IEEE pervasive COMPUTING
MOBILE AND UBIQUITOUS SYSTEMS
www.computer.org/pervasive

Cellular Census: Explorations in Urban Data Collection

Jonathan Reades, Francesco Calabrese, Andres Sevtsuk, and Carlo Ratti

Vol. 6, No. 3
July–September 2007

This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

IEEE COMPUTER SOCIETY

© 2007 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

For more information, please see www.ieee.org/web/publications/rights/index.html.

Cellular Census: Explorations in Urban Data Collection

Analysis of cell phone use can provide an important new way of looking at the city as a holistic, dynamic system.

Jonathan Reades
University College London

Francesco Calabrese, Andres Sevtsuk, and Carlo Ratti
Massachusetts Institute of Technology

Much of our understanding of urban systems comes from traditional data collection methods such as surveys by person or phone. These approaches can provide detailed information about urban behaviors, but they're hard to update and might limit results to "snapshots in time."

In the past few years, some innovative approaches have sought to use mobile devices to collect spatiotemporal data (see the sidebar, "Urban Analysis Using Mobile-Device Data"). But little research has been done to develop and analyze the much larger samples of existing data generated daily by mobile networks.

The most common explanation for this is that the challenge of data-sharing with the telecommunications industry has hampered data access. However, in early 2006, a collaboration between Telecom Italia, which serves 40 percent of the Roman market, and MIT's SENSEable City Laboratory (http://senseable.mit.edu) allowed unprecedented access to aggregate mobile phone data from Rome. Here, we explore how researchers might be able to use data for an entire metropolitan region to analyze urban dynamics.

The Real Time Rome platform

The TI and MIT collaboration, developed under the Real Time Rome label, was shown at the 2006 Venice Biennale. The installation incorporated both real-time and historical visualizations of mobile phone usage levels in central Rome during autumn 2006. The system architecture, including data collection, transfer, and processing, has been detailed elsewhere.[1]

TI supplied several different types of data, first and foremost of which was the *Erlang*, a measure of network bandwidth usage typically collected at the antenna level. Additionally, TI used its innovative Lochness platform to supply aggregate location and trajectory data on callers using the system for more than three minutes at a time. Two transportation companies—Atac-Rome (a public bus company) and Samarcanda (a private taxi company)—also provided supplemental GPS data to MIT for further processing. However, here we focus on the Erlang data collected over four months in late 2006 and covering a region of 47 km^2, considering how it can help us better understand urban dynamics.

An Erlang is one person-hour of phone use, so 1 Erlang could represent one person talking for an hour, two people talking for a half hour each, 30 people speaking for two minutes each, and so on. Consequently, Erlang data is both aggregate and anonymous, and deducing individual identities from the data collected and stored in the system is impossible. Additionally, because Erlang data is a standard measure used by most network operators, it's an accessible source for the analysis of typical GSM (Global System for Mobile Communication) networks. You can collect Erlang data without installing new applications or upgrading the base station controllers, both of which incur costs and operational risks for the networks.

Although Erlang data can't be linked to an individual subscriber and doesn't offer the locational

Published by the IEEE Computer Society ■ 1536-1268/07/$25.00 © 2007 IEEE

Urban Analysis Using Mobile-Device Data

The Massachusetts Institute of Technology's Reality Mining project (http://reality.media.mit.edu) successfully abstracted common behavioral patterns from the activities of 70 students and faculty issued with Nokia phones carrying specially designed logging software.[1,2]

Rein Ahas and Ülar Mark tracked the mobile phones of 300 users for a "social positioning method" analysis.[3] By combining spatio-temporal data from phones with demographic and attitudinal data from surveys, they created a map of social spaces in Estonia.

In the UK, the Cityware research group has taken a more readily scalable approach. They supplement the pedestrian flow data typically gathered as part of a space syntax analysis with data on Bluetooth devices passing through pedestrian survey "gates."[4]

However, approaches such as these can suffer from important limitations: they rely on the deployment of ad hoc infrastructure or require user consent. Consequently, sample sizes are necessarily more modest and might be limited in terms of the research's spatial extent.

REFERENCES

1. N. Eagle and A. Pentland, "Eigenbehaviors: Identifying Structure in Routine," 2006; http://vismod.media.mit.edu//tech-reports/TR-601.pdf.
2. N. Eagle and A. Pentland, "Reality Mining: Sensing Complex Social Systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, 2006, pp. 255–268.
3. R. Ahas and Ü. Mark, "Location Based Services—New Challenges for Planning and Public Administration?" *Futures*, vol. 37, no. 6, 2005, pp. 547–561.
4. E. O'Neill et al., "Instrumenting the City: Developing Methods for Observing and Understanding the Digital Cityscape," *UbiComp 2006: Ubiquitous Computing*, LNCS 4206, Springer, 2006, pp. 315–332.

specificity of GPS, it nonetheless remains attractive for urban research at scales where this level of resolution is unnecessary. In effect, Erlang data provides both a view of urban space as seen through network bandwidth consumption and, indirectly, insight into urban life's spatial and temporal dynamics. This aspect makes it an excellent jumping-off point for research supporting public-transport planning, health and safety, advertising, and other types of group-directed activity.

