Environment and Planning B: Planning and Design 2009, volume 36, pages 824 – 836

doi:10.1068/b34133t

Eigenplaces: analysing cities using the space – time structure of the mobile phone network

Jonathan Reades
Centre for Advanced Spatial Analysis, University College London, 1 – 19 Torrington Place, London WC1E 6BT, England; e-mail: j.reades@ucl.ac.uk

Francesco Calabrese, Carlo Ratti
SENSEable City Laboratory, Massachusetts Institute of Technology, room 10-485, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; e-mail: fcalabre@mit.edu, ratti@media.mit.edu,
Received 23 November 2007; in revised form 2 June 2008

Abstract. Several attempts have already been made to use telecommunications networks for urban research, but the datasets employed have typically been neither dynamic nor fine grained. Against this research backdrop the mobile phone network offers a compelling compromise between these extremes: it is both highly mobile and yet still localisable in space. Moreover, the mobile phone's enormous and enthusiastic adoption across most socioeconomic strata makes it a uniquely useful tool for conducting large-scale, representative behavioural research. In this paper we attempt to connect telecoms usage data from Telecom Italia Mobile (TIM) to a geography of human activity derived from data on commercial premises advertised through Pagine Gialle, the Italian 'Yellow Pages'. We then employ eigendecomposition—a process similar to factoring but suitable for this complex dataset—to identify and extract recurring patterns of mobile phone usage. The resulting eigenplaces support the computational and comparative analysis of space through the lens of telecommuniations usage and enhance our understanding of the city as a 'space of flows'.

Introduction

Sheller (2004, page 39) argues that contemporary networked cities are "constituted by flows of people, vehicles, and information", and yet data about these flows are increasingly difficult to collect and analyse using traditional social science research methods (Shoval, 2007, page 194). The underlying issue, identified by Batty (1990), is that the corporate infrastructures underpinning these flows are becoming both physically and legally invisible, forcing researchers to rely on surrogate data instead (Graham, 1997, page 107). This turn of events is unfortunate, since such networks are clearly having an enormous impact on the way that we inhabit spaces and places, not least in their ability both to mediate and to support increasingly complex patterns of physical copresence (Graham, 1998, page 182).

However, these very same networks are also ideal tools for tracking citizens and visitors across time and space, and thereby contributing enormously to our understanding of the city. In this paper we explore spatiotemporal activity patterns in the city of Rome, Italy, with data collected from the Telecom Italia Mobile (TIM) phone network. We attempt to connect telecoms usage to a geography of human activity using data on commercial premises advertised through Pagine Gialle, the Italian 'Yellow Pages'. We then employ eigendecomposition—a process derived from signal analysis, which is similar to factoring but more suitable for this complex system—to identify and extract recurring patterns of mobile phone usage in order to see if we can obtain a more detailed understanding of these spaces. The resulting eigenplaces support the computational and comparative analysis of space through the lens of telecommunications usage, and enhance our understanding of the city as a 'space of flows'.

The mobile network as a research tool

Several attempts have already been made to use advanced telecommunications networks for urban research, but the datasets collected have typically been neither dynamic nor fine grained. Because of proxy servers and the centralised setup of corporate infrastructure, the use of Internet Protocol (IP) address data for geographical analysis is extremely difficult, and perhaps even impossible (Hall et al, 2007, page 71). Townsend (2001) examined the geographical distribution of WHOIS [(1)] records to identify 'network cities', but companies may register dozens of domains to a back office, technical support site, or marketing department, which suggests that these results need careful handling. In contrast, although landlines are highly localised, they cannot be used to investigate activity in a public and mobile context.

Against this research backdrop the mobile phone network offers a compelling compromise between these extremes: it is both highly mobile and yet still localisable in space. Moreover, the mobile phone's enormous and enthusiastic adoption across most socioeconomic strata makes it a uniquely useful tool for conducting large-scale, representative behavioural research. By the end of 2006 many European countries had more subscribers than inhabitants: the United Kingdom had the highest penetration amongst large countries at 115 lines per 100 inhabitants, but this pales next to Luxembourg's remarkable 155 lines and Italy's 134 lines (Eurostat, 2007). Thus, the network's enormous spatial extent and continuous temporal coverage make it ideal for studying the functioning of cities and regions holistically, something that is functionally impossible, as well as unfeasibly expensive, using surveys or gate counts.

Naturally, a variety of innovative approaches have already sought to harness mobile phones for urban research, but they have all relied on sample sizes of, at most, several hundred simultaneous users, and have typically required end-user consent. Ahas and Mark (2005) performed a 'social positioning method' analysis, combining detailed demographic data with individual user 'trails' across Estonia. And the 'reality mining' project at MIT had success in abstracting individual behavioural patterns, termed eigenbehaviours, (Eagle and Pentland, 2006a) from handsets running a specially designed application. Using aggregate data from the entire network, however, we can hope to build maps based on the behaviours of *millions* of users spanning multiple spatial and temporal scales.

Rome in real time

Our analysis builds on aggregate mobile phone data supplied by TIM between July and September of 2006. Data were collected at fifteen-minute intervals at the transceiver level across a region of 47 km^2, and were transferred via secure-FTP (file transfer protocol) to servers at MIT. The three months' worth of data represents the cumulative daily phone usage—voice, text, and data—of more than one million TIM subscribers and visitors to Rome. In all, the dataset contains more than 3.5 million observations. More detail on the system architecture, including data collection, transfer, and processing, can be found in Calabrese and Ratti (2006).

