BIRKHÄUSER

DE-CODING THE CITY

How Big Data Can Change Urbanism

Dietmar Offenhuber, Carlo Ratti (Eds)

DECODING THE CITY

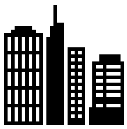
DECODING THE CITY
Urbanism in the Age of Big Data

Dietmar Offenhuber, Carlo Ratti (Eds)

Birkhäuser
Basel

Dietmar Offenhuber, Carlo Ratti

Introduction



How to explain the paradox that urbanism, as a profession, has disappeared at the moment when urbanization is everywhere – after decades of constant acceleration – is on its way to establishing a definitive global "triumph" of the urban condition.
Rem Koolhaas (1995)

•

In his 1995 essay "Whatever Happened to Urbanism," Rem Koolhaas diagnosed urbanism as a failed discipline. In the light of rampant global urbanization, the profession has failed to give shape or even to influence the physical, social, and economic realities of cities. While during the nineties, the main challenge was dealing with quantity, now, almost twenty years later, the situation presents itself as more ambiguous. Following more than fifty years characterized by suburbanization and the erosion of urban centers, cities in the developed nations have taken two different paths: They either have returned as global economic players of unprecedented power, or have become shrinking cities – hollowed out by deindustrialization and demographic changes (Ryan 2012). Meanwhile, in the emerging economies of the developing world, urbanization is continuing at an undiminished rate. Globally, it is expected that 67 percent of the world population will live in cities by the year 2050 (United Nations 2012, 2).

For the discipline of urbanism, the struggle continues. Both of these trajectories present challenges in terms of infrastructure provision, housing, and socioeconomic development. But planners, policy experts, and economists are no longer the only specialists responding to these challenges. New actors enter the stage and bring new approaches to the field. Perhaps the most significant developments have happened in the domain of data-intense methodologies. The term *big data* refers to the availability of massive amounts of machine-readable information. This information is generated by the sociotechnical systems in which humans are increasingly entangled, by choice or by necessity: cell phone networks, credit card systems, or social media networks. Since the "digital exhaust" generated by these systems is so closely connected to our daily lives, it becomes a valuable resource for observing the processes and interactions of society at almost no cost. However, because of the massive quantity of these data sets, which were not created and structured with research purposes in mind, new methods are required for analyzing them.

Following the increasing availability of such digital data sources during the past ten years, the social sciences have taken a quantitative turn, often labeled *computational social science* (Lazer et al. 2009). In combination with large data sets, new computational methods allow researchers to address topics such as environmental perception, sentiment, or social connection, which were previously limited to qualitative modes of inquiry. Computational social science has brought together sociologists with physicists, mathematicians, and computer scientists, who have recently discovered cities as a topic that provides numerous intriguing research problems to work on. In particular, the emerging field of network science – the study of complex networks – has made a significant contribution to urban research literature (Börner, Sanyal, and Vespignani 2007). Network science stands for a shift from an exclusively spatial perspective on urban data to a topological perspective, focusing on relationships and interactions between people, places, and institutions at any scale. In this sense, network science actualizes ideas that scholars such as Manuel Castells have introduced into the urban studies literature (Castells 1996). By abstracting cities as spatial social networks of interaction, network science helped to uncover structural commonalities shared by almost all urban systems, allowing researchers to describe and predict how cities evolve and will grow over time (Batty 2013). In engineering, the field of urban informatics stands for the instrumentation of cities with sensor networks (Foth 2009). This includes the ubiquitous integration of technologies such as Global Positioning System (GPS) into everyday devices, which have enabled a real-time representation of urban conditions. Smart cities, both an academic and an engineering discipline, is advanced by systems theorists and companies such as IBM, Siemens, or Cisco. The concept of smart cities promises to improve the management of cities by making its infrastructures more adaptive – able to collect information about its own state and to regulate itself based on the state of the whole system. Finally, perhaps most fundamentally, the role of the citizen in the governance of cities has changed in important ways. The rise of social media led to new forms of participation and social activism. Beyond traditional forms of participation in planning projects, citizens voluntarily fulfill increasingly sophisticated roles in monitoring, management, and governance of the city and its infrastructure – a phenomenon that Eric Paulos called the rise of the "expert amateur" (Kuznetsov and Paulos 2010).

In the light of these developments, we believe that it is necessary to reexamine the state of urban planning, and explore, while avoiding the trap of "Big Data Hubris" (Lazer 2014), how these approaches can lead to a new understanding of the city.

While the discussed approaches are relatively recent, they are not without precedents. The history of urban planning has many examples of paradigm shifts initiated by new technology.

By the 1960s, cybernetics, the science of dynamic feedback systems, had started to leave a mark in the domain of urbanism, with both good and bad results. On the positive side, cybernetic models brought a fresh perspective for investigating urban systems. With its focus on dynamic states, feedback, and systemic processes, cybernetics brought new attention to the temporal and ephemeral, and was in many ways conflicting with high-modernist planning theory with its strict compartmentalization of functions. But cybernetic models have also led to some catastrophic failures, as in the case of the 1970s redesign of New York's fire system by the RAND corporation, which left poor neighborhoods underserviced, resulting in rampant fires and social unrest (Flood 2010). The mismatch between complexity of the problem and the inadequacy of the means is self-evident in the perhaps most ambitious cybernetic experiment: project Cybersyn, designed to control Chile's national economy under the Allende presidency (Pickering 2010, 258).

Apart from the obvious failure to capture the sociopolitical dimension of urban systems, cybernetics can also be challenged from another perspective: it simply does not lend itself to good design theory. It is apt for simulating adaptive, complex dynamic systems, but it provides little guidance for future alternatives. As Andrew Pickering suggests, cybernetics is performative rather than representational: it operates in a black box that adapts to current states, but it does not provide an abstract image of the world in its current or desired state (2010, 19). Nevertheless, the idea that data opens up new territories for urban planning and design has a long tradition that can be traced back to the meticulous mapping efforts by Giambattista Nolli in eighteenth-century Rome or Ildefons Cerdà in Barcelona. In *The Sciences of the Artificial,* Herbert Simon calls for a rigorous design science, a "body of intellectually tough, analytic, partly formalizable, partly empirical, teachable doctrine about the design process" (Simon 1996, 113). Such a design science would serve two functions: first to evaluate the performance of a given design, and second, to guide the identification of alternative scenarios. The role of design is the reconciliation between the "inner" world of physical objects and an "outer" world of its goals and functions. "The natural sciences are concerned with how things are. [...] Design, on the other hand, is concerned with how things ought to be." (Simon 1996, 114)

About This Book

This is a book about models for capturing these phenomena for understanding and improving cities. While big data can improve our understanding of urban systems, little attempt has been made so far to think about the consequences for design. This book represents a cross-sectional view of the research agenda practiced at the SENSEable City Lab at the Massachusetts Institute of Technology. Situated in the Department of Urban Studies and Planning, the Lab is an interdisciplinary institution exploring how real-time technologies can help us to better understand our cities, as well as conceiving possibilities how these technologies can improve our cities. The authors of the individual contributions are current and past researchers at the MIT SENSEable City Lab, as well as frequent collaborators. They come from a range of different backgrounds, including architecture and urban planning, sociology, political science, mathematics, computer science, physics, and visual design.

The contributions in this book address the generation of urban data, their representation and analysis, and finally their relevance for urban planning and design. To that effect, this volume is structured into three parts. The first part focuses on case studies discussing the origin of urban data, their collection and generation, including their inherent gaps and biases. The second part focuses on questions of representation, either as visual or mathematical models. The third part finally focuses on the implications for urban design.

In his opening contribution, Fabien Girardin explains the notion of digital footprints, data left by humans using digital services. Girardin distinguishes between passive footprints, generated without the user's awareness, and active footprints, deliberately created and shared by the users. He illustrates how user-generated data from photo-sharing websites can be used to investigate the travel behavior of tourists. Since an increasing number of photos voluntarily uploaded to such sites contain explicit geographic and temporal information (Geotags), photo-sharing sites allow insights into how different groups of people travel and which interests and values guide them. Michael Szell and Benedikt Groß focus on the possibilities of public data sets collected for accountability purposes. They work with a data set of 170 million taxi trips in New York City over one year, and acquired from the city government through a Freedom of Information Act request. Identifying vast redundancies in the system, the authors explore the possibility of an alternative, sharable urban taxi system. In the third contribution, Tony Vanky investigates the sometimes-ambiguous relationships of trust of people toward urban data. Currently, there are few attempts to measure the relevance of urban real-time data: how they affect the interaction with urban infrastructure. Using Singapore as an example, Vanky describes such measures: whether there is use and appreciation of real-time urban data, and how the information affects the spatial decisions of citizens at an individual level.

David Lee discusses the data collection methodology of participatory sensing (Burke et al. 2006), a way to actively involve volunteers in targeted data collection using location-based technologies. Using the example of the Trash Track project, Lee explores how the experience of participating in such a project in return changes the behavior and perceptions of the participants themselves, for example, whether the helping to investigate the fate of trash might change their attitudes toward waste management and recycling. Finally, Francisca M. Rojas maps the cultural geography of New York City through aggregated cell phone data, contrasting the data set with the official census information in terms of validity and quality. In her analysis of the telecom data, she maps out not only the global activities of New York's economic centers; but also the realities of immigrants and migrant workers who remain in constant connection with their home countries.

The second section of the volume, dedicated to visualization and modeling, is opened by Kristian Kloeckl, who documents an initiative that explores the development of an urban real-time data platform for Singapore, facilitating the collection, combination, and distribution of multiple data streams from urban networks. These "urban demos" provide concrete examples of how meaningful visual representations of data open up possibilities for stakeholders from different domains, facilitating a crosscutting discourse about urban issues. The visualization experts Pedro Cruz and Penousal Machado reflect in their article on the use of metaphors and figurative approaches in urban data visualization. They focus on a dilemma that makes geospatial visualization a difficult problem – the dilemma between the spatial nature of urban systems and abstract nature of data to be represented. The authors explore the relationship between the properties of the underlying data and the representational strategy; distinguish between visualization as a "photograph" or a "caricature" of information. The network scientists Philipp Hövel, Filippo Simini, Chaoming Song, and Albert-László Barabási are concerned with the observation, formalization and prediction of human mobility behavior based on telecom data sets. These data sets from cell phone network providers include implicit information about the spatial movement of cell phone subscribers, making it possible to answer questions such as: how predictable are we in our daily routines? In their mathematical analysis, the authors discover a surprising regularity in the way people travel, and provide a mathematical model for describing human mobility behavior. Kael Greco explores representational strategies for urban data based on a mobility research case study in the city of Riyadh, Saudi Arabia. The text approaches the complexity and nuance of urban data from two related, but antithetical perspectives: first, using spatial data to develop new modalities of seeing the city, and second, using the structure and composition of the city to provide new ways of seeing and understanding social data.

The third part opens with a contribution by the urban planner Andres Sevtsuk, who contrasts the plan as the traditional representational tool of urban design with the representation of the city as a network. Sevtsuk, a researcher in the emerging discipline of configurational studies, shows how structural measures of urban path networks, such as "betweenness" or "reach" offer powerful approaches for explaining attractiveness and locational quality within an urban system. Markus Schläpfer's work also situated in the domain of scaling studies addresses the issue of polycentricism of urban structure in relation to the travel behavior of tourists and city dwellers. Using telecom data sets, Schläpfer investigated the destinations and temporal rhythms of hundreds of thousand people in Singapore, Lisbon and Boston. The physicist and pioneer of the emerging research area of scaling studies, Luís M. A. Bettencourt explains the many ways in which the overall scale of an urban system determines a range of urban qualities and measures, such as the average number of personal contacts, the per-capita economic output and innovation, but also the prevalence of crime. His mathematical theory describes how cities change as they grow, and how these changes affect the lives of their citizens. Stanislav Sobolevsky's contribution focuses on the ramifications of human communication for how regions and their boundaries are defined. Using examples of countries such as Great Britain, France, and Belgium, the author shows how Manuel Castell's concept of the space of flows (Castells 1996) can be charted in geographic space using cell phone data.

How can these methods lead to a new design practice for cities? Koolhaas concludes: "If there is to be a 'new urbanism' it will not be based on the twin fantasies of order and omnipotence; it will be the staging of uncertainty; it will no longer be concerned with the arrangement of more or less permanent objects but with the irrigation of territories with potential [...] it will no longer be obsessed with the city but with the manipulation of infrastructure for endless intensifications and diversifications, shortcuts and redistributions – the reinvention of psychological space." (Koolhaas 1995, 31)

Computational models driven by data are a powerful way to incorporate uncertainty, irrigate potentials, and capture subjective and invisible qualities generally associated with psychological space. Data allows us to model the highly dynamic nature of cities, their social life, and their infrastructure networks at an unprecedented level of detail.

Privacy and Surveillance

You imagine, as does everybody else for that matter, that our organization has for many years been preparing the greatest document center ever conceived, an archive that will bring together and catalogue everything that is known about every person, animal and thing.
Italo Calvino (1995)

•

Privacy and digital surveillance remain central concerns that are tightly connected to the technical nature of digital systems. Often, privacy issues arise as unintended consequences of these technical properties. In 2010 it was revealed that the two major mobile operating systems provided by Apple and Google store and collect the location information of its users. The immediate purpose was not surveillance or targeted marketing: both operating systems relied heavily on information about known Wi-Fi spots and cell towers for establishing the user's location. Since the location of private Wi-Fi spots is hard to come by and changes frequently, both systems relied on users' phones to collect this information through an invisible service, by automatically mapping every hotspot users encountered. Privacy concerns arise not from this original purpose, but from its consequence – the existence of a vast, dynamic database containing detailed information about each user's behavior.

In the early days of digital media, privacy was mainly discussed in terms of data generation – who is allowed to collect data, and who should have access? In the recent decade, this discussion has shifted toward a discussion of control. This means that the user should own and be able to control any data that concern her life, including being able to trade that data in exchange for money or services. In the simplest form, this control can be implemented via opt-in and opt-out mechanisms. Research on digital traces also leads to insights about how privacy can be protected in the age of big data – what works, what doesn't work, and how mechanisms for privacy protection can be improved.

The question of personal privacy is intimately linked to questions of government transparency. Protection of privacy requires mechanisms for controlling that those rules are actually followed by companies and governments, which can be addressed by a rigorous opening of government data sources. Full transparency for the government, full privacy for the citizen is a frequent demand. In reality, however, those two realms cannot be neatly separated, as private citizens and governments interact in many different ways.

However, none of the approaches sketched above are effective against the "deep state," the domain of government secrecy. The two faces of big data, the civic and the threatening, are palpable in current US legislation. The country has both one of the oldest and most highly developed implementations of a Freedom of Information law, which provides mechanisms for the mandatory public access to government documents. At the same time, the government circumvents this law by entertaining a vast network of agencies operating under strict secrecy outside of public accountability mechanisms. Bringing these issues under democratic control through formal and informal measures requires a public discourse that is informed by a high level of data literacy, a differentiated knowledge of technology and digital data. We understand this book as a contribution to developing data literacy by giving insights into the nature and methods of data-driven technologies. Finally, there is also a historical and cultural dimension of personal privacy. Plain text is perhaps the most persistent and explicit digital expression of human speech and thought, requiring only minimal digital storage space. Without a natural expiry date, the question of data life cycles arises – should personal data, including embarrassing messages from one's youth, be deleted at some point, or made inaccessible? On a larger scale, how much culturally relevant information would be lost in such a case about our mostly digitally documented world?

As Calvino's story about the attempt to preserve the world's memory illustrates, the "total archive" would be an insurmountable restraint for all human agency, which in his story ultimately leads to death.

Nevertheless, on closer inspection, the totalizing idea of big data turns out to be a myth. The data distribution is highly unequal; data is least available precisely where we need it most, for example in the hinterlands beyond the digital divide. The contributions in this volume illustrate the immense value contained in the traces generated by our digital lives for better understanding our cities, our cultures, and our society. We consider the variety of public data sets available from governments, research institutions, infrastructures, and voluntary data provided by citizens a basis for civic discourse and ultimately an integral part of public space.

Batty, Michael. 2013. *The New Science of Cities.* Cambridge, MA: The MIT Press.

Börner, Katy, Soma Sanyal, and Alessandro Vespignani. 2007. "Network Science," *Annual Review of Information Science and Technology* 41, no. 1: 537–607.

Burke, Jeff, Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani Srivastava. 2006. "Participatory Sensing," in *World Sensor Web Workshop,* 1–5.

Calvino, Italo. 1995. "World Memory," in *Numbers in the Dark and Other Stories,* translated by Tim Parks. New York: Vintage, 135–136.

Castells, Manuel. 1996. *The Rise of the Network Society.* Chichester: Wiley-Blackwell.

Flood, Joe. 2010. *The Fires: How a Computer Formula, Big Ideas, and the Best of Intentions Burned Down New York City – and Determined the Future of Cities.* New York: Penguin.

Foth, Marcus, ed. 2009. *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City.* Hershey, PA: Information Science Reference.

Koolhaas, Rem. 1995. "Whatever Happened to Urbanism?" *Design Quarterly* no. 164: 28–31.

Kuznetsov, Stacey, and Eric Paulos. 2010. "Rise of the Expert Amateur: DIY Projects, Communities, and Cultures," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries,* edited by Ann Blandford et. al. New York: Association for Computing Machinery, 295–304.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. "Computational Social Science," *Science* 323, no. 5915: 721–723.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis," *Science* 343, no. 6176: 1203–1205.

Pickering, Andrew. 2010. *The Cybernetic Brain: Sketches of Another Future.* Chicago: University of Chicago Press.

Ryan, Brent D. 2012. *Design After Decline: How America Rebuilds Shrinking Cities.* Philadelphia: University of Pennsylvania Press.

Simon, Herbert A. 1996. *The Sciences of the Artificial,* 3rd ed. Cambridge, MA: The MIT Press.

United Nations. 2012. "World Urbanization Prospects: The 2011 Revision." ESA/P/WP/224. New York: United Nations Publications.

Data – Source and Collection

Fabien Girardin

Catching the World's Eyes

Introduction

Visitors to a city have many ways of leaving voluntary or involuntary electronic trails: prior to their visits tourists generate server log entries when they consult digital maps or travel websites; during their visit they leave traces on wireless networks whenever they use their mobile phones; and after their visit they may add online reviews and photos. Broadly speaking, we propose two types of footprint: active and passive, also referred to in the literature as *volunteered* and *contributed* locational information (Harvey 2013). Passive tracks are left through interaction with infrastructure, such as the mobile phone network, which produces entries in locational logs, while active prints come from the users themselves when they share locational data in photos, messages, and sensor measurements.

The World's Eyes project **(figure 1)** investigated the active prints that reveal how people travel and experience a city. Particularly, we used the Application Programming Interface (API) of the photo-sharing platform Flickr to access publically available photos. In February 2009 Flickr broke the hundred-million-georeferenced-photos count (over a total of 3 billion photos in the repository).

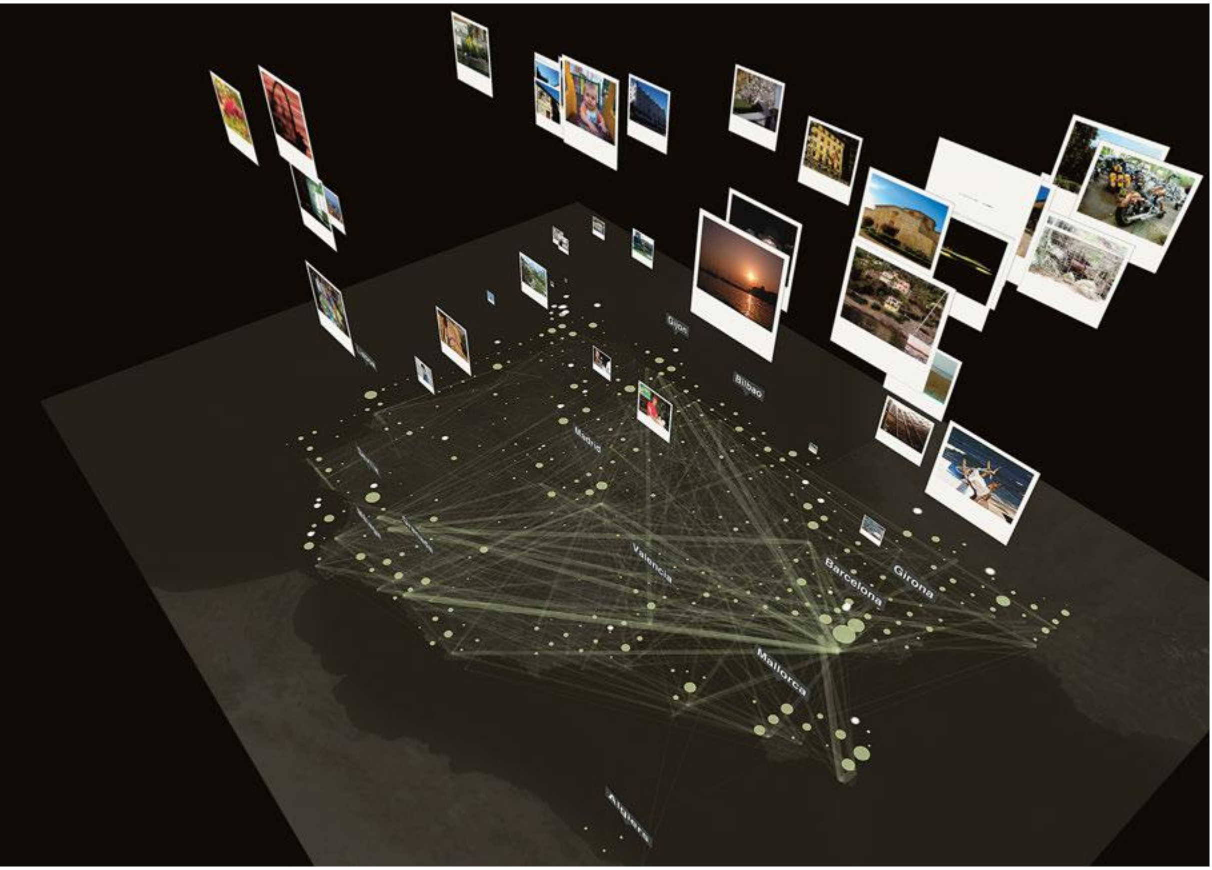


Fig. 1.
Screenshot of the World's Eyes project exhibited at the Design Museum in Barcelona in 2008. The visualization uncovers the evolutions of the presence and flows of tourists. As photos pile up to reflect the intensity of tourist activity, they uncover where tourists are, where they come from, and what they are interested in capturing and sharing from their visit.

This represents an unprecedented amount of publicly accessible data produced through people's interactions involving the Web and mobile devices. We caught these "eyes of the world" to investigate visitors' mobility in diverse locations such as the Province of Florence (Girardin et al. 2008a) and Rome (Girardin et al. 2008b) in Italy as well as New York (Girardin et al. 2009).

Each time a user anchors a photo to a physical location, Flickr assigns longitude and latitude values together with an accuracy attribute derived from the zoom level on a map. Unlike passive prints, we consider that user-generated content provides unique perspectives on mobility. Indeed, the effort of an individual to take a photo, select it, upload it onto a Web-sharing platform and geo-reference it can be more powerful than any survey or GPS log that researchers interested in human space-time activity could access in the past. There is a very real richness to the "intentional weight" that people attach to disclosing their photos, and the results clearly show that Flickr users have a tendency to point out the highlights of their visit to the city while skipping over the lowlights of their trip. This "I was here" brings a notion of subjectivity to the relation of people with space and place (Dourish 2006).

Our work suggests that exploiting this data set to know who visits different parts of the city at different times can lead to the provision of customized services (or advertising), the rescheduling of monuments' opening times, the reallocation of existing service infrastructure, or the evaluation of specific urban strategies. From a quite different perspective, tourists themselves could be aware of the current ways in which they populate the city, and adopt different strategies as a result.

From the Vision of Dynamic Maps of Human Processes to the Reality

The low cost and high availability of user-generated content now challenges any field that benefits from an in-depth understanding of large-group behavior. Indeed, only a few years ago, the possibility to produce fully dynamic time-space diagrams from the fusion of human activities data and novel forms of analysis was only discussed in the conditional. For instance, Zook et al. envisioned in 2004: "When many individual diagrams are aggregated to the level of cities and regions, these visualizations may provide geographers, for the first time, with truly dynamic maps of dynamic human processes. One might imagine them as twenty-first-century 'weather maps' of social processes."