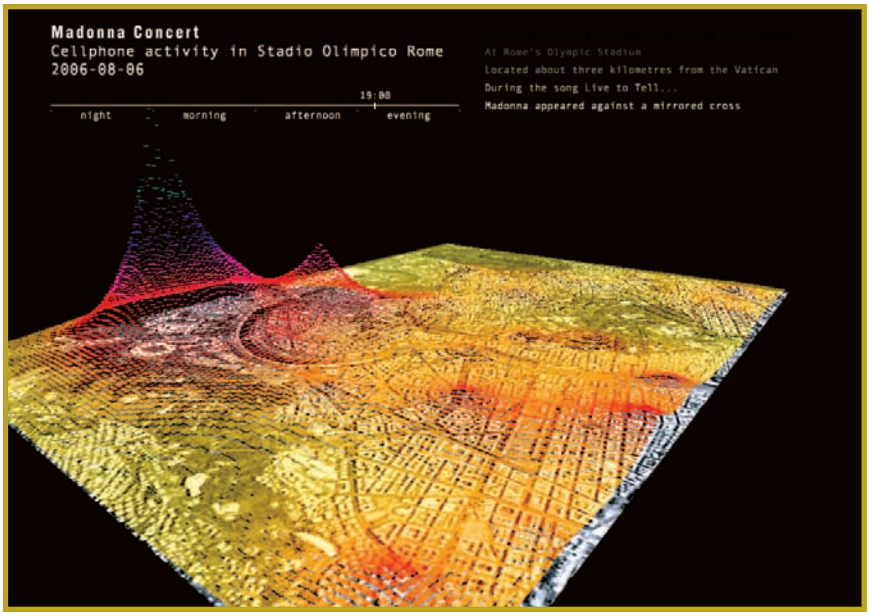


Figure 1. A 3D plot of telecommunications activity during a Madonna concert in Rome.

First visualizations and hypothesis

Figure 1 shows one of the simplest visualizations of Erlang data: a 3D plot of telecommunications activity during Madonna's controversial 6 August 2006 performance, when more than 70,000 people converged on the Stadio Olimpico for a concert condemned by the Pope. Generic Erlang maps such as this, which was presented at the Biennale, are graphically appealing and intuitively easy to grasp. However, they're actually quite difficult to interpret rigorously, and they provide little insight into local-area dynamics without additional processing.

We hypothesize that by employing various statistical techniques, we can use differences in Erlang data over time to derive clues to the types of activity in the immediate area of the mast. (A mast can carry multiple antennas, oriented in different directions or serving different frequencies.) This analysis is conceptually related to the idea of a chronotype (see the "Chronotypes and Space-Time Typologies" sidebar), except that we're characterizing spaces by their mobile-bandwidth use over time. By analyzing the bandwidth "signature" of each antenna, we try to envision how it might correlate with urban activities in the geographical vicinity.

Because Erlang data is an antenna-level measure, we needed an algorithm to spread the point data values across the

Chronotypes and Space-Time Typologies

To help conceptualize a city's complex hourly, daily, weekly, monthly, and annual rhythms, Luca Bertolini and Martin Dijst put forward Roberta Bonfiglioli's concept of the *chronotype*.[1] The chronotype is a useful conceptual handle for thinking about how different groups occupy the same space depending on the time of day. Bertolini and Dijst offer the example of a mixed-use area inhabited by young couples without children and by families. The young couples will likely work in another part of the city, returning perhaps only in the evening to socialize in bars and restaurants. In contrast, family members will go shopping and use other services during the day in this area. The same space can thus have two or more distinct uses and populations.

As another way to conceptualize these rhythms, Robbert Zandvliet and Dijst offer *space-time typologies*.[2] That is, they propose "a typology of urban, suburban, and rural municipalities … based on diurnal weekday variations in visitor populations"[2] as a way to understand how place works in Manuel Castells' "network society."[3]

REFERENCES

1. L. Bertolini and M. Dijst, "Mobility Environments and Network Cities," *J. Urban Design*, vol. 8, no. 1, 2003, pp. 27–43.
2. R. Zandvliet and M. Dijst, "Short-Term Dynamics in the Use of Places: A Space-Time Typology of Visitor Populations in the Netherlands," *Urban Studies*, vol. 43, no. 7, 2006.
3. M. Castells, "Grassrooting the Space of Flows," *Cities in the Telecommunications Age*, J. Wheeler, Y. Ayoma, and B. Warf, eds., Routledge, 2000, pp. 18–27.

area served, accounting for distance decay in signal coverage and multiple antennas on a single mast. Carlo Ratti and his colleagues took a center-of-gravity approach,[2] but to interpolate values for the entire metropolitan region, an alternative algorithm[3] was used to divide Rome into "pixels" measuring 1,600 m^2. We used an exponential distribution function to derive an Erlang point value based on a composite signal from the surrounding masts.

We use this mathematical notation:

- *Loc* is the set of 1,600 m^2 pixels.
- T_{96} is the set of times when we made observations each day of the week. Because we took measurements every 15 minutes, one day comprises 96 observations.
- *Day* is the set of {Weekday, Friday, Saturday, Sunday} (we discuss this in more detail later).
- $erlang(\delta, \lambda, \tau)$ defines the Erlang value at location $\lambda \in Loc$, at time $\tau \in T_{96}$, and $\delta \in Day$.
- $\underset{i \in I}{mean}\{a_i\}$ indicates the mean of the values a_i, $i \in I$.

(To preserve confidentiality, TI used a scaling factor to adjust the Erlang values transmitted to the SENSEable City Laboratory. This means that while the relative difference between any two observations is scaled consistently, the actual Erlang value at that point in time is unknown. So, it's helpful to focus on the relationships between points over time and space rather than the specific value at any one point in time.)

Using prior knowledge of the city, we arbitrarily selected eight locations that we expected to have markedly different signatures. Following an initial visualization exercise, we selected six for analysis:

- Termini, Rome's main passenger rail station and busiest subway station;
- Trastevere, a mixed-use area popular with Romans and tourists for its bars and restaurants;
- the Piazza Bologna, a residential area east of the city center;
- the area in front of the Pantheon (one of Rome's premier tourist attractions), which also contains many bars and restaurants;
- the Stadio Olimpico, a sports and major concert venue northwest of central Rome; and
- Tiburtina, a smaller rail and subway interchange.

