible. It is assumes that the accuracy of this information lies within the normal range of GPS accuracy of some 5 to 15 metre.

In a second step a social network is computed, based on the interactions of users in the dataset. To do this especially re-tweets (RT), twitter messages that have been resent by other users and at-tweets (@), messages specifically addressed at selected twitter users are employed to establish links between individual tweeters as well as a direction of interaction.

Together with the emerging social network and the location as well as the temporal information contained in the data a location based temporal social network can be visualised.

2.2 Data

The amount of data collected varies dramatically between the different locations. There are clearly the very actively tweeting cities such as New York and London with more than 800'000 location based messages sent over the course of one week. On the other hand there are a lot of places especially non-English speaking countries with far less activity, down to a few hundred. Additionally the total location based tweets and the actual GPS tagged messages diverge a lot. Furthermore there is not a simple, more messages result in more GPS tagged messages, equation that applies. It can well be that an very active place turns out very few Latitude/Longitude stamped tweets. As it appeared for example in the case of Sydney, Jakarta or Sao Paulo, where the percentage of geotweets is below 1 % of all location based messages.

Twitter is a relatively new service, being around some four years. The number of users is continuously growing dramatically. This fact put some constraints on the comparability of the data samples. Also the short-term usage of the service is loosely connected to large media events and it is expected that numbers fluctuate quite a bit.

3. RESULTS

The point cloud of twitter messages drawn from the database and mapped using a Mercator projection. This universal projection allows for recognition and readability of urban areas located around the globe.
For the mapping the individual point are being aggregated as a density surface.
Throughout the emerging landscape features have been renamed to reflect these conditions. The new names are fabricated using the real world names in combination with a landscape description of the virtual surface overlaid. This could be 'Mountain' or 'Peak' for high points, 'Slope' or 'Valley' for descending features or 'Desert' and 'Meadow' for average and consistent areas. Inactive areas are termed for example 'Quarry' or 'Ditch'. Together with the familiar real world element the locations become tangible and memorable points of orientation and maybe identification.

The defining landscape features in the virtual NCL map are the hot spots of twitter activity, the peaks. Here the morphology varies between the urban Areas dramatically. How the twitter traffic structures the NCL is unique for each city. There are however some characteristics that can be pointed out.
The different groups could be described as central, where one main location towers over the whole urban region, the multi, where different hotspots appear as peaks across the landscape and the featuring, where one or more features draw out as shapes, groups of peaks or ridges.

Linking this to the social network, see Figure 2, the activity pattern in the temporal sense gain of importance. The variety of

different pattern displayed by different groups is very distinct from activity pattern we normally see in everyday activities. There is more scope for the individual to jump in and out of activities, but connections on hold and reactivate others than what we know from real world interaction.

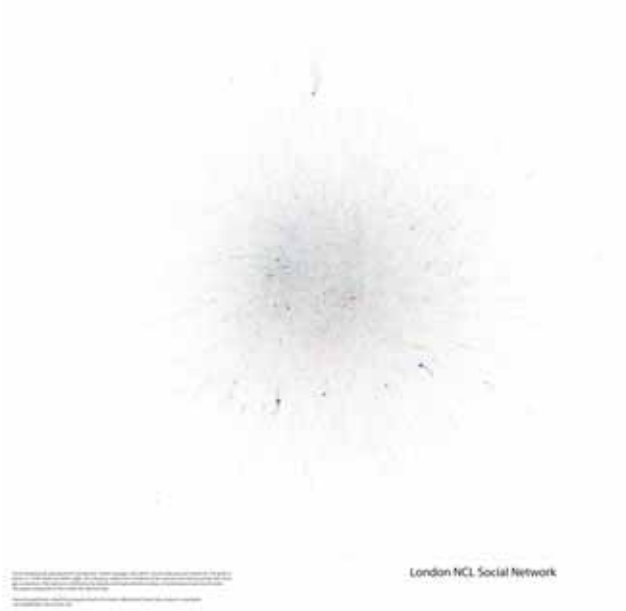

Figure 2 - Twitter activity based social network using the London NCL map data collected in a radius of 30 km within the London urban area.

On the city side the transformation of network activity hubs through out different time periods are striking and offer a new perspective on urban area usage as well as sense of place.
Application for this can be found many ranging from urban planning to transport management and modelling to health and safety as in the spatial spreading of information or infections.

4. BIBLIOGRAPHY

Boccaletti, S. et al., 2006. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5), pp.175-308.

Eagle, N., Pentland, A. & Lazer, D., Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*. Available at: http://www.pnas.org/content/early/2009/08/14/0900282106.abstract [Accessed January 28, 2011].

Hudson-Smith, A. et al., 2009. Mapping for the Masses. *Social Science Computer Review*, 27, pp.524–538.

Reades, J., Calabrese, F. & Ratti, C., 2009. Eigenplaces: analysing cities using the space – time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 36(5), pp.824 – 836.