Network data were supplied in Erlang, an aggregate measure of bandwidth usage that is often used in network capacity planning on GSM [(2)] networks and which can be easily collected by the operator. The Erlang is also completely anonymous, since it is defined as one person-hour of phone usage: one Erlang could be the result of a single

[(1)] WHOIS is a method used to find the owner of a domain name or IP address from an official database.

[(2)] Global system for mobile communications—the most popular global standard for mobile phone communications.

person talking for an hour, two people talking for a half-hour each, thirty people speaking for two minutes each, and so on. In effect, Erlang data provide a view of urban space as seen through network bandwidth consumption and, consequently, indirect insight into the spatial and temporal dynamics of urban life. More information on the tradeoffs involved in the use of Erlang for urban research can be found in Reades et al (2007).

The coverage areas of cellular transceivers overlap to ensure quality of service, and there are sometimes many transceivers on a single mast. Since we lacked the orientation data to untangle their relationships to one another, we used an exponential decay algorithm to divide Rome into 2115 'pixels', measuring 500 m $\times$ 500 m, rather than the more usual Voronoi polygon plot. The rasterising function, detailed in Calabrese et al (2007), enables us to derive an Erlang value for each pixel, which is a composite calculated from the surrounding antennas. Actual Erlang values were also modified by a secret scaling factor applied by TIM to preserve a modicum of commercial confidentiality.

From Erlang data to 'signatures'

Figure 1 shows a map of Rome to help orient readers for the subsequent analysis. We selected seven locations that we expected to have distinct patterns of behaviour and, consequently, mobile phone usage for in-depth study: (1) the Pantheon is a key tourist attraction and popular meeting point; (2) Tiburtina is a secondary interchange used by regional trains and buses as well as late-night intercity trains; (3) Olimpico is Rome's main concert and football venue; (4) Trastevere is known for its nightlife; (5) Piazza Bologna is predominantly residential; (6) Termini is Rome's main rail station; and (7) EUR is a purpose-built office area southwest of the downtown area.

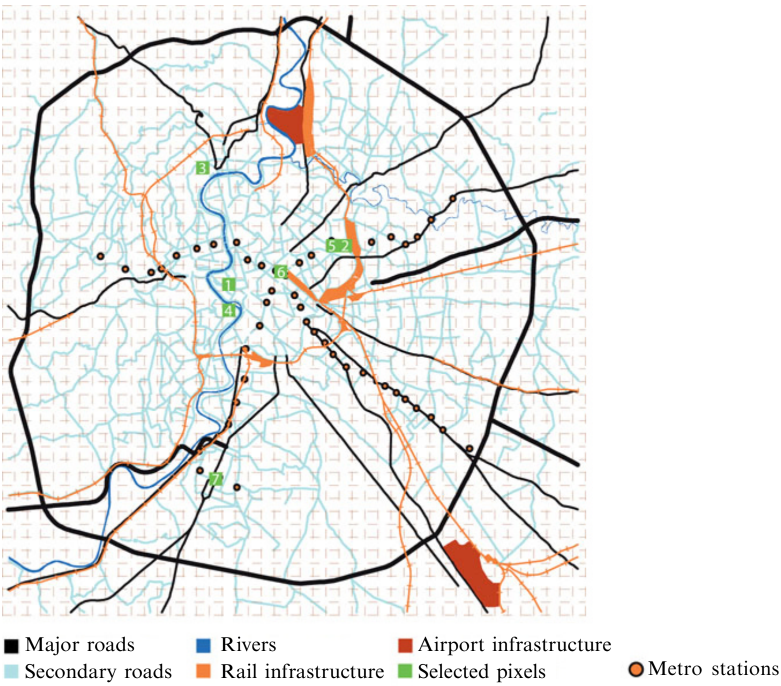


Figure 1. [In colour online, see http://dx.doi.org/10.1068/b34133t] Map of Rome and selected areas of special interest. See text for a description of the selected locations.

Projecting Erlang changes over time for some of the sites mentioned above yields the unique 'signature' of phone activity seen in figure 2. The area around the Olimpico stadium, a venue for sporting events and concerts, clearly shows a pattern of afternoon and evening activity not seen in other parts of the city. Comparing the tourism-driven area near the Pantheon with a more locally oriented site at Tiburtina suggests variable and distinct relationships between space, time, and aggregate phone usage, and we note in particular the variability during weekdays, and the contrasting drops in usage over weekends.

While figure 2 is helpful in grasping the coarser differences between areas of Rome, it is time consuming to perform this type of intense analysis operation for each cell. To explore the system as a whole we need to understand first the overall distribution of Erlang within the city. Figure 3 shows the distribution of peak Erlang, irrespective of time or day, for each of the pixels in Rome organised from greatest to least. The 'raw' values shown on the light green line in figure 3 follows an exponential distribution with its characteristic 'long tail'. Plotting the peaks on a logarithmic scale—the darker red line in figure 3—emphasises this distribution. We settled on $\log_{10}$ to generate the logarithmic distribution shown below, which we believe constitutes an

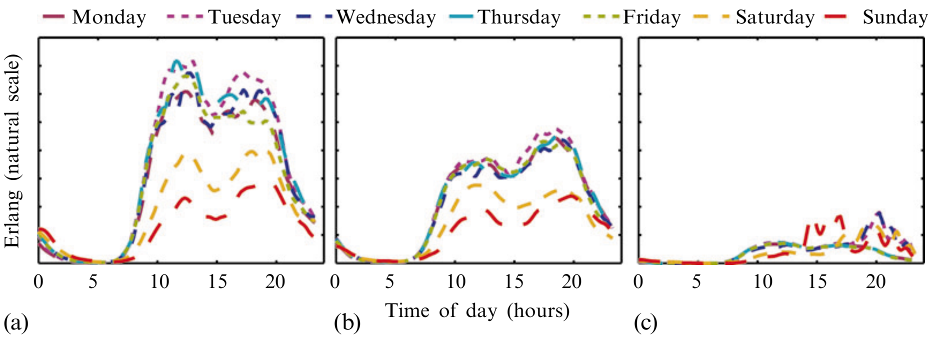


Figure 2. [In colour online.] Sample Erlang signatures from Rome: (a) Pantheon (tourist monument); (b) Tiburtina (rail/metro station); (c) Olimpico (sports and concerts). These three locations are identified by numbers 1–3, respectively, in figure 1.