Figure 2 shows the pixels for these six locations.

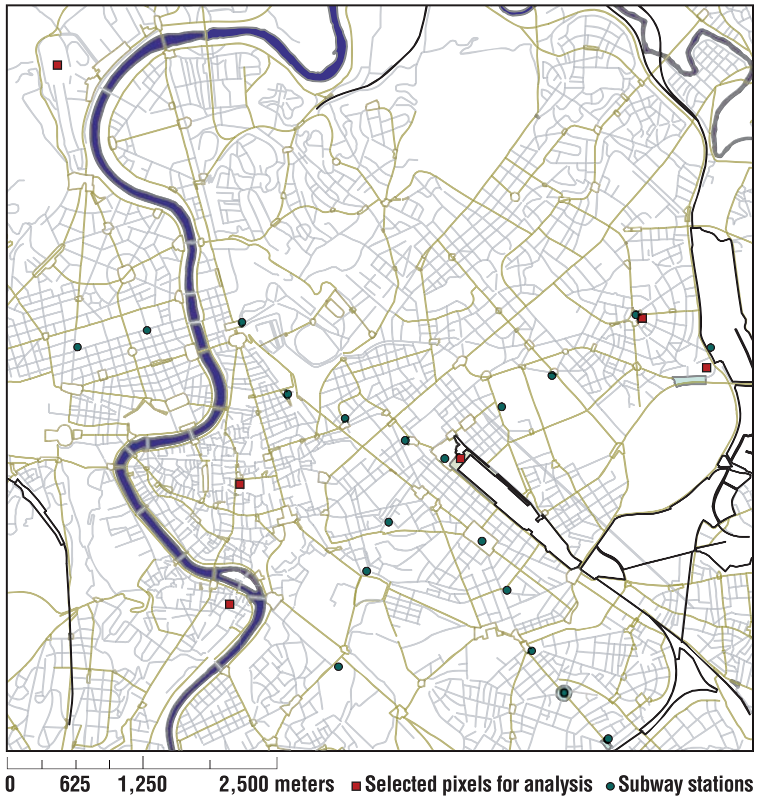
To minimize the impact of special events on the data set, we calculated an average Erlang value for each pixel at each 15-minute interval, using a 90-day period. So, for example, the data point for 9 a.m. Monday is an average of every 9 a.m. Monday value between 1 September and 30 November 2006. We excluded civic holidays from the calculation on the basis that they would introduce unnecessary noise.

Erlang data by day of the week

Beginning with a minimal level of processing, figure 3 shows how Erlang data changes over time at each of the six selected pixels. As the graphs indicate, Monday through Friday are broadly similar, except for a more rapid decrease in activity on Friday afternoon, suggesting a transition to the weekend. Even more strikingly, Saturday and Sunday values often drop below 50 percent of the typical weekday load, but the drop's magnitude varies dramatically from site to site. This finding indicates that weekday and weekend data should be treated separately in our analysis.

Intriguingly, areas more closely identi-

Figure 2. A map of Rome indicating the six locations ("pixels") selected for analysis of mobile phone usage.



fied with Roman residents, such as the Piazza Bologna and Tiburtina, display lower levels of day-to-day Erlang variance than those associated with more transient populations of commuters or tourists, such as Termini and the Pantheon. This peculiarity suggests that the greater the flux of people—or, possibly, nonresidents—through a site, the greater the variance in the signal. Conversely, predominantly "local" areas seem to have higher levels of routine or habitual activity and thus less variation between days.

However, in spite of the differences between weekdays, Fridays, and the weekend, figure 3 shows that all six locations demonstrate a broadly comparable rhythm—a rapid ramping-up of telecommunications activity between 6 and 10 a.m. on weekdays and a slower pace on weekends. Apart from the Stadio Olimpico, where the rhythms of concerts and football matches clearly show on the graph, patterns are quite uniform: a clear double peak and varying ratios of weekday-to-weekend activity. So, how can we make differences in signatures between different sites more evident?

Normalization

The magnitude of the differences between sites in figure 3 makes it hard to compare them in a more detailed way, so some type of data normalization is necessary. In figure 4, we plot the ratio of telecommunications intensity at one pixel against the average of every pixel in the system at that point in time (*normalization over space*). We then compare that to the daily pixel average (*normalization over time*). Using this approach, we can identify otherwise hidden shifts in the relative intensity of activity across Rome.

We employed these normalization steps:

1. For each location λ and day δ, we calculate

$$\operatorname*{mean}_{\lambda \in Loc}\left\{erlang\left(\delta,\lambda,\tau\right)\right\}$$

2. We then normalize the signature over space:

$$erlang_{norm\_space} = \frac{erlang\left(\delta,\lambda,\tau\right)}{\operatorname*{mean}_{\lambda \in Loc}\left\{erlang\left(\delta,\lambda,\tau\right)\right\}}$$

3. We then normalize the signature over time:

$$\tau \in T_{96} \rightarrow erlang_{norm\_space\_time} = \frac{erlang_{norm\_space}\left(\delta,\lambda,\tau\right)}{\operatorname*{mean}_{\tau \in T_{96}}\left\{erlang_{norm\_space}\left(\delta,\lambda,\tau\right)\right\}}$$

The differences in figure 4 are quite visible, and the radically different signature at the Stadio Olimpico indicates that you can readily recognize certain classes of urban activity by the unusual distribution of telecommunications activity. Such events will likely place a correspondingly high load on urban infrastructure and resources, and a similar spike in Erlang data is also likely during emergencies. This suggests that the real-time recognition of unusual concentrations of telecommunications activity might have relevance for public safety planning and transport scheduling.

