2

Brief Encounters: Sensing, Modeling and Visualizing Urban Mobility and Copresence Networks

VASSILIS KOSTAKOS
University of Madeira
EAMONN O'NEILL
University of Bath
ALAN PENN
University College London
and
GEORGE ROUSSOS and DIKAIOS PAPADONGONAS
Birkbeck College

Moving human-computer interaction off the desktop and into our cities requires new approaches to understanding people and technologies in the built environment. We approach the city as a system, with human, physical and digital components and behaviours. In creating effective and usable urban pervasive computing systems, we need to take into account the patterns of movement and encounter amongst people, locations, and mobile and fixed devices in the city. Advances in mobile and wireless communications have enabled us to detect and record the presence and movement of devices through cities. This article makes a number of methodological and empirical contributions. We present a toolkit of algorithms and visualization techniques that we have developed to model and make sense of spatial and temporal patterns of mobility, presence, and encounter. Applying this toolkit, we provide an analysis of urban Bluetooth data based on a longitudinal dataset containing millions of records associated with more than 70000 unique devices in the city of Bath, UK. Through a novel application of established complex network analysis techniques, we demonstrate a significant finding on the relationship between temporal factors and network structure. Finally, we suggest how our understanding

This research was supported by the UK Engineering and Physical Sciences Research Council grants EP/C547683/1 and EP/C54769111 (Cityware: urban design and pervasive systems).
Authors' addresses: V. Kostakos, Madeira Interactive Technologies Institute, University of Madeira, 9000 390 Funchal, Portugal; email: vk@uma.pt; E. O'Neill, Department of Computer Science, University of Bath, BA2 7AY, Bath, UK; A. Penn, Bartlett Graduate School, University College London, 1-19 Torrington Place, WCIE 6BT, London, UK; G. Rousos and D. Papadongonas, Department of Computer Science and Information Systems, Birkbeck College, University of London, Malet Street, London WCIE 7HX, UK.

DOI 10.1145/1721831.1721833 http://doi.acm.org/10.1145/1721831.1721833

and exploitation of these data may begin to inform the design and use of urban pervasive systems.

Categories and Subject Descriptors: H5.2 [**Information Interfaces and Presentation**]: User Interfaces—*User-centered design*; *Interaction styles*; *Theory and methods*; G.4 [**Mathematics of Computing**]: Mathematical Software—*User Interfaces*

General Terms: Design, Human Factors

Additional Key Words and Phrases: Pervasive, ubiquitous, mobile interaction, Bluetooth, sensing, mobility, encounter, complex network, social network, visualisation, copresence, virus, epidemic, trail, brief encounter, information diffusion, urban computing.

ACM Reference Format:

Kostakos, V., O'Neill, E., Penn, A., Rousos, G., and Papadongonas, D. 2010. Brief encounters: Sensing, modeling and visualizing urban mobility and copresence networks. ACM Trans. Comput.-Hum. Interact. 17, 1, Article 2 (March 2010), 38 pages.
DOI = 10.1145/1721831.1721833 http://doi.acm.org/10.1145/1721831.1721833

1. INTRODUCTION

A long-term goal of our research is to inform the design of pervasive computing systems that are deployed in urban environments. We regard such systems as more than a collection of technologies, but as a system that includes humans, technologies—mobile and fixed, wired and wireless—and urban space as its components. Developing understandings and ultimately requirements for applications and services that form part of such complex systems presents defining challenges for pervasive computing. The research reported in this article seeks to lay some of the groundwork by developing and applying empirical and analytical tools to begin a systematic understanding of some of the relevant phenomena, specifically patterns of urban mobility and encounter between technological devices and the people who carry and use them.

The richness of activity taking place in the city, along with the inherent difficulties of developing systems in real-world settings, makes it extremely difficult to approach the design of pervasive systems from a theoretically founded, top-down perspective. Furthermore, much of the information in a city does not necessarily flow through routers and cables, but rather through people carrying devices and those devices coming in range with one other, making it even more difficult to obtain data about such a dispersed system. To address these challenges, we combine a systemic view that includes the city, the people, and the technologies with a strongly empirical approach to systematically observing and recording in the city, and making sense of the collected data to build models of human behaviour that can begin to inform our understanding and design of mobile and pervasive systems.

In O'Neill et al. [2006] we described our development of novel methods for systematically observing and recording the city, physically, digitally, and socially. Our methods extended methods conventionally applied to understanding the traditional architectural features and uses of the urban environment. We showed how our methods help analyze and understand mobile and pervasive computing features and uses as integral aspects of that environment. As a central part of our approach, we automated the capture of longitudinal data on *mobility* and *encounter* of users and devices in the city.

This article makes several key contributions in terms of data visualization, complex network analysis, and constructing a model of urban encounter. We develop and apply novel techniques in mining the data collected using previously reported methods. Our novel data visualization and analytical techniques, and findings from applying those techniques, are a main contribution of this article. We show that mobility and encounter data exhibit significant patterns, both structural and temporal. Furthermore, through the use of modeling and emulation we examine the effects of these patterns, and derive a predictive model of urban encounters. Linking our findings to pervasive systems design, we suggest how mobility and encounter data may be utilized as contextual data, and discuss how our techniques might help inform the design of urban pervasive computing systems and enable their runtime self-adaptation.

In Section 2, we briefly review how we developed and applied our methods for collecting mobility and encounter data in a real urban environment extending previous work. Section 3 gives a brief overview of some of the data visualization techniques that we have developed to begin making sense of the captured data. In Sections 4 and 5, we move from visualization techniques to the development of increasingly more formal and systematic analytical concepts and tools, presenting a significant novel finding on the relationship between temporal and structural features of the dataset, and developing a predictive model of human encounter in the city based on our empirical data. In Section 6, we discuss the implications of mining these data for the design and use of pervasive computing systems.

2. CAPTURING MOBILITY AND ENCOUNTER DATA IN THE CITY

Mobility has received a lot of attention as a defining feature of the move from desktop-bound computing to pervasive computing. As is often noted, mobility introduces challenges of attention and orientation during use [Brewster et al. 2003; Nicol et al. 2004; Ker and Schiele 2006], and unreliable network connections [Kjaergaard 2006]. Strongly linked to mobility is the notion of encounter. The movement of people and devices through an urban environment brings them into contact with each other. In an urban pervasive computing system, there are additional patterns of encounter between diverse combinations of users, places, mobile devices, fixed devices, and services. This results in an enormously increased number of spontaneous interactions with consequent effects on security and privacy [Kindberg and Zhang 2001].

A number of projects have focused on capturing mobility data enabled by the popularisation of mobile and wireless technologies. For example, the Reality Mining project[1] collected proximity, location, and activity information, with proximity nodes being discovered through periodic Bluetooth scans and location information by cell tower IDs. Several other groups have performed similar studies [Eagle and Pentland 2006; Balazinska 2003; Chaintreau et al. 2006; McNett and Voelker 2005; Nicolai et al. 2005]. Most of these, such as Balazinska [2003] and Nicolai et al. [2005], use Bluetooth to measure mobility, while others, such as Chaaintreau et al. [2006] and McNett and Voelker [2005], rely on

[1]Reality Mining: http://reality.media.mit.edu, accessed 14/07/2007.

WiFi. The duration of such studies varies from 2 days to over 100 days, and the numbers of participants vary from 8 to over 5000 (see the Haggle[2] project and Crawdad database[3]). The MetroSense project[4] explores the use of people-centric sensing with personal consumer-oriented sensing applications such as Nike+,[5] and sensor-enabled mobile phone applications, which can potentially enable applications such as noise mapping and pollution mapping.[6] The Pervasive Mobile Environmental Sensor Grids (MESSAGE) project[7] aims to collect data at a metropolitan scale through smart phones carried by cyclists, cars, and pedestrians monitoring carbon dioxide values, with an ultimate goal of controlling traffic in the city of Cambridge. Similarly, the urban sensing project CENS[8] seeks to develop cultural and technological approaches for using embedded and mobile sensing to invigorate public space and enhance civic life.

A limitation of previous research in the relatively new field of pervasive computing is the absence of an established toolkit of concrete methods for recording, modeling, analyzing, and understanding salient properties of users and technologies in the urban context. In particular, our community lacks tools that can help us to combine and compare the individual user and aggregate city-scale perspectives. Aggregate models are sometimes based on either small samples of individual data that may not scale up reliably, or on simple software agents whose combined behavior may resemble human behavior at the aggregate level but whose individual behaviour bears no resemblance to individual human behavior in the city. Our major contribution is the development of such tools and techniques to help us analyze human behavior at both the individual and aggregate levels by considering mobility and encounter as two key inputs.

Recording mobility and encounters at the city scale is challenging. A key challenge is in acquiring large scale longitudinal data. In O'Neill et al. [2006] we described in detail the development of methods for automating such longitudinal observation through Bluetooth sensing. We employed these methods in collecting the data described in this paper.

Our central data collection technique exploits the characteristics of Bluetooth technology. Bluetooth is a proximity-based wireless communication technology, allowing devices within a short range (approximately 10, 100, or 250 meters) to communicate directly with each other. 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:

---

[2] Haggle Project: http://www.haggleproject.org, accessed 14/07/2007.

[3] Crawdad project: http://crawdad.cs.dartmouth.edu, accessed 14/07/2007.

[4] MetroSense Project: http://metrosense.cs.dartmouth.edu, accessed 14/07/2007.

[5] Nike+: http://www.nikeplus.com, accessed 14/07/2007.

[6] Noise Mapping England: http://noisemapping.org, accessed 14/07/2007.

[7] MESSAGE Project: http://155.198.92.106/pmesg.html, accessed 14/07/2007.

[8] Urban Sensing: http://research.cens.ucla.edu/projects/2006/systems/Urban_Sensing, accessed 14/07/2007.

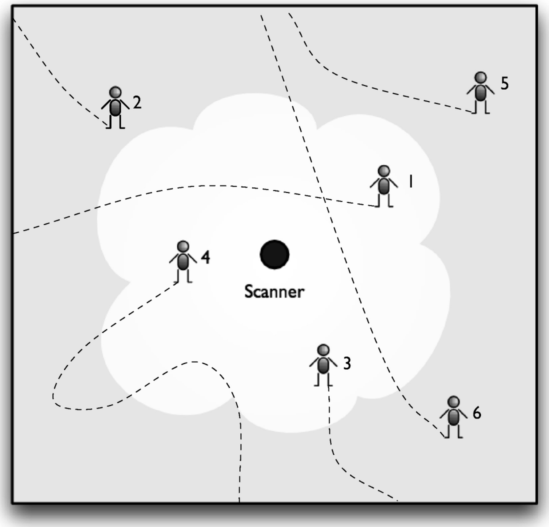


Fig. 1. At each scanning site, a computer uses Bluetooth to monitor the presence of discoverable mobile devices within an approximate 10-meter radius.

physical proximity, the explicit advertisement of *presence*, and the *unique identifiers* transmitted by each device.

We carried out our observations at 8 sites in the city of Bath (including six streets, a pub, and an office). These sites covered a good portion of the city center while at the same time representing different levels of pedestrian activity. Our technical setup involved installing a computer that constantly recorded the presence of discoverable Bluetooth devices within approximately a 10-meter range (Figure 1). The Bluetooth protocol enables any device to search its environment and identify nearby devices that support Bluetooth and are set to be discoverable. Each device has a unique 12-digit identifier, which makes it possible to differentiate between all discovered devices.

The Bluetooth discovery protocol consists of two stages. In the first stage, a list of nearby devices is retrieved, while in the second stage each of these discovered devices is queried for additional information such as device class and user-defined name. During the second stage, which can last minutes if many devices are present, a single-transceiver scanner cannot detect any new devices in the environment, thus failing to record passing Bluetooth devices that typically move into and out of scanner range within seconds at normal pedestrian speeds. In our scanners, one Bluetooth transceiver was dedicated to the stage one discovery of new devices, while the remaining transceivers were responsible for performing stage two, querying those devices for additional information. This method allowed us to capture more data in locations where there were high flows of pedestrian traffic.

It is worth noting that there is a clear distinction between Bluetooth device identity and a person's identity, and our system is able to capture only Bluetooth device identities. A device identity consists of the 12 hexadecimal digits that uniquely identify each device or, specifically, the Bluetooth chipset in the device. In the research reported here we are not interested in people's personal identities, but simply require a mechanism that helps us uniquely differentiate between individuals in space and time. Hence we use Bluetooth as an identification mechanism since Bluetooth is a good proxy for inferring an individual's presence and movement [O'Neill et al. 2006].

