Measuring urban mobility and encounter

[1]Eamonn O'Neill and [2]Vassilis Kostakos
[1]Department of Computer Science, Department of Computer Science, University of Bath, Bath BA2 7AY, UK, eamonn@cs.bath,.ac.uk
[2]Vassilis Kostakos, Dept of Mathematics & Engineering, University of Madeira, Madeira, Portugal, vassilis@cmu.edu

In [O'Neill et al., 2006] we described our development of novel methods for systematically observing and recording patterns of pedestrian mobility and encounter in the city. As a central part of our approach, we automated the capture of longitudinal data on mobility and encounter of mobile Bluetooth devices. Our recording of large-scale longitudinal data allows us to make two significant advances beyond traditional approaches to measuring behaviour in the city. First, we can inform aggregate level modelling and analyses with real world empirical data. This should help us to validate and improve upon the often simple approaches to such modelling. Secondly, we can investigate and analyse data that relate to a single user or a specific group of users, thus individualising our analysis in ways not possible with traditional aggregate approaches.

With no central servers to facilitate communication, Bluetooth devices rely on a discovery protocol to identify nearby devices. This protocol requires the initiating device to carry out an inquiry scan in a specific range of frequencies and wait for nearby devices to advertise their presence by transmitting their unique identifier. Thus, each inquiry scan provides information about which devices are in range at a discrete point in time. In our data collection we make use of the three key characteristics of Bluetooth: physical proximity, the explicit advertisement of the device's presence, and the unique identifier transmitted by each device.

Our combined methodology of manual and Bluetooth gatecounts allows us to estimate the penetration of discoverable Bluetooth in the urban population. As reported in [O'Neill et al., 2006], we found that for the city of Bath approximately 7.5% of observed pedestrians had discoverable Bluetooth devices. Our more recent measurements show a dramatically higher absolute number of Bluetooth devices. We are currently planning a new round of our combined observation methodology to investigate if this reflects a higher proportion of Bluetooth activity amongst the urban population.

The data record of Bluetooth activity is fundamentally a set of individual Bluetooth discovery events. In making sense of these data, we need to relate the individual events to a particular device and to its patterns of presence and absence across given scanner sites. In investigating encounter, we also need to relate these patterns across different devices. A temporal view allows us to begin making sense of the individual Bluetooth discovery records. Because of the use of unique identifiers in the Bluetooth protocol, each device can be associated with one and only one timeline across all our scanning locations in the city. A device moving past a scanner will generate a series of successive contact points on its timeline. Visualisations of the timelines reveal patterns of transience and persistence varying across times and spaces in the city and allow us to begin relating characteristics of those differing times and spaces to these data patterns. These data visualisations provide the foundation for an approach to making sense of our data in terms of three distinct abstractions: sessions, encounters and trails.

A session is defined as a set of contact points having no more than a threshold temporal distance $\delta 1$ between any two consecutive points. Thus, a session has an associated device, a start time, duration, and an associated location in the city (i.e. the scanner site). In the work reported in [O'Neill et al., 2006] we empirically derived appropriate values for $\delta 1$ by correlating human observations with Bluetooth observations. The concept of a session is central to our analyses, since it gives a time dimension to the discrete contact points generated by our scanners. Our next concept, encounter, builds on the concept of session. Encounter describes instances when two devices have been copresent. Thus, an encounter is defined by two devices, a location, a starting time and duration. To detect encounters we look for temporally overlapping sessions that took place at the same location. Our final concept, trail, extends the concept of a session with the spatial dimension. A trail is defined as a set of consecutive sessions for a given device, having no more than a threshold temporal distance $\delta 2$ between any two consecutive sessions. A trail, therefore, has an associated device, starting time, duration and number of hops (number of distinct sessions). Once again, $\delta 2$ has been empirically derived, and is based on our knowledge of the typical journey times between the physical locations we are observing.

In making sense of the patterns of movement and interaction of devices and people around the city, we first consider the distribution of session duration across our different scanning sites. We distinguish between persistent and transient devices using a threshold for session duration of 90 seconds. We empirically derived this threshold by measuring the session duration for individuals who walked past our scanners at a comfortable walking speed. This threshold of around 90 seconds allows us to establish empirically a conceptual distinction between transient and persistent devices and we can study how each conceptual group appears in different urban spaces.

We can represent trails as directed paths across a network graph. Each node can have metadata associated with it, such as duration of session, related semantic information (e.g. name, location co-ordinates, and so on), the identifiers of the devices that have visited it and various computed statistics such as frequency and average session duration. Thus, by preserving all the information recorded by each individual trail we can begin to analyse and compare trails. Graphs offer an effective way to inspect a set of trails and explore the relationships amongst them. For any given set of trails matching a set of criteria, we are able to inspect their layout and identify patterns. For instance, searching for the most popular trails late on a Friday night we can identify the taxi ranks as being the destination for many trails.

Visualising and analysing our raw Bluetooth activity data as sessions and trails allows us to begin making sense of the data in terms of people's behaviours in various forms of urban space (such as contrasting patterns of persistence between the pub and the street). Associating a unique timeline with every newly discovered device also allows us to trace the progress, or trail, of a device (and its user) by analysing the device's sequential presence at different scanning sites. A third crucial aspect of investigating the relationships between people, technologies and the city directly links the temporal and the spatial. Copresence or encounter requires that 2 or more devices are in the same space at the same time. It is in encounters that interactions occur: interactions between person and person, between person and fixed device, between mobile device and mobile device, between mobile device and fixed device, and so on. To study the patterns of encounter in the city, we first identify device sessions that overlap in time and were recorded at the same location.

Again we can represent these patterns as a network graph. Assuming that each device from our dataset becomes a node in this graph, the list of encounters describes the links between all nodes. Thus, we are able to generate social network graphs [e.g. Strogatz, 2001] that represent the patterns of encounter across our entire dataset. We can generate various graphs from our data, such as an individual social network graph per scanner site, or a graph of our entire dataset in one city-scale social network graph. Furthermore, we can generate these graphs over the entire lifetime of our scanning or over any specified period. An array of standard metrics such as closeness, clustering coefficient, etc [Freeman, 2004] enables us to identify meaningful subgroups of the population, and focus our attention on them. For instance, we might identify isolated individuals and examine the trails these individuals take across the city.

The approach presented here allows us to begin making sense of the data derived from measuring urban behaviour temporally and spatially. It reveals patterns of transience and persistence varying across times and spaces in the city and allow us to begin relating characteristics of those differing times and spaces to the data patterns. In related work, we have moved from visualisations to the development of more formal and systematic analytical concepts and tools that allow us to automate aspects of our data analysis.

References

1. Freeman, L. (2004) The Development of Social Network Analysis: A Study in the Sociology of Science. Vancouver, BC, Canada: Empirical Press.
2. O'Neill, E., Kostakos, V., Kindberg, T., Gen. Schieck, A.F., Penn, A., Fraser, D.S., And Jones, T. (2006) Instrumenting the City: Developing Methods for Observing and Understanding the Digital Cityscape. In *Proceedings of UbiComp* 2006, 315-332.
3. Strogatz, S.H. (2001) Exploring complex networks. *Nature*, **410**, 268-276.