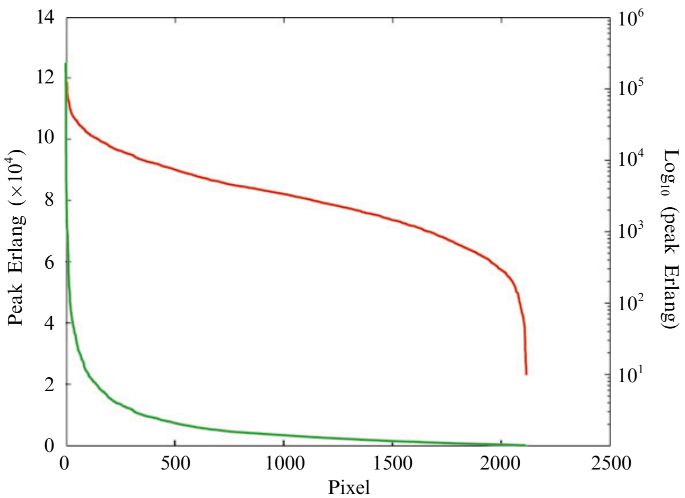
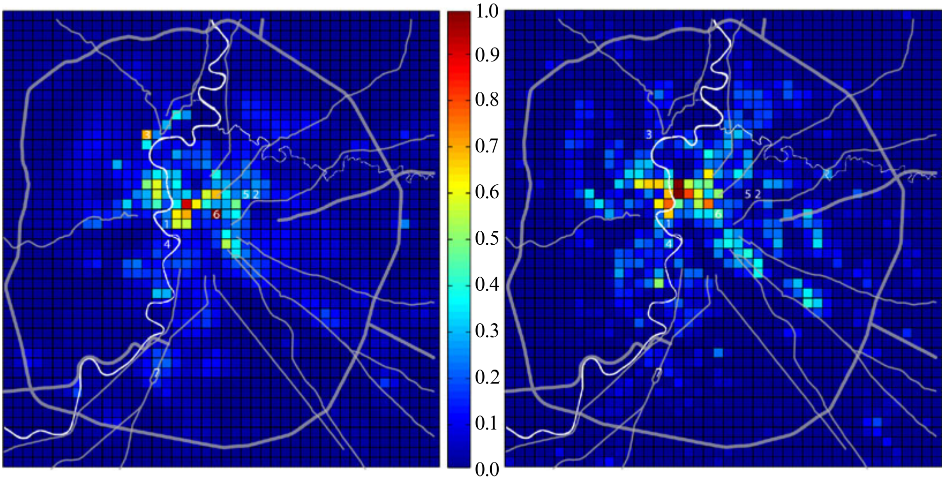


Figure 3. [In colour online.] Peak Erlang distribution. The green line (light grey in greyscale) denotes raw values; the red line (dark grey in greyscale) denotes $\log_{10}$ values.

appropriate transformation under the circumstances, and which also resolves the scaling issue identified above.

The use of a logarithmic scale also has important analytical benefits: the magnitude of the difference between the greatest and least peak Erlang values is so great that it would easily mask less immediately obvious features shared between cells with widely differing overall levels of telecoms usage. By taking the log of the dataset we are able to retain lesser features of the signatures. The normalised maximum Erlang distribution can be compared to a normalised distribution of businesses garnered from Pagine Gialle, which will be discussed in more detail below.

The two projections in figure 4 suggest a strong connection between these datasets. The highest levels of Erlang on the map are linked to people flows at Termini Station and a location near the Spanish Steps, though this second site also points to one of the challenges in using mobile data for urban analysis: it is a geographical high point and hosts a major telecommunications array that is readily 'visible' to phones at some distance from the area. Other points of interest include Piazza Navona, the Vatican, and Olimpico Stadium—the same site shown in figure 2. There is also the intriguing suggestion of a link to higher phone usage along Rome's major radial access roads and near their intersection with the circular. While this simple projection indicates a relationship between phone usage and socioeconomic geography, it is nonetheless important to quantify this connection.



(a) (b)

Figure 4. [In colour online.] (a) Peak normalised Erlang density map; (b) normalised business density maps. See text for a description of the numbered pixels.

Business location data

Fortunately, the support of Seat Pagine Gialle—who generously made available to us their directory of 50 000 businesses in Rome—enables us to explore the map shown in figure 4 in more depth. The database contains name and address information, latitudinal and longitudinal data, and an advertising category for each listed business. These categories ranged from the obvious ('articles for smokers'), to the amusing ('ornamental aquariums and accessories'), and the rather specialised ('eggs'). The data can be accessed through a web interface www.paginegialle.it. Coordinates were specified to four decimal places, which represents approximately 10 m accuracy. Using this positioning data we assigned each business to a pixel and organised the raw categories thematically into use-oriented groups.