If we assume that a single person carries a single Bluetooth device, typically a mobile phone, we can begin to draw some limited inferences about the movement of people from the movement of the devices. Thus, in O'Neill et al. [2006] we reported that for the city of Bath approximately 7.5% of observed pedestrians had discoverable Bluetooth devices. This number, which was derived by correlating manual counts of the number of pedestrians with Bluetooth counts of the number of discovered devices at the same location, will vary over time and place. However, for the analyses presented in this article we do not require knowledge of an exact figure. We should also point out that our data does not contain demographic information about those carrying Bluetooth devices, such as gender, age, social status, education, and where they live. To the extent that our data represents a sample of a larger population (in this case the city of Bath), demographics may be inferred from secondary sources such as local government records and statistics. Even in this case, however, it is a challenge to accurately establish which portions of the population actually carry a Bluetooth-enabled device and which do not.

As this was an observational study in the field rather than, for example, a controlled experiment, our scanners attempted to record any Bluetooth activity detected in the field at each site. At each location we attempted to minimize the amount of "noise" picked up by our scanners by appropriately locating the scanner in the physical environment. For indoor locations we placed our equipment near the centre of the enclosed space so that outside activity could not be detected. For monitoring streets, we placed our equipment on ground floor or first floor windows overlooking the street. However, given the inevitable vagaries of field observation studies and the properties of Bluetooth, in addition to these pre–data collection physical efforts, we also applied post–data collection analytical filters to reduce noise in the dataset.

Hence, due to the omnidirectional nature of Bluetooth, it was possible for our street scanners occasionally to pick up Bluetooth activity from nearby buildings. To address this issue, we can use filters to discard nearby static devices that are likely to be coming from a nearby building rather than the street. Another anomalous case is when a person carries two discoverable Bluetooth devices. In this case unrefined data analysis would falsely infer the presence of two people instead of one, although again post hoc filters could identify such pairs of devices that seem to be consistently discovered in unison. Another potential weakness of our data is the case where multiple people share a single Bluetooth device, in which case our analysis would falsely suggest one person making multiple visits. It is not possible to filter for this case; however, since mobile Bluetooth devices—most likely mobile phones—overwhelmingly tend to be owned and used by one individual, at least in Europe, it is a very unlikely case. For the Bluetooth data collection presented in this article we did not employ any of these filtering mechanisms as we have yet to establish their validity and robustness.

Our Bluetooth scanning equipment does not record anything other than Bluetooth discovery events. In O'Neill et al. [2006] we reported complementary data collection techniques that use manual field observations to correlate the publicly observable urban activities of nearby people with the Bluetooth data recorded by our scanners. For the research reported in this article we did not collect such data. The data visualization and analysis techniques and findings we describe here do not rely on such knowledge and do not—indeed cannot—make claims about the activities and interactions of individuals beyond their relative proximity to our scanner sites from time to time.

Over approximately 12 months we recorded more than 70,000 devices and 3,000,000 discrete records of their presence in the city. We stored our data in a number of database tables, with a single table containing raw data as recorded by our scanners. The records in this table contain a timestamp, the device Bluetooth ID, the device Bluetooth friendly name (if available), and a unique code for the location where the scanning took place. While the Bluetooth protocol allows for the distinction between different "classes" of mobile devices (such as mobile phones, smart phones, PDAs, etc.), we discarded such information.

Our recording of longitudinal and large-scale data allows us to make two significant advances that extend traditional approaches to modeling and understanding the city. Such approaches either rely on simulation-based analyses that often make crude simplifications about human behavior [e.g., Turner and Penn 2002] or analyze quantitative data that cannot be associated to individual behavior (for instance, automated counting of cars, aggregated demographic data, etc.). First, we can inform our aggregate level modeling and analyses with real-world empirical data. This should help us to validate and improve upon the often necessarily simplistic assumptions that are common in agent-based approaches to such modeling. Secondly, we can investigate and analyze data that relate to a single user or a specific group of users, thus individualizing our analysis in ways not possible with traditional aggregate approaches to modeling. In the next section we describe the data that we have collected and demonstrate some of the techniques we have developed for visualizing, analyzing, and making sense of these data.

3. DATA VISUALIZATION

The data record of Bluetooth activity is fundamentally a set of individual Bluetooth discovery events. A single discovered device typically generates multiple events while it is within range of a scanner. 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 critical feature of these data is the temporal aspect, a theme to which we will return in subsequent sections. A temporal view allows us to begin making sense of the individual Bluetooth discovery records. Figure 2 illustrates our

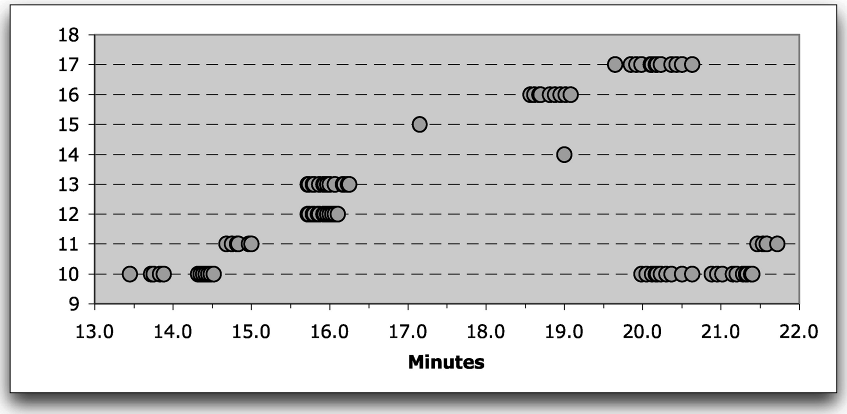


Fig. 2. A timeline visualization of our Bluetooth scanning. Each device is given its own timeline (dashed lines) and each Bluetooth discovery event is plotted as a circle on the timeline.

most basic temporal visualization of a snapshot of raw data generated by our Bluetooth scanners.

In Figure 2 we see a set of timelines, indicated as dashed horizontal axes. Whenever a new device (i.e., a device for which no Bluetooth discovery events are already recorded in our dataset) is discovered in the environment, it is allocated its own timeline above the last newly created timeline. 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. On each timeline we indicate with a blue circle the point in time at which a device was recorded by a scanner. We refer to these as contact points. Typically, a device moving past a scanner will generate a series of successive contact points on its timeline. So, for example, Device 16 (in Figure 2) appeared approximately 18.5 minutes after scanning began and was present for almost a minute.

Successively "zooming out" from our basic temporal visualisation provides a very useful tool for making sense of the data. Figure 3 illustrates the effect of zooming out from the timelines shown in Figure 2.

Our zoomed out timeline visualisation creates the cumulative effect of a diagonal line from bottom left to top right. This diagonal effect is created by the appearance of new devices over time. Any activity recorded below this main diagonal is attributable to persistent devices, this is, devices that remain in the vicinity of our scanner for relatively long periods of time. In investigating mobility, persistence is a crucial factor, since a device that is persistent at one scanner site is by definition not mobile, given a scanner radius of less than the area under consideration. Even more significantly, persistence is crucial since it increases the probability of encounters, as we discuss in Section 5.1.

Examining the "timeline diagonal" visualization allows us to compare patterns of transience and persistence across different scanner sites. Our knowledge of the urban spaces in which these scanners are situated then allows us to begin to relate features of the urban form to the observed patterns of transience and persistence. Thus, Figure 4 shows data from 3 Bluetooth gatecounts that

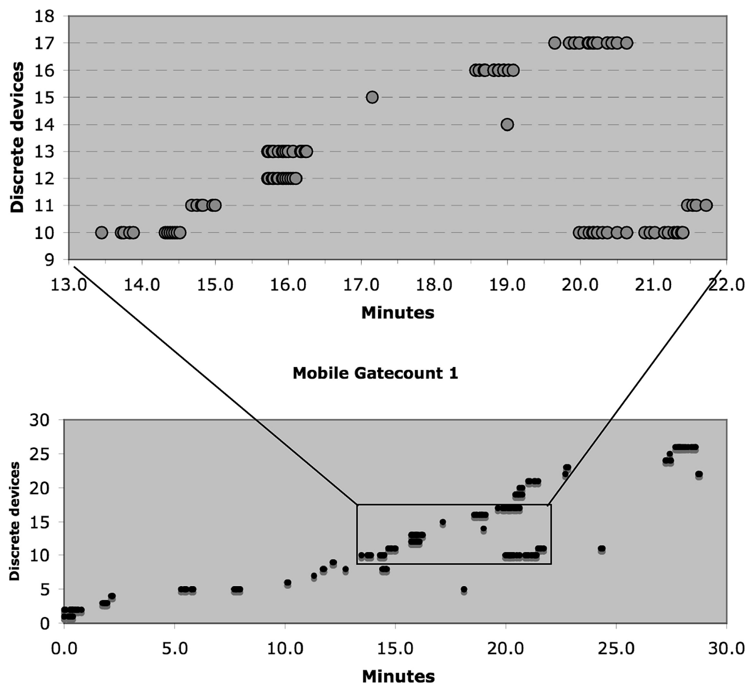


Fig. 3. A timeline visualization of our Bluetooth gatecounts. Each device is given its own timeline (dashed lines in top half) and each discovery event is plotted as a circle on the timeline.

took place at different sites and reflect contrasting patterns of Bluetooth presence. Gatecount 9, carried out near a tourist attraction, has a relatively high level of persistent devices, indicating that people are spending a few minutes at this location. We also observe bursts of Bluetooth activity recorded at 0, 5, 13, and 23 minutes, which correspond to groups of people arriving at the location. Gatecount 5, on a relatively quiet street in a residential area, recorded devices appearing at a rate of 22 per thirty minutes. These were mostly transient devices, with a few persistent devices including one that stayed for 17 minutes. This location was characterized mainly by individuals walking between their homes and the city center. Finally, Gatecount 10 was recorded on a busy street leading to the train station. This location has a much higher flow of devices, with 90 devices appearing in thirty minutes. These were almost entirely transient devices, correlating with pedestrians making their way between the train station and the city center.

Zooming out even further, we can begin to identify daily and weekly temporal patterns in the data. In Figure 5 we show a visualization from one location over 1 day, 1 week, and 2 weeks. In these graphs we can identify distinct days, which appear as distinct humps in the main diagonal of the timeline. From the length of the hump we can infer the number of new devices that appeared on that day. Furthermore, the slopes of sections of the main diagonal reflect the rate at which new devices appear. For instance Saturdays (the 6th and 13th days) have a much higher number of new devices appearing than any other day

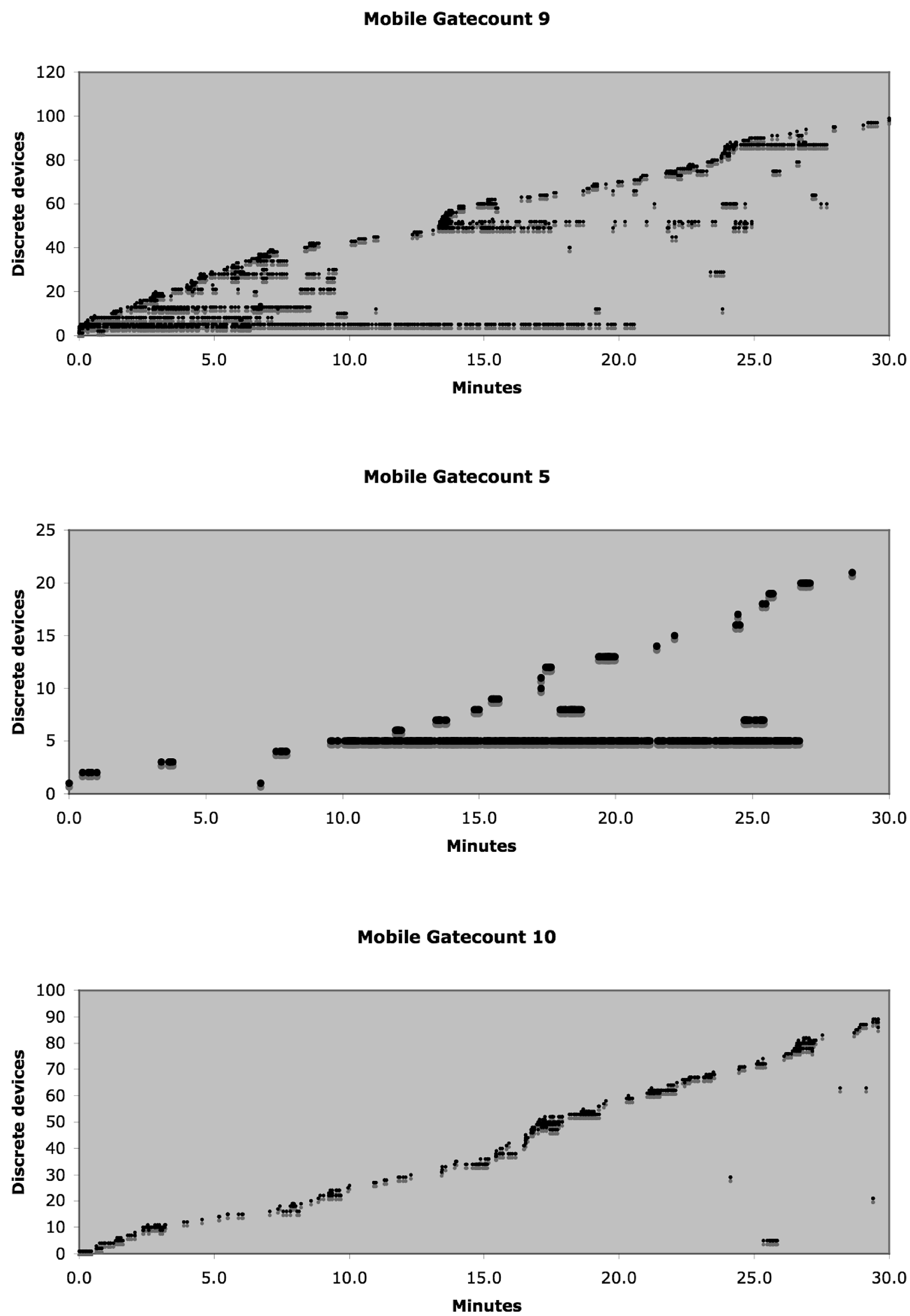


Fig. 4. Visualizing Bluetooth gatecount records. Activity below the main diagonal indicates persistent devices.