The presence of active prints suggests that we are at the end of the ephemeral; in some ways we have new means to replay the city and its processes. This potential to replay the city echoes very well with the recent interest of local authorities and urban planners in big data. For instance, tourism is hardly quantifiable, because tourists leave minimal tangible traces of their stay. In the World's Eyes project, the analysis and mapping of this user-generated content allowed the measurement of the attractiveness of leisure cities and their points of interest. In contrast it also reveals the unphotographed regions still free from the tourist buzz. As photos pile up to reflect the intensity of the tourist activity, they uncover where tourists are, where they come from, and what they are interested in capturing and sharing from their visit.

In order to explore that domain, we followed several steps that start from the collection of digital footprints. We used the Flickr API to retrieve the coordinates of photos and their accuracy, the time at which they were taken, and we also obfuscated the identifiers of their owners. Since we were particularly interested in the behavior of tourists, our analysis platform separated the photographers into two groups, locals and visitors, based on their disclosed presence in the city over time. For the study of Rome over a 3-year period, we collected a data set of 144,501 georeferenced photos that had been uploaded by 6,019 different users. With the accumulation of these data we extracted spatiotemporal characteristics such as seasonality, usage patterns, and spatial distribution, main flows of visitors (i.e., desire lines) and the main points of interest of the city.

Presence in Space and Time

To map the spatial distribution of users, data is stored in a matrix covering the entire study area. Each cell in the matrix includes data about the number of photos taken, the number of photographers present. In Rome, the analysis of visiting quickly uncovers the area's major visitor attractions such as the Coliseum and the main train station next to the Piazza della Repubblica (Girardin et al. 2008b). In addition, temporal signatures provide further evidence to the different types of presence that occur at tourist points of interest. In Rome, it can be further hypothesized that the Coliseum attracts sightseeing activities (i.e., photographers) over the weekend and that the neighborhood of the train station provides facilities for visitors on the move (e.g., people on business trips) during the weekdays.

Desire Lines

The study of digital footprints also enables us to uncover the digital desire lines embodied in people's paths through their visit of a city or a region. Based on the time stamp and location of photos, our analysis platform organized the images chronologically in order to reconstruct the movement of the photographers. More precisely, we start by revealing the most active areas obtained by spatial clustering of the data. Next, we aggregate these individual paths to generate desire lines that capture the sequential preferences of visitors. The location of each user activity (i.e., photo) is checked to see if it is contained in a cluster, and in the case of a match, the point is added to the trace generated by the owner of the photo. This produces multiple directed graphs that support better quantitative analysis, enabling us to obtain the number of sites visited by season, the most visited and photographed points of interests, as well as where photographers started and ended their journeys.

Places of Interest

Previous work has demonstrated that spatially and temporally annotated material available on the Web can be used to detect place- and event-related semantic information (Rattenbury et al. 2007). In a similar vein, analysis of the tags associated with the user-originating photos revealed clues about people's perceptions of their environment and the semantics of their perspective of urban space. For instance, the word *ruins* is one of the most-used tags to describe photos in Rome. Mapping the distribution of this tag for 2,866 photos uncovers the most ancient and "decayed" part of the city of Rome: the Coliseum and the Forum. We used this semantic information to define the main areas of photographic activities as part of an economic impact study of *The New York City Waterfalls* in 2008.

Case Study: Measuring the Impact of an Event

In a case study that took place in summer 2008 around *The New York City Waterfalls* public art project, we further explored the characteristics of explicit digital footprints to define indicators that measure the evolution of urban attractiveness. The objective for the local authorities was to compare the evolution of the attractiveness and popularity of urban places at the different vantage points of the project. Therefore, we measured the spatial distribution of locals and visitors and compared the evolution of the presence of digital footprints as evidence of the positive impact of *The New York City Waterfalls* on the attractiveness of the waterfront. Eventually, two main results enhanced the City's report on the event: the evolution of attractiveness based on the presence of photographers and the evolution of popularity based on centrality.

Evolution of Attractiveness Based on the Presence of Photographers

According to the relative presence of photographers, our investigation analyzed the variations of attractiveness indicator based on the presence of photographers during the summers of 2006, 2007, and 2008 (the year of the *Waterfalls*). It reveals a positive growth in the waterfront's attractiveness of 8.2 percent in summer 2007 and 20.7 percent in summer 2008 with respect to that of other areas of interest in New York City, such as Time Square and Central Park, providing an indication of the potential impact of the presence of the *Waterfalls* exhibition.

Evolution of Popularity Based on Centrality

The centrality of an area of interest determines its level of integration into the popular flows of photographers. Our PlaceRank indicator revealed that between 2006 and 2007 the vantage points lost their centrality by 15 percent while the other areas of interest increased their centrality by 10 percent. However, between 2007 and 2008 the vantage points gained 56 percent while the other areas of interest lost 30 percent. In 2008 the vantage points appear as central as other areas of interest, meaning that they are on the tourist path as much as other areas of interest in that section of the city.

This case study provided indications that the emergence of digital footprints creates an opportunity to evaluate in detail the use of space, the impact of events, and the evolution of the built environment. This approach could not only better inform urban design decisions and city management, but also enable local authorities to provide timely evidence to the public about the use of space and about the impact of interventions within the urban fabric. Indeed, the integration of our results in the official study of the economic impact of *The New York City Waterfalls* public art project shows that the indicators proposed by our analysis offered useful measures to complement traditional methods.

Discussion

The ubiquitous technologies that afford us new flexibility in conducting our daily activities are simultaneously providing the means to study our activities in time and space. The exploitation of user-generated content to better understand mobility in the urban environment led to several implications we would like to highlight.

Technical Implications

Besides our work on urban-attractiveness indicators, other research groups have been using a reality mining approach to derive specific characteristics of urban dynamics (Kostakos et al. 2008; Ratti et al. 2006). A major challenge in this type of approach is to draw a clear understanding of the boundaries and biases of the data. For instance not to confuse behaviors with endorsement, which can be considered as a limitation of *The New York City Waterfalls* case study, as it used the density of digital footprints as indicators of urban attractiveness. Therefore, future studies will need to rely on calibrations with ground truth information produced with proven techniques.

Additionally, some analyses suggest distinct profiles of georeferencing and geotagging photos. These profiles might be based on culture or nationality, the type of tourist in terms of their length of stay or familiarity with the city, their level of technical expertise or spatial orientation ability, and the type of task or type of environment visited. Other questions that should be considered relate to the types of situations during which users are more or less likely to use their mobile devices for data generation. Answers to these types of questions should allow us to define better the meaning of the data and to explore further their potential usage in social sciences and urban studies.

Methodological Implications

The ability to replay the city shows that there are opportunities for researchers to propose novel ways to describe the urban environment. However, there is a big assumption in seeing the world as consisting of bits of data that can be processed into information that then will naturally yield some value to people. It would lead to what we would call data-driven urbanism, as if urbanism could be driven by data. Indeed, the understanding of a city goes beyond logging machine states and events. In consequence, let us not confuse the development of novel maps from previously uncollectable and inaccessible data with the possibility of producing "intelligent maps." Our work precisely draws some critical considerations on the current state-of-the-art. At this stage we are still trying to figure out: (1) What parts of reality does the data reveal? (2) What we can do with them? (3) How can we communicate them to people for acquiring information? (Still a far stretch from "intelligent.")

Taking this caution into account, the application of our research approach seems promising to gain knowledge on the presence and flows of humans at a specific space and with particular technologies, leading to an approach we would coin as "human/data-based urbanism." It could consist in the use of:

- Qualitative analysis to inform the quantitative queries: This approach first focuses on people and their practices, without the assumption that something computational or a data process is meant to fall out from that. This qualitative angle can then inform a quantitative analysis to generate more empirical evidence of a specific human behavior or pattern. A few approaches in that domain address this perspective. Williams et al. (2008) for instance argue that our understanding of the city could benefit from a situated analysis of individual experiences within cities, rather than taking particular urban forms as a starting point for the study of urban experience.
- Quantitative data mining to inform the qualitative enquiries: In this approach, quantitative data help to reveal emerging and abnormal behaviors, mainly raising questions. The qualitative angle then can help explain the phenomena in situ. The qualitative approach actually requests to ask the right questions to learn anything meaningful about a situation. An example of the latter could have been applied to the context of the impact of *The New York City Waterfalls*. We used digital footprints to reveal the variations in spatial presence and abnormal patterns of temporal presence over the course of a 3-year period. In addition to this quantitative analysis we could have performed qualitative observation on the detected areas to reveal how the attractiveness evolved (e.g., Did people stay longer?).

This fosters the need for research and practitioners to develop a coherent understanding of the traces of the activity: both qualitative (e.g., audio and video recordings of action and interviews) and quantitative (e.g., user-generated content). With significant data on the actual use of the space, we can perform new types of postoccupancy evaluations, often overlooked in the practice of urban design and architecture (Brand 1995). However, the tools, metrics, and interpretation methods are still, for the most part, to be developed.

Societal Implications

Ubiquitous geoinformation is both immensely empowering (for the people and places able to construct and consume it) and potentially overpowering as institutional and state forces are able to better harness information with growing personal and spatial specificity. In consequence there are ethical and privacy implications to grapple with. In conjunction with people's own representation of traceability, there is a legitimate concern about the implications of research on geographically anchored digital footprints as presented in the World's Eyes project. Particularly our work exemplifies the shift from a large-scale, top-down Big Brother threat to privacy to a more local, bottom-up little sister types of people monitoring, which makes the whole notion of opting out of technology adoption one of whether to opt out of society.

In fact, these digital footprints have become inevitable in contemporary society and also necessary if we wish to enjoy many modern conveniences; we can no more be separated from it than we could be separated from the physical shadow cast by our body on a sunny day (Zook et al. 2004). The growth of our data shadows is an ambiguous process, with varying levels of individual concern and the voluntary trading of privacy for convenience in many cases. In summary, at the same time as ubiquitous geoinformation gives us new means to map and model human dynamics, it will also challenge current notions of privacy. The challenge is to appreciate and use the complexity and richness of ubiquitous geoinformation without crystallizing into authoritarian structures.

Brand, Stewart. 1995. *How Buildings Learn: What Happens After They're Built.* New York: Penguin.
Dourish, Paul. 2006. "Re-space-ing Place: 'Place' and 'Space' Ten Years On," in CSCW '06: *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work.* New York: Association for Computing Machinery, 299–308.
Girardin, Fabien, Andrea Vaccari, Alexandre Gerber, Assaf Biderman, and Carlo Ratti. 2009. "Quantifying Urban Attractiveness from the Distribution and Density of Digital Footprints," *International Journal of Spatial Data Infrastructure Research* 4: 175–200.
Girardin, Fabien, Filippo Dal Fiore, Carlo Ratti, and Josep Blat. 2008a. "Leveraging Explicitly Disclosed Location Information to Understand Tourist Dynamics: A Case Study," *Journal of Location-Based Services* 2, no. 1 (March): 41–56.
Girardin, Fabien, Francesco Calabrese, Filippo Dal Fiore, Carlo Ratti, and Josep Blat. 2008b. "Digital Footprinting: Uncovering Tourists with User-Generated Content," *IEEE Pervasive Computing* 7, no. 4 (October–December): 36–43.
Kostakos, Vassilis, Tom Nicolai, Eiko Yoneki, Eamonn O'Neill, Holger Kenn, and Jon Crowcroft. 2008. "Understanding and Measuring the Urban Pervasive Infrastructure," *Personal and Ubiquitous Computing* 13, no. 5 (June): 355–364.
Rattenbury, Tye, Nathaniel Good, and Mor Naaman. 2007. "Towards Automatic Extraction of Event and Place Semantics from Flickr Tags," in *SIGIR '07: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* New York: Association for Computing Machines, 103–110.
Ratti, Carlo, Riccardo Maria Pulselli, Sarah Williams, and Dennis Frenchman. 2006. "Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis," *Environment and Planning B: Planning and Design* 33, no. 5: 727–748.
Williams, Amanda, Erica Robles, and Paul Dourish. 2008. "Urbane-ing the City: Examining and Refining the Assumptions Behind Urban Informatics," in *The Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City,* edited by Marcus Foth. Hershey, PA: Information Science Reference, 1–20.
Zook, Matthew, Martin Dodge, Yuko Aoyama, and Anthony Townsend. 2004. "New Digital Geographies: Information, Communication, and Place," in *Geography and Technology,* edited by Stanley D. Brunn, Susan L. Cutter, and J. W. Harrington. Dordrecht: Kluwer Academic Publishers, 155–176.

Michael Szell, Benedikt Groß

Hubcab – Exploring the Benefits of Shared Taxi Services

Using Big Data to Research Human Mobility

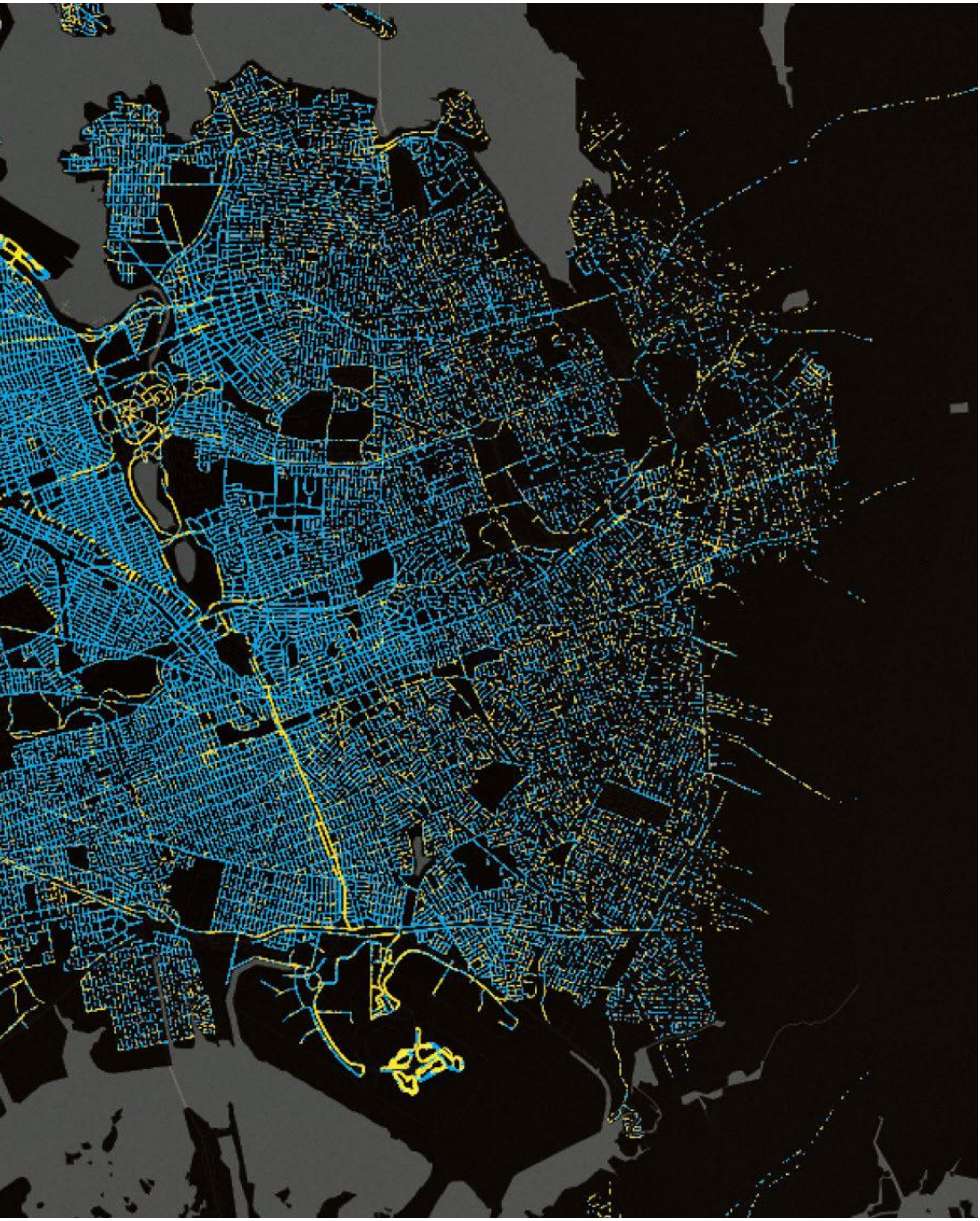
Human activity data has a huge potential for improving urban traffic systems. Data from human interactions such as phone calls, credit card transactions, or social networks, reveals regularities, high predictability (Song et al. 2002), and uncovers the hidden statistical "laws" behind our everyday behavior and movement patterns. The study of human mobility today is no longer limited to the social sciences; researchers from computer science, mathematics, or physics study our movement patterns as they would elementary particles of inanimate matter. Surprisingly, the aggregate and statistical behavior of highly complicated organisms such as humans can in many situations be understood and formulated in purely mechanistic terms (Ball 2003).

But how come all these terabytes of human mobility data now so easily available, and what are the consequences for designing transportation systems? This immediate availability is rooted in both technical and social advances: mobile phones have revolutionized the way in which we communicate: we are always available – and tracked. Telecommunication providers store the position of the nearest cell tower during every call we place or for every SMS we send. Global Positioning System (GPS) technology, equipped on every modern phone and mobile electronic device, allows an even more precise tracking of positions. Not only the paths of walking individuals can be tracked, but paths of individuals using any form of urban transportation. Installed on vehicles, trackers can paint powerful pictures of the flows of millions of vehicles within a city (figure 1) and increase our understanding of collective human behavior and the bottlenecks of our transportation systems.

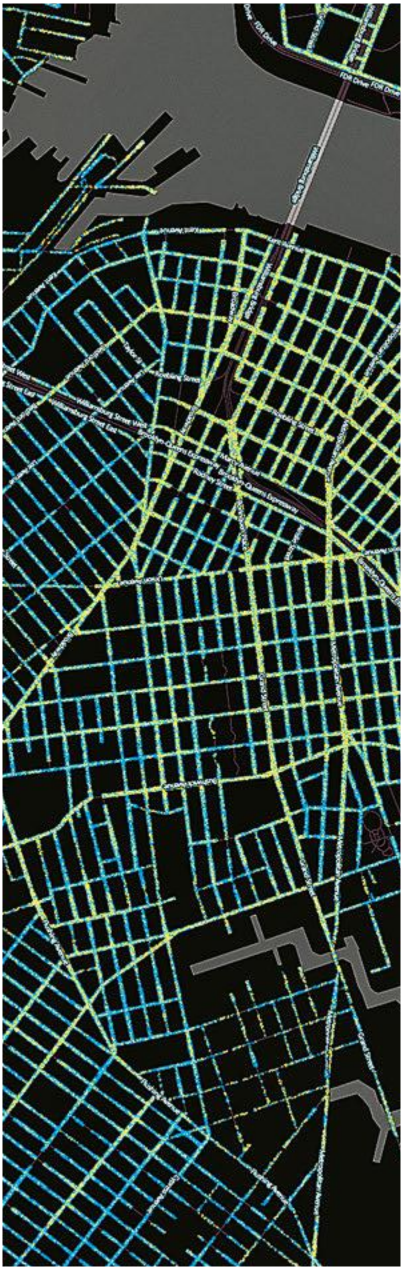
On the side of social developments, let us highlight the so-called open data or open government initiatives. Open data refers to the concept that certain data should be freely available to everyone to use as they wish, without restrictions from copyright, patents, or other mechanisms of control. Several administrative and governmental institutions have demonstrated in a number of recent cases that making data sets public can be a smart move: examples include open data from the public transportation system of cities, which have spurred the independent development of smartphone apps for live bus and train schedules; the Open311 initiative allowing citizens to more directly interact with their cities;[1] or various repositories aimed at use for scientists in various fields as listed in the Open Access Directory.[2] In a similar spirit, OpenStreetMap is a collaborative project mapping and publishing cartographic data under an open license.[3] In the United States, the modern idea of open government goes back to the year 1966, when a federal law known as the Freedom of Information Act (FOIA) was signed by President Lyndon B. Johnson, following ten years of congressional hearings championed by Congressman John E. Moss to gain access to deliberations of the executive branch under the Eisenhower administration (Blanton 2002). In his struggles, Moss was backed by high-profile journalists seeking to eliminate the bureaucratic hurdles and governmental secrecy of the time – as Moss remarked poignantly, “You had a hell of a time getting any information” (Kennedy 1978). The FOIA allows for the disclosure of previously unreleased documents controlled by the United States government and defines mandatory disclosure procedures. While it has been amended many times since, the main principle is still in effect today, allowing studies such as ours. Watergate and other far-reaching scandals have ensured the continued existence of FOIA and remain a catalyst for open information movements worldwide (Blanton 2002).

1 http://www.open311.org
2 http://oad.simmons.edu/oadwiki/Data_repositories
3 http://www.openstreetmap.org

Fig. 1.
Taxi traces in New York City, screenshot from Hubcab. The streets of New York City are partitioned into 40 m segments, all taxi activity of the city over the full year 2011 is visualized.
A segment is shown in yellow if it contains more pickups, in blue if it contains more drop-offs.



The thickness of street segments is proportional to the taxi activity; almost all roads are visited by taxis. Arterial roads tend to emerge as yellow while minor roads are blue, painting a beautiful picture of a complex urban system.



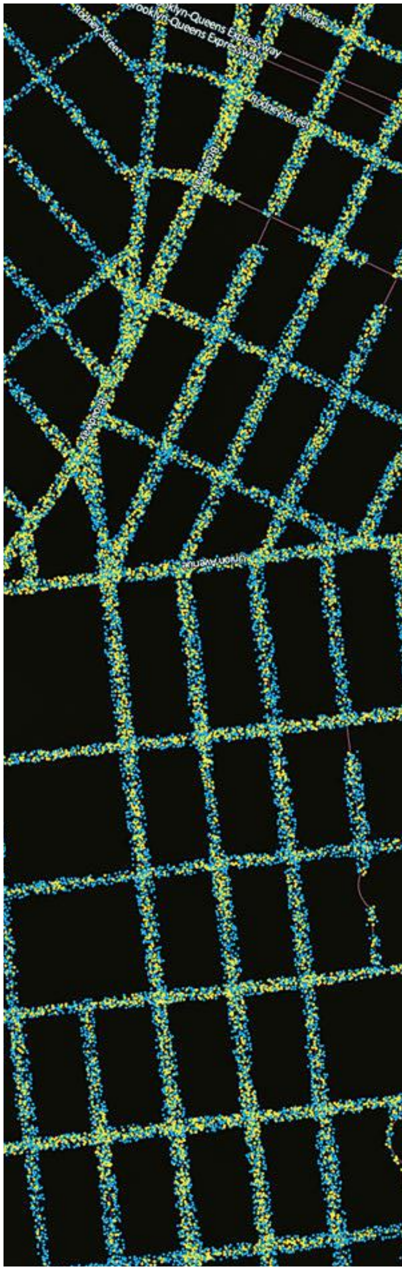


Fig. 2.
Different zoom levels of the Hubcab tool. The high spatial granularity of the data allows deep zoom levels, showing single taxi drop-offs and pickups in New York City on the street level.

Besides such desirable systematic initiatives for more available data in the public interest, data sources for researchers are still often available only through private or unplanned disclosures. For example, mobile phone information is usually proprietary data owned by telephone companies – research teams to persuade companies that data of public interest be disclosed for research purposes. A prime example for an unplanned disclosure is the Enron Corpus, a large database of over 600,000 e-mails written by 158 employees of the Enron Corporation, made public after the investigation of the company's scandal-ridden collapse in 2003.[4] Coordinated social network analysis of the corpus had uncovered previously undocumented corporate practices and gave insight into the nature of group dynamics in large organizations in general. In the context of mobility, the case of Apple iPhones' tracking and storing easily accessible logs of their owners' movements over several months, discovered in April 2011,[5] serves as a prime example for an unplanned disclosure of human movement data. The discovery of these logs led to initiatives such as crowdflow.net to create an open database of Wi-Fi and cell networks.[6]

4 John Markoff, "Armies of Expensive Lawyers, Replaced by Cheaper Software," *New York Times,* March 5, 2011, A1 (http://www.nytimes.com/2011/03/05/science/05legal.html).

5 Alasdair Allan, "Got an iPhone or 3G iPad? Apple Is Recording Your Moves," *O'Reilly,* April 20, 2011 (http://radar.oreilly.com/2011/04/apple-location-tracking.html).

6 Apple updated its operating systems a few weeks later, ending this possibility.

Improving Urban Taxi Systems

With so much data available, how can we use it to improve life in cities? We are interested in the workings of urban transportation systems. The rapid increase of urbanization and traffic volume in cities makes the smooth functioning of city-based transportation services more vital than ever before. The effect of congestion is a major constraint in urban environments and is felt significantly both economically and environmentally, with emissions impacting regional air quality.