On weekends, the higher levels of bandwidth use between midnight and 2 a.m. near the Pantheon and in Trastevere relative to work-oriented and residentially oriented pixels strongly suggest leisure activity. This feature suggests that we can also identify cultural and leisure areas on the basis of their telecommunications signature. Another feature with implications for the understanding of urban dynamics is the high level of activity at the transit hubs on weekday mornings, compared to residential sites.

Figure 5 shows a more diffuse pattern of spatial activity on weekends. This is

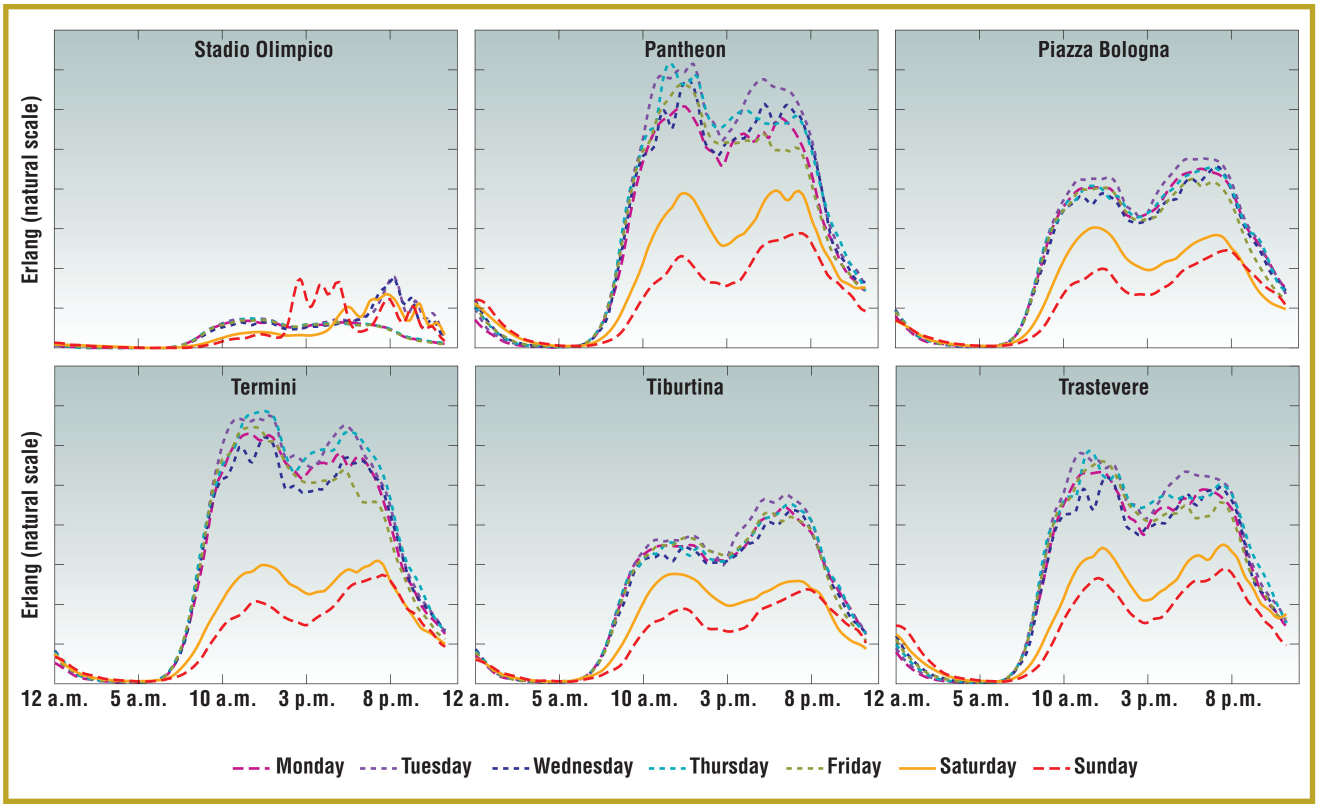


Figure 3. Erlang data for the six locations by day of week.

consistent with the idea that although weekday telecommunications activity at each site exhibits a more dynamic temporal pattern, weekend activity exhibits more spatial dispersal. From an urban-planning standpoint, this strongly suggests large commuter flows into the central business district during the week and more residentially oriented activity on weekends. Of course, planners are well aware of this spatial relationship, but spatial and temporal visualization of these features at this scale hasn't been possible before.

One caveat: the levels of activity between 3 and 6 a.m. throughout the week mean that any analysis using that period would be rooted in extremely low Erlang values. So, such a comparison might erroneously indicate excessive shifts in activity from site to site. Nonetheless, from this initial analysis, it seems that through normalized signatures we can reconstruct some of the functioning of the city using the invisible fingerprints of mobile phone infrastructure.