We hoped to demonstrate a measurable link between mobile phone usage and business activity, so as to provide a more solid empirical footing for the use of aggregate mobile data as a research tool. As a first step, we looked at the total number of businesses in each cell of the grid and attempted to correlate it to peak Erlang. The correlation yielded a standard deviation of 0.18001, and a fairly high incidence of pixels that departed radically from the mean. Similarly, plotting peak Erlang against the number of businesses indicated that there was no single regression that could account for the entirety of telecommunications activity in Rome. The lack of a simple correlation was to be expected, since mobile phone use is not structured solely around the location of businesses advertising in the Yellow Pages.

The enormous number of categories in the raw Pagine Gialle dataset, and the widely variable number of businesses in each, led us to select a smaller number of 'themes' from within the dataset for further study: (1) accommodation from *alberghi*, *motels*, and *pensioni*; (2) daily retail from *giornalai* and *tabaccherie*; (3) dining out from *bar e caffè*, *ristoranti*, *traittorie*, *osterie*, and *pizzerie*; and (4) food retail from *alimentari*, *panetterie*, and *frutta e verdura*.[(1)] These groups were selected so that there would be enough businesses across the city as a whole for a geographical projection to be meaningful, and so that they would be distributed evenly enough that their relative density provided insight into the city's overall makeup.

An appropriate way to measure the relative concentration of businesses in a region or city is the location quotient Q: the ratio of the local density of industry i in an area p to the overall density of that industry in the region. The importance of an industry is normally measured through either the number of employees or turnover. By dividing Rome into 500 m $\times$ 500 m pixels, we were adopting an approach similar to that taken by Goddard (1973); however, our dataset lacked any details on the scale of business operations for each advertiser because this information is not available free of charge to researchers in Italy. So Q was calculated in a nonstandard manner by taking a count of businesses from pixel p in category i (x_{ip}), and dividing by the total number of businesses in that pixel (x_{tp}). The result was, in turn, divided by the total number of businesses in all Rome from that category (x_{ir}) over the total number of businesses in all Rome (x_{tr}). This can be rewritten as:

$$Q_{ip} = \frac{x_{ip}x_{tr}}{x_{tp}x_{ir}} .$$

Clearly, this Q value is skewed towards pixels with a high density of small businesses, rather than larger, single-company facilities. However, since small-firm activity is known to be particularly important in Italy, we feel that, in the absence of employment data, this represents a suitable, though clearly not ideal, outcome.

The resulting spatial distributions are shown in figure 5, though attention should be given to the varying scales for each map. As might be expected, 'accommodation' is centrally located, near the areas frequented by tourists, but it also has significant peripheral clusters that might be connected to transport infrastructure (they are near to both of the airports and the orbital). The 'dining out' category suggests to some extent the well-known Italian passion for food—such establishments are quite literally everywhere, up to the edges of Rome's developed areas, and even beyond in places

[(1)] Translation: *alberghi* = hotels; *motels* = motels; *pensioni* = boarding houses; *giornalai* = newsagents; *tabaccherie* = tobacconists; *bar e caffè* = bars and coffee shops; ristoranti = restaurants; *trattorie* = restaurants (generally less formal than *ristoranti*); *osterie* = taverns; *pizzerie* = pizzerias; *alimentari* = grocery shops; *panetterie* = bakeries; *frutta e verdura* = fruits and vegetables.

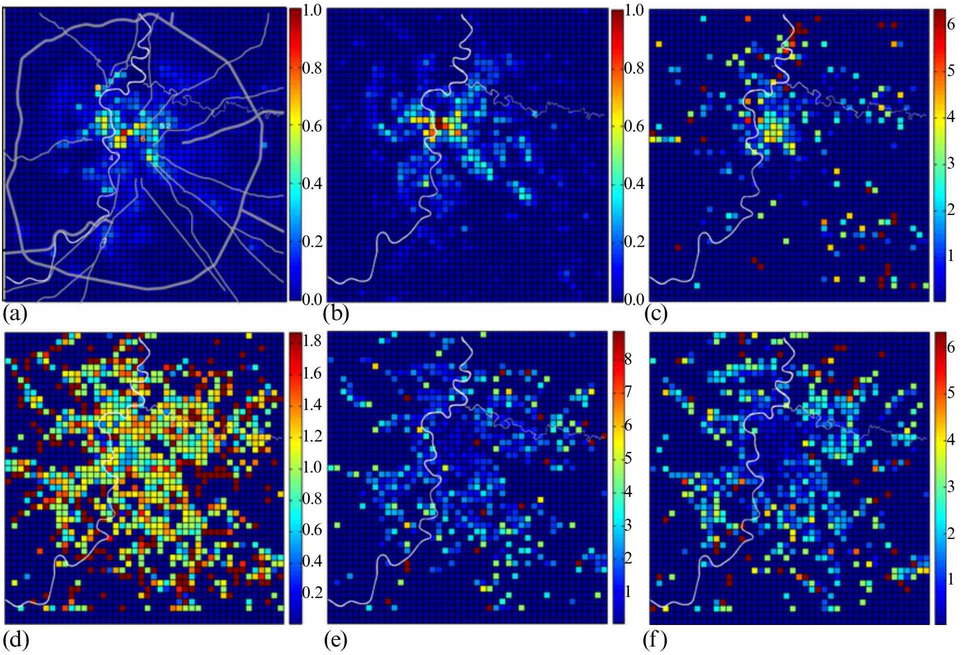


Figure 5. [In colour online.] Spatial distribution of businesses: (a) normalised peak Erlang, (b) normalised business density; (c) accommodation; (d) dining out; (e) daily retail; (f) food retailers.

overlooking the city. Finally, 'daily retail' and 'food retailing' are relatively scarce in the business core, but are more common towards the edges, which suggests that these are more residential areas.