As one of the major modes of urban transportation, let us here focus on urban taxi services. Despite its importance in urban transport, the taxi industry has drawn limited academic interest and has successfully resisted systemic changes and improvements in the past. For example, several New York City mayors had vowed to improve the system over the course of a number of decades but the industry remained essentially the same since the 1900s (Li 2006). However, it is of crucial interest to improve the efficiency of taxi dispatch systems for taxi companies and their clients, as well as urban planners – for example, to decrease fuel consumption using dynamic allocation processes, to improve service quality, and to improve the sustainability of urban traffic.

In a recent research project we studied and made use of a data set containing the positions of all 13,500 taxicabs of New York City (NYC) recorded during all 170 million taxi trips over the year 2011. The taxi business in NYC is regulated by the NYC Taxi and Limousine Commission (TLC) which issues strictly limited permissions (called "medallions") to cabs, granting the familiar yellow vehicles the exclusive right to pick up passengers in response to a street hail. Starting in 2008, all NYC taxicabs have been mandated by the TLC to be equipped with a so-called Taxi Technology System (TTS) including a GPS tracker with a live map of the vehicle's location shown to the passenger on a screen on the backseat. The GPS data collected by the trackers is submitted directly to the TLC. Installation of the GPS trackers was not without controversy – occasional strikes and lawsuits from taxi drivers against the TLC followed and have been ongoing for years, describing the installations in one class action complaint as "unconstitutional, warrantless invasions of their privacy," after a number of medallion licenses were revoked by the TLC for the alleged overcharging of passengers on the basis of the collected data.[7] Regardless, the TLC succeeded in its mandate and complete, anonymized data sets of cab positions can now be freely requested by anyone from the TLC via FOIA request.

[7] See Case 1:12-cv-00784-LAK, *Aka and Carniol v. Yassky et al.*

NYC is not the only city where positional data of its taxi fleet is available. In the Singapore-MIT Alliance for Research and Technology (SMART) initiative, the largest taxi provider of the city-state of Singapore has granted selected partners access to data collected from 16,000 out of the total 26,000 Singaporean cabs, for studies aimed to improve quality of life within the city. Similar studies exist with data from Shanghai, San Francisco, or Vienna, and more are expected in the coming years. Due to the high-grained resolution and completeness over one full year, we focused first on data from the taxi fleet of NYC; however, our results shall apply without loss of generality to urban taxi fleets in any arbitrary major urban zone.

Analyzing data from NYC immediately reveals that the taxi system as a whole is highly inefficient, as found in other cities as well. Inefficiency can be quantified by the total length or time of empty (i.e. no passengers) trips. Empty trip lengths typically display bimodal distributions: The first smaller peak corresponds to the expected average empty trip; the second peak can be found around the distance of city center to airport. It is these special zones of airports and city centers which play a major part in causing the inequilibria between supply and demand. Taxi drivers tend to accumulate at zones where pickups can almost always be expected, such as major roads or other high-throughput locations like airports. However, due to slow propagation of information, variations in actual pickups can become highly increased – especially at airports, queues of empty taxis can get very long at certain times when many drivers get the same idea and go there in unison. This inefficiency leads to adverse effects such as higher-than-necessary emissions, congestion, bottlenecks, etc. The consequences might lead to a reconsideration in urban planning or a strategically superior placement of these areas in future cities.

Several improvements of taxi systems have been attempted in the past, most often in a self-organized manner. Ride-sharing of taxis is a well-known phenomenon in many Asian cities, where passengers split the costs for giving up some comfort and travel time. More recently, commercial attempts at intermodal transport, i.e., the combination of different transportation options, including taxis, as in Daimler's Moovel,[8] or car-sharing services have been introduced, however often suffering from acceptance or similar efficiency problems. Further, a number of smartphone apps have been developed to easily sense taxis, hail taxis, and pay for taxi trips with the tap of a button. These apps are now being employed in cities worldwide with varying success. Urban taxi systems as a whole do not seem to have shown substantial changes, however.

[8] http://www.moovel.com

It is due to this rigidity and inefficiency that we propose a new system, which may replace old ones or work in parallel. For this purpose we use the available data and mathematical modeling, designing the new system to be more efficient, causing fewer emissions, and being more affordable to customers than existing ones. The trips in the NYC data set cover over 99 percent of all the city's streets (except for parts such as Staten Island which are not serviced by taxis). Each street segment has a unique set of destination points and origin points (street segments were taxis are going to or coming from). In particular we noticed that a large number of trips could potentially be combined and therefore be saved. In dense areas of Manhattan many pairs of street intersections exist where up to 2 million trips are started and ended in the immediate vicinity of the start and the end point (over one year). A large number of these trips start and end at similar times, making them redundant if the passengers were willing to share a cab. Therefore we focused our approach on developing a new dispatch algorithm with the basic idea of combining trips.

This new process aims to minimize the total cost of the service while satisfying certain comfort criteria, e.g., customers should not have to wait for an additional percentage of time than before. In the simplest case a taxi has a capacity of two passengers. When two passengers are found at the same starting point wishing to be delivered to the same end point, both at the same time, the taxi is able to pick up and deliver both. Some assumptions may of course be relaxed, i.e., the start and end points do not necessarily have to match exactly, or may just lie along the route of the longer trip. In any case, we first show via simulations that the current taxi system may be optimized to solve the demand more efficiently with a cab capacity of two. It is then natural to expand the cab capacity to arbitrary larger values. If models, simulations, and analytical investigations show that a larger capacity is even more efficient, practical implementation of large-capacity cabs may resemble "taxi limousines," where passengers have their own private compartment inside a larger vehicle, similar to buses but with individual, dynamic routes.

Mathematically, our approach defines a trip shareability network where nodes are trips, and links between nodes mean that two trips can potentially be combined following some rules and limitations. When more than two trips are to be shared, the network becomes a much more complex entity with links of higher dimension. The algorithm then solves the so-called maximum matching problem, which leads to an optimal solution for the trip combination task. Exact solutions are computationally not feasible for large graphs, however algorithms exist which are able to approximate optimal solutions in reasonable run time. In this process we define a prolongation parameter that stands for the amount of time up to which a passenger is willing to prolong her trip. If this time is chosen low, quality of service is higher since the passenger has to wait for shorter time, but fewer trips can be combined. The situation is reversed for long prolongation times.

The aim of the study is to assess the potential for shareability without substantially lowering the quality of service. Preliminary results on the set of trips within Manhattan are encouraging: For the case of two trips sharing, if passengers are willing to prolong their trips up to five minutes, more than 90 percent of trips can be shared, reducing total travel distance – thus pollution generated – by more than 40 percent. In the case of sharing three or more trips, the potential for shareability seems even larger. The trade-off of an up-to-five-minutes-longer trip for a halved cost seems a reasonable offer. Note that the trip-sharing algorithm appears straightforward to implement practically in an online dispatch system. After hailing a taxi, a user can receive feedback on her shareability options almost instantly.

While the technical details of a shareable urban taxi system seem feasible to solve algorithmically, achievement of the potentially major political task of convincing both authorities and consumers of its viability remains an open question. The overhead of extending existing technologies could be perceived as high, and incentives for overcoming fears of contacts between alien passengers may have to be created (apart from split costs), such as physically separated, private compartments within each vehicle.

Hubcab – The Tool

Parallel to our research efforts we have developed a complementary online tool, named Hubcab, to give a larger audience the possibility to explore the studied phenomena in an easy and fun-to-use way, available at www.hubcab.org. The tool visualizes the same data set of NYC taxi trips, displaying all 170 million taxi trips and allowing interactive exploration of the city from a fresh perspective, demonstrating the potential of a smart dispatch algorithm to a worldwide audience.

Particular attention was paid to represent the amount and direction of the trips in an intuitive way. Drop-offs and pickups are represented coherently and in a dynamic, hierarchically structured way, revealing the maximum information of single drop-off and pickup points in higher zoom levels. This approach has the advantage that the user can see both the direction and an approximation of the amount of the trips at the same time. For cartographic data, we used maps from openstreetmap.org. A script was developed to subdivide NYC's vast road network into more than 200,000 segments of 40 meters' length to provide a high-resolution experience. The vast majority of these segments contains at least one pickup or drop-off, some up to thousands (over a full year). Hubcab allows users to zoom into the map of drop-offs and pickups, showing single points of pickups and drop-offs in zoom levels of unprecedented magnitude (figure 2). The user can switch between different time slices, showing the full data or data segmented in time highlighting different hot spots of pickup or drop-off zones during different times of the day. The radius of this vicinity can be changed on the fly. All the trips between pickup and drop-off points are selected dynamically; Hubcab allows users to explore and visualize the flows between 40 billion (200,000 × 200,000) pairs of street segments. The Hubcab tool provides a unique insight into the inner workings of the city from the previously invisible perspective of the taxi system with a never-before-seen granularity. Hubcab allows one to investigate exactly how and when taxis pick up or drop off individuals and to identify zones of condensed pickup and drop-off activities. Most important, the visualization of flows between any two points of the city allows citizens to experience the redundancy of a large number of trips and the vast potential for improvement. Hubcab expands and changes the perception of urban space, and potentially the behavior of its inhabitants and visitors. Societal and political implications are apparent; urban planning may significantly change through the use of such tools. The sensed, visualized data can be utilized to better design cities, prototyping new urban futures superior on a systemic level. In our particular case this improvement is expected to lead to less congestion in road traffic, lower running costs, and a less polluted, cleaner environment.

In our approach we have been using methods involving optimization algorithms on graphs. Besides this logistical optimization approach, what other types of solutions to the problem would be possible? Previous research using mobile phone data has shown that the specific space syntax can constitute an important influence on the geography of human activity (Reades et al. 2009), and therefore on the functioning of the transportation system. It might be viable to consider rearranging urban structures for shifting land-use patterns, with the goal of improving urban transportation systems in an integrated approach. In any case, an efficient, systematic improvement will be accomplishable only after measuring the status quo and using rigorous analysis on the obtained data sets.

The authors thank research collaborators Paolo Santi and Giovanni Resta (Institute of Informatics and Telematics of CNR, Pisa, Italy), Steven Strogatz (Dept. of Mathematics, Cornell University, Ithaca, NY), Stanislav Sobolevsky and Carlo Ratti (SENSEable City Lab), the Hubcab Web developers (47nord), as well as the National Science Foundation, the AT&T Foundation, the MIT SMART program, the MIT CCES program, Audi Volkswagen, BBVA, Ericsson, Ferrovial, GE, and all the members of the MIT SENSEable City Lab Consortium for supporting the research.

Ball, Philip. 2003. "The Physical Modelling of Human Social Systems," *ComPlexUs* 1, no. 4: 190–206.

Blanton, Thomas. 2002. "The World's Right to Know," *Foreign Policy* 131 (July/August): 50–58.

Kennedy, George. 1978. "Advocates of Openness: The Freedom of Information Movement." Ph.D. diss., University of Missouri, Columbia.

Li, Sonny Heng. 2006. "Multi-Attribute Taxi Logistics Optimization." Ph.D. diss., Massachusetts Institute of Technology.

Reades, Jonathan, Francesco Calabrese, and Carlo Ratti. 2009. "Eigenplaces: Analysing Cities Using the Space-Time Structure of the Mobile Phone Network," *Environment and Planning B: Planning and Design* 36, no. 5: 824–836.

Song, Chaoming, Qu Zehui, Nicholas Blumm, and Albert-László Barabási. 2010. "Limits of Predictability in Human Mobility," *Science* 327, no. 5968: 1018–1021.

Anthony Vanky

Data Availability / Data Relevance: Evaluating Real-Time Urban Information Usage in Singapore

Real-time urban technologies have flourished.[1] Traffic and transportation information, environmental quality measures, even information on the emotions of cities (mined from social media): all these can be monitored in real time (Golder and Macy 2011). Traditionally used for system optimization and management, these systems are increasingly accessible to the general public. Mobile devices and personal technologies can inform users of the current conditions of the infrastructures and resources with which they interact on a daily basis.

The availability of this data via new personal technologies allows individuals to change behaviors in response to the current situation of their environment. A person might leave home at a certain time to beat traffic, timing departure according to real-time information rather than routine or habit, for example. Or an individual might remain at work just a bit longer to avoid rain during their bus commute, drawing information from real-time meteorological stations.

While the breadth and volume of information is growing, there are few public measures of the relevance of this data, how it alters human interaction with urban infrastructure. We note the

[1] In this chapter, *real-time urban technology* and *real-time technology* are defined as situated technologies that provide dynamic data and monitoring in real time or near real time. Examples include traffic congestion reports, GPS-enabled public transportation arrival times, and air-quality monitoring. Data related to open government is not included in this study as it is not made available in real time.

number of users, but neglect the quality and manner of use. This research is aimed at describing those measures: whether there is use and appreciation of real-time urban data, and how the information affects the decision-making processes of citizens at an individual level. Here, we take a certain population in Singapore, the subject of this case study, for our exemplar.

To borrow a term from engineering, this research is concerned with the actuation of the data. Technological measures of real-time data usage, such as records of API requests or the amount of data transferred, have already been established. This study looked beyond the movement of data, and endeavored instead to examine the end user: focusing on the perceptions and actions, if any, that resulted from use of the data. Further, while some studies of real-time information focused on isolating realms of technology, such as transportation alone, there is still an opportunity to understand its implications broadly, both from the direction of access, and from the ultimate use of that data for action.

For many planners, these technological services are the means to empower the individual citizen to use and move through the city more smoothly. People can depend on concrete information, rather than intuition or memory to enhance their interaction with the urban environment. In addition, they can contribute feedback on the performance of the services they use. More and more cities are equipping their public transit with real-time data feeds. Energy companies are installing so-called smart meters with mechanisms for customers to monitor their energy use in real time. The sources of technological innovation are creating more products to sell to cities for their own monitoring and evaluation purposes. While the use of these technologies is unprecedentedly widespread, current analysis fails to measure the end-user experience. Studies rely on remote measures, interpreting general use by the number of API calls onto the network by software applications, or number of reports submitted, rather than the experience of the end user, the citizen.

The assumption often made with regard to this technology follows the mantra of "If you build it, they will come"– the more opportunities for open and real-time data, the more citizen users (Williams et al. 2008). The growth of these technologies is indisputable.

It is expected by decade's end, more than $ 108 billion will have been invested in the implementation of smart urban technologies (Navigant Research 2010). However, without an understanding of the user's experience, we cannot understand how and where these tools have leverage, and whether the investments have actual impact on the everyday lives of citizens. These measures can better inform the process by which data may empower individuals to live more efficiently, by understanding how much they want to be.

Singapore as Case

Singapore is a unique, potentially prophetic case. Cultural and economic factors made the nation well suited for this study in particular. It was possible to test how individuals used and appreciated available real-time data.

The country's focus on technology has roots in Singapore's historical role in trade and commerce with the rest of the world. The city-state became a fully independent republic in 1965 upon its separation from the newly founded Malaysia, formed by the union of Malaya, Sabah, Sarawak, and Singapore in 1963. Independence was followed by the departure of the British military and the largest employers from the island, and the movement toward a heavy reliance on foreign direct investment and multinational corporations seeking cheap land and cheap labor for economic growth. This lead to the emergence of Singapore as one of the four Asian economic "tigers." In the latter half of the 1980s, Singapore's economy began a shift away from cheap manufacturing toward a strong service sector in addition to "value-added" advanced manufacturing, finance, and trade. To facilitate this shift, economic planners realized the importance of information technologies as a strategic thrust for development.

The government established the National Computer Board (NCB) in 1981 "to drive Singapore to excel in the information age by exploiting IT extensively to enhance our economic competitiveness and quality of life" (Mahizhnan 1999). The NCB published the first National IT Plan in 1986, which aimed to develop public and private sector IT capabilities, all the while enhancing the country's overall economic competitiveness. The subsequent plan, *A Vision of an Intelligent Island: IT2000 Report,* released in 1992, examined how the results of a more pervasive and connected IT could make Singapore a major economic hub of the world: "In our vision, some 15 years from now, Singapore, the Intelligent Island, will be among the first countries in the world with an advanced nationwide information infrastructure. It will interconnect computers in virtually every home, office, school, and factory." (National Computer Board, Singapore 1992)

More than simple export-oriented economic plans, the ambitions were seen as infrastructure, an integral part of the holistic development of Singapore. The government has estimated public sector spending on urban information and communication technology (ICT) infrastructure will reach US $1.2 billion in 2013, up 4.8 percent from five years before, with an additional US $2.6 billion as part of a larger technology master plan (Infocomm Development Agency 2005). Concomitant to the development of Singapore's economic growth, however, were the increasing pressures on its urban infrastructure. Quite literally, Singapore could not easily grow. Much of the technological investment as a result went toward real-time urban infrastructure for the management and operation of the island's urban infrastructure. Together, the milieu of economic and urban conditions led Singapore to becoming a "smart city."

While the government has made a strong commitment to technology and the availability of data, there has not yet been an evaluation on the adoption and perceptions by the public that is targeted. Singapore offers a good environment to test the potential of real-time information; this is a technologically savvy population, where the concerns of the availability of technology for adoption are muted as a result. As such, it provides a unique location for one to evaluate the perceptions of real-time information as the traditional barriers to adoption and use are of less concern.

Already, the number of mobile phones owned by the survey respondents exceeded the number of individuals responding. This figure is not unsurprising for the island, as the mobile phone subscription rates are 145.24 per 100 individuals,[2] and nearly 7 in 10 had access to the Internet (Singapore Department of Statistics 2012). This research also found that many were familiar with real-time information. This research found that when individuals were asked

[2] The United States, by comparison, has 90.78 mobile subscriptions per 100 and Hong Kong 179.39 per 100.

to evaluate their own familiarity of real-time information and real-time data, 59.6 percent noted some familiarity with the concept. A country in some ways ahead of the times, Singapore may prove representative of what is to come across the globe, as smart urban technologies continue to proliferate.

Methodology

An initial survey tested perceptions concerning data and the reasons behind its use. From the larger Singaporean population, government employees were selected for study based on their familiarity with technology and their access to various types of technology through their daily professional work. A total of 3,221 employees were solicited to respond to the survey, representing about 20 percent of the Singaporean government. A qualitative analysis was also performed to provide a contextual understanding of real-time data usage at the time of decision making. This component of the project focused on observing how individuals used and sought information at several transit stations across the island. This research provided insight into how the responses from the mobility section of the survey translated into actions.

On the Move

Real-time data for mobility is perhaps the domain where one can find the most opportunities for access. In this regard, a majority of respondents "use real-time information to inform [their] mobility and transportation plans"; more than two-thirds have sought real-time transportation information. Individuals are currently seeking information regarding their daily transportation plans – both public transportation and road and traffic conditions. The reasons for which individuals seek data and change modalities fall into three general categories: (1) money; (2) time; and (3) ease. For most people, the primary motivation to change travel plans is to save time. Other reasons frequently given by subjects were to make the trip easier to manage, to optimize their tasks and errands, and to make the whole trip feel "smoother." Saving money as a reason to change plans came fifth on the list. We might speculate that this is because of the relative low cost of the transportation system in Singapore. These findings were also reinforced by the respondents' written comments; the majority of participants stated that they use real-time information to change their commute to save time or to reach their destination faster.

Item	Before Data			With Data			Δ, in x̄
	x̄	n	SD	x̄	n	SD	
On Foot	3	54	0.70	3.7	45	0.81	0.7
Public Bus	2.4	91	0.78	3.2	75	0.93	0.8
Public Rail (MRT/LRT)	2.7	106	0.70	2.85	89	1.02	0.2
Taxi	2.8	47	0.77	3.64	39	0.73	0.8
Private Car	3.2	66	0.83	4.04	56	0.69	0.8

Note: 1=Not pleasurable, 2=Somewhat not pleasurable, 3=Neutral, 4=Somewhat pleasurable, 5=Pleasurable

Table 1.
Comparing between users of real-time data, and those who have not: How pleasurable is your average trip?

To evaluate a qualitative response to real-time transportation information, individuals were asked to rate how pleasurable their commutes were before and after their use of real-time information across transportation modalities. If an individual did not use real-time data, they were asked to speculate their emotional response to gauge the level of expectation. As seen in prior literature, individuals responded that the use of the data made their commutes more pleasurable, although the perceived improvement was slightest with the public rail (MRT and LRT) systems. This is likely due to the regularity with which trains arrive, often at an interval of less than five minutes. In the case of pedestrian traffic, the expected improvement was more than 100 percent greater than the actual perceived shift. This imbalance of expectations could lead to distrust of the open-data initiatives, even though they operate as planned. The one notable exception is regarding the use of public bus information; those who did not currently use the data expected less improvement than seen with users.

While there is an effect of real-time information on how pleasurable the commute is, which has been seen in prior research as noted above, we find an interesting phenomenon that individuals do not often change their routines all the time as a result of the insight gained by the data. One can speculate that individuals are enjoying their wait more. When asked if the real-time information gathering has resulted in a "[purposeful change in] the time you spent waiting for your mode of transportation," 57.5 percent responded "sometimes" and 16.5 percent responded "never," while 19.7 percent responded "often" and only 6.3 percent regularly changed their wait time. Similarly, individuals infrequently changed their choice in travel mode (e.g., switching from the MRT to a bus or to a taxi). A majority said they sometimes change modes (65.4 percent), while 21.3 percent said they never do. Of the remaining, 11.8 percent said they often change modes, while only 1.6 percent would often change.

These findings, however, seem to contradict the earlier finding of people wanting to reduce commute and wait times. As for the question of mode shifts, one reason for this may be due to the population's general reliance on the bus system over other modes, and the duplication of buses serving certain routes in the central area of the island. For instance, the People's Park Complex bus stop is serviced by eight buses that run along Eu Tong Sen Road, which means that it is convenient to stick with bus services, even if current information informs commuters of a delay in regular service.

Conveying the Message

With regard to transportation, much literature highlights the mixed success of at-stop real-time information signage in affecting the perceived reliability and wait times for public transportation (Al-Deek et al. 1988; Boyce 1988; Dziekan and Kottenhoff 2007; Dziekan and Vermeulen 2006; Schweiger 2003). But with the costs involved with installing and implementing various mechanisms for access, the issue of where people access the information is salient when trying to make the data as useful as possible for the general population. The difficulty in creating access plans is the variety of ways in which people use the data in their decision-making process.

Literature noted the importance of where information is seen in changing behavior related to energy consumption; it has to be regularly seen in order to change our autonomic actions related to resource usage. That is to say, the importance of both location and access to information are relative to when we make decisions. For transportation data, the challenge is that a large percentage noted that their data use came only after their plans had been made; the data served to help in deciding mode shifts or smaller-wait-time changes. Individuals during the qualitative analysis said the same. The data came later in the decision-making process.

Fig. 1.
At Harbourfront, both at the MRT station and the bus stop, individuals can be found using their mobile devices as means of distraction, but also information gathering.

The importance of smartphones and tablet computers in this society should not be underestimated. In both the survey and the observations, individuals have a clear relationship with their devices (figure 1). At any bus stop, or on the MRT, a significant number of individuals can be found using their devices to chat with friends, play games, surf the Internet, etc. Ito et al. (2005) recognized in Japan a similar culture, where digital devices increasingly mediate social life in public transportation. Specifically, device use is visually based within this environment; phone calls are increasingly infrequent and often considered rude.

Various agencies in the government are now sharing their data over APIs to allow the public to access the data and to develop tools at large. The one exception is the agencies related to public transportation. In May 2011 SBS Services, one of two public bus transportation providers, closed its API to developers for use in their applications (Yap 2011). This action rendered several independent apps and services unusable. What remained was SBS's own Iris application, with full functionality of real-time bus information. It does get a lot of use by the public as a tool for trip status and planning, but many individuals chided it as unintuitive, without relevance, and difficult to use. With the recognition that respondents overwhelmingly see their smartphones as their preferred access point for data, a larger discussion about the openness of data and its availability across spectra for multiple user groups is warranted: the preferred method of access by users might be worthy of greater focus than is in place at present.

The creation and sharing of these sources of real-time information may come from a variety of sources, and each comes with its own benefits and drawbacks. It is unsurprising that this survey population would have a high degree of trust in government-sourced information. In the domains of mobility, environmental quality, and safety and security, the respondents noted, on average, that they trusted the real-time information a majority of the time (table 2). There is trust of nonpublic-sourced information, such as on television, or broadcast news as well, but to a lesser degree. Further, there is trust of crowdsourced information, which in many cases can be quite accurate, but to the smallest degree of the three.