Cluster analysis

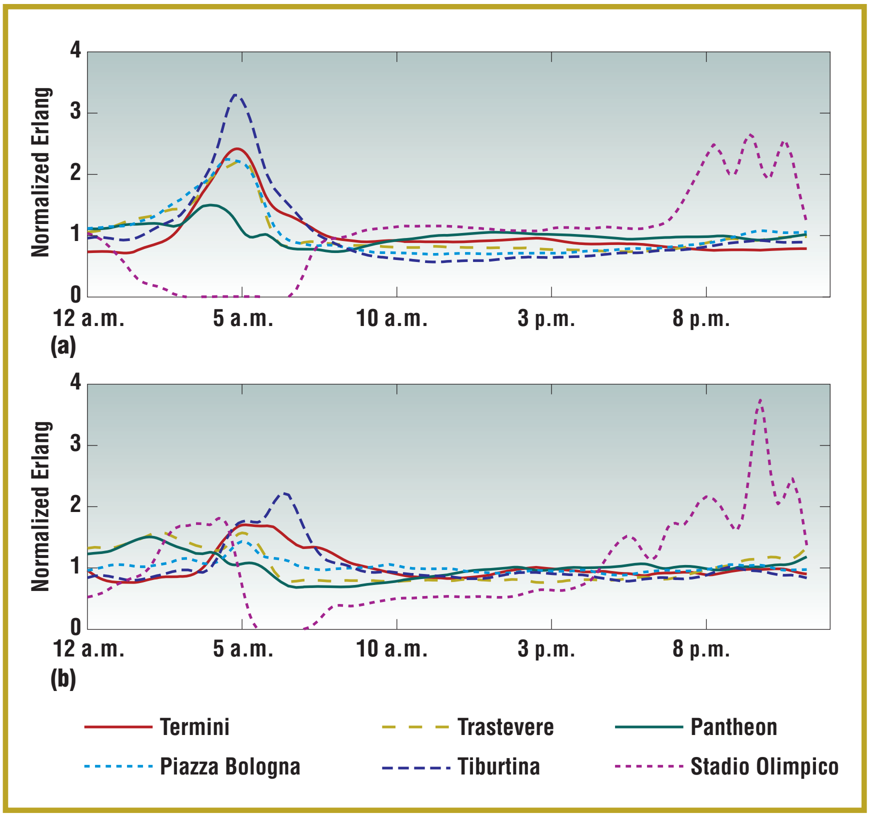
So far, we've focused largely on individual pixels, and we've identified some interesting features at a fairly detailed spatial level. Our preliminary analysis indicates that residential areas, commuter hubs, nighttime hot spots, and even special-event venues demonstrate features consistent with our contextual, anecdotal knowledge of Rome. However, validating our hypotheses requires a more rigorously quantitative study. The ultimate goal is to take the derived signatures, group them by degree of similarity, and map them to urban spatiotemporal structures.

As a proof of concept, we created a simplified vector—required for computational manageability—to feed pixel data for each of Rome's 262,144 pixels to a clustering algorithm. An examination of our six selected pixels suggested that six times in the daily cycle of Erlang activity are particularly significant: 1 a.m., 7 a.m., 11 a.m., 2 p.m., 5 p.m., and 9 p.m. Each of these points lies toward the middle of a period of rapid change or significant variation between sites—the early morning rise in activity, late morning peak period, early afternoon lull, afternoon peak, and evening drop. The six normalized Erlang values thus make up the coordinates of a vector that describes, in a limited way, each pixel's signature.

We could use many clustering techniques to create segmentations based on the affinity between vectors. We chose a K-Means approach, such that every observation in a cluster is as much like other members of that cluster and as different as possible from members of any other cluster. With six coordinates from each day, and separate sets of coordinates for Monday through Thursday (one set of averaged observations), Friday, Saturday, and Sunday, the K-Means algorithm used a 24-dimensional space.

We employed two clustering steps. First, for each pixel, we calculate *fea-*

Figure 4. Erlang data normalized over space and time by site: (a) Monday through Thursday and (b) Saturday.



ture(loc) = {*erlang*(δ, λ, *j*)}, *j* = 1 a.m., 7 a.m., 11 a.m., 2 p.m., 5 p.m., and 9 p.m.

Second, the K-Means clustering algorithm partitions the pixels into mutually exclusive clusters. Each cluster is characterized by its centroid, and the algorithm aims to minimize the error function:

$$\sum_{k=1}^{nClusters} \sum_{loc_j \in Cluster_k} \left|\text{distance}\left(loc_j, centroid_k\right)\right|^2$$

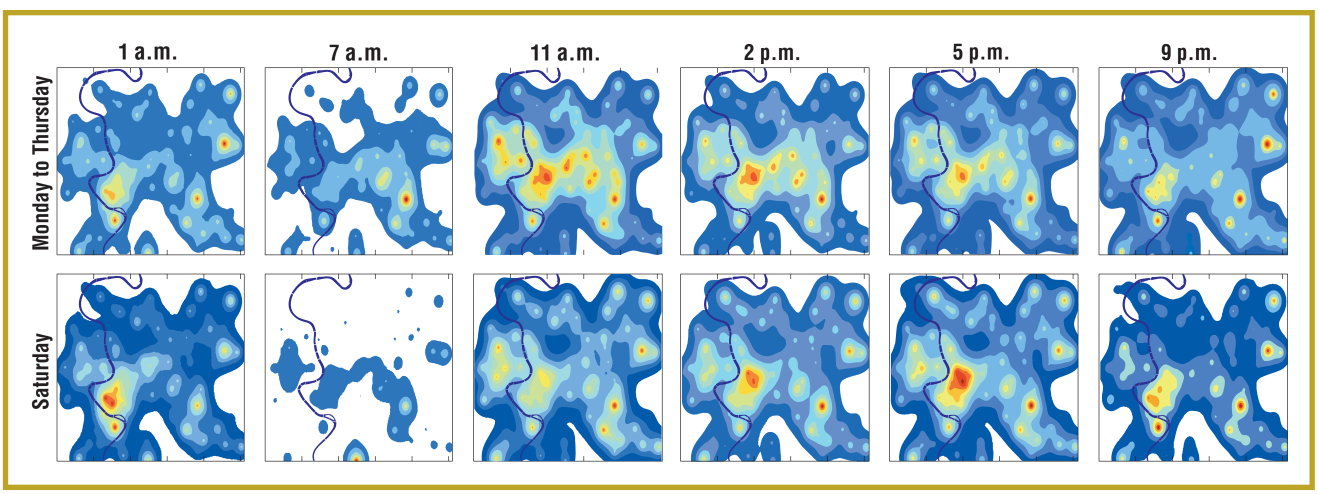
where $cluster_k$ is the set of objects related to the cluster *k*, and $centroid_k$ is the mean of all the points in $cluster_k$. We calculated the distance between pixels using the squared Euclidean distance:

$$distance\left(loc_1, loc_2\right) = \left(\sum_{\tau=1}^{24} \left|feature\left(loc_1\right)_i - feature\left(loc_2\right)_i\right|^2\right)^{1/2}$$