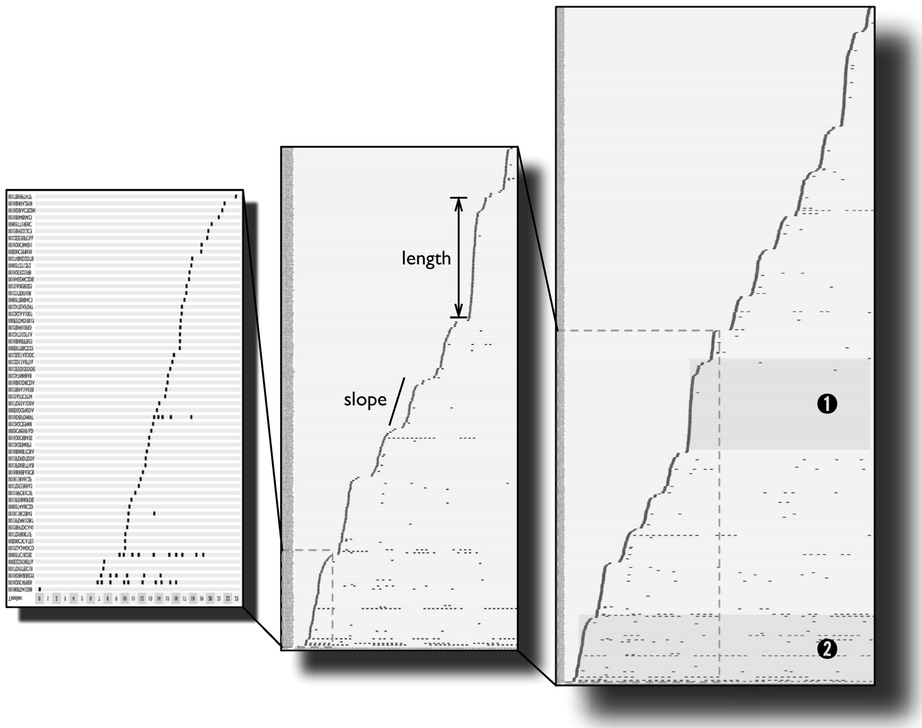


Fig. 5. Visualizing Bluetooth gatecount records over 1 day (left), 1 week (middle), and 2 weeks (right).

(giving a long hump), and at a higher rate (giving a steeper slope). We also see in shaded region 1 that the devices appearing on Saturday do not reappear during the following week, or indeed during the following Saturday. Additionally, in shaded region 2 we see that many devices that appear on the first day, a Monday, then reappear during the following weeks at regular intervals. These patterns in the data reflect a daily routine observed during weekdays, coupled with a relatively distinct pattern at weekends [Department for Transport 2005].

The visualizations presented in this section allow us to begin making sense of the data temporally and spatially. They 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. While these visualizations are themselves a very useful toolkit in making sense of the data, visual inspection offers only one approach to analysis of our data, and typically assists in hypothesis generation. In the next section, we describe how we built on our visualizations to develop more formal and systematic analytical concepts and tools. These allow us to automate aspects of our data analysis. We then go on to demonstrate the application of these concepts and tools to produce important novel results in understanding mobility and encounter in the urban environment.

Table I. The Characteristics of Our Basic Concepts: Session, Encounter, and Trail

Concept	Characteristics
Session	Device, location, start time, duration
Encounter	Device x 2, location, start time, duration
Trail	Device, location x n, start time, duration

4. DATA ANALYSIS: SESSIONS, TRAILS, AND ENCOUNTERS

In this section we introduce the basic concepts of our approach to our data analysis, and then proceed to describe some of the findings resulting from applying our tools to mine the dataset collected from our scanner sites in Bath. To illustrate our analysis and results, we draw on examples from four contrasting scanning sites. These sites are: a passageway on our university campus, a street in the city centre, a city center pub, and an office, also in the city centre.

The data visualizations discussed in the previous section provide the foundation for an approach to making sense of our data in terms of three distinct abstractions: sessions, encounters and trails (Table I and Figure 6). A session is defined as a set of contact points having no more than a threshold temporal distance of $\delta 1$ between any two consecutive points, that is, $\delta 1$ is a time-out threshold. 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. We can inspect a visualization such as Figure 2 to identify distinct sessions. For instance, device 13 generated a session between 15.5 and 16.5 minutes. Similarly, device 10 recorded a session between 13.5 and 14.5 minutes and between 20 and 21.5 minutes. 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. Visually inspecting Figure 2, we see that devices 12 and 13 encountered each other, since their sessions overlap, and they recorded these sessions at the same location in the city of Bath. We calculate the duration of each encounter as the period of overlap between the sessions. Thus, devices 12 and 13 encountered each other for just over 30 seconds, while devices 10 and 17 encountered each other for about 45 seconds.

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. Note that a trail may span a number of locations across the city. 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.

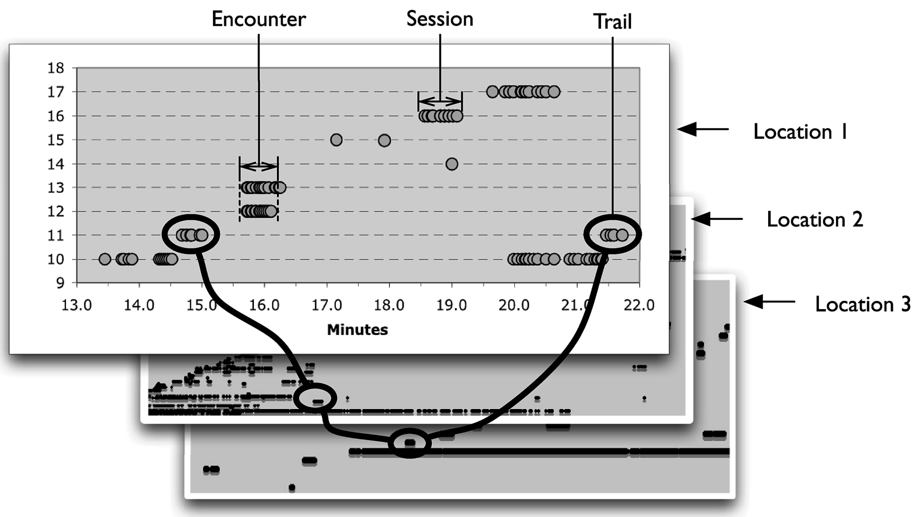


Fig. 6. A visual representation of sessions, encounters, and trails.

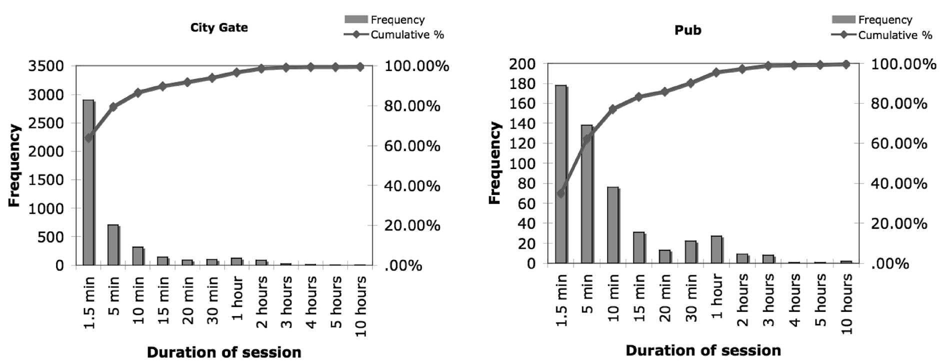


Fig. 7. Duration of presence of Bluetooth devices at two locations with distinct characteristics.

4.1 Sessions

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 [O'Neill et al. 2006].

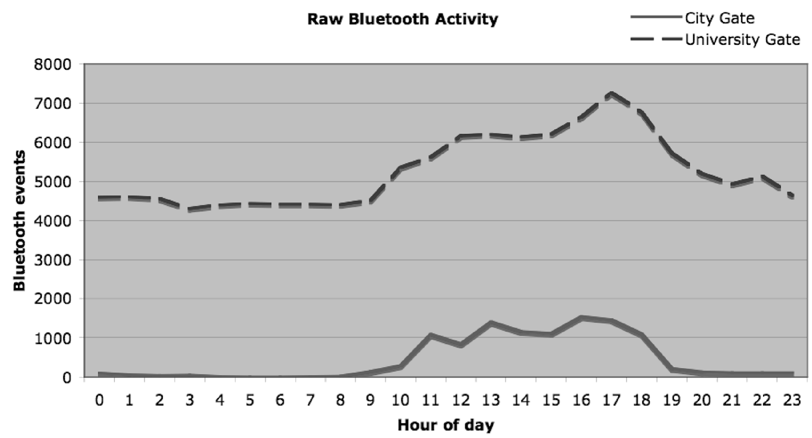
In Figure 7 we compare the distribution of session duration for two locations: the street in the city center (left) and the pub (right). From these distributions we can see that in the city centre street the vast majority of devices is seen for less than 90 seconds per session, indicating highly transient devices and people. On the contrary, the distribution of session duration in the pub suggests that people are much more static, with about 65% of sessions lasting more than 90 seconds.

This threshold of around 90 seconds allows us to establish empirically a conceptual distinction between transient and persistent devices. From a static observer's (or scanner's) point of view, transient devices come and go in less than 90 seconds, while persistent devices remain relatively static. Considering the pub in Figure 7, we see that while the number of sessions with duration up to 20 minutes gradually decreases, there is a rise in the number of sessions lasting half an hour and one hour. A Mann–Whitney U test showed that the distributions of the two samples differed significantly ($U = 27177$, $p < 0.0001$). Relating this observed pattern in the data to our knowledge of the space in which the corresponding scanner is situated, we attribute this pattern to the actual activities taking place in a pub, as opposed to the street. People have a reason to spend an hour in the pub rather than on the street.

If, conceptually, we distinguish between slow-moving (persistent) and fast-moving (transient) devices, we can study how each group flows through different spaces in the city. Note that transient devices are those that have sessions of up to 90 seconds, and hence are shown in the leftmost bar of each bar chart in Figure 7. In Figure 8 we show on an hourly basis the number of transient devices that passed along the street in the city center (solid line) and the passageway on our university campus (dashed line). Examining transience here shows us how raw Bluetooth data can be misleading. From the top of Figure 8 it appears that our university scanner gate was much busier than the city center gate, with much more Bluetooth activity recorded at the university gate (Figure 8 top). However, filtering out multiple records per device and persistent devices (indicating nearby static Bluetooth devices), we can identify the transient Bluetooth devices, shown in the bottom half of Figure 8.

So the city center gate peaks at 15 unique transient devices per hour, while the campus gate peaks at 6 devices per hour. The 2 graphs have a very similar profile despite recording Bluetooth traffic at sites with very different characteristics. This reflects the temporal pattern of activity *across* the city, with Bluetooth traffic at both sites peaking in mid-afternoon. The peak of 15 devices per hour for the city centre gate refers to the period around 1 p.m., corresponding to 7.8% of the pedestrian traffic for that location [O'Neill et al. 2006]. Again, these temporal patterns in the data reflect the periodicity of urban life as observed in other studies, for example of urban transport [Department for Transport 2005].

Our measure of transience, based on our empirically derived threshold for session duration, describes how people's movement may be viewed from the perspective of a specific location or a static observer. For instance, a smart poster placed in our city center street would expect to see about 15 highly transient devices between 12 p.m. and 1 p.m. Similarly, we can calculate the number of unique devices with sessions of, say, more than 30 minutes in the pub, and so on. In addition to understanding the patterns of persistence and transience from the perspective of a single space, we can also begin to explore how transient devices move through a sequence of urban spaces. To make sense of the data from this perspective, we use the concept of trails.