Item	x̄	n	SD
Mobility			
Official Public Sources (e.g., government)	3.19	124	0.46
Nonpublic Sources (e.g., private news television, nongovernment apps)	2.88	124	0.45
Crowdsourced	2.56	124	0.59
Environmental Quality			
Official Public Sources (e.g., government)	3.22	120	0.48
Nonpublic Sources (e.g., private news television, non-government apps)	2.96	120	0.44
Crowdsourced	2.49	120	0.57
Safety and Security			
Official Public Sources (e.g., government)	3.20	121	0.50
Nonpublic Sources (e.g., private news television, non-government apps)	2.96	121	0.46
Crowdsourced	2.54	120	0.62

Note: 1=Always distrust, 2=Distrust majority of the time, 3=Trust majority of the time, 4=Always trust

Table 2.
How much do you trust the following sources of information?

Trust of this information is necessary for one to change their habitual or intuitive actions, and repeated incorrect data could lead to resentment and distrust. However, as this survey population may be employed at agencies that also create government-sourced data sets, or the data used may be the same data these individuals have access to in a professional capacity, it is unsurprising there is a higher level of trust placed on this government-produced data.

The issue of trust of the data becomes important when one is seeking behavior change; and trust is difficult to obtain. Overwhelmingly, the survey respondents noted their trust of government-provided real-time data over that from other sources. This could be due to the familiarity these government employees have with the source of the data, or the data itself from regular professional work with it. Repeatedly, however, respondents questioned from the qualitative research noted a distrust of the real-time transportation data. Some mentioned that it was merely propaganda from the government, while others could not reconcile the schedule shown with their intuitive knowledge of a bus's performance.

A major challenge for the acceptance of real-time information in planning and decision making is precisely due to the fact that it is real-time. The fact that waiting times can count upward, or hold for several minutes due to unpredictable transport conditions, can be confusing and frustrating for the average user, and appear to be unreliable or untrustworthy. In a way, it is like the general distrust of the weatherman: We listen to the forecast, but know that it is likely to be inaccurate. The irony is that many individuals, during the qualitative research, used the at-stop real-time signboards in tandem with referring to printed bus schedules – measuring static schedules against the actual real-time information. This was particularly true with the older bus riders at People's Park (figure 2).

Fig. 2.
A man checks a printed schedule for the next bus arrival at People's Park, even though a digital signboard is just above the sign he is looking at.

Fig. 3.
Locations of Qualitative Research

This sentiment of distrust works against the heavy investments into signage and infrastructure to support the availability of real-time information. Surprisingly, there were signs describing the real-time information on the at-stop signboards, even though printed maps tend to have a small key to interpret that information. Education about this information to the larger public may be a worthwhile consideration as this information becomes more available across the island.

Experience with data is a key to trusting it. One could speculate that demographic characteristics associated with "tech savvy," like educational attainment or age, would be predictors of trust. However, no significant association between trust and education or age could be found.

Conclusions

The challenge of measuring perceptions is that they change over time. Changes in technology, the increased familiarity of certain devices, and the ever-changing economic and societal pressures on individuals can drastically alter the findings of a similar research endeavor in the future. This research does offer a yardstick, and a different perception on data within the milieu of the present, with current technology and contemporary citizens.

Comprehensively, the mentality of "if we build it, they will come" with regard to data has served Singapore well – largely because Singapore has the capital to take risks in creating tools for data access across modalities. However, with the insight from this research, it may be both more efficient and effective to target specific habits and routines as places of intervention, focusing on how users actually use the breadth of real-time information available on the island. The general trend is toward openness about the data, but especially with ambitions toward augmenting behavior, a specific focus toward targeted and focused interventions may be more effective than the current “shotgun” approach to data availability.

Al-Deek, Haitham, Michael Martello, Adolf May, and Wiley Sanders. 1988. *Potential Benefits of In-Vehicle Information Systems in a Real-Life Freeway Corridor Under Recurring and Incident-Induced Congestion*. Berkeley: Institute of Transportation Studies, University of California, Berkeley.

Boyce, David E. 1988. “Route Guidance Systems for Improving Urban Travel Demand and Location Choices,” *Transportation Research Part A: General* 22, no. 4 (July): 275–281.

Darby, Sarah. 2006. “The Effectiveness of Feedback on Energy Consumption: A Review for DEFRA of the Literature on Metering, Billing and Direct Displays.” Oxford: Environmental Change Institute, University of Oxford.

Dziekan, Katrin, and Arjan Vermeulen. 2006. “Psychological Effects of and Design Preferences for Real-Time Information Displays,” *Journal of Public Transportation* 9, no. 1: 71–89.

Dziekan, Katrin, and Karl Kottenhoff. 2007. “Dynamic At-Stop Real-Time Information Displays for Public Transport: Effects on Customers,” *Transportation Research Part A: Policy and Practice* 41, no. 6 (July): 489–501.

Golder, Scott, and Michael Macy. 2011. “Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures.” *Science* 333, no. 6051: 1878–1881.

Infocomm Development Authority of Singapore. 2005. *Technology and You: Singapore Infocomm Foresight 2015*. Singapore: IDA, Republic of Singapore.

Ito, Mizuko, Daisuke Okabe, and Misa Matsuda, eds. 2005. *Personal, Portable, Pedestrian: Mobile Phones in Japanese Life.* Cambridge, MA: The MIT Press.

Mahizhnan, Arun. 1999. “Smart Cities: The Singapore Case,” *Cities* 16, no. 1 (February): 13–18.

Mehndiratta, Shomik Raj, Christopher Cluett, Michael Kemp, and Jane Lappin. 2000. *Transit Watch – Bus Station Video Monitors: Customer Satisfaction Evaluation; Metropolitan Model Deployment Initiative: Seattle, Washington*. Seattle: Battelle Memorial Institute.

National Computer Board, Singapore. 1992. *A Vision of an Intelligent Island: The IT2000 Report*. Singapore: SNP Publishers.

Navigant Research. 2010.“Smart Cities.” http://www.navigantresearch.com/research/smart-cities (accessed April 18, 2014).

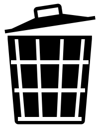
Singapore Department of Statistics. 2011. *Census of Population 2010 Statistical Release 2 – Households and Housing.* Singapore: Ministry of Trade and Industry, Republic of Singapore.

Williams, Amanda, Erica Robles, and Paul Dourish. 2008. “Urbane-ing the City: Examining and Refining the Assumptions Behind Urban Informatics,” in *The Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City,* edited by Marcus Foth. Hershey, PA: Information Science Reference, 1–20.

Yap, Jimmy. 2011. “The full ST [*Straits Times*] Interview with SG Buses Developer Muh Hon Cheng,” *iMerlion.* http://www.imerlion.com/2011/05/full-st-interview-with-sg-buses.html (accessed August 6, 2012).

David Lee

Tracking Waste to Reduce Waste



Introduction

One of the most promising applications of urban sensing is to present data to citizens in a way that reveals normally unseen processes in the city. These could be short-term ephemeral processes like air-quality variation and traffic congestion on nearby roads, or they could be long-term processes like growing crime incidence and climate change. Each application is an opportunity to improve awareness of a particular problem and evoke some sort of response from citizens, either through a change of opinion or a shift in behavior, which could spark larger debate and positive action.

For most cities, building a new infrastructure of electronic sensors remains prohibitively expensive. A cost-effective alternative to this is participatory sensing, where many volunteers perform sensing tasks independently and contribute their data to the whole picture (Paulos et al. 2009). For example, citizens blogging photos of dying street trees could cost less than installing sensors to monitor every tree in real time.

Such approaches engage the greater community and make sensing more transparent and perhaps less threatening (Burke et al. 2006). Additionally, while data collected from nonexperts might be less robust for science or policy research, it could be more appropriate for feedback to citizens (Dutta et al. 2009). For instance, if we want to report air pollution to pedestrians, we should measure it at street level where they are densest, rather than atop streetlights and buildings. We still lack a good understanding of how we might perceive data differently if we participated in collecting them. Having citizens help measure phenomena that are not plainly visible (like rising ozone pollution in the lower atmosphere) could prime them to trust the results more than if they were simply broadcast by experts. Citizens who have already invested time in revealing a problem might be more likely to change their behavior as well. However, participating might also make us less confident in data, perhaps by showing firsthand the practical limits or flaws of the process. Any of these could have profound implications for urban sensing.

Trash Track

Trash Track began with a simple idea, to understand where our garbage goes once it has left our sight. Technology now enables us to track the movement of any man-made object from its creation to its disposal, and the systems governing where our trash ends up are both incredibly complex and rarely understood. Waste travels an elaborate network of transfer stations, landfills, and reprocessing plants by truck, train, boat, and plane, and is constantly rerouted by regulations and markets. We wanted to understand how well this infrastructure functions and how individual actions, such as throwing away a mobile phone, translate into wide-scale consequences.

In October 2009 the MIT SENSEable City Lab studied the movement of trash through the Seattle waste-removal system using remote tracking sensors. Researchers from the Lab traveled to Seattle to deploy 2000 GPS tracking tags; each tag was attached to a separate trash item, then thrown away at some location in the city (Boustani et al. 2011). We asked Seattle residents to volunteer through the Trash Track website; selected volunteers provided trash items, helped catalog and attach tags, and deployed trash into the removal stream. **Figure 1** shows how we worked with volunteers at their homes to prepare the trash items.

Fig. 1.
Volunteers help with distributed sensing by identifying trackable trash items, attaching the tags using protective quick-setting foam adhesive, and throwing these items away.

Over the following two months, we remotely tracked the location of each tagged piece of trash. We then visualized these paths in maps, animations, and videos, presenting them in public exhibitions and websites. This allowed both volunteers and the general public to see the "second life of trash" in real time, after it was thrown away. From such maps it was clear that trash could end up traveling far beyond the city limits, across state and country borders, especially electronic and hazardous waste. The method and direction of transport could thus mean a great difference in the energy and emissions impact of that waste.

By having their trash "talk" to them, we sought to connect people with an urban infrastructure that is both ubiquitous and inscrutable. Revealing the unseen processes of waste removal could increase awareness of the environmental impact of individual waste generation, and encourage citizens to act more sustainably. This information would inform infrastructure planning at the city, regional, or international scale, as well as trash-disposal decisions at the individual, human scale.

During our study, we surveyed volunteers from this experiment to measure how they understood and reacted to the real-time maps of their trash's movement. We also surveyed nonvolunteers to determine if they reacted differently. We wanted to know: Did the experience of helping to deploy the sensors alter how the volunteers viewed the results? If we want to use Trash Track to change attitudes or behavior on a large scale, does it help to have citizens participate in the tracking?

Methodology

A total of 84 volunteers participated in the Trash Track project, over two weekends in the month of October 2009. We remained in contact with each of the volunteers since completing the tag deployment; they constituted group A of the study.

Group B comprised of members of the general public who visited the Trash Track website. Because they did not participate in Trash Track in any way prior to viewing the data, they acted as the control group. They were recruited through e-mail announcements of the website launch through the same social networks used to recruit the original volunteers.

Tracking Results Website Interface

We collected the trajectories of each trash tag in real time over three months, in the form of time-stamped geographic coordinates. We overlaid these raw trajectories on a Google Maps mash-up, with custom icons showing the first and last known locations and a line tracing out all points in between. Users could view the paths of each trash item individually, or see the latest known locations of a batch of trash tags. The user could choose to filter by type of trash, or to see only their own trash (for volunteers who provided items). Figure 3 illustrates the two basic views of the Web interface. Since the data shown was generated in real time, the actual results shown would change during the months following initial deployment.

We asked each participant in these two groups to complete two surveys, one prior to viewing the tracking results website and one afterward. The survey questions attempted to capture the ways individuals responded to the Trash Track data. Both the presurvey and postsurvey asked Likert-scaled questions on the subject's current attitudes and beliefs in the impact of their decisions on the environment. Subjects were asked about their attitude toward waste generation and disposal, as well as how often they took certain sustainable actions like recycling and bringing e-waste to designated disposal sites.

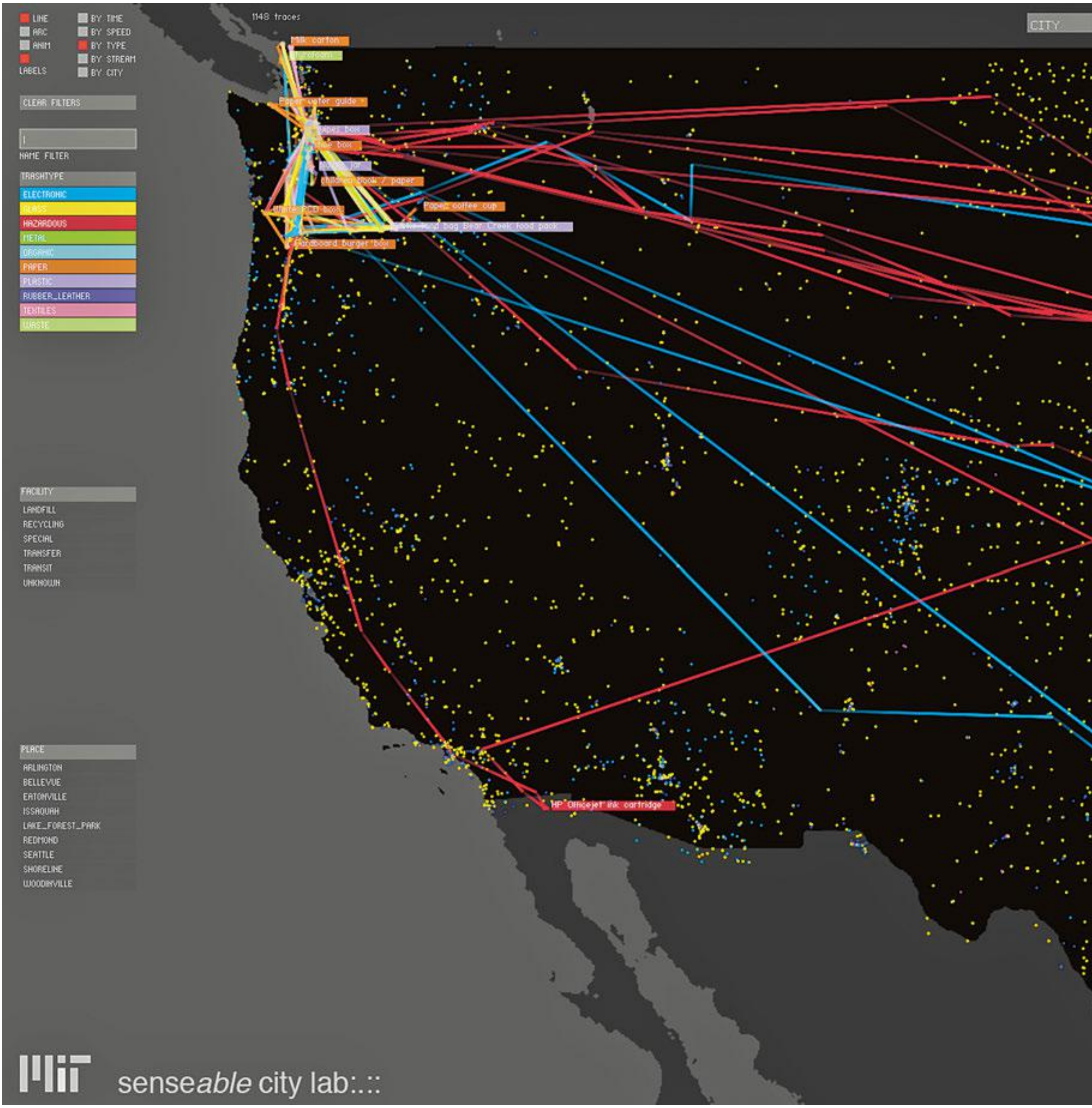


Fig. 2.
Visualization of the trajectories of every trash item tracked in the experiment.
Visualization: Dietmar Offenhuber, SENSEable City Lab 2010

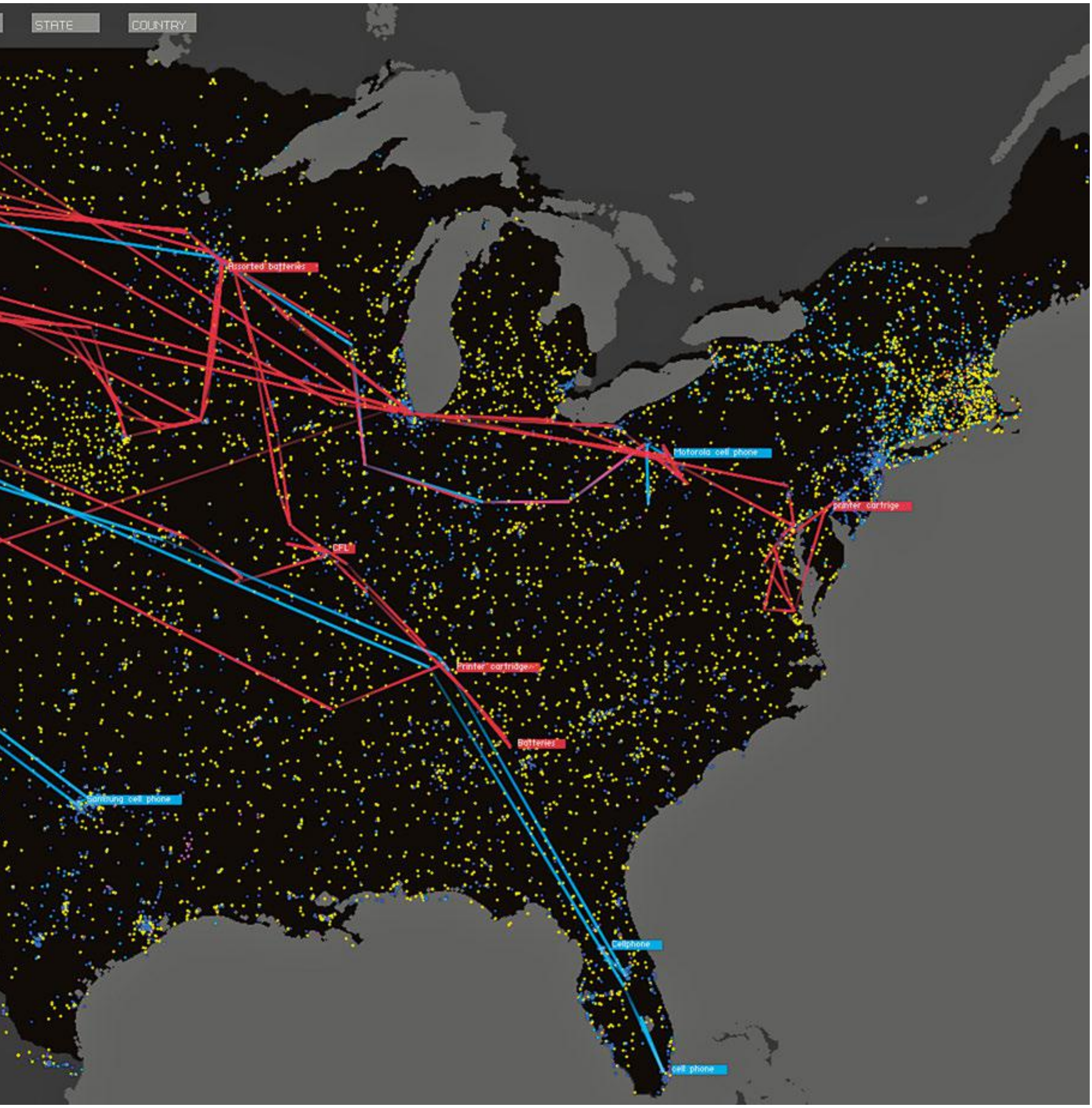

STATE
COUNTRY
Assorted batteries
Motorola cell phone
printer cartrige
CFL
Printer cartridge
Batteries
Samsung cell phone
Cellphone
cell phone

Fig. 3.
An example of the Web interface presenting the data results of Trash Track to the public. On left is the view of a single item's trajectory; on the right is an aggregated view of where many tags have ended up. White markers represent transfer stations and disposal sites local to Seattle.

We tested two hypotheses:

(1) Both before and after viewing the Trash Track data traces, volunteers and nonvolunteers held different attitudes and reported different behaviors related to trash disposal.

(2) Viewing the Trash Track data led to a change in attitudes and behaviors. The nature of this change differed for volunteers and nonvolunteers.

Results

Using an online survey tool, we polled more than 200 visitors to the Trash Track website before they were able to view the data traces from the tag deployment. A month later, we contacted every participant with a request to take the postsurvey, with 70 of the original participants completing this follow-up questionnaire. From these 70 responses, 32 were volunteers from the Seattle deployment, and the remaining 38 were website visitors who had not volunteered.

We first compared average responses among the two groups. Volunteers placed higher faith in local disposal and recycling systems, and understood where their trash went after disposal. They also sought to reduce their waste footprint and recycle more after viewing the Trash Track data, significantly more so than nonvolunteers. In contrast, viewing the data traces closed the gap between nonvolunteers and volunteers, who at first seemed more confident knowing the locations of hazardous-waste disposal and how Trash Track worked.

Ordinal logistic regression models further revealed how volunteering correlated with stronger responses to sustainability questions. Volunteers were more likely to consider the amount of packaging waste that came with purchased products. They were more likely to believe in the efficiency of waste-removal systems and the effectiveness of curbside recycling. The same models showed that in general, subjects only changed their responses to two of these questions after viewing the data traces. When controlling for the different aspects of the Trash Track traces that they investigated, both volunteers and nonvolunteers felt an improved understanding of where their trash went and how the Trash Track system worked.

A final regression model tested whether volunteers changed their responses between surveys in significantly different ways from the nonvolunteers. This model revealed three things:

(1) On the presurvey, volunteers were more likely to indicate knowing the locations of hazardous disposal sites than nonvolunteers.

(2) Nonvolunteers were more likely to indicate knowing the locations of hazardous disposal sites after seeing the Trash Track traces.

(3) Volunteers experienced the opposite effect upon seeing the Trash Track traces. They were less likely to indicate knowing the locations of hazardous disposal sites on the postsurvey.

Concluding Remarks

In summary, the models show that, while the information feedback from Trash Track did significantly change some people's understanding of trash issues and sensing technologies, it did not significantly change their attitudes or behavior toward trash generation and disposal.

Volunteers involved in executing the experiment reported engaging in more frequent sustainable behavior. On average, they tried to reduce their waste impact by reusing items, minimizing packaging, and recycling when outside the home more often than nonvolunteers. They also expressed more confidence in the effectiveness of the waste-disposal system and their own understanding of it. Yet, these effects did not completely linger in the long term.

Past the “high” of participating in the experiment and seeing its results, volunteers lost or gained little in the following months. The memory of the event might have faded over time, without new information or ways to participate beyond that initial engagement. The three significant changes between the presurvey and post-survey were in knowledge-based questions. Both volunteers and nonvolunteers reported significant differences in how well they understood where their trash went, how the Trash Track experiment worked, and where hazardous-waste disposal sites were located. As an experiment in creating awareness, then, Trash Track achieved some success in explaining how the waste stream works and how tracking technology can be used to reveal its path.

At the same time, the opposite ways that volunteers with the project and nonvolunteers changed their responses to the latter question (knowing where hazardous-waste disposal sites were located) points to another, potentially positive outcome of viewing the traces. For some, the traces revealed previously unknown facts about where these materials are disposed. For others, the traces demonstrated that the system was more complex and unpredictable than they had previously thought, forcing them to reevaluate exactly what they knew and did not know about practices in their city.

We must be careful about drawing broad conclusions from this study, considering the inherent problems of self-selection and ceiling effects in using voluntary surveys. In the context of waste disposal, we must study these effects further, in more controlled settings with a larger sample population, in order to understand how participatory sensing supports citywide behavioral change. From free-response questions on our survey we found anecdotes that indicated some discrete attitude and behavioral changes:

- “[Trash Track] made me more aware of details on how to dispose of certain items. Before, I had been too lazy to look up where to take an old laptop or light bulbs. I didn’t want to throw them away, so they just stay around. Trash Track made me motivated to learn the proper disposal process for such items.”
- “Drawing the connection between my garbage and where it goes creates a sense of responsibility. Hopefully it will help me

bridge the gap between my consumer choices and making the planet a little cleaner."

We believe such quotes reflect the potential impact of participatory sensing projects, and future projects should try to reliably effect lasting change. For instance, a focus on electronic waste items and their sometimes long, unpredictable trajectories could engage participants over a longer period of time. The risks presented by unsafe e-waste disposal are more immediate and severe, and citizens often have more options for disposal that can be tracked and compared. A crowdsourced e-waste tracking project, with independent deployments in homes across the world, could yield valuable data for the collective, while revealing best practices to the individual.

Fortunately, the technology behind sensing trash locations is advancing and becoming increasingly scalable, allowing for more accurate, real-time feedback to citizens. How this feedback can encourage sustainable behavior, and transform social norms, is worth further careful study.