From eigenvectors to eigenplaces

With the aim of understanding better the relationship between business and residential areas, and their cumulative impact on mobile phone infrastructure, we then turned to a technique that has been employed by radio astronomers as well as by researchers using an individual handset context for social group analysis (Eagle and Pentland, 2006b). The process of eigendecomposition extracts patterns from large datasets, and its outputs are, in a sense, the roots of equations and their associated coefficients. Thus the electromagnetic signature of a pixel is decomposed into an arbitrary number of components, enabling us to sketch out the basic dimensions of the relationship between mobile bandwidth usage and urban activity.

The number of eigenvectors needed to reconstitute the source with an acceptable margin of error depends on the complexity of the raw input data. Quite simply, the more random the original signal, the more information we need to rebuild it. Multiplying each component vector by its coefficient and summing the result enables us to reconstruct the source dataset. Thus, if S is the observed signature for a randomly selected pixel, C is a coefficient, and V an eigenvector, then:

$$S = C_1 V_1 + C_2 V_2 + \ldots + C_n V_n \ .$$

In our case, data for one day from one pixel can be expressed as a vector with one row and ninety-six columns—one column for each of the Erlang values recorded at fifteen-minute intervals. Lining up the ninety days' worth of data that we collected between 1 September and 30 November yields a matrix of ninety rows and ninety-six

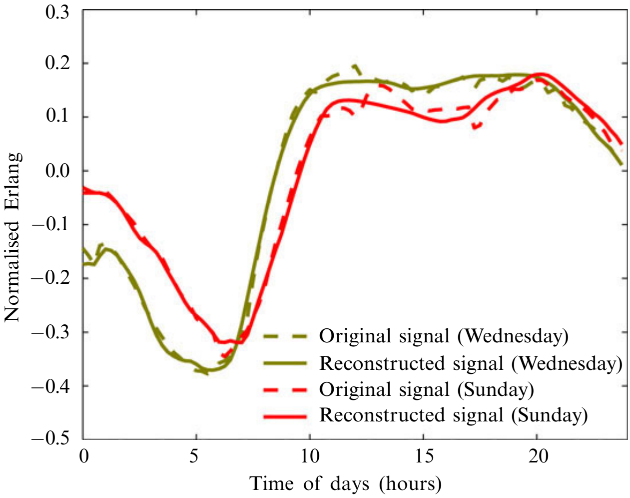


Figure 6. [In colour online.] Eigenvector reconstruction of signals from the Pantheon.

elements, thereby allowing us to calculate a covariance matrix and then decompose it into a set of vectors, each modified by the appropriate coefficient. Figure 6 shows how well the original signal can be approximated using only the first four eigenvectors from the covariance matrix.

We termed the output of this process an 'eigenplace', since it connects the spatio-temporal patterns of mobile phone usage to human activity spaces. By extracting the principal components of each pixel separately, we are able to abstract the daily and weekly routines embodied in an individual place. In effect, the more strongly 'routinised' activity in a pixel becomes, the larger the principle eigenvector's coefficient becomes. Conversely, spaces with unpredictable types of usage will attach greater importance to the lesser vectors.

Since the process of eigendecomposition may be new to many readers, it is helpful to consider our implementation of this method in more detail. The sequence of mathematical operations that generates these vectors is given by:

(1) L_{2115} is the set of 2115 pixels in Rome, each measuring 250 000 m^2.

(2) T_{96} is the set of fifteen-minute intervals at which observations were made each day.

(3) D_{90} is the set of days for which observations were made during the study period of three months.

(4) $E(\delta, \lambda, \tau)$ defines a unique Erlang value for the triplet referenced by a location $\lambda \in L_{2115}$, a time $\tau \in T_{96}$, and a day $\delta \in D_{90}$.

(5) $E_{\log} = \log_{10} [E(\delta, \lambda, \tau)]$.

(6) Normalisation over time—which benchmarks a pixel's Erlang value at one point in time against the mean of all times—scales the signal so that we can compare pixels without considering the magnitude of their signals:

$$E_{\text{normalised}}(\delta, \lambda, \tau) = \frac{E_{\log}(\delta, \lambda, \tau)}{\underset{\tau \in T_{96}}{\text{mean}}[E_{\log}(\delta, \lambda, \tau)]} .$$

(7) For any given location, a 90 × 96 correlation matrix can be calculated as:

$$C(loc) = \sum_{\text{day}=1}^{90} E_{\text{normalised}}(\lambda, \delta, \tau) E^{T}_{\text{normalised}}(\lambda, \delta, \tau) .$$

(8) From here we can calculate the eigenvalues (α_i) and eigenvectors [eigenvector(λ, τ)] of the correlation matrix. We then reorder the ninety-six eigenvectors so that the largest eigenvalue is first.
(9) Once the eigenvectors have been derived, it is possible to reconstruct a generic signature $E_{\text{normalised}}(\delta, \lambda, \tau)$ as the weighted sum of the eigenvectors [eigenvector(λ, τ)]. The weights are simply coefficients, which can be calculated as follows:

$$E_{\text{normalised}}(\delta, \lambda, \tau)\text{eigenvector}_i(\lambda, \tau)^T, \qquad \text{for } i = 1, \ldots, n \text{ eigenvectors} .$$

In figure 7 we show the eigenvectors for the seven pixels selected on page 4, in order to develop our understanding of how these areas behave over time. The first eigenvector clearly suggests a single, underlying, structure to mobile phone usage that varies little across space and is driven primarily by the time of day. Consequently, it seems reasonable to connect the first eigenvector to the diurnal cycle of Rome—people waking up, going to work, breaking for lunch, and then heading home in the evening. Conversely, the high rate of change in the fourth eigenvector suggests that it is largely composed of 'noise'. So if the first vector is of little use for distinguishing between locations, and the fourth describes only minor aspects of the original signal, this leaves the second and third eigenvectors to account for the majority of the difference between pixels, and here we find some intriguing features.