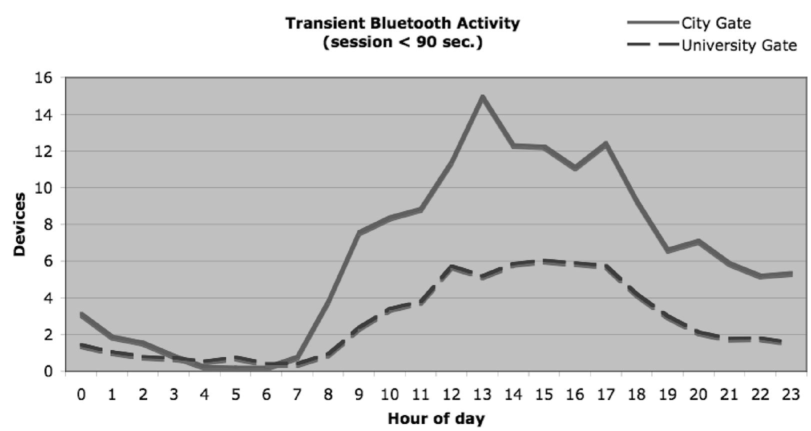


Fig. 8. Raw Bluetooth activity (top) and transient Bluetooth devices (bottom) for our campus and city center street gates.

4.2 Trails

A trail is a sequence of sessions recorded for a specific device over time across a succession of urban spaces in which there are scanners. Thus, 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 coordinates, 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 analyze and compare trails.

To carry out trail analysis, we have built a query engine which can be used to search, retrieve, and rank trails based on any of the properties of the trails (device, hops, start time, duration) as well as their combinations. For instance, we can search for trails between any two specific locations, which enables us to identify the range of routes people take between those two locations. We can also search for trails that start at a specific location and last more or less than a fixed period of time, thus identifying the routes that people take when leaving, for example, a central square or a train station. Similarly, we can identify all trails that pass through a specific location, such as a coffee shop, and thus begin to understand the routes that the coffee shop serves. We can also search for cyclic trails, thus identifying locations that may act as entry and exit points to a site, such as the entrance/exit to a zoo. Furthermore, we can query for trails that occur within a specific time period (e.g., in the morning, in the afternoon, on Tuesdays, or during a particular event such as market day or a music festival), thus identifying how people's trail preferences change over time and seasons.

All of these analyses are possible on complete datasets, or may be applied to a subset of the data focusing, for example, on a specific device, or comparing males *vs* females, teenagers *vs* adults, or locals *vs* tourists, provided that such data has been recorded.

Having retrieved all trails that match our search criteria, our system ranks these trails according to their significance. A trail is significant if it matches one or more of the following criteria.

—It is one of the top *n* trails in terms of trail popularity. Given all trails that fulfill the search criteria, popularity for each trail can be defined as the number of times that the trail was recorded in the data, or the number of distinct devices that recorded the specific trail.
—It is one of the top *n* trails in terms of average time spent on the trail, or some other time-related statistical measure.
—It is one of the top *n* trails in terms of relevance of the locations to a particular subject (e.g., shopping the high street, or taking the scenic route).

The results of our search queries can be viewed as text or visualized as a network graph, as illustrated in Figure 9. Here, each node represents a location, while links between nodes represent transitions between locations made by a specific mobile device and, by extension, its user. For visualization purposes, we assign each trail a distinct color, and use a layout algorithm that avoids overlaps between graph edges. Graphs can offer an effective way to inspect a set of trails, and explore the relationships among them. For any given set of trails matching a set of criteria, we are able visually 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.

Figure 9 highlights that our trails are not topographic descriptions of routes between points in the city. Clearly, if we were interested simply in a trace of the movements of a device through urban space, we could use various technologies, for example, GPS, to provide trails. However, the approach described here provides a means of analyzing trails in which the nodes are particular spaces in the city with particular meanings—the pub, the museum, the doctor's surgery, the suspected safe house. We can then begin to mine the data in terms of these meanings associated with the network, typically by looking for patterns and deviations from patterns. So, for example, the Tourist Information Centre in Bath would like to know which sites people visit, in what order and over what

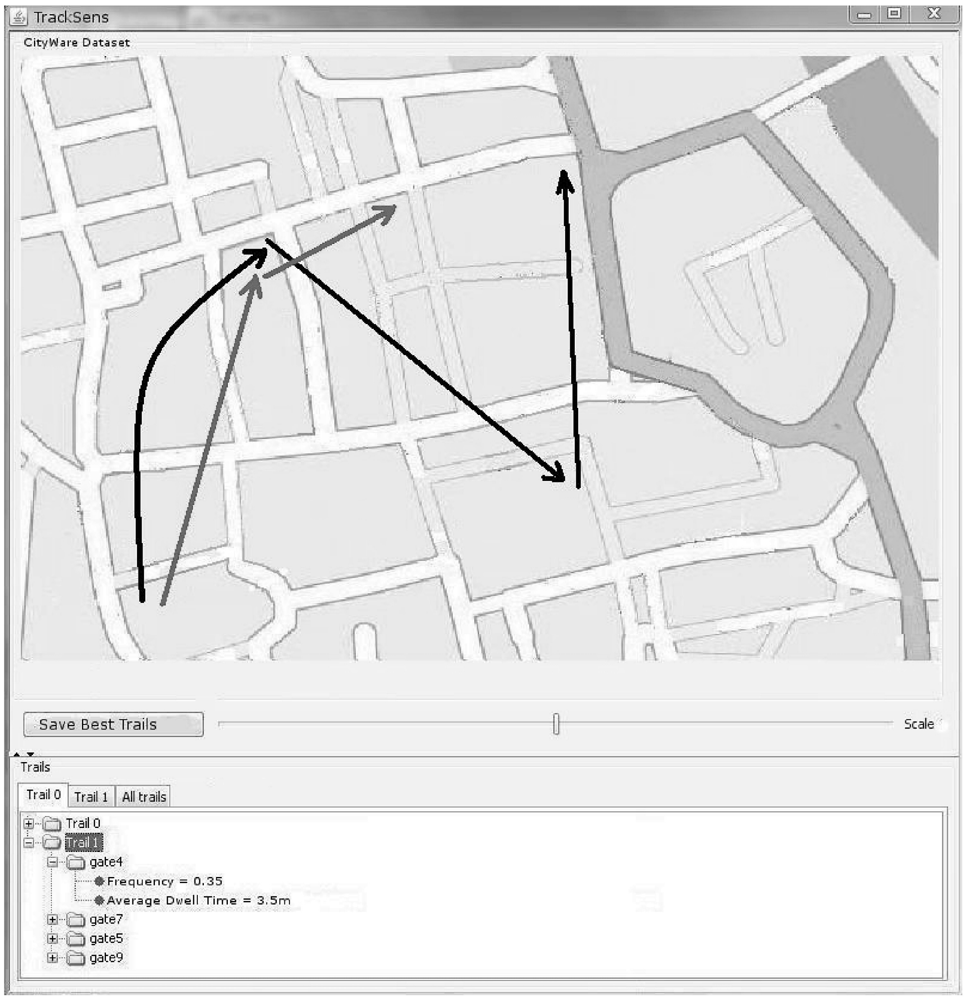


Fig. 9. Visualization of trails in the city of Bath as a network graph. In this graph, nodes represent locations, while edges represent transitions between locations made by a specific mobile device (and its user).

period, but they do not particularly care about precisely which routes people chose between those sites. In this case, by identifying patterns in visitors' trails, as well as deviations from these patterns, the tourist agencies can provide more appropriate services and better adaptation to visitors' needs.

While our trail visualizations are amenable to visual inspection much like our earlier visualizations, they represent a crucial move to using network graphs. From each location we have a number of transitions, each of which goes to a distinct location. Hence, the totality of locations is connected via a tree of transitions. Within this tree we choose to analyze trails that represent linear chains of transitions between locations. In the following sections we describe how we move from analysing network graphs visually to analysing them algorithmically.

4.3 Encounters

Visualising and analyzing our raw Bluetooth activity data as sessions and trails allowed 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 allowed 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 be in the same space at the same time. It is this notion of encounters between the diverse combinations of mobile and fixed devices and people in the urban environment that provides the richest ground for mining our dataset. 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. As previously noted, however, our Bluetooth scans do not record the details of any real-world interactions. They simply record that 2 or more devices were simultaneously within range of one of our Bluetooth scanners. Hence, our notion of encounter in our Bluetooth data does not describe real-world relationships between humans, but rather opportunities that may arise for networking, both social and digital, amongst humans and devices.

To study the patterns of encounter in our data, we first need to identify instances where two or more devices were present at the same time in the same place, that is, within range of one of our scanners. For example, from Figure 2 we can see that devices 12 and 13 encountered each other at 15.5 minutes and were together for approximately 1 minute. In observing encounters, we can once again make use of relatively simple timeline visualisations, as illustrated in Figure 10. In this figure each unique device's timeline is represented as a dotted line, while yellow bars indicate the sessions actually recorded by each device at this scanner site. We can therefore record encounters by identifying device sessions that overlap in time and were recorded at the same location. For example, in Figure 10 we can see that device 1 encountered devices 3, 4, and 6, while device 3 encountered devices 1 and 4.

However, the timeline visualization is not an appropriate tool for examining encounter over the entire dataset of millions of events. Hence, we required additional tools to analyze encounter. We developed filters that analyzed our dataset and gave us instances of devices encountering each other across every location in our study. For each encounter, we identify the two devices involved, the location of the encounter, the time when the encounter began, and the duration of the encounter. These results are stored as records with the following form:

device1_id, device2_id, location, start time, duration.

At this stage in our analysis we have a long list of such records, describing which devices encountered each other at which locations and times. This list

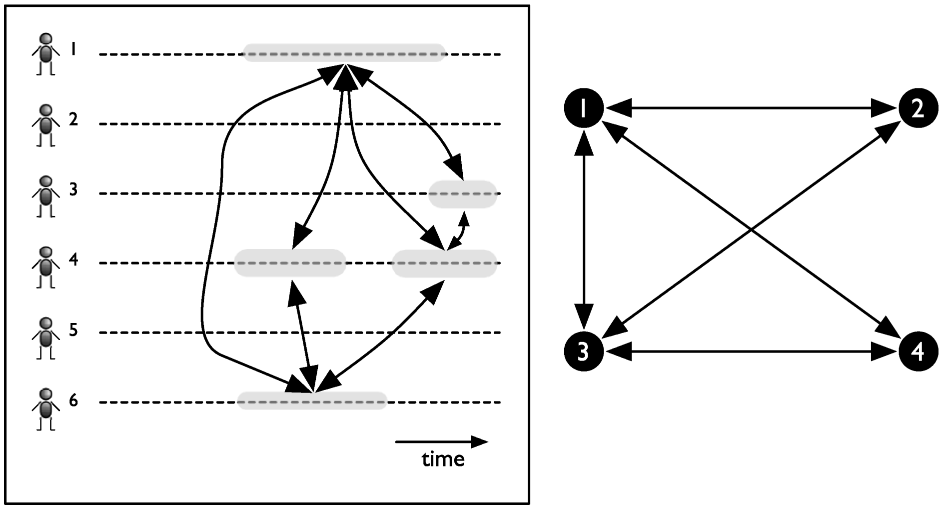


Fig. 10. By identifying device sessions (yellow bars) that overlap in time at a given scanner site, we can record encounters (arrows) between devices.

of encounters is a textual representation of the patterns of encounter in our dataset. We then transform this textual representation to a network graph. Assuming that each device from our dataset becomes a node in this graph, then the list of encounters describes the links between all nodes. Thus, we are able to generate "complex network" graphs [Strogatz 2001] that represent the patterns of encounter across our entire dataset. We can generate various graphs from our data, such as an individual graph per scanner site, or a graph of our entire dataset in one city-scale graph. Furthermore, we can generate these graphs over the entire lifetime of our scanning or over any specified period.

In Figure 11 we show complex network graphs from the pub in our study for periods of 1 day, 1 week, 1 month, and 6 months. In this graph, each device is represented as a node in the graph, and connected nodes indicate that these devices encountered each other at some time in this period. We observe that most devices are linked to a single cluster, while few devices are isolated. The latter indicate cases where a device was seen only by itself and never in the presence of other recorded devices. The size of a node represents the total amount of time that the corresponding device has spent in this location. Similarly, the color of the nodes (blue to red) indicates the "betweenness" of a node (from 0 to 1 respectively), which describes the extent to which an individual acts as a bridge between other people in the network [Freeman 2004]. Hence red individuals are those who are central to the graph, while blue individuals are relatively isolated. This metric, along with an array of further 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 (blue nodes) and examine the trails these individuals take across the city.