We would like to acknowledge the Trash Track team and volunteers in Seattle who helped execute the original experiment; Qualcomm, Waste Management, Sprint, and City of Seattle (Office of Arts and Culture Affairs, Seattle Public Utilities, The Seattle Public Library), and all the SENSEable City Lab Consortium members for material and technical support.

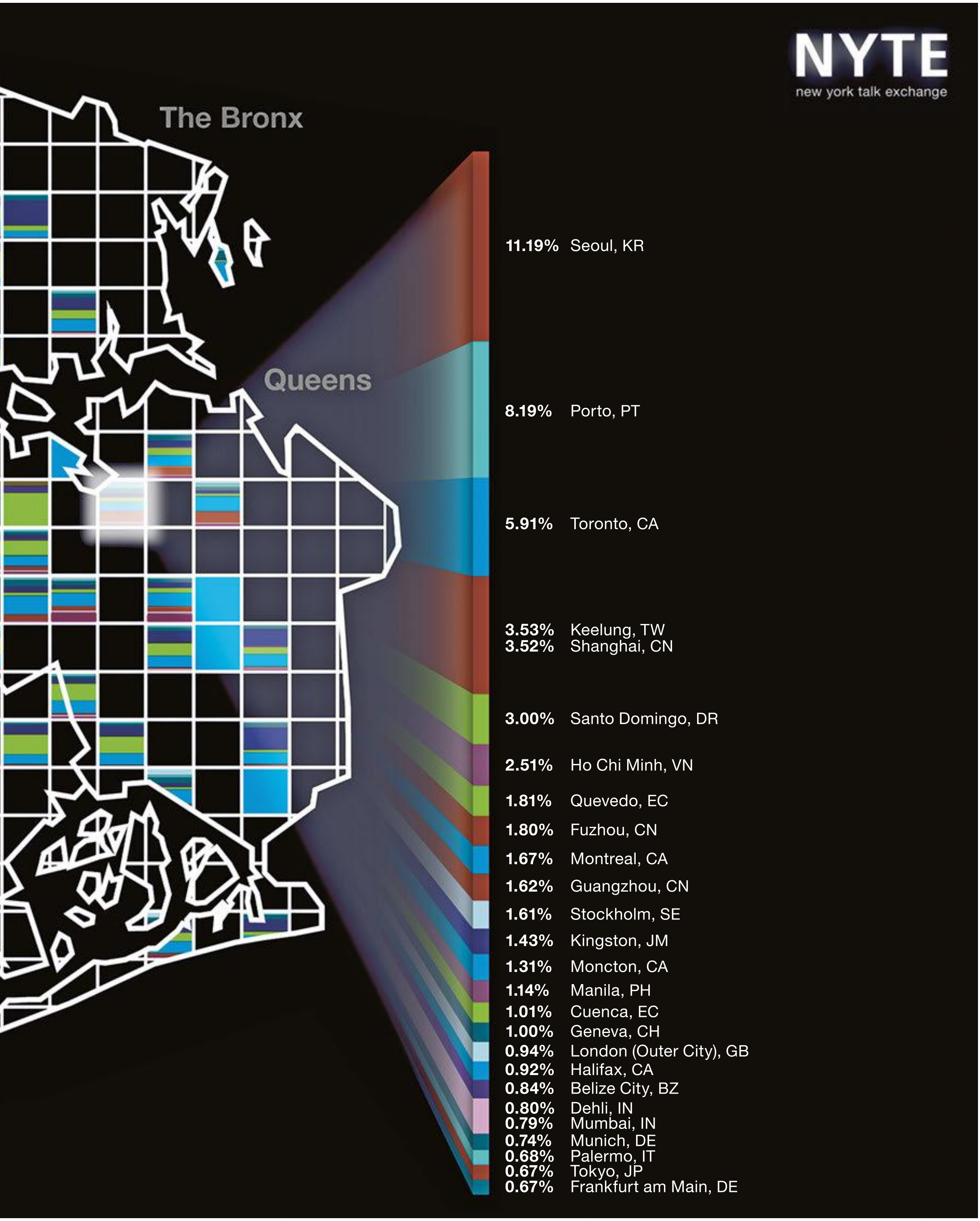
Boustani, Avid, Lewis Girod, Dietmar Offenhuber, RRex Britter, Malima I. Wolf, David Lee, Stephen Miles, Assaf Biderman, and Carlo Ratti. 2011. "Investigation of the Waste-Removal Chain through Pervasive Computing." *IBM Journal of Research and Development* 55, nos. 1/2 (March): 119–129.

Burke, Jeffrey A., Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B. Srivastava. 2006. "Participatory Sensing," World Sensor Web Workshop, ACM Sensys 2006, Boulder, CO, October 31, 2006.

Dutta, Prabal, Paul M. Aoki, Neil Kumar, Alan Mainwaring, Chris Myers, Wesley Willett, and Allison Woodruff. 2009. "Common Sense: Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors," *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*. New York: Association for Computing Machinery, 349–350.

Paulos, Eric, R. J. Honicky, and Ben Hooker. 2009. "Citizen Science: Enabling Participatory Urbanism," in *The Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City,* edited by Marcus Foth. Hershey, PA: Information Science Reference, 414–436.

Fig. 1.
The visualization shows how different neighborhoods reach out to the rest of the world via the telephone network. The city is divided into a grid of 2-kilometer square pixels where each pixel is colored according to the regions of the world wherein the top connecting cities are located.



The widths of the color bars represent the proportion of world regions in contact with each neighborhood. Visualization: Aaron Koblin, SENSEable City Lab 2008

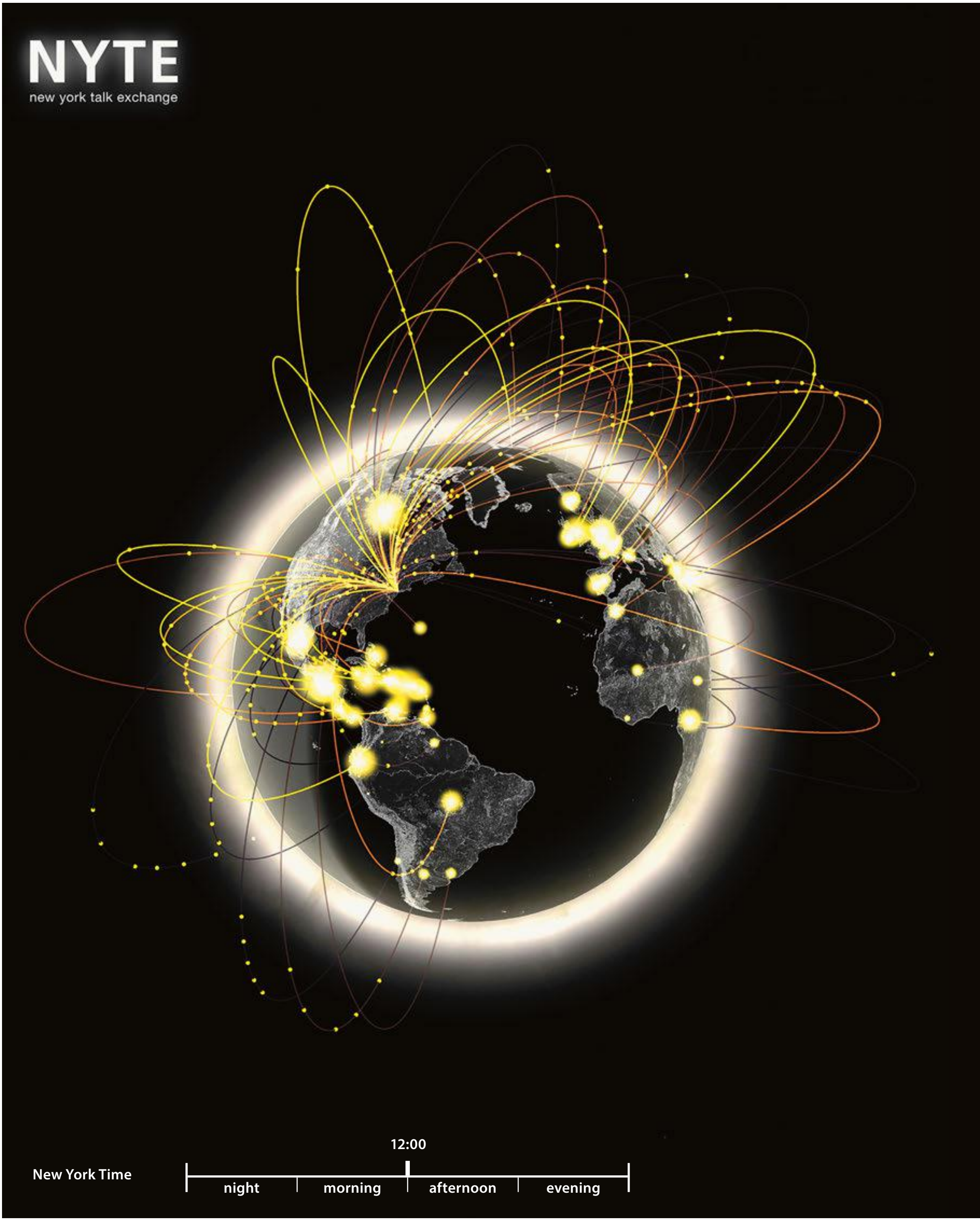


Fig. 2.
The volumes of Internet data flowing between New York and cities around the world. The size of the glow on a particular city location corresponds to the amount of IP traffic flowing between that place and New York City. A greater glow implies a greater IP flow. Visualization: Aaron Koblin, SENSEable City Lab 2008

Francisca M. Rojas

New York Talk Exchange: Revealing Urban Dynamics through the Global Tele-communications Network

Introduction

This study is based on the MIT SENSEable City Lab's *New York Talk Exchange* project, which created several visualizations of telephone calls between New York City and the rest of the world using data from September 2008 (Rojas et al. 2008). This study analyzed that data further to better understand how telecommunications flows can reveal otherwise invisible global dynamics of urban places. To which areas of the world is New York most intensely connected through the telecommunications network? Which neighborhoods are the most global when measured through international telephone calls? And what may be the purposes of those global telecommunications links?

The use of telecommunications data to understand urban dynamics and inform urban planning and policy is not a new approach. Ithiel de Sola Pool notes that as early as 1915, "telephone companies were the principal source of systematic demographic information on urban and neighborhood growth trends and characteristics." (1983, 42) Almost a century ago, the availability of these data led to the growth of the field of urban planning. In the mid-1950s the geographer Jean

Gottmann also recognized the value that could be extracted from flows of information for studying cities. Gottmann analyzed AT&T domestic long-distance call data from New York in *Megalopolis* (1961), where he proposed that by understanding the density and directionality of information links between places, we could also understand something about the activities rooted in those locales. "The density of the flow of telephone calls is a fairly good measure of the relationships binding together the economic interests of the region. Telephone calls represent not only economic and governmental relationships but also social and family links." (Gottmann 1961, 590) A half-century later, the digital traces generated by the city allow us to pick up on this approach and investigate links on a global scale.

The Data

The analyses in this study rely on three sources of information: (1) aggregated long-distance telephone call records provided by a major US telecommunications carrier; (2) neighborhood demographics drawn from the US Census; and (3) fieldwork consisting of semistructured interviews and observation in select neighborhoods. The call data includes long-distance telephone traffic between New York City neighborhoods and more than 200 countries around the world. The number of calls and the sum of call minutes between two locations are aggregated into ten-minute increments for every day of September 2008. This includes calls made via landlines, wireless phones, and prepaid calling cards. Within New York, the calls are geographically located and aggregated on the scale of a wire center, or telecoms switching station, dividing the city into 67 distinct areas for analysis. The demographic data is joined to the call data on this same geographic scale. We estimate that the data analyzed in this study represents about 25 percent of the total international call traffic for the city.[1] This is the figure used to normalize call data in the analyses presented for this study.

Twenty-four semistructured interviews were carried out in upper Manhattan and central Queens, two areas of the city with a high intensity of international calls. Interview questions mirror the call data by asking subjects about the frequency of calls in a month, length of conversation, mode of call, and cost. This approach

[1] The share of the international telecommunications market in New York City held by the telecoms provider involved in this study is not publicly available, but according to the FCC's 2007 International Telecommunications Data report, in 2004 this company generated more than one quarter of the total billed international minutes in the United States.

facilitates the interpretation of the quantitative analyses, helping to determine, for example, whether the calculation of call minutes per capita looks reasonable and if the weekly call pattern coincides with people's lived experience.

A Dual Geography of Talk

Prior analyses of international practices in global cities suggests that New York City's counterparts in globalization should be primarily its business partners: London (UK), Tokyo (Japan) and Frankfurt (Germany) (Sassen 2001, 2002; Derudder 2008). A few studies on global cities discuss how information and communications technologies (ICTs) also enable transnational practices among migrants, with cheap phone calls in particular acting as the "social glue" that binds immigrants to their countries of origin (Levitt and Glick Schiller 2004; Vertovec 2009). By analyzing calling destinations from one of the world's most emblematic commercial and immigrant cities we should expect telecommunications flows from New York to link both to the city's global business counterparts and to its corresponding migrant-sending countries.

In examining New York's long-distance calls according to their international destinations we see that among the top ten calling counterparts for New York City, seven are migrant-sending countries to the city and only three are notable business partners. Table 1 shows the share of call volume by country alongside figures on New York's foreign-born population and US trade partners.[2] The Dominican Republic and Mexico are ranked first and second and are also the top two sending countries within New York's foreign-born population. Meanwhile, the three countries with obvious business relationships to the city, using US trade data as a proxy, are the UK, Canada, and Germany. A third type of country emerges from this list, those that have both a trade relationship with the US and that send migrants to New York: Mexico and India. A look at per capita GDP for each of these countries suggests that, with the exception of Mexico and India, strong economies tend to trade goods with the US, while weak economies tend to export labor to the United States.

Examining call volumes by country is informative but may obscure the gravitational pull of New York toward a particular

[2] The analyses presented here use outgoing call minutes as the standard base of analysis instead of the number of calls. This is the protocol by which the Federal Communications Commission and TeleGeography base their measures and analyses.

Rank	Country	World Region	Call Destination from NYC (%)	Call Destination from USA (%)	NYC Foreign-Born (%)	USA Import Partner (%)	USA Export Partner (%)	Per Capita GDP (USD)	Type of Link
1	Dominican Republic	C. America / Caribbean	11.7	2.3	13.1	0.2	0.5	4,179	M
2	Mexico	North America	9.1	17.8	6.4	10.8	12.0	9,516	M/T
3	United Kingdom	Europe	7.5	4.0	1.2	2.9	4.4	45,510	T
4	Canada	North America	7.0	12.2	0.8	16.1	12.9	43,396	T
5	Guatemala	C. America / Caribbean	5.5	3.2	0.8	0.2	0.4	2,548	M
6	Ecuador	South America	5.3	0.8	4.9	0.3	0.3	3,432	M
7	Jamaica	C. America / Caribbean	3.3	1.1	6.3	0.0	0.2	4,565	M
8	India	Asia	3.0	10.8	2.7	1.3	1.6	947	M/T
9	Germany	Europe	2.5	2.0	0.8	4.9	4.4	40,308	T
10	Philippines	Asia	2.4	3.5	2.0	0.5	0.7	1,624	M

Note: M=migrant, T=trade partner, M/T=migrant and trade partner

Sources: compiled by author from data from large US telecommunications provider, US Census 2006–2008 American Community Survey (ACS), Federal Communications Commission 2009, IMF Direction of Trade Statistics 2008, International Telecommunications Union Basic Indicators 2007

Table 1. Comparison of top ten destination of calls from New York City ranked by percentage of total international call minutes for September 2008 with the corresponding percentage of US total international calls (2007), NYC percentage of foreign-born by country of origin (2006–2008), US trade partners (2007), and Per Capita GDP (2007).

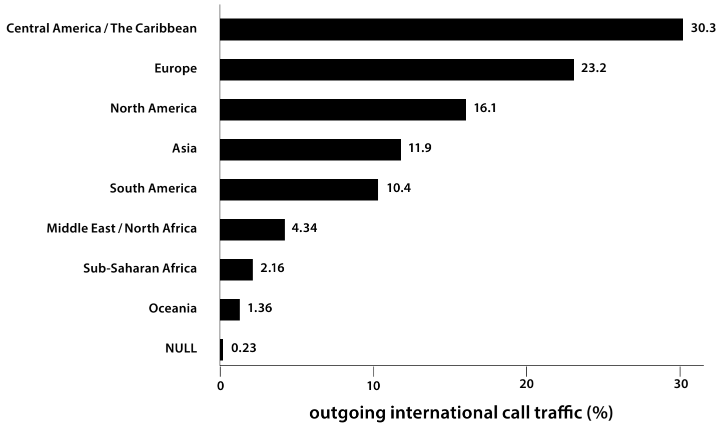
region of the world. Measuring the volume of calls by world region (figure 3) reveals that the city has a strong relationship with Central America and the Caribbean, reflecting the strength of migrant links to the city.[3] By being principally oriented to Central America and the Caribbean (30 percent of total international calls), followed by Europe (23 percent) and North America (16 percent), New York diverges from the national US trend in terms of its global links through telecoms. International calls from the United States as a whole are generally destined for North America (29 percent), Asia (20 percent), and Europe (15 percent) (FCC 2008), consistent with the country's principal trading partners (IMF 2008).

New York's calling destinations represent a mix of commercial partners and immigrant-sending countries and thus reveal a dual geography of international telecommunications links. In terms of engaging telecommunications as a tool for transnational activity, migrants appear to be on equal footing with global corporate interests in generating considerable international call volumes out of New York.

Migrants and International Calls

Migrants' influence on the volume of New York City's international calls can be examined further by seeing if the concentrations of

[3] The Central America and Caribbean region does not include Mexico, which is considered to be part of North America.



Source: Large US telecommunications provider

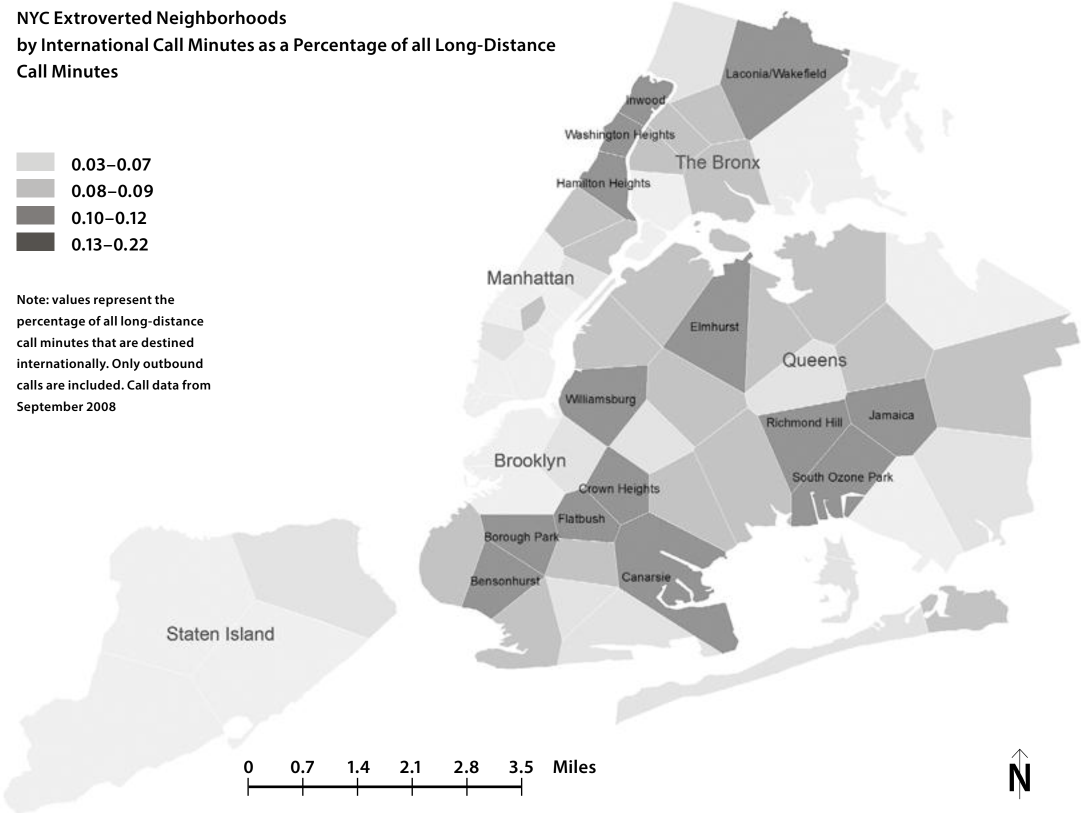
Fig. 3.
International calls from New York by world region (September 2008)

particular foreign-born groups in the city's five boroughs match up with the volumes of call traffic destined for different countries.[4] This analysis employs a simple regression by borough (Bronx, Brooklyn, Manhattan, Queens, Staten Island), where the destination countries of phone calls from that borough are regressed on the country of origin of each borough's foreign-born population. This regression model log transforms the two variables to normalize the distributions and linearize their relationship.

The fit of the regression model is strongest in Manhattan, where the foreign-born population model accounts for 67 percent of the variability in international call volume. We find that for a 10 percent increase in Manhattan's foreign-born population from a given country, on average we can expect 10 percent more minutes of talk to that country from Manhattan per month. This interpretation is also valid for Brooklyn and Queens. While these results imply an almost one-to-one association between the presence of foreign-born groups and the volume of calls to their countries of origin on the borough scale, by not controlling for income by country of origin these results may be less precise than desirable. Nevertheless, finding that foreign-born groups and calling destinations by country statistically coincide supports the premise that communications flows can inform us about urban dynamics on a subcity

[4] Not all 223 countries represented in the call data are included in the regressions reported in figure 3 because the 2006–2008 American Community Survey does not list country of origin data for as many countries as New York's neighborhoods connect to over the telephone. Nevertheless, up to 69 countries are represented by data for call minutes and foreign-born population in each borough.

NYC Extroverted Neighborhoods
by International Call Minutes as a Percentage of all Long-Distance Call Minutes



Sources: US Census 2007–2008 American Community Survey
US telecommunications provider

Fig. 4.
New York City's extroverted neighborhoods by international minutes as a percentage of all long-distance calls.

scale. It suggests the potential value for urban planning in being able to establish a relationship between migrants' countries of origin and their calling destinations.

Finding a relationship between the origin and destination of telecoms flows on a neighborhood scale presents the possibility of updating the century-old planning practice of using telephone data as an indicator of demographic change and growth in the city. In an age when the function and management of urban areas is increasingly digitized – GPS-enabled transit systems already communicate the whereabouts of a bus in real time – and where powerful computation can easily analyze these resulting data sets, the information cycle between city function and city manager could become instantaneous. It is conceivable that the

Population Division of New York City's Department of City Planning may be able to use data about the origin and destination of international phone calls to generate estimation algorithms that predict migrant flows into and out of the city on a neighborhood scale on a regular basis. Given the importance of migration to the city's vitality, such an approach could both expedite the city's enumeration efforts without relying entirely on the US Census's efforts, which notoriously undercount foreign-born people, particularly those who are undocumented migrants, and provide information on a scale of spatial aggregation below that of the Community District level.

Extroverted Neighborhoods

To determine which areas of the city are the most oriented abroad, or "extroverted" (Massey 1994), we scale further down from the borough level to the city's neighborhoods. Extroverted neighborhoods are those that have the highest ratio of international total call minutes to domestic (US) call minutes. This analysis therefore involves separating the long-distance call minutes by domestic and international destinations as shares of the total long-distance volume. We can then compare the New York City neighborhood rates of international calls to the US-wide share of international calling minutes, which serves as a threshold figure to determine whether to designate a particular neighborhood as extroverted. According to data for 2008 from the US Federal Communications Commission and the International Telecommunications Union, 12 percent of all long-distance calls made in the United States were destined for international points in that year. By this threshold, 14 out of the 64 wire centers analyzed in New York City record a higher percentage than the US average in outgoing international calls.

The extroverted neighborhoods are all primarily residential areas of the city in the outer boroughs **(figure 4)**, "...where the city's working and lower-middle classes reside, and where its new immigrants settle" (Sanjek 1998, 29). Indeed, the percentage of foreign-born people in all but one of the neighborhoods is well above the city mean of 34 percent. These global links become apparent when walking through places like Elmhurst in Queens, one of the more extroverted neighborhoods of the city. On Whitney Avenue you are transported to Indonesia; Bexter Avenue takes you to Colombia; coexisting on Roosevelt Avenue are India, Pakistan, and Mexico; and you can visit a cross section of Asian locales along Broadway.