As a result of the clustering process, we can group all pixels in the city into any arbitrary number of groups based on the affinity of their composite Erlang signature. In our tests, we found a mix of clusters that suggest a complex set of relationships between signatures. Given the sheer complexity of cities, this is hardly surprising. However, the existence of several small clusters with much stronger levels of affiliation or differentiation indicates that the overall data set includes some quite distinct signatures. These signatures will likely map to distinct types of urban activity.

For this initial research, we worked with eight clusters as a compromise between simplicity and specificity. Doing this gave us a fair cophenetic correlation value of 0.7704. Cophenetic correlation is one way to gauge the clusters' fit to the original data set—values approaching 1.0 suggest a good fit—by comparing pairwise linkages between observations.

Figure 5. Erlang data for Rome normalized over space and time. Intensities range from low (blue) to high (red).



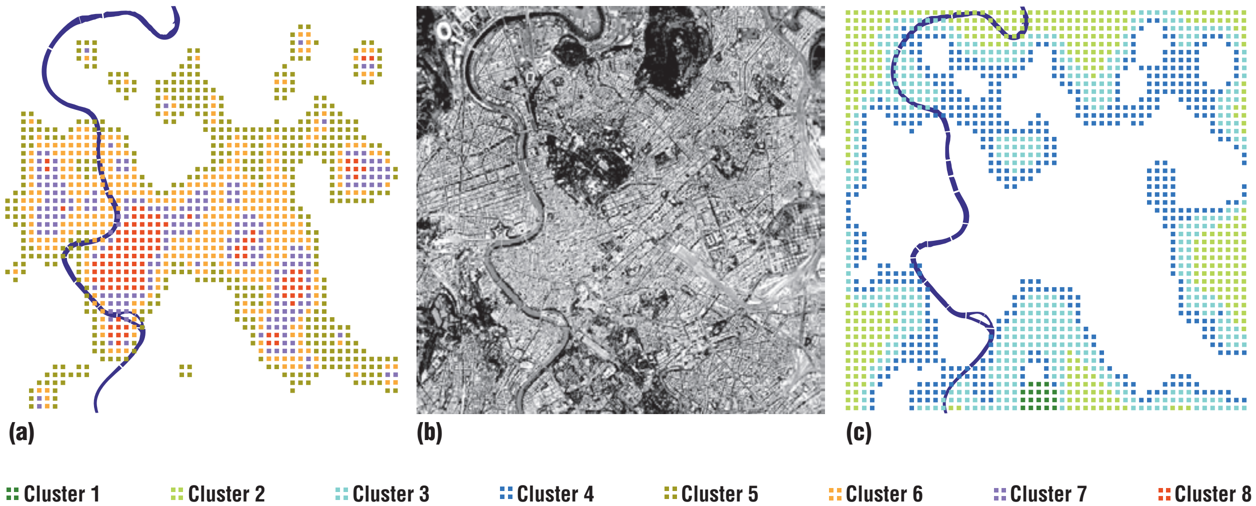


Figure 6. Analysis of eight clusters of Erlang data: (a) clusters 1–4; (b) a satellite view of Rome, for comparison; (c) clusters 5–8.

Projecting these clusters onto a map of Rome (see figure 6) naturally indicates that they're closely linked to the normalized Erlang signatures. The edges of Rome's urban core are clearly visible, as are the hot spots of urban activity straddling the Tiber River. The map suggests an overall structure to the city, with a correspondence between levels of telecommunications activity and types of human activity. At this point, however, we can't verifiably connect cellular signatures to specific types of human activity.

We then adjusted the metric to favor the two most distinctive types of use seen in the normalized graphs: early morning use suggestive of commuting behavior and late evening use suggestive of nighttime leisure activities. For these clusters we obtained cophenetic correlations of 0.7630 and 0.8508, indicating that the clustering approach has substantial promise.

The red nighttime-leisure cluster in figure 7a shows two discrete spatial groupings that map anecdotally to known areas of evening activity: Trastevere and the area ranging to the west and south of the Piazza Navona, and the vicinity of the Piazza Spagna. The red commuter clusters in figure 7b quite astonishingly map to the most important points of entry to the city by car and train: Termini station, Tiburtina, the end of the Corso d'Italia, the Porta Maggiore, and the Porta San Giovanni.

Discussion

Our preliminary findings suggest that signature analysis can provide an important new way of looking at the city as a holistic, dynamic system. In particular, the mobile phone network lets us develop a real-time representation of those dynamics at the city and city-region scale. This approach can complement traditional collection techniques, which are often outdated by the time they're available to policy makers and the general public. Of course, because our hypotheses so far are based on anecdotal evidence, our findings will require additional validation, which we outline below.