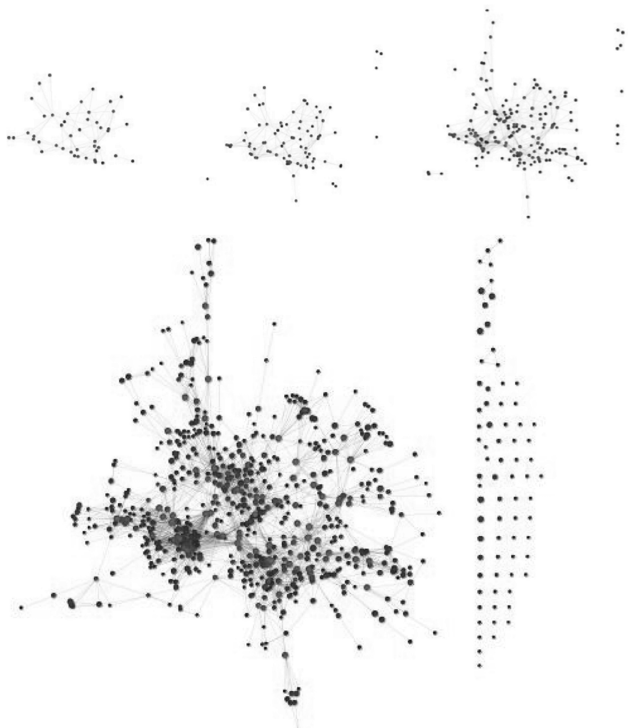


Fig. 11. A network graph visualization of the encounters at one of the locations in our study. The graphs represent our data after a day (top left), a week (top center), a month (top right), and six months (bottom).

Once again, network graphs prove to be a useful tool in making sense of the data. However, the complexity of our network graphs makes visual inspection cumbersome or impossible for all but the smallest and simplest subsets of our data. In the next section, therefore, we describe our application of more algorithmic approaches in order to make tractable and to automate the mining of our entire dataset.

5. COMPLEX NETWORK ANALYSIS

An advantage of using network graphs to represent our data is that we can use well-established algorithms for analyzing the structure and properties of such graphs [Kostakos and O'Neill 2007]. To enable this kind of analysis, we encode our data as graphs by representing each device as a node and creating links between those devices (nodes) that encountered each other. Furthermore, depending on where we set the boundaries of our dataset we can derive a number of different networks. In our case we consider 5 distinct networks (Table II). The "Bath" network represents all the devices captured in the city of Bath, and all the encounters between those devices. The other four networks (campus, street, pub, office) were generating by only considering the data captured at each site respectively. For instance, the "pub" network is a graph where nodes

Table II.
Structural properties of our complete dataset for 12 months and some of its subsets. For each subset we show size of the graph, number of edges in the graph, density of edges (total edges / size * size-1), size of largest cluster (core), average number of edges per node (k), diameter of largest cluster ($\lambda_{\max}$), average distance (λ), and average clustering coefficient (C)

	Size	Edges	Density	Core	k	$\lambda_{\max}$	λ	C
Bath	70516	652446	0.03%	69655	18.53	11	3.45	0.47
Campus	3109	120273	2.5%	3101	77.37	6	2.57	0.44
Street	11853	58111	0.08%	10584	9.80	12	3.23	0.28
Pub	13476	126768	0.1%	13383	18.81	9	2.61	0.69
Office	321	2419	4.7%	318	15.21	4	2.04	0.82

represent the people that where observed in the pub, and links between nodes represent encounters between those people that took place in the pub. We have focused our analysis on these 4 locations is that they represent a diverse range of spaces that can be found in a city.

Using network representations for data coding, we are able to derive a variety of properties about each node (hence user) of our dataset. These properties are commonly used in assessing the role of a node in the context of the whole network or in the context of its immediate neighbourhood in the graph. Their use as indicators of a node's role and position in a graph has lead to these metrics being termed "structural." Examples of such properties include "degree" (the number of links that a node has), or the clustering coefficient (also known as "transitivity": the extend to which a node's neighbors know each other). In addition, we can derive metrics for a graph as a whole. One such metric is the size of the graph (number of nodes) or the average shortest path (average number of hops required to go from one node to another).

Having established these metrics, we can compare distinct networks in two ways. First we can directly compare metrics that describe the graphs as a whole (e.g., number of nodes). In addition, we can compare the distributions of node-specific variables. For instance, given two graphs we can consider the distribution of nodes' degree (number of links) or clustering coefficient. These are two of the types of analysis that we present in this section.

In Section 5.1 we present a structural analysis of the network graphs captured by our Bluetooth scanners, which gives us insights into the patterns of encounter of Bluetooth devices across different locations in a city. In Section 5.2 we present an analysis of the temporal properties of our complex networks. Finally, in Section 5.3 we develop and test a model of urban encounter. We show how this model describes very well our observed data on encounters, and provides a basis for explaining the temporal patterns we have observed, as well as the structural properties of our data.

5.1 Structural Analysis of Encounters

In Table II we present structural properties of our entire dataset for Bath over 12 months, as well as four distinct subsets. We note that our graphs do not represent explicit social relationships between individuals, but rather represent the encounters between individuals' mobile devices in a city. These encounters may provide opportunities for networking (both social and digital) that arise due to the movements and copresence of people and their devices.

All the networks in Table II exhibit small average distance λ, that is, average shortest path between pairs of nodes, and high clustering coefficients C, that is, the probability that x connected to y and x connected to z implies y connected to z. We point out that a random graph with the same number of nodes and edges as the Bath graph in theory should have a clustering coefficient of 2.6×10^{-4}. The small λ and large $C$for all graphs are indicative of small-world networks [Watts and Strogatz 1998], suggesting that in our networks most nodes are not neighbours of one another but most nodes can be reached from every other by a small number of links. The corollary in the city is that while most people in Bath directly encounter only a small portion of Bath residents, they are only a small number of hops away from everyone else in Bath. This is in contrast to a network such as, for example, the national road network where an average journey requires a large number of changes of direction, or hops [Strogatz 2001].

Furthermore, the distribution of connections per node across the whole dataset follows an approximate power law equation y = x -α [Newman 2005] with $\alpha \approx 2.5$ Figure 12(a)). This suggests that in our networks a few nodes act as "highly connected hubs", while most nodes are not well connected. In other words, over time a few devices collect an extremely large number of encounters, while the vast majority of devices have a small number of encounters. This characteristic is "*scale-free*," meaning that it is independent of the number of nodes in the network [Barabasi 1999]. Therefore, even if the size of the community changes drastically, the property will persist.

Finally, clustering in our dataset follows the approximate relationship $C(k) \approx 1/k$ (Figure 12(b)), which suggests an underlying modularization of our data [Dorogovtsev et al. 2002; Ravasz et al. 2002]. This property, which has been observed in fractal-like structures resulting from recursive self-replication, suggests that as the number of a node N's links grows, the density of links between the nodes encountered by N decreases proportionately. Thus, if device A has a small number of encounters and it encounters devices B and C, there is high probability that B has encountered C, and this probability decreases as the number of A's encounters increases.

The presence of small world, scale-free, and clustering properties is a crucial finding. It indicates that our data are not random but rather follow a pattern seen often in nature. Specifically, power law distributions (straight lines on a graph where both axes are set to a logarithmic scale) have been identified in a number of natural and social phenomena, including the behaviour of water molecules near boiling temperature, the size of craters on the moon, the magnitude of earthquakes, citations of scientific papers, Web site hits, and the structure of the internet's backbone [Newman 2004].

The structural properties of the subnetworks in our dataset tell a story that intuitively makes sense in terms of the places where data recording took place. For instance we see the strongest clustering$C$and smallest distance λ in the office rather than on the street, indicating a more clustered, tightly knit network of encounters in the office. Correspondingly, we would expect an office environment to be much more clustered than a street. Furthermore, we see that

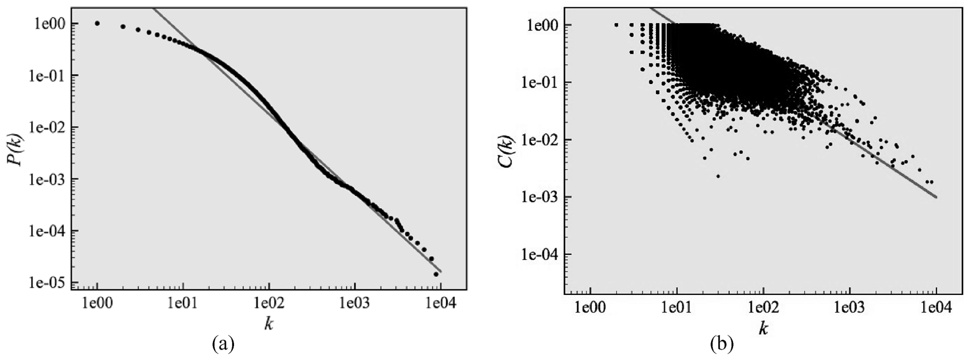


Fig. 12. Structural properties. On the left (Figure 12(a)), the Pareto distribution of degree k follows an approximate power law with $\alpha \approx 2.5$ (red line). On the right (Figure 12(b)), we observe that degree k versus clustering coefficient $C(k)$ follows the relationship $C(k) \approx 1/k * c$ shown in the red line ($c = 9.8$).

the street network has far fewer encounters than the campus network, even though it has many more nodes than the campus network. In fact, the campus network has only a quarter of the nodes but twice as many encounters as the street, suggesting a small yet active community. On the other hand, the street appears to be a large sparse community with relatively less activity.

To get further insight into the different types of communities and their structures as recorded at our scanning sites, we can calculate the degrees of separation between all possible pairs of nodes at each location. An often cited, although controversial, paper examining degrees of separation is Milgram's work suggesting that people are connected by around six degrees of separation [Milgram 1967]. More recent work by Leskovec and Horvitz [2008] looking at the global social network of the Microsoft instant messaging service found that the average distance between people was 6.6. Just as Milgram's work considered social relationships as opportunities for forwarding letters and Leskovec and Horvitz's work considers conversations as opportunities for passing along information, we consider encounters in our dataset as opportunities for forwarding information or even viruses, as described below. In Figure 13 we plot the probability distribution of graph distance across the four subnetworks of Table II. Both the pub and office networks are heavily shifted towards the left, indicating "tight" networks, while the campus and street networks are shifted towards the right, indicating relatively "loose" networks. Practically, the distinction between tight and loose networks is that people in locations with tight networks are familiar with each other, as for example the regular customers in a pub or the employees of an office. On the other hand, one is likely to be surrounded by strangers in locations with loose networks such as the pub.

The average distance we have observed is smaller that the values suggested by previous work (e.g., 6 or 6.6). We believe this is due to the fact that our data collection mechanisms are operating on relatively small, location-bound communities. Sending letters and having instant messaging conversations can potentially be done with anyone on the planet, but physical copresence is typically more difficult and expensive to establish. These limitations, we argue,

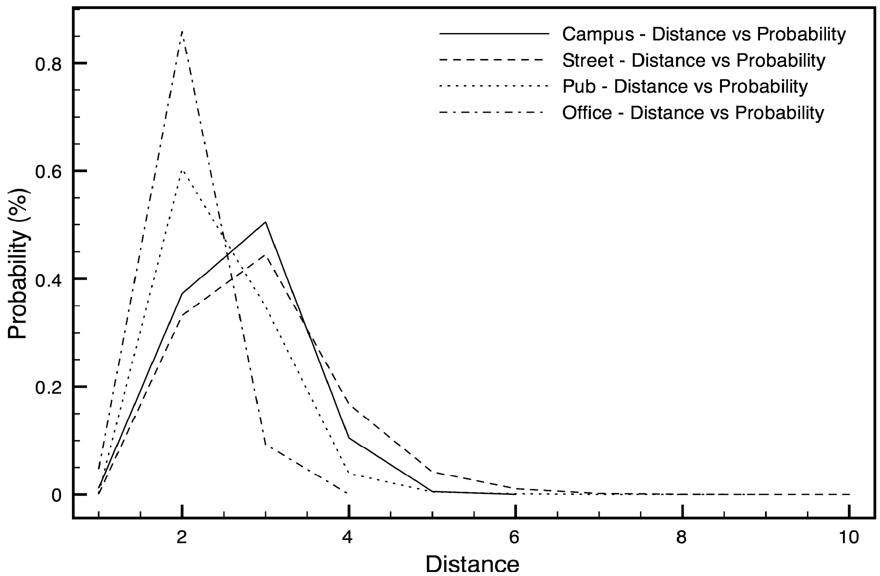


Fig. 13. Probability distribution of distance between any two nodes in the network graphs of four scanning sites.

act against the spreading of the network, and effectively act in a way so as to minimize the average distance between nodes.

In Section 5.1 we have investigated what are known as structural or static properties of network graphs. In other words, we have derived statistics and metrics from a representation of our data that completely discards temporal and sequential information. While useful in many respects, this popular type of analysis has considerable limitations, as it assumes a static 2D representation of the data (i.e., a complex network graph) from which it derives metrics. In the following section we present a novel analysis that provides insights on the temporal properties of our data, taking into account the sequence in which encounters take place.