In the second eigenvector the higher values during late night and early hours of the morning in Trastevere and the Pantheon strongly imply nighttime activity. The much smaller drop in phone usage around midday at Piazza Bologna suggests an area with a different usage profile from the other six pixels, and it is the only site connected to a known residential space and, hence, to family. The third eigenvector is much more difficult to interpret on the basis of anecdotal knowledge, and it should be noted that the coefficients associated with this vector are not of the same

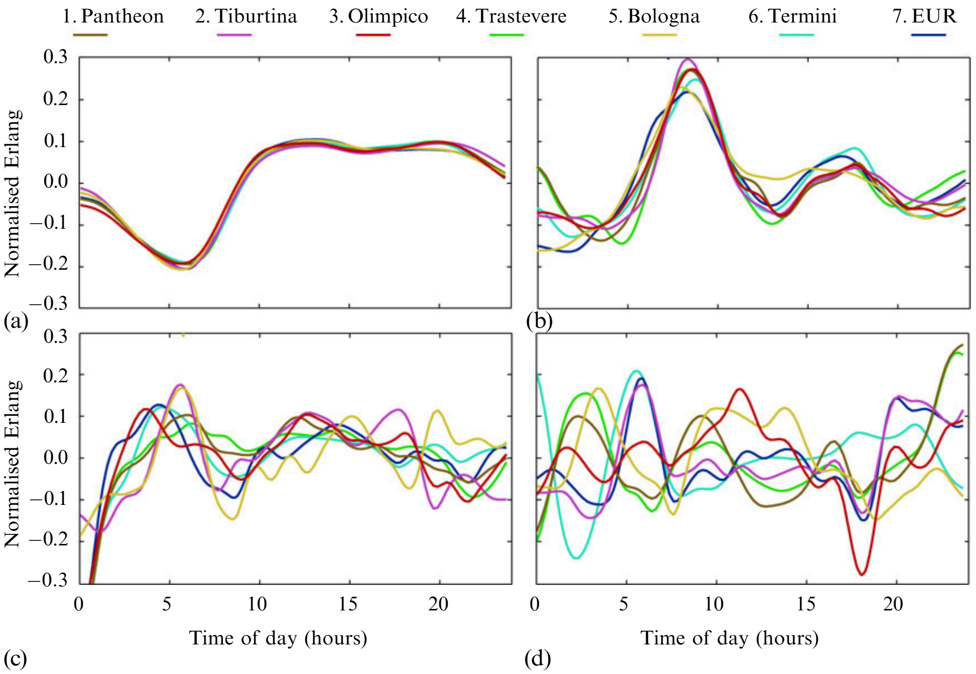


Figure 7. [In colour online.] Eigenvectors for selected pixels: (a) first eigenvector, (b) second eigenvector, (c) third eigenvector, (d) fourth eigenvector. The numbers 1–7 refer to specific locations indicated in figure 1.

magnitude as the first and second; but it is interesting to note that usage at the adjacent Piazza Bologna and Tiburtina falls in parallel during the morning as mobile users flow into central Rome, but follows a markedly different trajectory in the evening as people head home. In contrast, trends in phone usage at the comparatively more distant Trastevere and Pantheon seem to fall more naturally in sync.

While these eigenvectors can be interpreted intuitively, they should not be considered separately from their corresponding eigenvalues, particularly since in some cases the coefficients may be negative. In figure 8 we show how the coefficients for three of the seven sites vary over the course of a week. As is clear from the figure, the coefficients of the first eigenvector are often an order of magnitude larger than those of the second, third, or fourth eigenvectors, and this clearly squares with our understanding from figure 7 that the first vector describes an underlying rhythm to life in Rome.