Notably, none of the business centers of the city – not the Financial District or even Midtown – can be considered an extroverted area by measuring the ratio of international to domestic call volume. Nevertheless, in absolute and relative terms, some of the business-oriented wire centers, including Chelsea and the Financial District, are the areas of the city that generate the most international call minutes **(table 2)**.[5]

[5] Not included in this table nor in further analyses is the Tribeca/Chinatown wire center since the high call volumes associated with this node is not representative of conditions in its respective neighborhood but rather is a function of New York's citywide telecommunications infrastructure. This wire center contains within it at least two major long-distance switching stations that route much of the city's long-distance traffic for the telecoms provider involved in this study: 33 Thomas Street and 32 Avenue of the Americas. These exchange buildings collect and route long-distance traffic for areas beyond their immediate vicinity and thus represent a case where we see many more calls aggregated than can be explained by the demographics that correspond to that wire center's geographic boundaries.

Neighborhood Characteristics

To what extent does neighborhood context help predict the strength of international ties in that location? The principal variables that may explain variation between neighborhoods in terms of how extroverted they are via telecoms are: rate of foreign-born residents, rate of people who speak a foreign language, and median household income. If international calls are the social glue of migrant transnationalism, then high rates of foreign-born residents and foreign-language speakers should serve as strong predictors for a neighborhood's extroverted calling activity. We find that indeed, the percentage of foreign-born people in an area is a significant predictor of international calls as a percentage share of total long-distance minutes.

A regression analysis of international calls as a percentage of all long-distance calls by median household income indicates an inverse relationship between income and percentage of international calls. **Table 3** reports the coefficients for regressions that employ dummy variables to compare across income categories – low (below $ 38,320), medium ($ 38,320–$ 71,850) and high (above $ 71,850) – while holding constant the percentage of foreign-born, the percentage of foreign-language speakers, and the number of households per neighborhood.[6] The omitted variable represents the middle-income neighborhoods in New York.

All models indicate that low-income neighborhoods will generate a predicted volume of international calls that is 4 percent higher than middle-income neighborhoods. Low-income neighborhoods differ significantly from middle-income neighborhoods in this respect. But high-income neighborhoods are predicted to make 4 percent fewer international calls than middle-income areas of New York, a result that is only significant in Model 1, where no control variables are included in the regression.

This regression analysis indicates that the rate of international calling minutes is influenced in divergent ways by the rate of foreign-born people in a neighborhood and an area's median household income. While finding a positive association between the foreign-born and international calling is to be expected given the importance of ICTs in immigrant life – a finding that is further supported in this study by field interviews – it is nevertheless surprising to discern an inverse association between income and international calling. This means that the people living in neighborhoods with the fewest economic resources are the ones most engaged in transnational practices through telecommunications. While this can be attributed in part to the dramatic fall in prices for international calls, it is also a testament to how vital global communications links are to the immigrant communities that make up the list of extroverted neighborhoods, almost all of which have median household incomes that fall below the city's median figure of $47,900. In fact, we can even expect the middle layers to be more engaged in global talk than the highest-income neighborhoods of the city, including its commercial areas. Nevertheless, as shown in table 2, commercial areas do dominate global talk in terms of absolute volume, but in considering international talk as a share of total calls, we find that lower-income immigrant areas of the city are the most engaged.

[6] To divide income into low, middle, and high for New York City, I use the median household income for New York as drawn from the wire center data provided by the large US telecoms provider, which is $47,900. I determine low to be at or below 80 percent of the median (<$38,320), middle to be 80 to 150 percent of the median ($38,320 to $71,850), and high to be at 150 percent of the median and above (>$71,850). This method comes from Booza et al. 2006. Excluded in the regression analyses of this section are the wire centers that correspond to Tribeca/Chinatown since it is not representative of its geographic vicinity; JFK Airport, since it does not have corresponding demographic information, and an unnamed wire center in Manhattan that does not have corresponding call data. Other outliers with international calls above two standard deviations from the mean were also removed for this analysis.

Patterns of Talk

What is the purpose of this intense engagement across boundaries through telecommunications? Analyzing patterns of call data by day and by hour allows us to infer the purposes of international calls. A comparison of the daily volume of long-distance calls by domestic and international destinations for the most extroverted neighborhood in New York City, Elmhurst in central Queens, illustrates a striking cadence of call patterns by day of the week (figure 5). The volume of domestic calls, indicated by a black line, is high on Mondays and Tuesdays and decreases as the week

Commercial Areas of New York

Neighborhood	Borough	International Minutes	Domestic Minutes
Chelsea	Manhattan	5.53	4.95
Financial District	Manhattan	2.24	2.67
Murray Hill	Manhattan	1.46	2.99
Times Square	Manhattan	0.93	0.63
Midtown West	Manhattan	0.72	1.40
Midtown East	Manhattan	0.49	0.55
Battery Park City	Manhattan	0.38	0.50
Kips Bay	Manhattan	0.32	0.41
Bowling Green	Manhattan	0.28	0.33
Clinton	Manhattan	0.20	0.41

Residential Areas of New York

Neighborhood	Borough	International Minutes	Domestic Minutes
Flushing	Queens	1.18	0.83
Washington Heights	Manhattan	0.62	0.19
Inwood	Manhattan	0.57	0.16
Elmhurst	Queens	0.57	0.16
Corona	Queens	0.53	0.40
Flatbush	Brooklyn	0.50	0.15
Williamsburg	Brooklyn	0.50	0.19
Richmond Hill	Queens	0.49	0.16
Astoria	Queens	0.47	0.28
Jamaica	Queens	0.47	0.18

Note: International call minutes are normalized by population and estimated market share of the US telecommunications provider and indexed by the mean per capita volume of calls for all wire centers.

Source: Author's analysis of data from a large US telecommunications provider

Table 2.
Rank of top extroverted areas of New York by normalized call volume for international and domestic calling and type of wire center/neighborhood: predominantly business or residential

progresses. Domestic calling dips dramatically on Saturdays and falls even further on Sundays. Yet the rhythm of international calls, shown by the gray line, reveals an inverse effect where calling dips during the week and peaks on the weekends, implying the social purpose of calls. The difference in calling patterns between weekdays and weekends suggests that international calls from immigrant neighborhoods principally serve to maintain links to family and serve affective purposes. By contrast, call volumes in the Financial District in Manhattan are consistently high during the week and low on weekends, with more domestic calls than international calls (figure 6).

Explanatory Variable	Model 1	Model 2	Model 3	Model 4
Low Median HH Income (below $ 38,320)	0.04** (0.02)	0.04** (0.02)	0.04** (0.01)	0.04*** (0.01)
High Median HH Income (above $ 71,850)	-0.04* (0.02)	-0.00 (0.02)	-0.00 (0.02)	-0.00 (0.02)
Foreign-Born Population (%)		0.20*** (0.04)	0.23*** (0.06)	0.22*** (0.06)
Foreign-Language Speakers (%)			-0.04 (0.05)	-0.04 (0.05)
Number of Households				2.59^{-07} (2.11^{-07})
Constant	0.09** (0.01)	0.02 (0.02)	0.03 (0.02)	0.02 (0.02)
R-squared	0.22	0.44	0.44	0.46
N	60	60	60	60

Note: Numbers in parentheses are standard errors. *p < 0.05, **p < 0.01, ***p < 0.001.

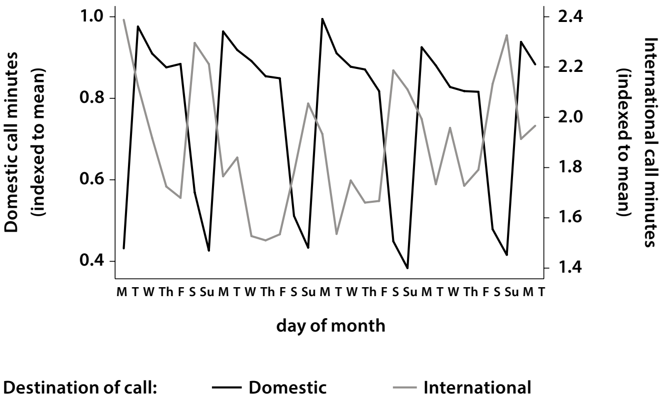
Source: Author's analysis of data from a large US telecommunications provider, September 2008 and 2007–2008 American Community Survey (ACS)

Table 3.
OLS Regression coefficients for international call minutes as a percentage of total long-distance calls on dummy variables for low and high median household income, percentage of foreign-born, percentage of foreign-language speakers, and number of households for neighborhoods in NYC

A comparison of calling patterns by hour of the day supports the assessment that international calling from immigrant neighborhoods reflects the social nature of these transnational exchanges. Figure 7 overlays the pattern of calls aggregated by hour for the month of September in four areas of the city: Elmhurst and Washington Heights (solid lines) as representative of extroverted immigrant neighborhoods, and the Financial District and Midtown West (dashed lines) as representative of commercial areas of the city. As in earlier analyses, the figures are indexed to the mean value for international call minutes by wire center. The predominantly business-oriented neighborhoods generate their highest call volumes during the day, with peaks between 9:00 a.m. and 12:00 noon. Meanwhile, the predominantly immigrant neighborhoods reach their peak calling hours very late into the evening, consistently between 11:00 p.m. and 2:00 a.m.

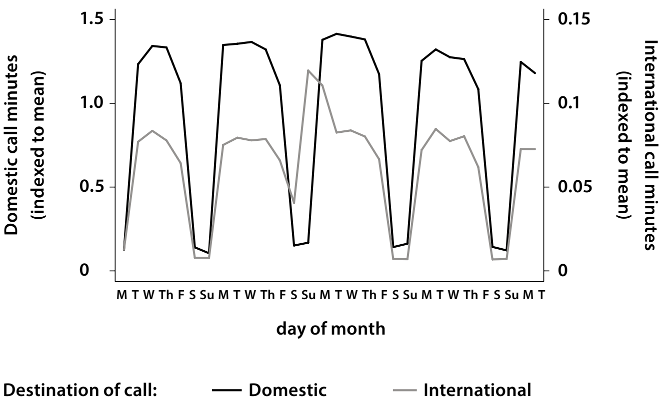
A manager at the La Nacional call center in Elmhurst – where people make international calls from booths, or *cabinas* – corroborated these patterns in the data: the weekends are the busiest days for the *cabinas,* while Tuesdays and Wednesdays are the slowest days. Interviews supported this finding, with subjects noting that weekends were the days where they called most and spoke the

Fig. 5.
Long-distance calls during September 2008 from Elmhurst (Queens)



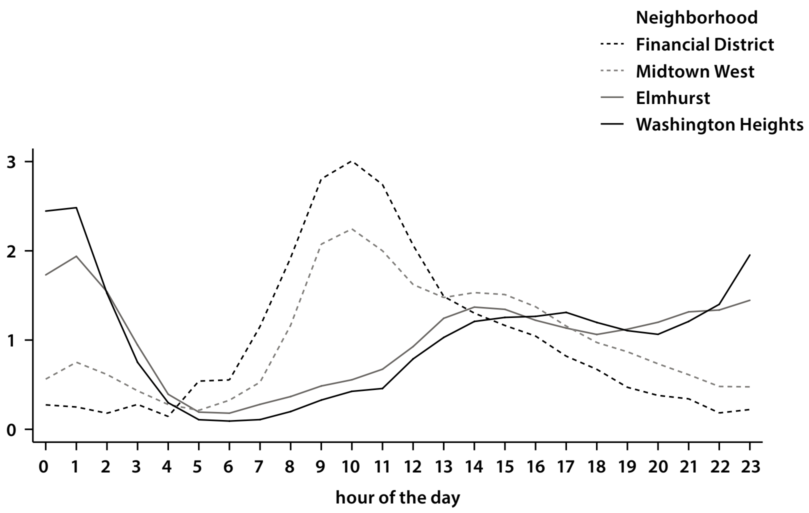
Source: Author's analysis using data from large US telecommunications provider

Fig. 6.
Long-distance calls during September 2008 from the Financial District (Manhattan)



Source: Author's analysis using data from large US telecommunications provider

Fig. 7.
Call minutes by hour from extroverted neighborshoods, residential versus commercial



Source: large US telecommunications provider
Note: Figures from September 2008, indexed to mean call minutes

longest not only because they had the time to do so but also because they were more likely to find their calling counterparts at home. Calling to the fixed lines of home phones is an important strategy since a popular unlimited international calling plan for mobile phones allowed outgoing calls only to fixed telephone lines. Also with prepaid calling cards, rates tend to be substantially lower when connecting to landline phones than when calling to mobile phones.

For migrants from the Dominican Republic in particular, “Sundays are about three things: church, family and rest,” according to an elderly Dominican woman who arrived in New York City in the late 1960s.[7] Generally, people in Latin America do not work on Sundays and most commercial establishments are closed on that day. Calling family at home on a Sunday gives the caller the opportunity to speak with the maximum number of family members at one time. After church, people come together at home for a long lunch and calling on Sundays ensures that New Yorkers will connect with family in their country of origin, and that both calling counterparts will have enough time to get news, manage business, and share gossip over the phone.

7 Translated from Spanish by the author.

Conclusion

While it is evident from previous research that advances in ICTs over the past twenty years have facilitated the processes of globalization for corporate functions, particularly in a command-and-control center such as New York, the question of how the other leg of globalization – immigration – has fared due to the increased affordability and accessibility of ICTs remains to be explored more deeply. This work has sought to open a window into the nexus between migration and telecommunications by showing that migration is indeed an influential force in shaping New York's global space of flows (Castells 1996), and that ICTs such as the telephone are fundamental tools in forging the transnational processes that today help define the twenty-first-century migrant experience.

This study is a snapshot in time of how New York relates to the world through the global telecommunications infrastructure with a focus on how these invisible global dynamics are reflective of life in immigrant neighborhoods. The global destinations of calls from New York are more diverse than existing research would have us expect and include not just trade partners but also some of the city's more significant migrant-sending countries. A strong predictor of the variation in volume of international calls generated from city neighborhoods is the rate of foreign-born residents and median household income. This approach is supported by evidence that migrants' countries of origin coincide well with the destination countries of international calls on the scale of the borough, adding another dimension to the strong association between foreign-born groups and international call volumes. And finally, international calls in immigrant neighborhoods follow a marked pattern of high volumes on the weekends and during the evenings, reflecting the purpose of calls as being social and affective in nature. Having established that some of the city's most globally engaged or extroverted neighborhoods have high concentrations of immigrants, ethnographic fieldwork sought to answer how and why migrants engage in global talk.

Perhaps the connection to how an analysis of intangible and invisible flows of information on a global scale can inform the practice of urban planning and policy on the city and neighborhood scales may appear tenuous. But conceptualizing the hybrid, transnational experiences of migrant groups in the twenty-first century – which are largely facilitated by ICTs – is fundamental to understanding urban areas in the United States and in other developed countries moving forward. Foremost in this respect is the fact that New York City has reversed its decades-long population decline and instead has experienced growth in the past thirty years solely as a function of foreign immigration (NYC Department of City Planning 2004). The city is projected to continue growing due to international migration into the coming decades and we can expect that telecoms links to sending countries will likely deepen.

This article is based on the author's doctoral dissertation (2010) at MIT's Department of Urban Studies and Planning, produced through research done in collaboration with the MIT SENSEable City Lab and funded through a research fellowship by the AT&T Foundation. MIT SENSEable City Lab researchers Francesco Calabrese and Jon Reades built the database of aggregated call records employed in this research and Jon Reades developed the geospatial approach employed in the mapping of call volumes. Special thanks to Francesco and Jon for their contributions to this work, and to the author's doctoral committee – Professors Frank Levy, Manuel Castells, and Carlo Ratti – for their guidance.

Castells, Manuel. 1996. *The Rise of the Network Society.* Oxford: Blackwell.

Derudder, Ben. 2008. "Mapping Global Urban Networks: A Decade of Empirical World Cities Research," *Geography Compass* 2, no. 2 (March): 559–574.

Federal Communications Commission. 2008. *Trends in Telephone Service.* Washington DC: Federal Communications Commission.

Federal Communications Commission. 2009. *2007 International Telecommunications Data.* Washington DC: Federal Communications Commission.

Gottmann, Jean. 1961. *Megalopolis: The Urbanized Northeastern Seaboard of the United States.* Reprint. Cambridge, MA: The MIT Press, 1967.

International Monetary Fund. 2008. *Direction of Trade Statistics.* Washington DC: International Monetary Fund.

Levitt, Peggy, and Nina Glick Schiller. 2004. "Conceptualizing Simultaneity: A Transnational Social Field Perspective on Society," *International Migration Review* 38, no. 3 (September) 1002–1039.

Massey, Doreen. 1994. *Space, Place and Gender.* Minneapolis: University of Minnesota Press.

New York City Department of City Planning. 2004. *The Newest New Yorkers 2000: Immigrant New York in the New Millennium.* New York: Department of City Planning, Population Division.

New York City Department of City Planning. 2010. *Current Population Estimates.* http://www.nyc.gov/html/dcp/html/census/popcur.shtml (accessed July 5, 2010).

Pool, Ithiel de Sola. 1983. *Forecasting the Telephone: A Retrospective Technology Assessment.* Norwood, NJ: Ablex Publishing Corporation.

Robinson, Jennifer. 2002. "Global and World Cities: A View from Off the Map," *International Journal of Urban and Regional Research* 26, no. 3 (September): 531–554.

Rojas, Francisca M., Clelia Caldesi Valeri, Kristian Kloeckl, and Carlo Ratti, eds. 2008. *NYTE: New York Talk Exchange.* Cambridge, MA: SA+P Press.

Sanjek, Roger. 1998. *The Future of Us All: Race and Neighborhood Politics in New York City.* Ithaca, NY: Cornell University Press.

Sassen, Saskia. 2001. *The Global City: New York, London, Tokyo,* 2nd ed. Princeton, NJ: Princeton University Press.

Sassen, Saskia, ed. 2002. *Global Networks, Linked Cities.* New York: Routledge.

Vertovec, Steven. 2009. *Transnationalism.* New York: Routledge.

Representation – Models and Visualization

Kristian Kloeckl

The City as a Digital Public Space – Notes for the Design of Live Urban Data Platforms

Think of an encounter with a friend on the street. You both stop and exchange information – greetings, personal anecdotes, the weather, etc. In the conversation you choose your words based on your knowledge of the subject matter, as well as your knowledge of your friend's background. When he frowns, you understand that he disagrees; his gesture informed you he wanted to intervene and respond. You stop to give him an opportunity to speak up. You note that your friend is getting impatient when he looks at his watch. He sees that you have noticed his impatience and explains he has an appointment to catch. All of this happens in real time, based on the exchange of many tiny signals. We are able to capture the dynamics of such interactions effectively, given the signals exchanged are perceivable, strong enough to be registered, and are understood by the recipient.

We adapt to environmental information similarly; however, it becomes more difficult with growing distance to the phenomenon of interest, or when its signals cannot be registered by the human senses. Now, consider being able to read all these sources of environmental information and base decisions upon them, just as effectively as in the conversation with your friend.

Urban space today is pervaded by digital networks and systems, creating information that represents human activity. While most digitally managed urban systems generate operational data for their own purposes, they normally do not share those data directly with other systems or the public. As a result, digital information representing human activity in the city exists in many different places, locked within their specific domain. In the LIVE Singapore! project, we developed an open platform for the collection, combination, and distribution of large numbers of such real-time data sources, encouraging developer communities to use these data streams for the creation of meaningful and beneficial civic tools.

LIVE Singapore! acknowledges that vast amounts of data are being generated as a by-product of urban system operations, and that this data contains meaningful information beyond the boundaries of the original system. In this sense, LIVE Singapore! does not attempt to solve a specific urban issue, but rather explores the development of new digital tools that enable fundamentally new approaches to how urban issues can be addressed.

Singapore: An Ideal Real-time City?

Singapore is in many ways an ideal setting for the exploration and experimentation of an open real-time data platform for multiple reasons. Singapore is a city-state and as such all territory is one city, both in terms of its dynamics as well as in terms of political administration. Singapore's 5,076,700 inhabitants (based on the 2010 Census) live in an area covering only 710.2 km² of surface area. Its territory is an island connected with the Malaysian mainland via only two bridges for motorized traffic; all other passenger and freight transportation is managed either via sea or air, with Changi Airport being a significant hub in the Southeast Asian region. Singapore is the world's second-largest container port as well as largest transshipment container port in the world. Singapore is a highly developed city, with advanced infrastructures and networks and a tech-savvy population with a cellphone penetration rate of 143.6 percent in 2010 (Source: IDA – Infocomm Development Agency Singapore). Additionally, Singaporeans are familiar with radical changes in their city structure considering the country's rapid growth since its independence in 1965 both in economic as well as in infrastructural terms: Singapore's subway system, opened only in 1987, has now expanded to four lines, 87 stations, and 130 km in length (Source: Land Transport Authority Singapore).

For all these reasons, Singapore is an ideal context for setting up an experimental urban real-time data platform, with an investigation of all the research and development aspects connected to it. Given these special circumstances, the biggest risk is that not all results may be generalizable into other urban contexts.

Open Data Initiatives and Platforms

LIVE Singapore! is inspired by recent open-data initiatives. In May 2009 the US government launched the first of what has since been referred to as data.gov (data.gov 2011) initiatives, making nonsensitive historical data sets publicly available "to easily find, download, and use data sets that are generated and held by the Federal Government." Some of the motivations for "giving data about people back to the people" include fostering economic, scientific, and educational innovation and promoting citizen participation, as well as reducing government costs. This initiative has been replicated in various other countries and cities (e.g., UK, London, San Francisco) and multiple applications have been created based on open data sets available through these platforms.

Beyond public data, platform projects such as Xively allow sharing sensor data over the Web ("data brokerage"), as an infrastructure for the Internet of Things.

An Open-Platform Project

The LIVE Singapore! project builds upon the above-described initiatives. It is a flexible and scalable open platform enabling the collection, combination, and distribution of real-time data urban streams. The project understands itself as an enabling platform for developer communities, who can build applications using combinations of the data feeds available through the platform. As in the case of the data.gov initiatives, we firmly believe in the creative potential of cities and their inhabitants, and let citizens, city planners, companies, and authorities work with live data streams, setting the base for an emerging economy based on new ways to extract value from data.

While our previous projects with real-time data from a single source have yielded interesting results, it becomes increasingly clear how the combination of very different types of data can bring about even richer insights. The data used in this project can be divided into three groups: (1) data as a by-product of existing networks; (2) data collected with tags or sensors; and (3) data actively shared by people.

As such, the platform has to resolve the conflicting requirements to provide data connectors that require little preparation and standardization on the provider side, not to discourage participation, and at the same time to make data streams as interoperable and coherently structured as possible. In this sense, LIVE Singapore! increases the complexity of the urban system, since "complexity is directly related to connectivity. As connections or interconnections proliferate, complexity expands and, correlatively, information increases." (Taylor 2002, 139) This leads to another of the platform's declared requirements, to provide easy access and easy programmability in order to attract a large active developer community.

Meaningful Relations Between Different Data Types Are Created by Users Themselves

Traditionally, the main value of online platform projects is the data that these platforms accumulate and which give them deep insights into their user's activities and preferences. LIVE Singapore! is fundamentally different – it deals with stream data and therefore does not store any data. It is always the data provider who has control over what data are made available, simplifying questions of data ownership, privacy, and security. The platform is therefore structured in a way to make data streams available in a coherent form without storing them. Historic data are therefore hosted on the providers' side and made available through connectors via the platform.

The main value of LIVE Singapore! lies therefore in keeping track of the connections made between real-time data streams by developers and users creating applications. This also leads to a question of semantics related to the connections: what is the value and meaning of the combination of any group of data streams? Our approach lets the users make and describe these connections in order to gradually grow a knowledge management system informing the evolution of the platform.

Feedback Loops

Human activity in cities generates data in multiple technology systems. For example, when riding the subway and tapping an electronic ticket to enter and exit the station, a data trace containing information about your ticket number, the trip start and end time, and the line number is generated. Taken together, this data stream can give insight into how many passengers are on a train at any given time, where train delays effect most passengers, and so on. By presenting this information back to people, they can take decisions that make the system more efficient. As cybernetics pioneer Jay Forrester explains, a feedback loop is created: "In an information-feedback system, conditions are converted to information that is a basis for decisions that control action to alter the surrounding conditions. The cycle is continuous. We cannot properly speak of any beginning or end of the chain. It is a closed loop." (Forrester 1961, 61) While it is certainly still the case that in such a system "it is always the presently available information about the past which is being used as a basis for deciding future action" (Ibid., 15), working with real-time data today as described does move that past significantly closer to the present. While Forrester points out that delays in systems do not necessarily have a negative effect, rather a stabilizing one, the feasibility of shorter delays increases the flexibility and can help to use resources and energy in a more effective way.