5.2 Temporal Analysis of Encounters

Although network analysis algorithms are readily available and there is a large body of research on representing and analyzing many complex human and natural phenomena as networks of relationships amongst individuals [Strogatz 2001; Onody and de Castro 2004; Holme 2003; Holme et al. 2007], previous research has used network representations that aggregate the relationships at discrete intervals [Snijders 2001; Leskovec et al. 2005]. Time has been explicitly considered by [Kumar et al. 2006], but only to describe when individuals join a network as opposed to when individuals interact with each other. However, links between network nodes such as the encounters between mobile devices represented in our network graphs are typically intermittent, and their aggregation does not take account of important temporal information, inhibiting understanding of the network's dynamic behavior and evolution.

Previous techniques have been developed to describe the dynamics of networks such as the Brazilian soccer network [Onody and de Castro 2004], online dating networks [Holme 2003], and student affiliation networks [Holme et al. 2007]. However, such work typically relies on the analysis of a limited number of discrete snapshots of the complex networks [Snijders 2001], often because of the impossibility of collecting large-scale longitudinal data. Our data, on the other hand, consist of a chain of events that allows for a minute-by-minute evolving description of the network as people move into and out of contact with each other and our scanners. Here we explore the temporal properties of our network by focusing on three key aspects: presence and frequency of nodes, presence and frequency of links, and temporal order of events. (There are alternative approaches to analyzing our data without explicitly taking into account the relationships between users, for example, by building models of user activity that utilize multiple sources of contextual information such as location and time of day [Eagle and Pentland 2006] resulting in the ability to predict or infer user activity.)

While in Figure 11 all nodes are visible simultaneously in the graph representation, in the city they were available intermittently, and only when the corresponding individuals were at a scanning location. We call this availability *node presence* (n_p), calculated as the total amount of time an individual spent near one of our scanners during the study, that is, the sum of their sessions. This measure is effectively the sum of all sessions for each device. In Figure 14(a) (black solid line) we see that n_p follows a power law with $\alpha \approx 1.9$. Thus, while most individuals were seen only for a few seconds, a few accumulated a presence of more than a month during their visits to our scanner sites. A further temporal aspect, *node frequency* (n_f), describes the number of distinct sessions recorded for each unique device. In Figure 14(b) (black solid line) we see that n_f follows a power law with $\alpha \approx 2.6$. Thus, most individuals were seen only once, while a few recorded more than 1000 sessions. Finally, we have observed that n_p and n_f are not correlated across all the whole dataset, suggesting that the total amount of time someone spends at a location is not related to how many times they visit it.

Previous work has suggested that power laws occur when nodes with high values for connectedness k, that is, many links or previous encounters, are more likely to attract new links, an explanation known as preferential attachment or "the rich get richer" [Barabasi 1999]. A classic example of this scheme is the way in which Web sites link to one another: a new page appearing on the Web is much more likely to link to already very popular pages (such as www.google.com) rather than to unpopular pages such as the authors' home pages. The analysis of our data, however, suggests an alternative explanation. When considering our urban encounters, node availability is a prerequisite for establishing links: a device gains links by being near a scanner and "waiting" for others to show up. As illustrated in Figure 10, an individual encounters *all* other copresent individuals, regardless of how many encounters they have previously had, that is, regardless of their value of k. Nodes' behavior, however, varies in n_p and n_f— some are more "persistent" or "frequent" than others, and thus are more likely to gain new links. For example, a customer entering the pub is more likely to

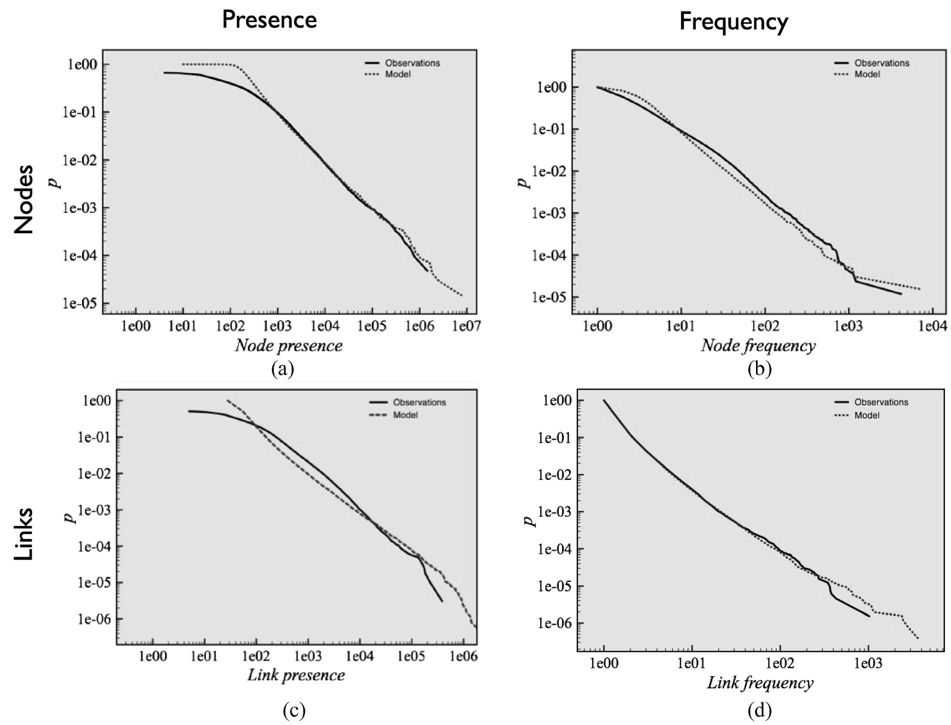


Fig. 14. Temporal properties. Figure 14(a) (top left) shows the distribution node presence follows a power law with $\alpha \approx 1.9$. Figure 14(b) (top right) shows the distribution of node frequency follows a power law with $\alpha \approx 2.6$. Figure 14(c) (bottom left) shows link presence follows a power law with $\alpha \approx 2.3$. In Figure 14(d) (bottom right), we observe that link frequency follows a power law with $\alpha \approx 2.7$. Black solid lines represent the recorded empirical data, dashed lines represent the model we have developed.

encounter the barman, not because the barman has had many encounters, but because the barman is more likely actually to be in the pub. Thus, our analysis suggests that a power law distribution of k can also result from an attachment process driven by nodes' temporal availability,rather than connectedness per se. We find that for any frequency this model is a good approximation of our real world data.

Availability and connectedness are completely independent in our analysis. In theory, a node may be highly available (i.e., a person spending a lot of time at a location), yet have very low connectedness because few other nodes are available at that location (e.g., few others go to that location). On the other hand, a node of high availability can have high connectedness due to the busy nature of the location. Similarly, low availability nodes can have either low or high connectedness.

We investigated the effects of n_p and n_f by exploring the temporal properties of encounter: *link presence* (l_p) and *link frequency* (l_f). These are properties that allow us to focus our analysis on the encounters per se as opposed to the devices. Link presence is the sum of the durations of all encounters between two specific devices, while link frequency is the number of times two devices

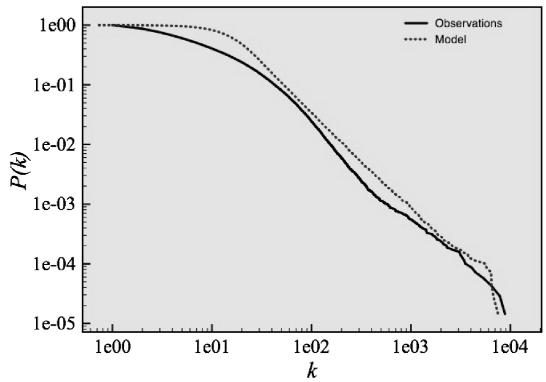


Fig. 15. Degree distribution. Our model (dashed line) results in a degree distribution that approximates our observations (black solid line).

have encountered each other. In Figure 14(c) (black solid line) we see that l_p follows a power law with $\alpha \approx 2.3$. Thus, while during our study overwhelming numbers of individuals had brief encounters, a few devices accumulated a total of up to a few days next to each other during their various encounters. Similarly, in Figure 14(d) (black solid line) we see that l_f follows a power law with $\alpha \approx 2.7$. Thus, whilst most individuals met only once or twice, a few met up to 300 times.

One of our main goals was to improve our modeling of city-scale phenomena, such as patterns of encounter, by informing our modeling and model-based analyses with real-world empirical data. Our initial approach was a fairly conventional one of developing a simulation model and comparing the results with our empirical dataset. This approach can facilitate cross-validation of the assumptions underlying the simulation with the results of observing real world phenomena in the field. Thus, to test our hypothesis that nodes' temporal availability, their presence and their frequency drive the temporal and structural properties of networks, we first developed a simulation model of encounter driven by n_p and n_f. At each step of our model, a node was activated with probability n_f, and if activated, stayed active for a duration of n_p multiplied by the number of steps it was previously inactive. In our model, attachment between nodes was nonpreferential: all simultaneously active nodes were linked to each other. We analyzed the behavior of our model with varying population sizes, each time resulting in a scale free distribution of k, l_p and l_f.

In Figure 14 (dashed lines) we see the resulting scale-free distributions of presence and frequency of a modeled population of 70,000 nodes using the n_p and n_f derived from our observations. These are clearly a very good fit to the corresponding observed distributions from the city (black solid lines in Figure 14). In Figure 15 we see the resulting scale-free distribution of connectivity k, again showing a very good fit to the observed distribution from the city.

Our results suggest that unlike networks where nodes are, in normal operation, constantly available (e.g., the Internet, the electricity grid, or the road network), when we consider encounters between mobile devices in the city, availability follows a scale-free distribution, which determines the distribution of encounters, or links, between individuals. Hence, whereas on a desktop computer with a fixed network connection we can, for example, connect to a given Web site on the Internet at any moment even if some intermediate routers and lines fail, when we consider device-to-device communication between mobile wireless devices in the city streets this availability is not constant. The availability of device-to-device communication in this environment does not depend on technological infrastructure such as routers and cables, but rather it depends on the patterns of movement and encounter between devices in urban space. Our findings describe and predict how these patterns affect device-to-device availability. Having robust models of this availability would bring pervasive systems developers closer to taking full advantage of the networking mesh that individuals' mobile devices can collectively create. This mesh can be used as an infrastructure for free communication between individuals in cities, or between devices [Chaintreau et al. 2006]. An aspect of our work that we are currently pursuing is to consider how our findings can help us build predictive models that can extrapolate a population's behavior and suggest the state of the social network at some point in the future.

5.3 Modeling Information Diffusion Through Encounters

An important effect of encounter networks where the timing and availability of encounters is critical is in the dissemination of information across networks. There are many examples of the importance of such information dissemination, including message passing in peer-to-peer human or computer networks [Chaintreau et al. 2006; Jones et al. 2007] and the spread of viruses, both digital and biological [Pastor et al. 2001; Vazaquez 2006; Newman et al. 2002].

In a further effort to improve modeling using empirical data, we next created an emulation environment in which we studied the diffusion patterns of information packets or digital viruses. We deliberately describe our system as an *emulation*, as opposed to a *simulation*, because the underlying mechanisms are not probabilistic as in a conventional simulation model but reflect real-world events as recorded by our Bluetooth scanners. In our system, each encounter provides an opportunity for two devices to exchange information or a virus. Our emulation environment is novel in that it effectively offers a dynamic bond percolation model [Newman et al. 2002] for understanding virus spread without relying on probabilistic operations since we know the precise time and duration of each encounter.

To initiate diffusion in our emulations, we used as seeds all possible devices in our dataset, at all possible points in time. At each iteration we injected a single device with information, and our emulation then worked through our records of encounters, recreating every encounter in the exact order our scanners recorded them in the real world. During an emulated encounter, if one of the two devices had the information, then it passed to the other device. Finally, we aggregated the results of all trials to derive our final results. This process is in contrast to what researchers have traditionally done, which involves generating encounter patterns in a probabilistic manner because they have no other means of generating valid and realistic temporal patterns of encounters. Our use of Bluetooth scanning to automate the real-time recording of patterns of encounter over extended periods allows us to record and retain the detailed temporal properties of the data.

To better understand how the temporal structure of our data affects information dissemination, we injected two types of information into our network: nonexpiring and expiring information. Nonexpiring information can be passed along indefinitely, while expiring information can be passed along only up to a limited number of days after it was received. Thus, in the world, expiring information is information that is invalid after a few days (for example, "knowledge that the circus is in town"). It resembles viruses that are active in their host only for a few days. Nonexpiring, or delay-tolerant, information on the other hand is exemplified by knowledge of facts such as the launch of a new shop in the city center or the opening of a new motorway between two local towns.

Initially, we ran emulations using our complete dataset. We then selectively removed from the dataset the links between our network nodes, and observed the cumulative effect of removing these links on the transmission of information through the network. In our first trial, we removed links as follows: given the complete set of links ordered by l_p, that is, by duration of the encounter represented by the link, we began removing links at the top of the order, that is, by removing the briefest encounters first. At each step, we removed 10% of the links in our total ordered set of links and observed the cumulative effect on information transmission across the network.