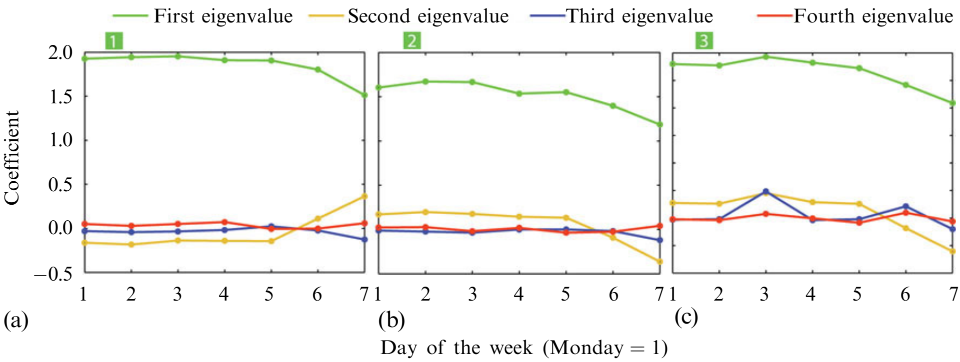
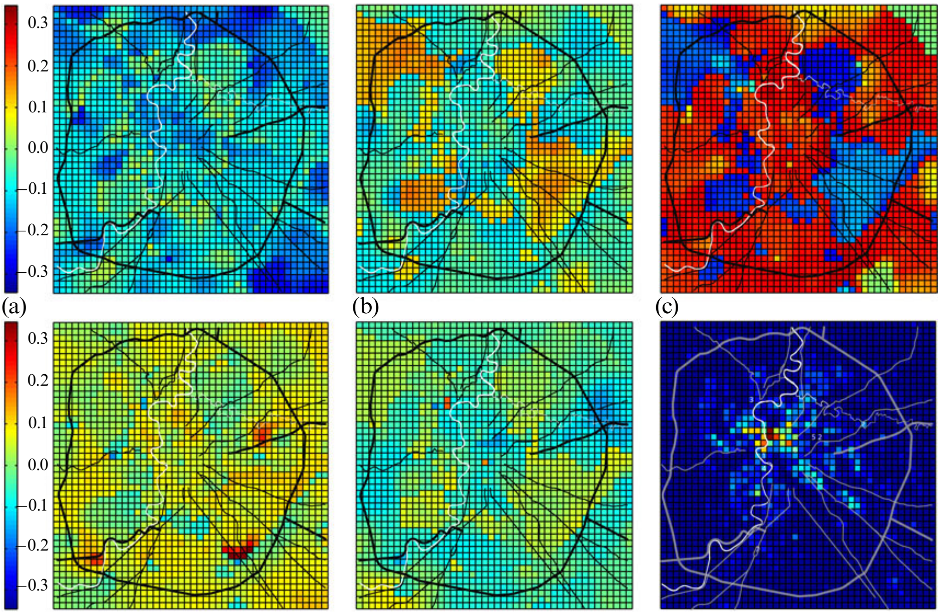


Figure 8. [In colour online.] Primary eigenvalues for selected pixels: (a) Pantheon (tourist monument); (b) Tiburtina (rail/metro station); (c) Olimpico (sports and concerts). These three locations are identified by numbers 1 – 3, respectively, in figure 1.

We see also that the relationship between vector and value remains fairly constant over the course of the week, but changes significantly over the course of the weekend and especially on Sunday. This drop in the first coefficient corresponds to the lower levels of overall usage observed for weekend activity in figure 2. The plot for Olimpico is particularly interesting because of the impact of Wednesday night football matches and the stronger resemblance between Wednesday and Saturday—which both have nighttime events—than with Sunday.

Taking the eigenvectors and eigenvalues together, the first and fourth eigenvectors can be removed to leave only the most distinctive elements of the signal, in a process tantamount to normalisation. This process still cannot be used to support the direct comparison of pixels as the eigenvectors are not comparable, but it does enable us to take different points in the day and examine the rank of each pixel as measured by the magnitude of its second and third eigenvectors. This type of profiling enables us to see how the relative focus of human activity moves around the city over the course of the day.

Consequently, we returned to the idea behind the normalised peak Erlang plot in figure 4. Now, however, we are able to focus on the difference that is unexplained by the daily rhythm of urban life. Figure 9 shows several distinct groupings of pixels. Three pixels, two adjacent to Olimpico and one in the very centre of Rome at Termini Station, stand out for the relatively high levels of usage at 9 pm in the evening [figure 9(e)]. In contrast, the map for 9 am [figure 9(c)] shows comparatively high usage across almost all of Rome except for three distinct areas—between Strada dei Parchi



(a) (b) (c) (d) (e) (f)

Figure 9. [In colour online.] Second eigenvector map: (a) 1 am; (b) 5 am; (c) 9 am; (d) 5 pm; (e) 9 pm; (f) normalised business density

and Via Toscolano, to the East of Urbe airport, and to the northwest of Via del Foro Italico. It is intriguing that these same areas also have higher levels of use in the evenings, which would be consistent with residential use.

It is also interesting to examine the map from 1 am [figure 9(a)], from which we can attempt to deduce some areas popular for late night activity: to the west of the Tiber is Trastevere and to the south of central Rome are Ostiense and Testaccio, all of which are areas known for their clubs and bars. Perhaps the most interesting plot, however, is that for 5 pm [figure 9(d)]: three areas are particularly prominent, and they all lie along major radial routes out of the CBD. Since these active areas are located just before motorists would reach the orbital highway, the most likely explanation at this time is that these indicate a popular calling point for commuters leaving the city.

Limitations

Rome's long pedigree guarantees a high degree of overlap between activities, which makes it very difficult to deduce the predominant land use directly. Moreover, unlike Britain, Italy does not make low-level census and business data available for free to academics. In addition, because we are examining the entire urban system, unlike Eagle and Pentland (2006b), we cannot assign our pixels to three a priori categories: nowhere in Rome is simply work, home, or other. Consequently, it has proved difficult to map the derived eigenplaces on to discrete types of land use or human activity—it is likely that the signal is simply too complex.

However, we might reasonably expect that similar data from North American or Northern European cities—because of the stronger spatial segregation of activity between the CBD and the suburbs—would show stronger signal differentiation. This suggests that an eigenplace analysis of Rome is one of the more challenging undertakings. We are endeavouring to gain access to comparable datasets for other cities, but the requirements of negotiating with each operator separately and of convincing them of the value of this research has made this process challenging. Indeed, data access remains the principle obstacle to the expanded use of telecoms data in urban planning and modelling.