This raises important questions: Will real-time feedback loops decrease the stability of the system, or instead might they establish stability? We might take some cues from Donella Meadows when she points out that the interconnections between the system's parts play a special role: "Changing interconnections in a system can change it dramatically." (Meadows 2008, 16)

System Pictures

When dealing with real-time data from urban systems and networks, the SENSEable City Lab has always placed emphasis not only on platform software development and the analysis, but also on the communication of insights embedded in data collected and studied within the urban context. Data visualizations have proven particularly useful in this undertaking. As Meadows describes so well, "Words and sentences must, by necessity, come only one at a time in linear, logical order. Systems happen all at once. [...] To discuss them properly, it is necessary somehow to use a language that shares some of the same properties as the phenomena under discussion. Pictures work for this language better than words, because you can see all the parts of a picture at once." (Ibid., 5) This characteristic of providing the whole picture at once is also the reason why control rooms dealing with large amounts of data show sophisticated diagrammatic representations rather than just numeric information. A second reason for the strength of data visualizations is the additional processing power that can be tapped into on the spectator's side. Seeing an animated visualization of urban real-time data involves the spectator's capability of interpretation. She will bring in prior knowledge and experience for analyzing and filtering the information in the process of observation, similar to a person standing on top of a high-rise building to observe the traffic situation before deciding which way to drive off – in this case the spectator brings in her knowledge of past processes and integrates them with what she sees in real time.

LIVE Singapore! Urban Demos

LIVE Singapore! focuses on making urban real-time data available to citizens, and in this spirit, at iterative steps, prototypes are being developed and shared with the city's public to inspire public discussion about the potential of harnessing this data for the planning, management, and living of our cities.

To this point the first two of these urban demos have been developed: the "LIVE Singapore! exhibition" at the Singapore Art Museum in April 2011 and the "Visual Explorations of Urban Mobility" project showcased at the Future Urban Mobility Symposium organized by SMART (Singapore-MIT Alliance for Research and Technology) at the National University of Singapore in January 2012.

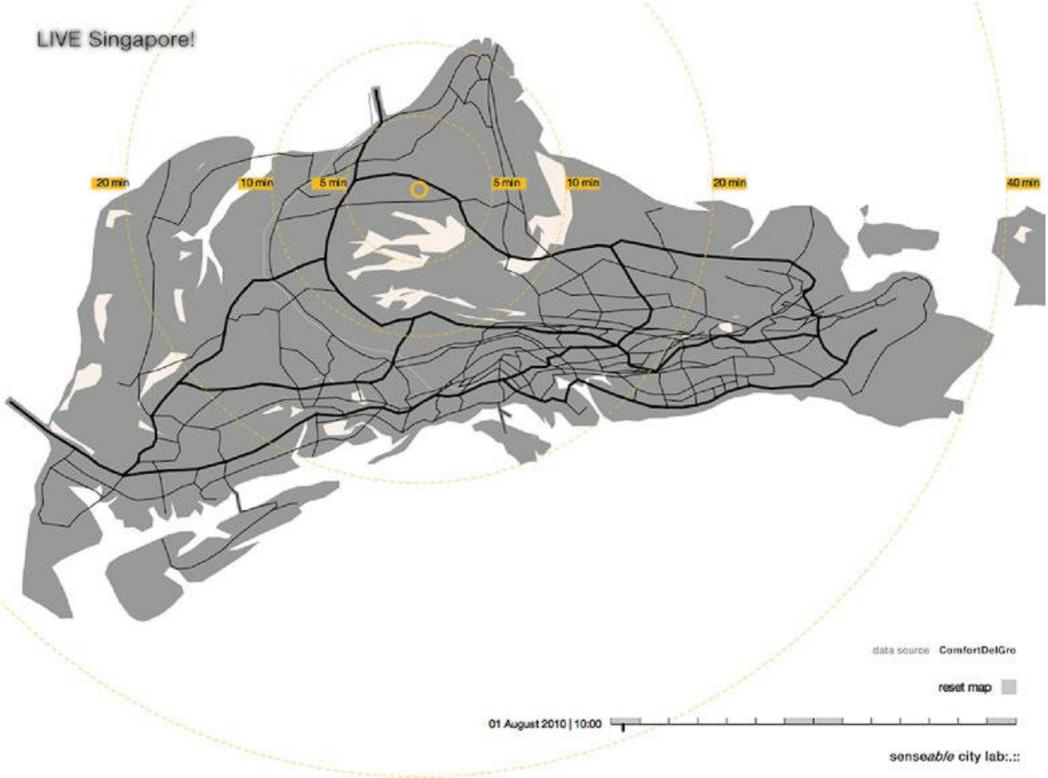


Fig. 1.
Isochronic Singapore: Using GPS location and speed data from the city's largest taxi fleet operator (16,000 taxis) we use this data as an indicator for overall road travel times. As vehicular traffic opens up and jams during the course of the day, the time we need to move in Singapore shrinks and expands. How long will it take you to go from home to any other destination? In this isochronic map the deformations are proportional to travel time – and reveal the changes in the course of a weekend/weekday. Visualization: Xiaoji Chen, SENSEable City Lab

First Urban Demo: LIVE Singapore! Exhibition

The LIVE Singapore! exhibition, staged in the former Saint Joseph's Institution, now an art museum, comprised five projections and one LCD screen showing dynamic, interactive visualizations driven by combinations of data sets and live streams from several key urban networks, including the airport, taxi fleet, environmental agency, container shipping port, cellphone network, electricity network, and the wind speed sensor network.

Second Urban Demo: Visual Explorations of Urban Mobility

Following the showcasing of the prototypes at the Singapore Art Museum and in collaboration with Singapore's Land Transport Authority, we developed three interactive applications that provide insight into the wealth of information that the data generated by Singapore's transportation infrastructure offer. Prototypes of these applications were presented as Visual Explorations of Urban Mobility in 2012 in Singapore.

One of the applications, called Touching Bus Rides, developed within the Visual Explorations of Urban Mobility project, focuses on the public transportation bus network. Singapore's public transport system requires passengers to tap their smart card passes when boarding and alighting subways and buses. While these actions translate into different, distance-based fare prices for travelers, they also offer perspective on the passengers aboard those vehicles at any given time.

An interactive multitouch interface enables users to actively explore Singapore's bus network and see where most passengers get on and off buses, how people connect between the island's stations, and the way these patterns change throughout the day.

Users can switch between various visualization modes to get different perspectives on the same data set. The selected bus line, the highlighted bus stops, and the currently chosen time stay consistent in all modes, while the application seamlessly animates through the different visualizations. The application displays multiple bus service charts with symbols indicating the boarding and exiting of passengers at each station, and allows comparison over time. A user may also select and isolate a single bus route on the map of Singapore. The timeline of this visualization doubles as a histogram representing passenger load. The user can quickly browse through time, altering the range dynamically from 30-minute intervals up to an entire day.

Observations from the Urban Demos

The two urban demos described above aim at showing the value of abundantly available real-time operations data generated in urban space. In the context of our research, the role of such urban demos is manifold: (1) to bring together the various fields of research in the development of iterative working prototypes capable of conveying the project vision; (2) to enable citizens to experiment with the project and prototypes and provide valuable feedback for further development; (3) to provide a concrete milestone for external partners to join in making the project become a reality.

Making these urban demos happen involves the formulation of data-sharing agreements with urban system operators and the coordination and collaboration with their in-house departments

Facing page, top: Fig. 2. **Raining Taxis: Singapore's mobility is heavily reliant on taxis, but what happens when it rains? Getting hold of a cab is not the easiest thing in the world. This interactive visualization helps explore how Singapore's transportation system behaves by combining taxi and rainfall data, and investigating how in the future the system can be streamlined in order to better match taxi supply and demand. Visualization: Kristian Kloeckl, Aaron Siegel, SENSEable City Lab**

Facing page, bottom: **Fig. 3. Urban Heat Islands: As heat is generated by energy usage, high-energy consumption can translate into small local temperature rises of up to a few degrees (called *man-made* or *anthropogenic heating*). Combining data on the energy consumption of the city's different zones with the wind speed, local temperature rise can be estimated. In this sense, a potential addition of measured urban temperatures can provide the basis for a future city condition-monitoring program. Visualization: Xiaoji Chen, SENSEable City Lab**

throughout the process – work that goes beyond our group's usual research focus. This process has provided us with valuable insights into the comprehensive spectrum of challenges and opportunities involved in working with urban data.

Unlocking Data Silos

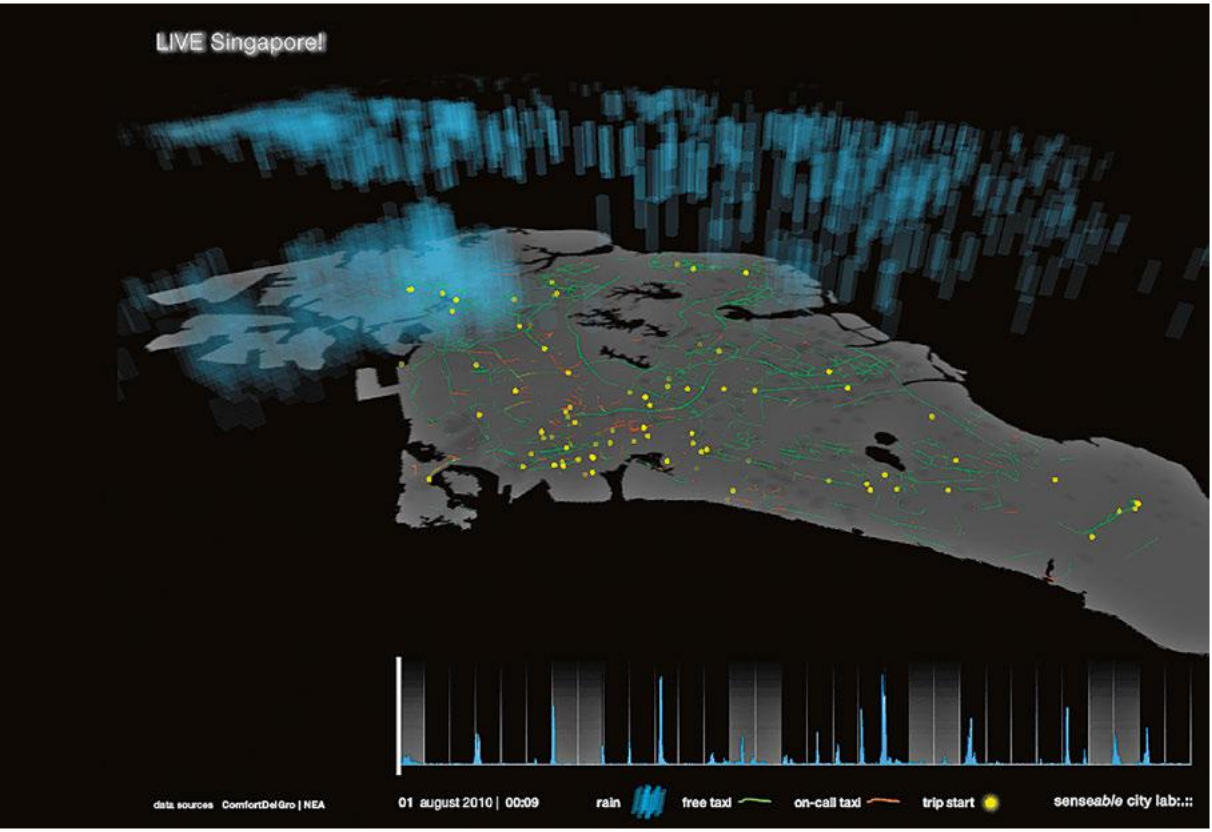
Since the systems have not been designed with the intention of sharing generated data, an understanding of these systems is necessary in order to use and share their data in a meaningful way. Often this creates steep technological challenges – meaning either the data cannot be shared outside the system or it would require a considerable investment to do so. Furthermore, data interoperability is another key issue, not only for sharing data between different organizations and systems, but also within single organizations that use multiple distinct systems. The combination of data generated by systems within a single organization often rivals the complexities of combining data beyond institutional boundaries.

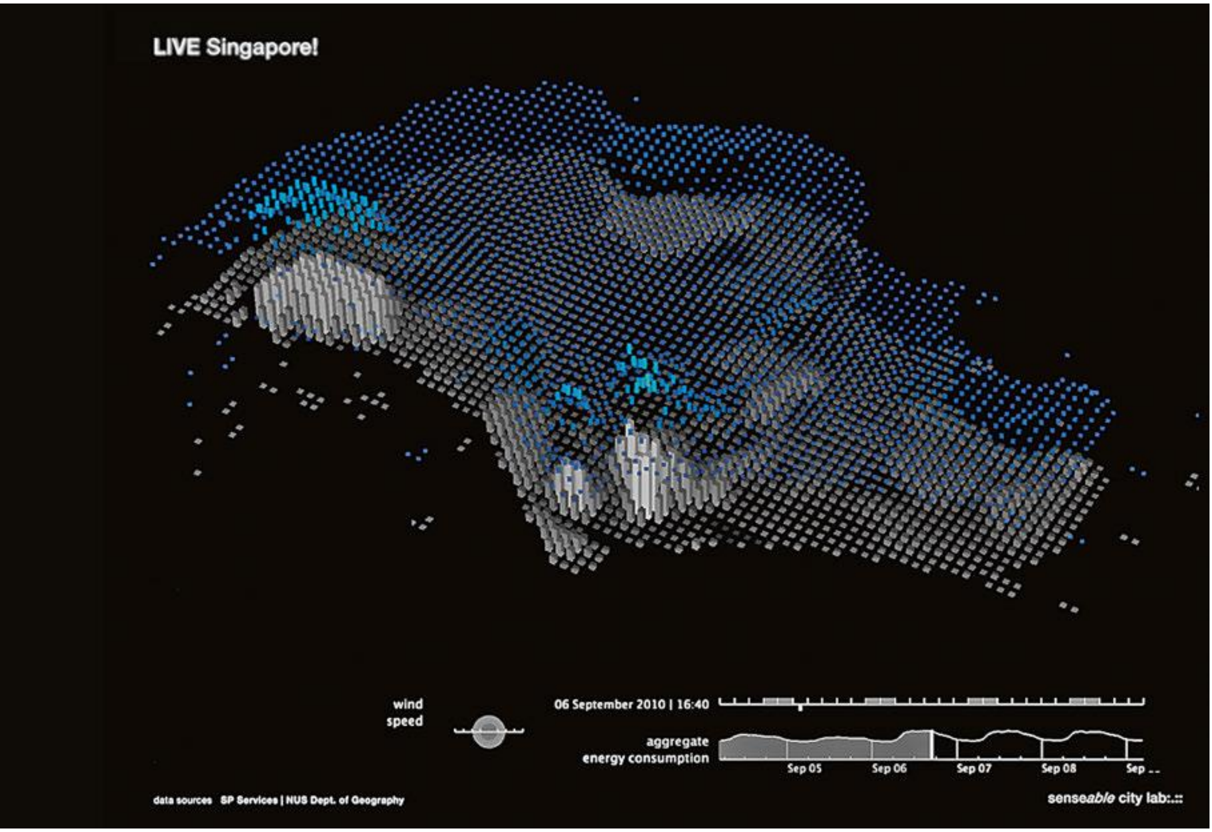
Using data beyond its original purpose requires a thorough understanding of how and in what context that data are generated.[1] For example, if a sensor provides temperature readings on the outside of the building, it matters to know whether it is located near a ventilation outlet, exposed to sunlight, or in the shade. Projects such as LIVE Singapore! can often be an incentive for data providers to dig deep in the modalities of data generation within their organization to fully exploit the data's potential. System data is usually managed by IT departments, which often are even physically distant from the departments of planning and operations. Consequently, those responsible for managing data systems tend to focus on the system's flawless operation from a technology perspective. However, those involved in planning and operation decisions often have little or no access to the information concealed in that data.

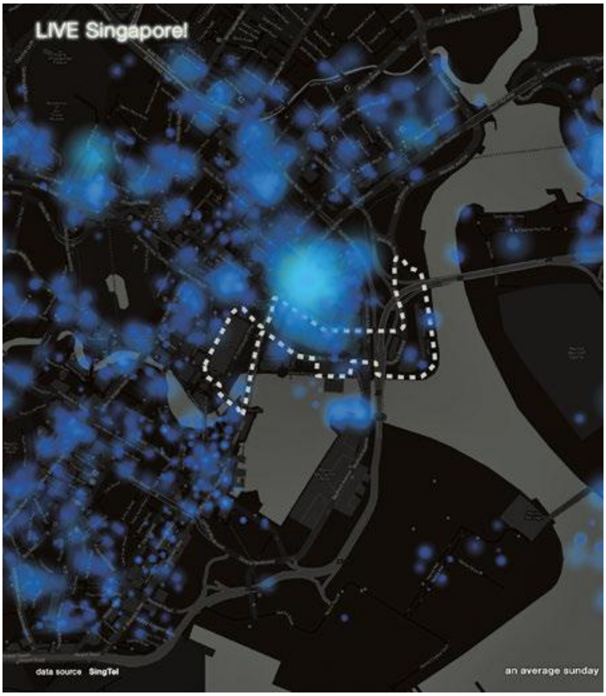
Enabling the Codevelopment of Ideas for New Urban Solutions

Developing meaningful visual representations of urban system data is a critical component of creating access to information concealed within. This enables bringing together stakeholders from a wide range of disciplines and competences to open up a new domain of discussion and creative thinking.

[1] For example in the case of data generated by electronic ticketing in public transportation networks the primary purpose is to grant a passenger access to the transportation system upon the verification of the validity of his ticket.


LIVE Singapore!
data sources ComfortDelGro | NEA
01 august 2010 | 00:09
rain
free taxi
on-call taxi
trip start
senseable city lab:.::


LIVE Singapore!
wind speed
06 September 2010 | 16:40
aggregate energy consumption
Sep 05
Sep 06
Sep 07
Sep 08
Sep ...
data sources SP Services | NUS Dept. of Geography
senseable city lab:.::



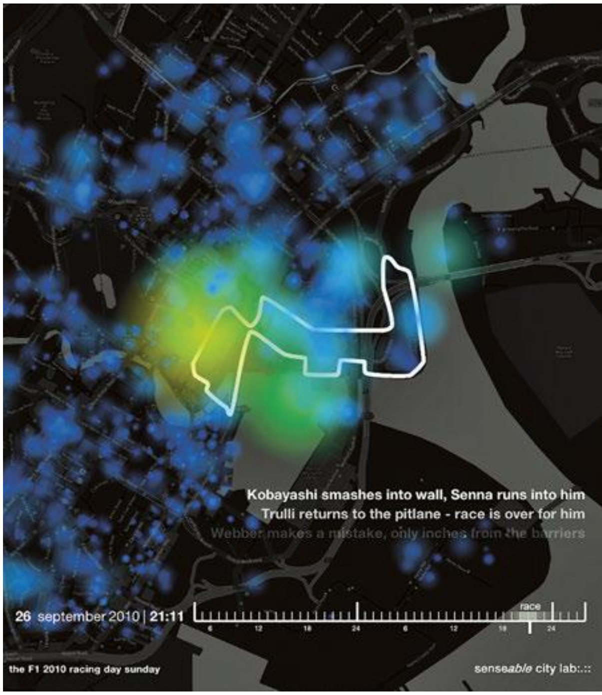
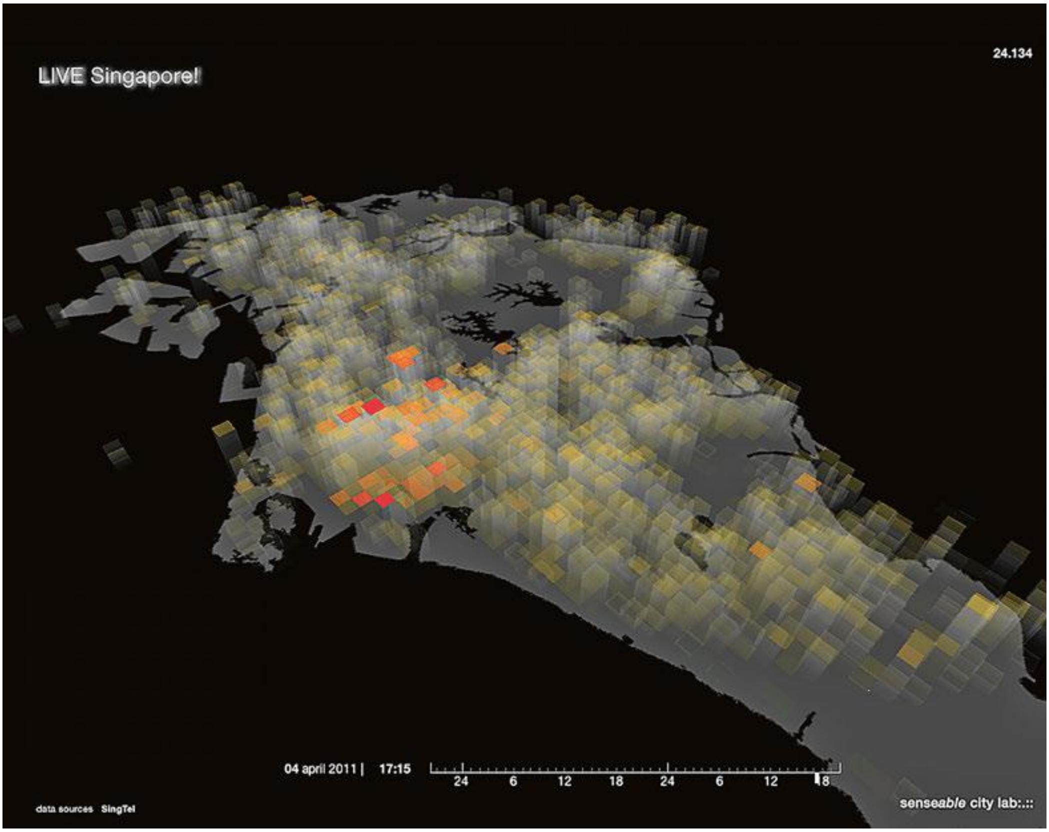
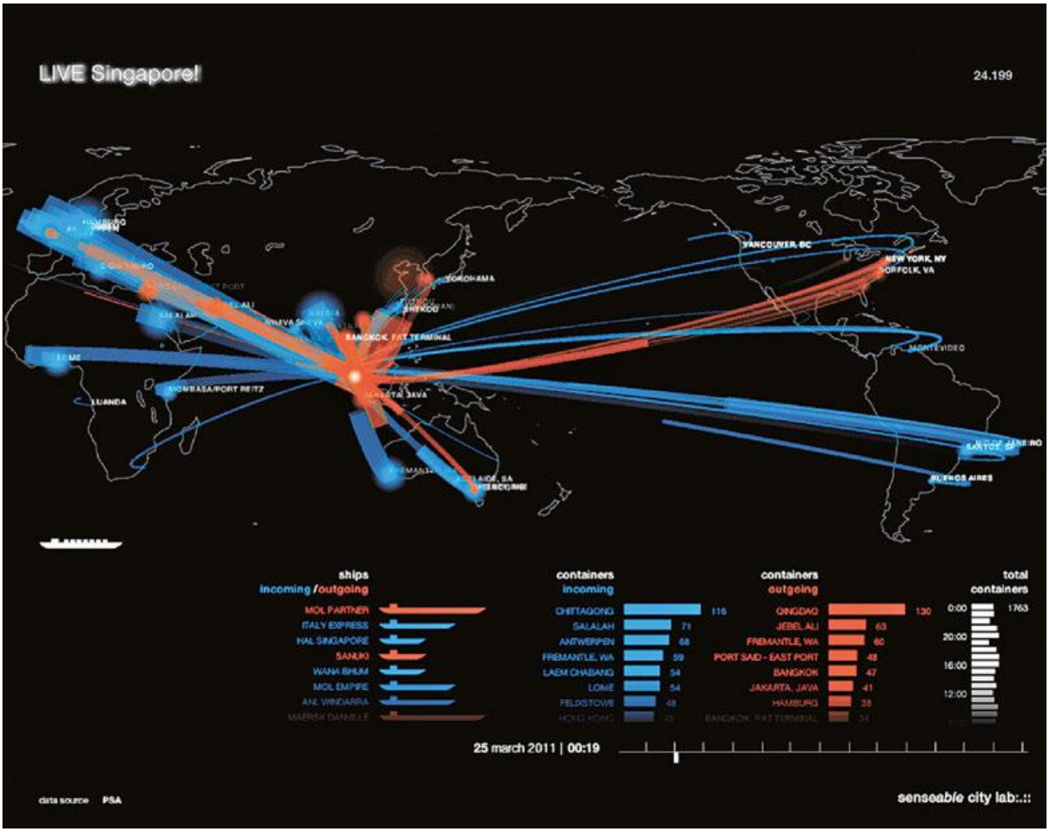