We next ran emulations in which we began removing links from the bottom of the list working our way to the top, thus removing the most persistent encounters first. In terms of city life, brief encounters predominantly represent encounters with strangers, such as passing in the street, while the most persistent encounters predominantly represent encounters with friends, family and coworkers, such as meetings, meals, and so on. Encounters with "familiar strangers" [Paulos and Goodman 2004; Milgram 1977], such as waiting at a bus stop or train platform at around the same time every weekday, are typically represented more towards the persistent end of our order.

As a measure of information dissemination across the network, in Figure 16 we show the number of devices that *nonexpiring* information can reach over time. We observe that the removal of brief encounters (e.g., passing encounters between strangers on the street) substantially diminishes the ability of nonexpiring information to spread through the network, while the removal of the same number of persistent encounters (e.g., longer meetings between friends) has a much smaller effect. For example, we see that up to 40% of persistent encounters can be removed and 98% of devices still receive the information. If, on the other hand, we remove 40% of brief encounters then only 65% of devices get the information.

Granovetter's account of the importance of weak social ties [Granovetter 1973] suggests that more new information flows to individuals through weak rather than strong ties. This is because our close friends tend to move in the same circles that we do, hence the information they receive overlaps considerably with what we already know. Weak ties such as acquaintances, by contrast, know people that we do not, and thus receive more novel information.

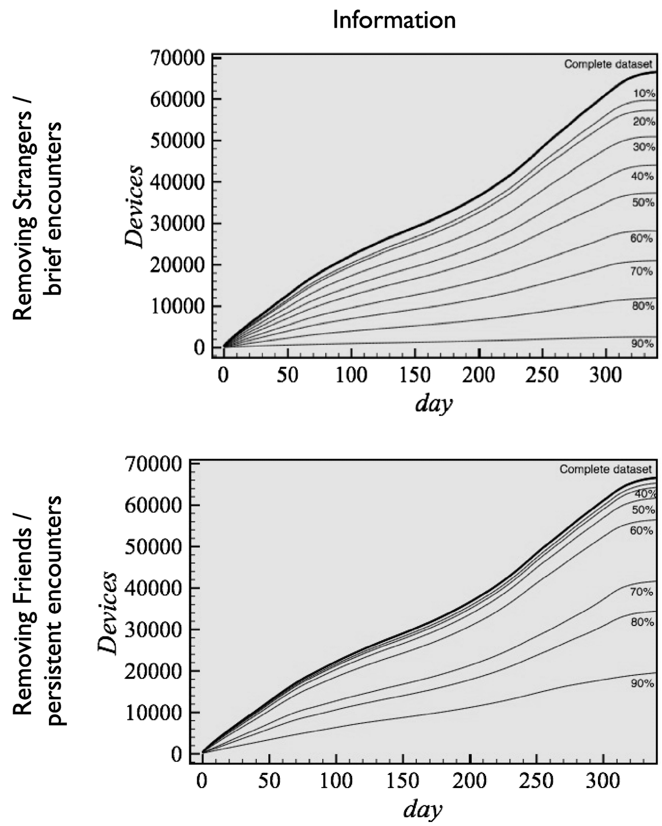


Fig. 16. The effect of selectively removing the briefest encounters (with strangers) or most persistent encounters (with friends) in the propagation of nonexpiring information.

Our emulations suggest that brief encounters may be viewed as the temporal equivalent of weak social ties in that they are crucially important to network cohesion [Bohannon 2006].

In our second set of trials we again ran exhaustive emulations, but this time we injected devices with short-expiry information packets, and observed changes in their diffusion by varying the amount of time they remained active in a device. Such packets represent time-sensitive information, messages, or viruses, as they remain active in their hosts for only a few days. Again we removed the most brief or most persistent encounters from our emulations in cumulative steps of 10% of the links, and observed the propagation of the packets. In Figure 17 we show the results of our emulations with information packets that remain active for 3 days.

We used the Susceptible-Infected-Susceptible model [Anderson and May 1992], which means that once a device has recovered from a virus, the device is susceptible to contracting the virus again. It is interesting to note that the small periodicities visible in these graphs are 7 days long, reflecting underlying weekly patterns in the encounters in the city. In marked contrast to nonexpiring

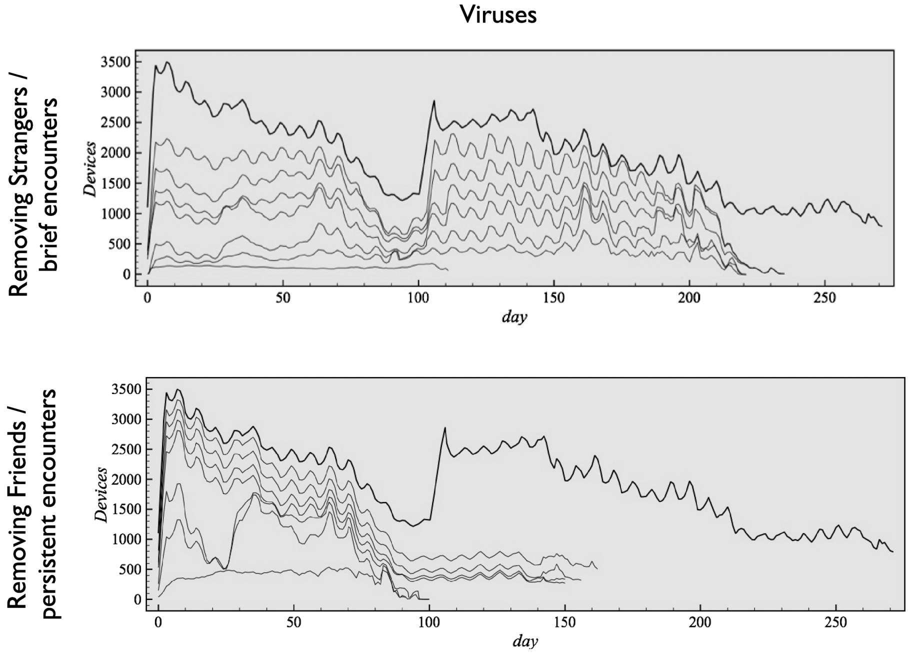


Fig. 17. A 3-day virus (transmission probability of 1) is injected in our dataset, and we selectively remove the most persistent encounters (friends) or most brief encounters (strangers).

information (Figure 16), we found that the removal of brief encounters has a small effect on diffusion across the network, while persistent encounters have a large effect, with the removal of as little as 10% of the most persistent encounters immediately diminishing the network's capacity to sustain the virus for long periods of time (Figure 17, bottom). Removing the 10% of most persistent encounters immediately reduces the virus' expectancy by 110 days, while in the top of Figure 17 removing the 10% of briefest encounters reduces the epidemic by only 30 days. Thus, we have made a second extension to Granovetter's [1973] theory of weak social ties. Our findings suggest that when considered in a temporal domain, nonweak ties (i.e., persistent ties) sustain the flow of time-expiring information.

An interesting reinterpretation of our results is that viruses last longer on networks with high proportions of persistent encounters. The removal of even a small proportion of persistent encounters quickly brings an epidemic to a stop. This is consonant with epidemiologists' suggestion that most infections take place between family members [Ferguson et al. 2006].

In this section we have examined both the structural and temporal properties of encounters between mobile devices (and by implication their users) and our fixed scanners (and by implication other fixed devices and digital services) across an urban environment. We found that the structural properties of these encounter networks are both scale-free and small world. We showed that their temporal properties are also scale-free, and we characterize such networks as Brief Encounter Networks due to the overwhelming proportion of short and infrequent encounters between individuals. We have shown that the temporal behavior of the nodes gives rise to the observed power-law connectivity. We have demonstrated that brief encounters are fundamental to the propagation of nonexpiring information while, conversely, persistent encounters are needed to sustain time-expiring information through a network of urban encounters. In terms of urban pervasive systems, people may rely on their persistent encounters with other people or location-based services to exchange time-sensitive information such as coordination feedback, traffic news, or limited-time marketing promotions, while brief encounters are likely to provide access to gossip, innovations, or news that a new restaurant has opened.

6. IMPLICATIONS FOR URBAN PERVASIVE SYSTEMS

In this article we have presented data collection methods and developed analysis and modeling techniques for making sense of the captured data on patterns of movement and encounter in the city. Our applications of these techniques have provided novel findings including the importance of temporal properties in the structure of encounter networks and their capacity to disseminate information. These methods and findings have implications for specific current challenges within pervasive computing such as the design of Delay and Disruption Tolerant Networks (DDTNs), more generally for our understanding of mobility and encounter of devices and their users within the urban environment, and more generally still for our fundamental understanding of network properties that are found throughout the natural and artificial worlds. While our methods and findings may have potential to help inform pervasive systems design, a key characteristic of our techniques is that both the data collection and the analyses of the collected data can be completely automated. This implies that our analyses may be used by the pervasive systems themselves at run time. This approach could help in enabling pervasive systems dynamically to adapt their run time operation to the changing urban environment with its flows of people and devices.

6.1 Potential to Inform Pervasive Systems Design

The tools we have presented in this article form the basis for deriving consistent and operationalized means of instrumenting and comparing different urban spaces. Given a dataset from each environment where our system is to be deployed, our analytical tools can highlight their differences and constraints, and represent them to designers for consideration early in the design process. For instance, considering the installation of a pervasive system such as a smart bus stop system, our tools can give designers an insight into the differences between, for example, Bath and Tokyo, or between 2 different places in Bath. In addition, our model of encounter can provide the means to stress-test a design by considering the expected number of visitors, the frequency of their visits, and the duration of their visits. To study such differences, ideally one would have to take samples from “equivalent” locations in the city. There are at least two approaches to measuring the equivalency of two locations in different cities: use metrics derived from architectural theory such as Space Syntax [Hillier 1996], or choose functionally equivalent locations such as the train station exit, a central parking lot, or a shopping mall. While the first approach is more consistent and theoretically founded, the second can be more relevant and can leverage knowledge of a city's context.

There are further advantages in an automated approach to data collection exemplified by our Bluetooth scanning, in that simply installing more hardware can instrument large geographic areas. This means that, for example, we can begin to automate tasks such as analyzing the requirements of 150 different bus stops in a city center by installing a scanner at each bus stop. Whereas many approaches require human observers to monitor sites and record data, as in conventional space syntax methods such as gatecounts and static snapshots [Hillier 1996], using our techniques we can automate the collection of some forms of data. Automated data collection cannot replace all forms of human data collection but, as we noted in O'Neill et al. [2006], the combination of human observer and automated data collection methods can capture a wide range of data that neither approach alone can encompass. And of course, automated data collection can be active 24 hours a day over very long periods and can thus begin to capture a consistent longitudinal picture.

To build up this picture, designers can use our analysis techniques to look for patterns of presence, encounter and trails across a city, or across sites of interest (i.e., Figures 7, 8, 9, and 13). These patterns offer concrete metrics for comparing and differentiating between locations, and identifying appropriate design approaches. To continue our bus stop example, our analysis can show differences in visiting patterns at each bus stop. Thus, while some bus stops may be continuously busy 9 a.m. to 5 p.m., other bus stops may experience bursts of groups of people. Similarly, some bus stops may host passengers for an average of 2 minutes, while other bus stops may host passengers for more than 30 minutes. Additionally, designers can identify those bus stops that have mostly regular passengers, likely commuters, or those bus stops that have a small proportion of regular passengers, and design each bus stop appropriately.

In terms of encounter, our analysis can help designers identify the structure of the community observed at a specific location. Considering systems beyond our bus-stop examples, designers can be made aware of the presence of numerous tight-knit groups appearing at a specific location, how often they come in contact with each other, and even begin to anticipate the chance that these groups interact directly. All these are means of understanding the differences between locations, which subsequently may be interpreted and instantiated as design decisions.

Finally, designers can look for trail patterns among locations of interest, such as a set of bus stops under consideration. This can offer valuable insight towards creating a consistent and overarching experience in terms of, for example, an advertising campaign. For example, our analyses can aid designers in identifying the origins to, and destinations from a specific bus stop. Additionally, designers can begin to anticipate the context in which a specific bus stop will be used. For instance, a bus stop may be used regularly on a Friday night, acting as a final social hub before everyone goes home.

While our techniques and models can be used to provide insights about some of the relevant features of the urban environment, our techniques do not provide a panacea for pervasive systems designers. Our techniques might be used to explore whether the majority of users at a bus stop are daily users, but they cannot be used alone to determine the specific design consequences of such an insight. Such design decisions depend on a much broader set of requirements and contextual constraints. Design remains in the hands of the designers and their understanding of the city context, users' cultural values and even basic usability principles. Our toolkit does not seek to replace but rather to complement designers' existing tools.