Discussion

It is helpful to place this work within the context of a broader shift in urban studies from intentional to behavioural research (O'Neill et al, 2006, page 317). The intimately personal nature of the mobile phone and our growing ability to track it across the urban landscape raises new questions regarding personal privacy at the same time as it opens an unprecedented window into how millions of people, each pursuing their individual interests and responsibilities, use the city. Although the cellular network cannot offer the locational specificity of GPS, its scalability means that it nonetheless remains very attractive for research at scales at which a higher resolution is unnecessary. In particular, this class of data would seem to be particularly relevant at the city-region level, for which the system is much too large to explore using traditional analytical methods.

The lack of existing research in this field has led us to test a range of analytical approaches co-opted from fields such as computer science and astronomy in the hope that these will shed light on the dataset. We have shown that there is a clear, identifiable relationship between human activity and Erlang, but that the nature of this relationship is neither simple nor linear. A straightforward regression therefore appears insufficient to account for the way that bandwidth usage changes in areas of intensive activity, such as the CBD. Nonetheless, we have provided modest empirical support for our experiential knowledge of the city's terrain—although indirect, the distribution of businesses provides important evidence that phone usage is connected to local land use in very real ways.

In this eigendecomposition the primary eigenvector clearly indicates a common underlying pattern to mobile phone usage in Rome, while the secondary and tertiary vectors indicate spatial variation that is very suggestive of temporally-related and activity-related influences. One powerful effect of this process is that it essentially allows us to factor out the diurnal signal, leaving the key points of difference between each pixel or set of pixels. Thus we have early evidence that the concept of the 'eigenplace' provides a new tool with which to distinguish between areas on the basis of their telecommnications signature.

This surprisingly fine-grained computational approach appears quite sensitive to time-based and place-based influences, and so this new type of data clearly has the potential to significantly expand our understanding of urban dynamics. As a result, we may be able to develop new methods for improving the delivery of public services to all inhabitants—rather than viewing the city in terms of static infrastructures, we can visualise it as an interconnected set of flows (and immobilities), and can try to reach out in ways optimised to the nature of the problem from the human, and not the purely physical or mechanistic, standpoint.

Acknowledgements. The Real Time Rome exhibit at the 10th International Architecture Exhibition at Venice, Italy was sponsored by Telecom Italia, with technical partners: Biennale de Venezia, City of Rome, Google, Atac-Rome buses, Samarcanda Taxi, and Microtek. We gratefully acknowledge the significant contributions of Burak Arikan, Assaf Biderman, Filippo Dal Fiore, Saba Ghole, Daniel Gutierrez, Sonya Huang, Sriram Krishnan, Justin Moe, Francisca Rojas, and Najib Marc Terazi. The authors would also like to thank Sir Peter Hall and Michael Batty of UCL, and Larry Vale of MIT for supporting a collaborative research environment. And a final thank you to the Balzan Foundation, whose support for young researchers through the Balzan Prize helped to make this trans-Atlantic collaboration possible.

References

Ahas R, Mark Ü, 2005, "Location based services: new challenges for planning and public administration?" *Futures* **37** 547 – 561

Batty M, 1990, "Invisible cities" *Environment and Planning B: Planning and Design* **17** 127 – 130

Calabrese F, Ratti C, 2006, "Real time Rome" *Networks and Communication Studies* **20** 247 – 258

Calabrese F, Colonna M, Lovisolo P, Parata D, Ratti C, 2007, "Real-time urban monitoring using cellular phones: a case-study in Rome", SENSEable working paper, http://SENSEable.mit.edu/papers/pdf/CalabreseRatti2007SCLWorkingPaper.pdf

Eagle N, Pentland A, 2006a, "Eigenbehaviors: identifying structure in routine", http://vismod.media.mit.edu/tech-reports/TR-601.pdf

Eagle N, Pentland A, 2006b, "Sensing complex social systems" *Personal Ubiquitous Computing* **10** 255 – 268

Eurostat, 2007, "Mobile phone subscriptions (per 100 inhabitants)", http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1073,46870091&_dad=portal&_schema=PORTAL&p_product_code=ECB11024

Goddard J B, 1973, "Office employment in central London: a spatial analysis" *Progress in Planning* **1** 111 – 232

Graham S, 1997, "Cities in the real-time age: the paradigm challenge of telecommunications to the conception and planning of urban space" *Environment and Planning A* **29** 105 – 127

Graham S, 1998, "The end of geography or the explosion of place? Conceptualising space, place and information technology" *Progress in Human Geography* **22** 165 – 185

Hall P, Pain K, Green N, 2007, "The informational geography of Europolis: mapping the flow of information", in *The Polycentric Metropolis: Learning from Mega-city Regions in Europe* Eds P Hall, K Pain (Earthscan, London) pp 70 – 90

O'Neill E, Kostakos V, Kindberg T, Fatah gen. Schiek A, Penn A, Stanton Fraser D, Jones T, 2006, "Instrumenting the city: developing methods for observing and understanding the digital cityscape", in *Lecture Notes in Computer Science* number 4206 (Springer, Berlin) pp 315 – 332

Reades J, Calabrese F, Sevtsuk A, Ratti C, 2007, "Cellular census: explorations in urban data collection" *Pervasive Computing* **6**(3) 30 – 38

Sheller M, 2004, "Mobile publics: beyond the network perspective" *Environment and Planning D: Society and Space* **22** 39 – 52

Shoval N, 2007, "Sensing human society" *Environment and Planning B: Planning and Design* **34** 191 – 195

Townsend A M, 2001, "The Internet and the rise of the new network cities, 1969 – 1999" *Environment and Planning B: Planning and Design* **28** 39 – 58