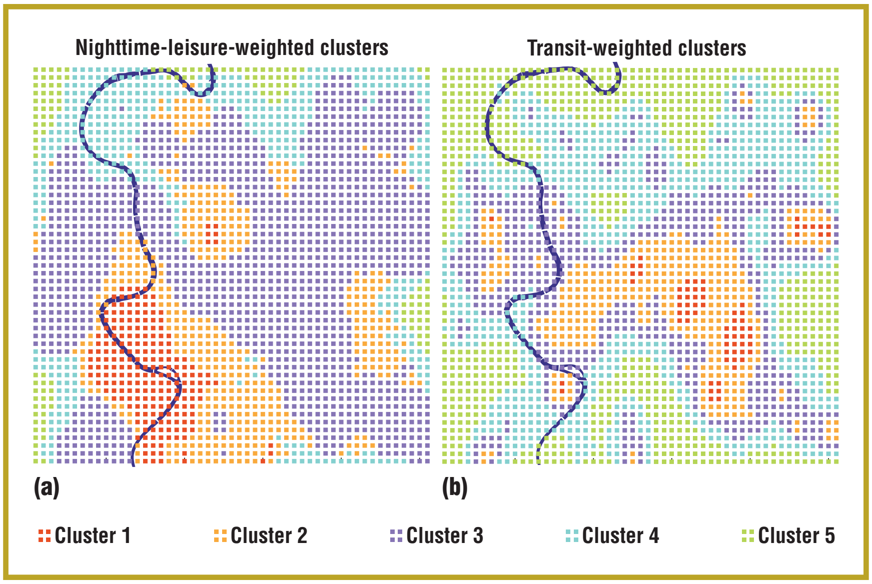
What's most promising about this early research is the extent to which our findings seem to parallel those of other European researchers[4,5] as well as more conceptual research into telecommunications' impacts on urban behaviors.[6,7] In particular, we can characterize areas on the basis of flows and dynamics rather than on the basis of comparatively static physical or demographic features.

Moreover, we've recently received data from Pagine Gialle (the Italian Yellow Pages) with which we intend to validate our initial findings by linking the signatures to spatial data on business types and densities. In so doing, we can build on the processing requirements we discussed earlier in this article:

1. Antenna and pixel values must be normalized over both space and time to provide a measure of relative telecommunications intensity.
2. The substantial differences between weekdays and weekends require treating them separately in a classification algorithm.
3. The key time periods intimated in this initial analysis appear to be 12 to 2 a.m., 5 to 8 a.m., 10 a.m. to 12 p.m., 2 to 6 p.m., and 8 to 11 p.m. However, as our initial cluster analysis makes clear, these aren't the only factors.

We expect several other analytical approaches to yield insights into network usage patterns. One of the most promising approaches is Eigenbehavior analysis.[8] Because we can easily map the signature to a vector representation of the sort already used in the cluster analy-

Figure 7. Analysis of the five clusters covering Erlang data for the two most distinctive types of cell phone use: (a) nighttime leisure, (b) early morning commuting.



sis, deriving the Eigenvectors should be quite straightforward. Applying other analytical techniques such as Fourier and wavelet transform plots might reveal new, distinct characteristics. Ideally, each of these analyses will eventually feed into a single categorization process that can discriminate between discrete types of behavior at the antenna and pixel levels.

Limitations of research

As we mentioned before, TI's masking function meant that we weren't able to work with true Erlang values. An additional constraint is that owing to operational requirements and planning restrictions, GSM masts are irregularly distributed and oriented. To manage the computational and mathematical complexity of calculating point values for 262,144 pixels over a three-month period, our algorithm spreads Erlang data through all 360 degrees, producing a possible skew in the overall distribution.

Finally, not all masts handle both the 900- and 1,800-Hz bands used in Europe. So, some network activity might gravitate toward more physically remote base stations with the hardware to process calls in a particular band. We don't have data that would let us compensate for these possible biases. So, without adopting an entirely different approach to data collection—one that the network operator would have been reluctant to support at this development stage—localizing phones more accurately is impossible.

Although the data to which we currently have access has clear, substantial limitations, we believe our approach represents an appropriate trade-off between locational specificity and implementational feasibility. Fortunately, analysis at the city and city-regional scale doesn't depend on the high level of accuracy that location-based services typically require. So, we feel that there's plenty of opportunity to gain valuable insights into urban dynamics using Erlang data at smaller scales, and we intend to move forward with it.

It would be exciting to compare the signatures collected from Rome with similar data from other major European cities such as London, Paris, or Frankfurt. For instance, it's reasonable to expect that cities with more distinct spatial patterns of human activity might display correspondingly more distinct patterns of network use and more readily classifiable signatures. Unfortunately, at this time commercial considerations appear to preclude using data from other network operators.

This issue highlights the extent to which research using cellular networks must take nonscientific factors into account. First, a policy framework at the national or European level that encourages networks to share nonidentifiable data with planning and policy researchers would be immensely helpful. Clearly, there are important considerations from the standpoint of commercial confidentiality, personal privacy, and possibly even national security. However, in the absence of clear regulatory guidance, further research using cell phones—the most widely deployed device with locational capabilities—won't be possible at the city or city-region scale.

Without encouragement, far more detailed data sets held by the networks will never see the light of day. For instance, paging data—generated by polling the phones in a cell to obtain a list of IMEI (International Mobile Equipment Identity) numbers at the mast level—could provide unmatched detail on travel origins and destinations, and on population densities. By scrambling handset identifiers with changing encryption schemes, reporting only partial trajectories, and never reporting on cells or paths containing fewer than an agreed minimum number of users, you would be able to perform this kind of research without compromising personal privacy.

This data would also assist enormously in understanding how individual and group behavior changes over time and space. This would not only shed further light on the rhythms of urban life but also address the fact that you can't derive metrics on activity and population densities from Erlang data alone.