6.2 Dynamic Runtime System Adaptation

Since an urban pervasive system could itself collect data in a similar way to our approach and could then analyze the data, the results of the analyses could be used by the pervasive system dynamically to adapt its runtime operation. This could be reflected at many levels from choosing routing algorithms for messages to dynamically adapting user interfaces.

For example, pervasive systems can be made more usable by deploying what we term attention span interfaces. By interface we refer to any delivery medium, such as smart posters, public displays, speakers, and mobile devices. All such interfaces can adapt their presentation of information based on knowledge of how much time they expect users will have to receive the message. For example, a display at a train station can detect when people are going past it very quickly, or very slowly, and present the information in an appropriate way. Note that Figures 7 and 8 respectively represent a sample distribution of how much time people are expected to spend at a location, and how many transient visitors a location can expect over the course of a day. In other situations, a system can adapt its behavior based on how long a person has remained in a specific location. Thus, users may receive particular information on their device only if they have spent more than 20 minutes at a specific location. Similarly, appropriate action may be taken if a display detects users who have not been seen before; in this case the display may present, for example, extended instructions.

An analysis of encounters can also be utilized at runtime by pervasive systems, enabling information and operation to be adapted based on inferred relationships and habits of users. Thus, a system can react differently to a couple seen repeatedly as opposed to a large group of people never seen before. In this fashion, pervasive systems can optimize the delivery of information to suit individuals, small groups, or large groups. For instance, a smart poster may utilize Near Field Communication (NFC) for delivering information to small groups of people thereby preserving power, but switch its operation to Bluetooth when large groups of people attempt to use it.

The analysis of trails has the potential to have a big effect on developing city-wide usable pervasive systems. Our approach, which we call trail-mnemonic interfaces, relies on adapting interfaces based on an understanding of the trail-based context in which an interface is used. This involves an understanding of where users have been before coming to a location, and where they are expected to go after visiting a location. Analyzing the transition tree between city locations using our trails engine can produce such information. Each node in our transition tree has associated pre- and post-visit nodes, signifying locations that users visit before and after a given location. At the aggregate level, such an analysis yields a model of movement in the city that could be used to do predict where users may go to next, or from where users may be coming. Such models may be used to adapt applications to specific cities. At the individual level, personalized transition trees can be used to adapt our engine to each user's behavior patterns, by independently considering each user's movement in the city. This can allow applications to adapt to their own users' mobility patterns.

Furthermore, an important concern in city-scale systems is the efficient transmission and dissemination of information. While many devices can be directly connected to a global network such as the Internet, there are compelling reasons for having a complementary device-to-device network throughout a city. First, such networks can potentially be free since they involve no central infrastructure and no network operator costs. In addition such a decentralized structure has potentially more network capacity than a centralized network; hence, it can sustain much higher bandwidth between its users. A key contribution of our work is in providing a methodology for studying communities and understanding the networking opportunities that arise between individuals and how these could contribute to the community's ability to maintain and diffuse information. Our theoretical model of encounter can be used to study expected patterns of dissemination, and to test possible strategies for accelerating or stopping the dissemination of information through the population.

Finally, a crucial issue in developing pervasive systems is privacy. An interesting paradox we must deal with is that the smarter our systems get, the more they can infer about their users, raising further privacy concerns. The techniques we have developed in this article offer a three-stage approach to dealing with privacy. Users can completely opt out of our system by setting their Bluetooth to "invisible" either temporarily or permanently. This will make our system completely oblivious to them. Alternatively, users can set their Bluetooth to "visible" or "discoverable," potentially contributing to the analyses we have discussed in this article. These analyses operate at an aggregate level and derive models of aggregate behavior. In turn, the benefits in terms of novel services offered can only be at the aggregate or anonymous level, since little or no information about the individual user is available to the system. Finally, users may wish to provide the system with personal information, typically through an online registration process, in exchange for a more personalized experience and services. We are beginning to experiment with the latter approach, where users who register with our system can use personalized services based on our Bluetooth scanning [Kostakos et al. 2008].

7. CONCLUSION

In this article we describe a unique approach to recording, analyzing, and understanding mobility and encounter in the urban environment. We describe a

Bluetooth-based method for capturing data on urban pedestrian behavior, and discuss a number of analytical tools we have developed for making sense of these data. Specifically, we discuss the concepts of sessions, encounters, and trails, in the context of deriving them from raw data, as well as utilizing them as part of the design process and run-time implementation of a pervasive system.

Our focus throughout the article has been to inform the understanding, design, and use of pervasive systems. Usability for pervasive systems entails much more than using the correct icons and appropriate menu structures. While designers of desktop systems can draw on an array of tools and theories, pervasive systems designers have no such luxury. The complexity of pervasive systems, which include people, spaces, and technologies, makes it extremely difficult to adopt a top-down theoretical approach. Here we describe our development of a bottom-up, observation-driven data mining approach.

Data mining typically requires large amounts of data, which is something that human observers cannot adequately produce. Thus, our first step has been the automation of observation, such that large amounts of data can be collected to begin with. As we have shown, however, the raw data requires further refinement and analysis (e.g., Figure 8). In fact, consecutive refinement of our data has resulted in increasingly useful and generalizable insights into patterns of mobility and encounter in the city.

Our passive observation methods coupled with our analysis techniques offer suggestions for how location-based services could automatically capture, analyze, and exploit data on flows of users, visiting patterns for specific locations, and trail patterns across locations. Further analysis of the complex networks of encounters leads us to identify power-law distributions commonly found in nature, which suggest that natural laws underlie the apparent complexity and "randomness" of movement and encounter in cities. Mining our observation data even further, taking into account temporal properties, we identified the importance of Brief Encounter Networks, complex networks whose structure is heavily influenced by well-defined temporal patterns of encounter. Our large-scale longitudinal data collection and analyzes have revealed emergent properties of the city as a system that cannot be captured through interviews, interventions, or controlled studies.

ACKNOWLEDGMENTS

The authors thank Tim Kindberg, Ava Fatah gen. Schiek, Danae Stanton Fraser, and Tim Jones.

REFERENCES

ANDERSON, R. M. AND MAY, R. M. 1992. *Infectious Diseases in Humans*. Oxford, UK. University Press, Oxford, UK.

BALAZINSKA, M. AND CASTRO, P. 2003. Characterizing mobility and network usage in a corporate wireless local-area network. In *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services (MobiSys '03)*, ACM Press, New York, 303–316.

BARABÁSI, A.-L. AND ALBERT, R. 1999. Emergence of scaling in random networks. *Science 286*, 509–512.

BOHANNON, J. 2006. Tracking people's electronic footprints. *Science 314*, 914–916.

BREWSTER, S. A., LUMSDEN, J., BELL, M., HALL, M., AND TASKER, S. 2003. Multimodal 'Eyes-Free' interaction techniques for wearable devices. In *Proceedings of ACM CHI*. 463–480.

CHAINTREAU, A., HUI, P., CROWCROFT, J., DIOT, C., GASS, R., AND SCOTT, J. 2006. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proceedings of the 25th IEEE Conference on Computer Communications (INFOCOM)*. IEEE CS Press.

DEPARTMENT FOR TRANSPORT 2005. Focus on personal transport. TSO, London, UK.

DOROGOVTSEV, S. N., GOLTSEV, A. V., AND MENDES, J. F. F. 2002. Pseudofractal scale-free web. *Physical Review E*, 65, 066122.

EAGLE, N. AND PENTLAND, A. 2006. Reality mining: sensing complex social systems. *Pers. Ubiq. Comput. 10*, 4, 255–268.

FERGUSON, N. M., CUMMINGS, D. A., FRASER, C., CAJKA, J. C., COOLEY, P. C., AND BURKE, D. S. 2006. Strategies for mitigating an influenza pandemic. *Nature 442*, 448–452.

FREEMAN, L. 2004. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, Vancouver, BC, Canada.

GRANOVETTER, M. 1973. The strength of weak ties. *Am. J. Sociol. 78*, 1360–1380.

HARRISON, S. AND DOURISH, P. 1996. Re-place-ing space: the roles of place and space in collaborative systems. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. ACM Press, 67–76.

HILLIER, B. 1996. *Space Is the Machine*. Cambridge University Press.

HOLME, P. 2003. Network dynamics of ongoing social relationships. *Europhy. Lett. 64*, 3, 427–433.

HOLME, P., PARK, S. M., KIM, B. J., AND EDLING, C. R. 2007. Korean university life in a network perspective: Dynamics of a large affiliation network. *Physica A 307*, 821–830.

JONES, E. P. C., LI, L., SCHMIDTKE, J. K., AND WARD, P. A. S. 2007. Practical routing in delay-tolerant networks. *IEEE Trans. Mobile Comput. 6*, 8, 943–959.

KERN, N. AND SCHIELE, B. 2006. Towards personalized mobile interruptibility estimation. In *Proceedings of the 2nd International Workshop on Location and Context Awareness*. Lecture Notes in Computer Science, vol. 3987, 134–150.

KINDBERG, T. AND ZHANG, K. 2002. Validating and securing spontaneous associations between wireless devices. In *Proceedings of the 6th Information Security Conference (ISC'03)*. Lecture Notes in Computer Science, vol. 2581, Springer, 44–53.

KJÆRGAARD, M. B. 2006. Automatic mitigation of sensor variations for signal strength based location systems. In *Proceedings of Location and Context Awareness*, Dublin, Ireland. Lecture Notes in Computer Science, vol. 3987, 30–47.

KOSTAKOS, V. AND O'NEILL, E. 2007. Quantifying the effects of space on encounter. In *Proceedings of Space Syntax Symposium*. 9701–9709.

KOSTAKOS, V., O'NEILL, E., AND JONES, S. 2008. Social Networking 2.0. In *Extended Abstracts of the Conference on Human Factors in Computing Systems*, 3555–3560.

KUMAR, R., NOVAK, J., AND TOMKINS, A. 2006. Structure and evolution of online social networks. In *Proceedings of the 12th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)* (poster track).

LESKOVEC, J., KLEINBERG, J., AND FALOUTSOS, C. 2005. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*. 177–187.

LESKOVEC, J. AND HORVITZ, E. 2008. Planetary-scale views on a large instant-messaging network. In *Proceedigs of the International World Wide Web Conference*.

MCNETT, M. AND VOELKER, G. M. 2005. Access and Mobility of Wireless PDA Users. *SIGMOBILE Mob. Comput. Commun. Rev*. *9*, 2, 40–55.

MILGRAM, S. 1967. The small world problem. *Psycho. Today*. 60–67.

MILGRAM, S. 1977. The familiar stranger: An aspect of urban anonymity. In *The Individual in a Social World*, S. Milgram, 51–53. Addison-Wesley, Reading, MA.

NEWMAN, M. E. J. 2005. Power laws, Pareto distributions, and Zipf's law. *Contemp. Phy. 46*, 323–351.

NEWMAN, M. E. J., JENSEN, I., AND ZIFF., R. M. 2002. Percolation and epidemics in a two-dimensional small world. *Phy. Rev. E*, *65*, 021904.

NICOL, C., BREWSTER, S. A., AND GRAY, P. D. 2001. A system for manipulating auditory interfaces using timbre spaces. In *Proceedings of CADUI 2004*. 366–379

NICOLAI, T., YONEKI, E., BEHRENS, N., AND KENN, H. 2005. Exploring social context with the wireless rope. *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*. Part I, Lecture Notes in Computer Science, vol. 4277, 874–883, Springer.

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 the Conference on Ubiquitous Computing*. 315–332.

ONODY, R. N. AND DE CASTRO, P. A. 2004. Complex network study of Brazilian soccer players. *Phy. Rev. E*, 037103.

PASTOR-SATORRA, R. AND VESPIGNANI, A. 2001. Epidemic spreading in scale-free networks. *Phy. Rev. Lett. 86*, 3200–3203.

PAULOS, E. AND GOODMAN, E. 2004. The familiar stranger: anxiety, comfort, and play in public places. In *Proceedings of the Conference on Computer-Human Interactioin.* ACM, 24–29.

RAVASZ, E., SOMERA, A. L., MONGRU, D. A., OLTVAI, Z. N., AND BARABÁSI, A.-L. 2002. Hierarchical organization of modularity in metabolic networks. *Science 297*, 1551–1555.

SNIJDERS, T. A. B. 2001. The statistical evaluation of social network dynamics. In *Sociological Methodology*, M. E. Sobel and M. P. Becker Eds, Basil Blackwell, 361–395.

STROGATZ, S. H. 2001. Exploring complex networks. *Nature*, *410*, 268–276.

TURNER, A. AND PENN, A. 2002. Encoding natural movement as an agent-based system. *Envir. Plann. B*, *29*, 4, 473–490.

VAZAQUEZ, A. 2006. Epidemic outbreaks on structured populations. *J. Theor. Biol. 245*, 1, 125–129.

WATTS, D. J. AND STROGATZ, S. H. 1998. Collective dynamics of small-world networks. *Nature 393*, 440.

Received November 2007; revised September 2008, March 2009; accepted November 2009