The challenge is that as the data becomes more useful, it also becomes more sensitive to both operators and end users. An all-or-nothing approach to privacy has hampered this discussion.

It would be helpful to move toward a more nuanced understanding of how to preserve reasonable expectations of privacy by the network user while creating mechanisms to permit future research. We need to establish the extent to which certain types of data and analysis create either the perception of a privacy invasion[9] or the real risk of trail reidentification,[10] and to set out the trade-offs for public review and discussion.

ACKNOWLEDGMENTS

The Real Time Rome exhibit at the 10th International Architecture Exhibition at Venice, Italy was sponsored by Telecom Italia, with technical partners Biennale di Venezia, the City of Rome, Google, Atac-Rome, Samarcanda, 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.

We also thank Matthew Jull, Tim Kindberg, and our anonymous reviewers for their valuable feedback; Giovanni Celentano of the University of Naples and Lucio Pinto of the Silvio Tronchetti Provera Foundation for their generous suggestions; and Peter Hall and Michael Batty of University College London and Larry Vale of the Massachusetts Institute of Technology for fostering a collaborative research environment. And a final thank-you to the Balzan Foundation, whose support for young researchers through the Balzan Prize helped enable this transatlantic collaboration.

REFERENCES

1. F. Calabrese and C. Ratti, "Real Time Rome," *Networks and Communication Studies*, vol. 20, nos. 3 & 4, 2006, pp. 247–258.
2. C. Ratti et al., "Mobile Landscapes: Using Location Data from Cell-Phones for Urban Analysis," *Environment and Planning B: Planning and Design*, vol. 33, no. 5, 2006, pp. 727–748.
3. F. Calabrese et al., "Real-Time Urban Monitoring Using Cellphones: A Case-Study in Rome," SENSEable Working Paper, SENSEable City Laboratory, Massachusetts Inst. of Technology, 2007; http://senseable.mit.edu/papers/pdf/CalabreseRatti2007SCLWorkingPaper.pdf.
4. L. Bertolini and M. Dijst, "Mobility Environments and Network Cities," *J. Urban Design*, vol. 8, no. 1, 2003, pp. 27–43.
5. R. Zandvliet and M. Dijst, "Short-Term Dynamics in the Use of Places: A Space-Time Typology of Visitor Populations in the Netherlands," *Urban Studies*, vol. 43, no. 7, 2006.
6. S. Graham, "Cities in the Real-Time Age: The Paradigm Challenge of Telecommunications to the Conception and Planning of Urban Space," *Environment and Planning A*, vol. 29, no. 1, 1997, pp. 105–127.
7. M. Sheller and J. Urry, "The New Mobilities Paradigm," *Environment and Planning A*, vol. 38, no. 2, 2006, pp. 207–226.
8. N. Eagle and A. Pentland, "Eigenbehaviors: Identifying Structure in Routine," 2006; http://vismod.media.mit.edu/tech-reports/TR-601.pdf.
9. M. Langheinrich, "Privacy Invasions in Ubiquitous Computing," 2002; http://guir.berkeley.edu/pubs/ubicomp2002/privacyworkshop/papers/uc2002-pws.pdf.
10. B. Malin, "Betrayed by My Shadow: Learning Data Identity via Trail Matching," *J. Privacy Technology*, 2005, pp. 1–22; www.jopt.org/publications/20050609001_malin_abstract.html.

the AUTHORS

Jonathan Reades is an MPhil and a PhD candidate at the Bartlett School of Planning at University College London. His research interests are the application of mobile phone data to topics in urban planning such as business clustering and communication, and the spatiotemporal structures of European cities. He previously spent eight years in data analytics, helping telecom firms use their data for targeted marketing. His first degree was in comparative literature at Princeton University; he's a student member of both the Royal Town Planning Institute and the Town and Country Planning Association. Contact him at the Planning Dept., 4th fl., The Bartlett School, Wates House, 22 Gordon St., London WC1H 0QB, UK; j.reades@ucl.ac.uk.

Francesco Calabrese is pursuing a PhD in informatics and automatics engineering from the University of Naples Federico II and is a visiting doctoral student at the Massachusetts Institute of Technology's SENSEable City Laboratory. His research interests include hybrid control systems, embedded control systems, computer–numerical-control machines, finite-time stability, and real-time monitoring and analysis of urban dynamics. He received his Laurea in informatics engineering from the University of Naples Federico II. He's a member of the IEEE and the IEEE Control Systems Society. Contact him at the SENSEable City Laboratory, MIT 10-485, 77 Massachusetts Ave., Cambridge, MA, 02139; fcalabre@mit.edu.

Andres Sevtsuk is a PhD candidate in city design and development and urban information systems at the Massachusetts Institute of Technology. His research interests include accuracy and reliability in using mobile phone data for estimating the distribution of people in cities, the relationship between urban form and movement, and the study of new possibilities in street parking and one-way vehicle rentals using distributed computation and sensing. He received his MS in architecture studies from MIT. Contact him at 7 Dodge St., Cambridge, MA 02139; asevtsuk@mit.edu.

Carlo Ratti is an associate professor of the practice of urban technologies at the Massachusetts Institute of Technology, where he directs the SENSEable City Laboratory. He's also founding partner and director of carlorattiassociati, an architectural firm. He received his PhD in architecture from the University of Cambridge. He's a member of the Ordine degli Ingegneri di Torino and the Association des Anciens Elèves de l'École Nationale des Ponts et Chaussées, and is a UK Registered Architect. Contact him at the SENSEable City Laboratory, MIT 10-485, 77 Massachusetts Ave., Cambridge, MA 02139; ratti@media.mit.edu.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.