Fig. 4.
Formula One City: Large-scale events disrupt cities daily routines. What better opportunity to explore this effect than Singapore's Formula One Grand Prix? How does this event impact our daily life? How do we respond to it? How do we share our excitement via cell phone? On this dynamic map the color and size of the glow are proportional to the amount of text messaging during the Formula One race. While the right side of the screen shows the race day, the left side compares the same day of a nonrace week. Visualization: Kristian Kloeckl, Oliver Senn, SENSEable City Lab

Facing page, top:
Fig. 5.
Real-Time Talk: Singapore's mobile phone penetration is above 140 percent; many own more than one device. How do we make use of the island's cell phone network via voice calls and text messages? How can this inform us about the usage of urban space in real time? In this visualization height (logarithmic scale) and color (linear scale) intensity indicate the level of cell phone network usage. Visualization: Aaron Siegel, Kristian Kloeckl, SENSEable City Lab

Facing page, bottom:
Fig. 6.
Hub of the World: Singapore is the world's largest transshipment container port and one of the busiest airport hubs in the world. The visualization addresses questions such as: How does this constant stream of people and goods passing through affect Singapore? Where do these flows come from and go to and how many of them are here to stay? Visualization: Kristian Kloeckl, Aaron Siegel, SENSEable City Lab

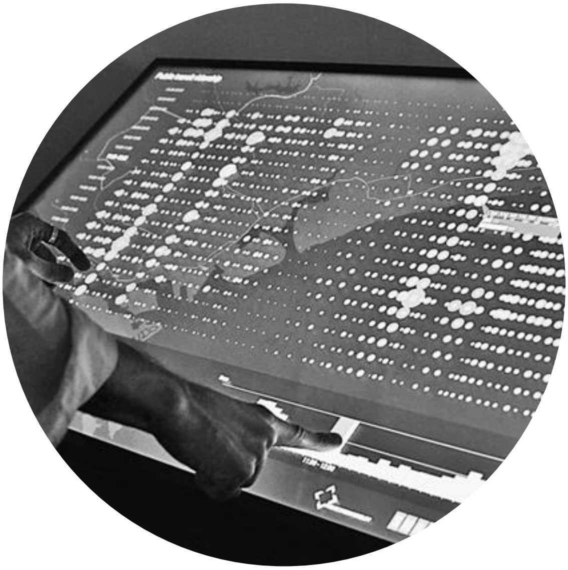

LIVE Singapore!
24.134
04 april 2011 | 17:15
24 6 12 18 24 6 12 18
data sources SingTel
senseable city lab:.::


LIVE Singapore!
24.199
VANCOUVER, BC
NEW YORK, NY
NORFOLK, VA
YOKOHAMA
SHEKOU
BANGKOK, PAT TERMINAL
MONTEVIDEO
LOME
LUANDA
MOMBASA/PORT REITZ
JAKARTA, JAVA
BUENOS AIRES
ships
incoming /outgoing
MOL PARTNER
ITALY EXPRESS
HAL SINGAPORE
SANUKI
WANA BHUM
MOL EMPIRE
ANL WINDARRA
containers
incoming
CHITTAGONG 116
SALALAH 71
ANTWERPEN 68
FREMANTLE, WA 59
LAEM CHABANG 54
LOME 54
FELIXSTOWE 48
containers
outgoing
QINGDAO 130
JEBEL ALI 63
FREMANTLE, WA 60
PORT SAID - EAST PORT 48
BANGKOK 47
JAKARTA, JAVA 41
HAMBURG 38
total
containers
0:00 1763
20:00
16:00
12:00
25 march 2011 | 00:19
data source PSA
senseable city lab:.::

An example for this is the ChangiNOW initiative, which was born out of the first of the urban demos described above. Once we had initiated the work with data from a taxi company and the airport, the idea was born to bring airport data of actual passenger arrivals together with free taxi availability, to better match demand and supply at an airport's taxi stands – an often overlooked example of intermodal trip connections.

As another example, the Touching Bus Rides touch-table brought multiple stakeholders together. Their discussions resulted in ideas for three applications: (1) creating dynamic bus lanes, activated based on the number of bus passengers stuck in slow-moving traffic; (2) the establishment of dynamic express-bus lines that would make direct connections between distant stations whenever real-time demand patterns would indicate such a need; and (3) the creation of dynamic information boards at bus stops to inform waiting people of the occupancy of the next bus in addition to its arrival time.

Fig. 7.
Fig. 8.
Photos from the LIVE Singapore! exhibition opening on April 7, 2011 at the Singapore Art Museum (SAM) and the 2012 Future Urban Mobility Symposium



This interdisciplinary way of working with urban systems data reflects the role of data as a boundary object, in that "they have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation" (Star and Griesemer 1989). As an example, data generated by a telecommunications network might be investigated to inform about diverse aspects such as the system's status, existing social networks among the subscriber base, spatial mobility dynamics, and economic activity, to name but a few. In this way, working with urban data through interactive visual interfaces and across disciplinary boundaries enables the formulation of new questions about urban dynamics that go beyond the resolution of mere optimization problems. While substantial research and development has been carried out in recent years to make this become a reality, some of the key challenges in moving forward will be the definition of ownership of data as well as of any platform capable of facilitating its consultation and exchange in order to successfully represent a city's public dimension in the digital domain.

Forrester, Jay Wright. 1961. *Industrial Dynamics.* Cambridge, MA: The MIT Press.

Meadows, Donella H. 2008. *Thinking in Systems: A Primer,* edited by Diana White. White River Junction, VT: Chelsea Green Publishers.

Murty, Rohan, Geoffrey Mainland, Ian Rose, Atanu Roy Chowdhury, Abhimanyu Gosain, Josh Bers, and Matt Welsh. 2008. "CitySense: An Urban-Scale Wireless Sensor Network and Testbed." In *Proceedings of the 2008 IEEE International Conference on Technologies for Homeland Security, May 12–13, Waltham, MA*. Stoughton, WI: The Printing House, 583–588.

Paulsen, Hinrich, and Uwe Riegger. 2006. "SensorGIS – Geodaten in Echtzeit," *GIS-Business* no. 8: 17–19.

Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39," *Social Studies of Science* 19, no. 3 (August): 387–420.

Taylor, Mark C. 2002. *The Moment of Complexity: Emerging Network Culture.* Chicago: University of Chicago Press.

Zardini, Mirko, and Wolfgang Schivelbusch. 2005. *Sense of the City: An Alternate Approach to Urbanism.* Baden: Lars Müller Publishers.

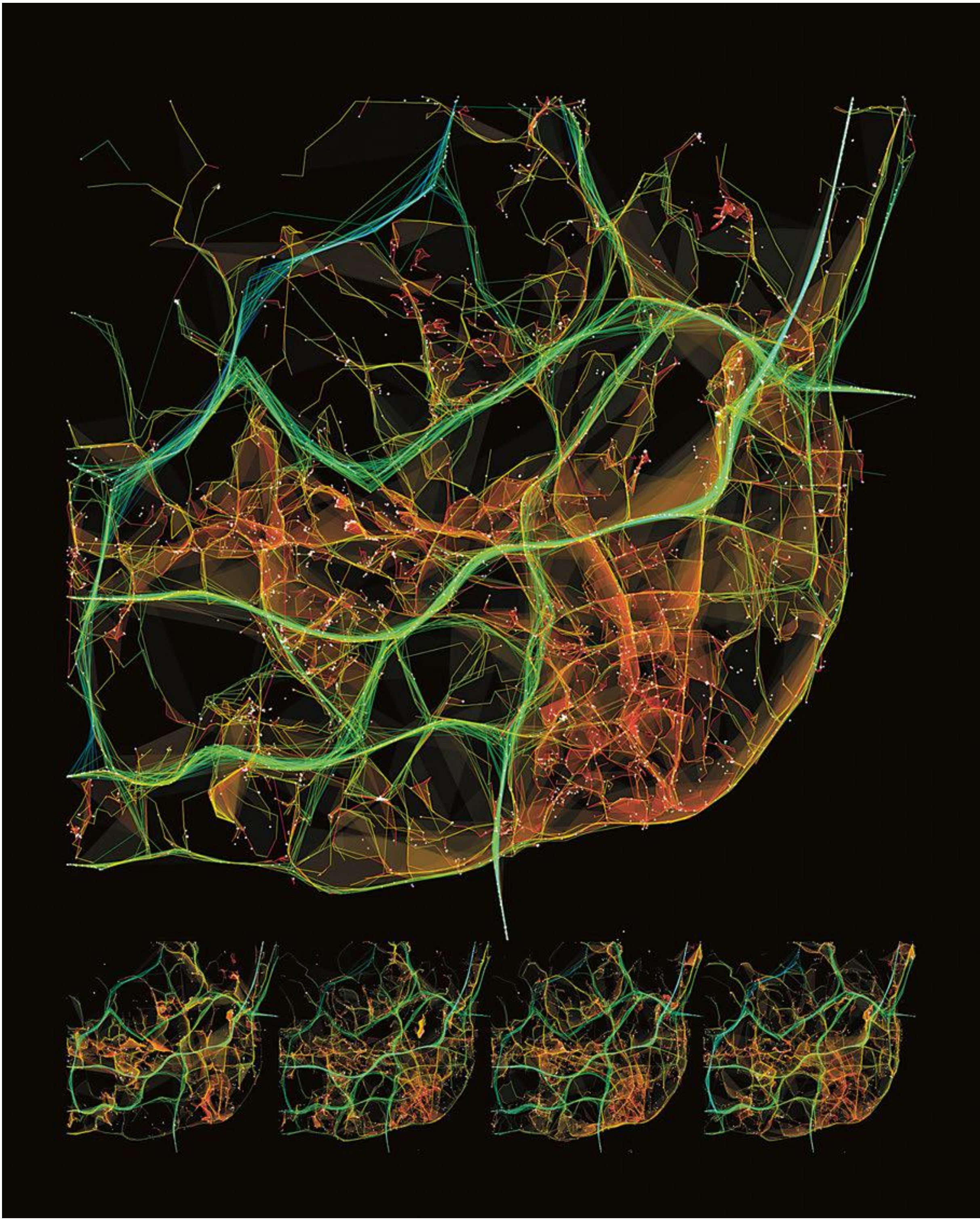
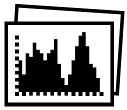


Fig. 1.
Top: **Photograph of the traffic of Lisbon at 9:00 a.m., rush hour**
Bottom: **Evolution of the city during the day: different areas are illuminated, and the color of the arteries changes. Visualization: Pedro Cruz, Penousal Machado**

Pedro Cruz, Penousal Machado

City Portraits and Caricatures



Introduction

Cities are a very rich source of data in order to understand how people organize themselves on different scales. The discourse on the visualization of cities offered different approaches, from more abstract to illustrative and figurative ones. In this chapter, we approach the visualization of cities in a less abstract, puristic way. Distancing ourselves from scientific visualization, we embrace what we call a *figurative approach to visualization.* With this approach we take a step back from the idea that visualization has to represent data only in the most direct and abstract way. The figurative approach extends the concept of portraits and caricatures into the realm of data, using visual metaphors and introducing visual distortions for emphasizing certain aspects of the data. This permissiveness enables us to build visualization artifacts that use playful analogies to better communicate the complex nature of cities to a nonacademic, general audience.

This article will describe the process that took two visualization cases from a direct to a figurative approach, not entering into detailed technicalities, but tying it with a cogent discourse that presents and applies portraits and caricatures in the context of data visualization.

A Figurative Approach to Visualization

Information visualization is an interdisciplinary area that intersects graphic design, human-computer interaction, computer graphics, and data mining. Its objective is to synthesize large amounts of data for broad audiences, creating and clarifying messages from data. The figurative approach described herein uses a taxonomy of photographs, portraits, caricatures, and disfigurements to describe distinct approaches to visualization. The main concept behind such taxonomy is degrees of authorship. We refer to authorship as the varying intent with which a particular perspective over the data is made visible by its creator. Authorship flourishes in contemporary information visualization, mostly under the term "information art" and is also embedded into scientific visualizations. It is important to notice that visualization artifacts that go beyond the sole purpose of analyzing patterns in data are often seen as distortions of the truth, especially in the context of academic visualization. While this is a legitimate argument under a puristic perspective, the caricature of data is conceptually inevitable, as Fernanda Viégas notes: "traditional analytic visualization tools have sought to minimize distortions, since these may interfere with dispassionate analysis. Is it possible that this focus on minimizing 'point of view' is misguided? For one thing, it is generally impossible to create a visualization that is truly neutral, just as it is impossible to create a flat map of the Earth's surface without distorting distances."
(Viégas and Wattenberg 2007, 190)

Considering that authorship is always present to some degree in data visualization, our taxonomy identifies four major ways to undertake data visualization characterized by an increasing degree of authorship: photographs, portraits, caricatures, and disfigurements. But authorship is not the only characteristic to describe these concepts and tie them together in a continuous fashion. The figurative approach focuses on the idea that visualization does not have to invest mainly in abstract aesthetics, but can carry strong visual metaphors and exaggerate certain data features in its communication language. Since information visualization is part of the broader data visualization field, which also encompasses scientific visualization and, arguably, artistic visualization, this seemingly unscientific taxonomy describes this spectrum within visualization from the more scientific to the more artistic.

Data Photographs

In our taxonomy, photographs are the most direct reproduction of a data set – models that are the closest to a one-to-one mapping of data. This notion refers to Lev Manovich's concept (2011) of visualization without reduction or direct visualization. Instead of mapping data into a symbolic abstraction via Jacques Bertin's visual variables of position, size, value, texture, color, orientation and form (Bertin 1967), direct visualization preserves the nature of the data. What is text in a data set remains text on the representation space; what is an image remains an image; what is speech remains speech, and so forth. Although not deprived of authorship, because a photograph can be taken from different angles, they intend to faithfully reproduce the subject while maintaining every characteristic of it, in this case, data.

Data Portraits

In data portraits, the author has a more pronounced role than in photographs, by recurring to visual metaphors in the representation. Visual metaphors herein do not account for merely decorative visual elements, described as "chartjunk" by Edward Tufte (1983), but refers to graphical elaborations that have a close semantic meaning toward the data set and the author's message. These semantic metaphors are closely related to Donna Cox's visaphors (Cox 2006) in the sense that they are approximations of data, more dependent on subjective interpretations. A semantic visual metaphor is a figurative evidence of certain characteristics in data in addition to the ones directly mapped. Such evidences are elaborated on a graphical level and result in less abstract, often more familiar, natural and expressive artifacts.

In addition, visual metaphors carry a sense of novelty and uniqueness, that nevertheless can get lost once embedded into our culture. For example, the pie chart, first introduced by William Playfair (1801)[1] was at that time a novelty that has been lost with their pervasive use. Pie charts carry a tender visual metaphor that conveys a sense of unity of the parts in a whole.

[1] William Playfair, *The Statistical Breviary; Shewing the Resources of Every State and Kingdom in Europe* (London, 1801).

Data Caricatures

We regard data caricatures as an extension of data portraits since they also take advantage of semantic visual metaphors. They have, though, its distinctive features that we explain below.

A caricature is often regarded as a figurative representation of a subject that exaggerates its prominent features in order to improve recognizability (Redman 1984). This augmented recognizability is a consequence of what is called the "peak shift effect": an individual trained with a visual representation *r,* shows a response to a similar representation *b* that increases with its differences compared to *r* (Ramachandran and Hirstein 1999). Such a reference model *r* is an essential part of a caricature, which is always present, either physically or as a mental image.

The concept of caricature in a data visualization context comes with certain restrictions. As explained, caricatures depend on a mental image as a reference. Such a reference can generally not be taken for granted in data visualization since even the most direct visualizations introduce a new form for previously unvisualized data. Therefore, the application of caricatures in a visualization context depends on the familiarity of the reference model for its caricatured representation to allow for comparing the differences.

As suggested above, caricatures are tied with the concepts of exaggeration and recognizability. Exaggeration in the context of data means an overemphasis of one data dimension over the others. This can be graphically attained in several ways beyond the amplifying numerical differences in the data, as done in the work of Peter Rautek (Rautek et al. 2006), resulting in distortions of form, position, size, or color of a visual element toward the reference. Recognizability means, in the context of data, the clarification of a caricature's intent, highlighting messages derived from the data. While a data caricature is not necessarily superior in clarifying a message compared to a photograph or a portrait, we nevertheless believe that this is the case for many visualization solutions. To summarize, a data caricature is a visualization model that graphically distorts a reference representation of a data dimension in order to emphasize that same or other depicted dimension.

An obvious way to implement data caricatures in visualization is through the distortion of geographical positions. This principle has been applied since the nineteenth century with cartograms. Cartograms distort geographic maps in order to represent other data dimensions. For example, area cartograms resize countries in a world map proportionally to their population or GDP. This idea was simplified by Daniel Dorling, famous for the Dorling Cartograms (1996) – these cartograms do not maintain the shape of geographical objects, but usually replace them with circles of a size proportional to the mapped dimension. Although they oversimplify geographical shapes, Dorling Cartograms have proven to be very effective. The caricatural property of such methods lies on the emphasis given to a certain data dimension (population, for example) against a geographic map. This map is the reference model that is either fully represented or that can be mentally reconstructed.

Data Disfigurements

Data caricatures introduce distortions in the representation of data. When such distortions are exaggerated beyond certain levels, they can result in data disfigurements. Data disfigurements overemphasize certain dimensions to a point of communicating erroneous messages toward the data. They corrupt the semantic value of visual metaphors, undermining the clarification intent of visualization, and even producing unintelligible artifacts. Nevertheless, the exaggeration of distortions gives more room for authorship and potentially more memorable artifacts.

Figurative Visualization of Cities

The case studies in which we implement direct and figurative approaches to visualization refer to mobility systems of the cities of Lisbon and Singapore. Lisbon's data set[2] contains GPS traces of vehicles in the city during one month, with position and current speed indication. The Singapore data set includes information about people boarding and leaving buses (so-called tap-ins and tap-outs) at bus stations with its respective paid fares in the city of Singapore over one week.[3]

Photographing Lisbon

The spatial and temporal resolutions of Lisbon's data set did not allow a visualization that clearly depicted traffic patterns for each

[2] Data provided by the CityMotion project/MIT Portugal. Visualization project partially supported by the project PTDC/EIA-EIA/108785/2008 COSMO – COllaborative System for Mobility Optimization.

[3] Data provided by the Land Transport Authority in Singapore. Visualization developed for the LIVE Singapore! initiative, supervised by Kristian Kloeckl at MIT SENSEable City Lab and SMART (Singapore-MIT Alliance for Research and Technology).

day separately. Therefore, the information was condensed into a single virtual day, grouping the data by the second and displaying it as an animation. In order to enhance the temporal patterns in traffic even more, each vehicle, represented by a small white dot, leaves a trail that remained visible for 30 minutes in simulated time. The trail is almost transparent and colored accordingly with the speed of the vehicle. Using a restricted color scale of pure hues, the colors red and orange indicated slower velocities, pure green for 50 km/h, and the cyan tones for higher velocities. The trails tend to visually cluster into thicker lines for major arteries that emerge with mixed hues and opacities, representing traffic density (by thickness and opacity) and average speeds (by hue) at that point in time. For example, narrow streets tend to be thinner and red, while highways are thicker and green. The color of the highways that cross Lisbon changes to more yellow hues at rush hours.

Visualizing a data set usually involves looking for problems in the system being depicted: the most apparent salient features in Lisbon's data set are the congested areas. In order to emphasize those, another visual component was added that draws with a very low opacity the area covered by each vehicle in 30 minutes. The covered area is defined by closing the route, connecting its origin and destination points; its opacity is higher as the velocity lowers, and the hue is the same as the corresponding trail. This representation emphasizes the areas with slower traffic, being more opaque orange and red (speeds that get closer to green are almost transparent in this representation). Drawing such areas pinpoints the most problematic areas in Lisbon during the day, which is an easier perceptual task rather than trying to delineate those areas by only extracting colors of lines – it is notorious, for example how downtown keeps being illuminated during daytime, and how peripheral areas activate before the rest of the city. More than directing the audience's attention, it also adds visually to the artifact, making it more detailed and at the same time more compelling (figure 1). All the superimposition of semitransparent lines and shapes leaves little room for the visual extraction of concrete data points, such as the temporal velocity of a certain vehicle, but it is a direct representation, a photograph, of data and provides an overall picture of the traffic evolution during a day in Lisbon.

Due to the high level of detail, the artifact had to be rendered offline and assembled into a single animation afterward. However, it is sometimes of interest to have a real-time visualization, making it possible to pinpoint problematic issues as they unfold. The following project on Singapore addresses such intentions.

Photographing Singapore

The data set concerning Singapore's bus network describes stops at bus stations, but provides no information about the trajectories between stations. Nevertheless, we consider that the representation of such trajectories could provide a much more tangible artifact and an interesting visualization challenge: representing buses moving on a network, and not just buses at bus stations. To attain this, buses were implemented as programmed automatons that simply respond to inputs from a simulated environment (reactive agents), where data constitutes the environment. This also permitted the fulfillment of another objective: creating an interactive visualization that runs in execution time, enabling the rapid visualization of the outcomes of the applied techniques.

Each bus, as an agent, only knows its next stop, its arrival time, and the current simulation time. Data is injected on execution time into the simulation environment, redirecting each tap to the corresponding bus. Each bus has a buffer of the next stops, each in charge of retrieving the next bus stop when the time has come, as well as removing passed bus stops. Knowing only the arrival time, the next stop, the current position, and the current time, the bus travels to the next stop. Many approaches could be used to model the movement of each bus between bus stops: the first approach implemented a nonlinear movement, varying with the time that the bus has to reach the next stop. It was named "lazy buses" because the agents, in spite of knowing that they had to be at a certain stop at a certain time, start their travel slowly, but rush as the time urges (the distance traveled between two stations is inversely proportional to the squared time for the destination).

Beyond the portrayal of buses' travels, it was also interesting to depict which were the more congested. The density of bus stations in the visualization canvas did not allow for the labeling of each of them at the same time. A geographical clustering algorithm

Fig. 2.
Results of the clustering algorithm with a search radius of 44 meters and a minimum of 5 buses per cluster

was used to pick which stations were the most congested, with more than *N* buses stopped at the same time. Our approach to such an algorithm (Ester et al. 1996; Finkel and Bentley 1974) was capable of accommodating a good performance, running in execution time for the amounts of data being used. The first results of the clustering as well as the lazy buses' behavior implementation are depicted in **figure 2**. Small gray crosses represent all the bus stops, and triangles the actual buses, with the triangle indicating the current direction. Only the bus stations that correspond to clusters of buses at that instant are labeled.

Caricaturing Lisbon

In order to build a caricature of Lisbon's data set, we had to first work on its portrait and therefore a semantic visual metaphor. We see cities as systems with complex behaviors, but which have their own identity by shape, normal activity cycles, and abnormal activity bursts that should be diagnosed. On that account, and also on an authorial aesthetic choice, the traffic in Lisbon is portrayed exploring the metaphor of a living organism with circulatory problems by building a system of blood vessels.

A computational structure, a form-preserving "skeleton" subject to simulated elastic, spring-based forces, was generated for each major road in Lisbon. The skeletons are connected when roads intersect and also affect other neighboring areas when they move. This allowed the city to change its shape accordingly with the traffic on the arteries. Springs, as elastic devices, are able to change their shape with some resistance when reacting to a force. If changes on data values act as forces for those springs, then we are able to attain smooth transitions between shapes, as well as to naturally return to the initial shape configuration when the forces are null. Those shape mutations translate the actual perceived distances within a city, contrasting with the common geographical mapping (that is, for our caricature, the reference model). This type of mapping is a classic caricatural visualization since it is a type of distance cartogram – such cartograms are common in the form of isochronic cartograms, mapping all the geographic distances in function of travel time from a well-defined origin point. In our case we wanted to present how the travel times evolve for each artery at the same time and also how changing these distances alters the overall shape of the city.

We are dealing with a rather complex spring-based physics system – as previously mentioned, each spring distends or compresses according to the data; in this case, each spring is also connected to others, forming a skeleton. As a consequence, each spring cannot change its shape freely, being also influenced by connected springs and other data values. With so many frequently opposite forces being applied to the same elastic devices, there is a high risk of rupture – the skeletons collapsing in configurations that make unintelligible images of the city, and by extension, disfigurements. Therefore, we had to ensure temporal stability in order to attenuate abrupt variations, leaving time for the system to adapt to new variations and avoiding ruptures. That way, data was averaged to a single day, and only aggregations of one hour with intervals of

10 minutes were used to influence the system and produce the visualization (i.e., from 3:20 to 4:20, from 3:30 to 4:30, and so on). The system is then excited in the following ways: a greater number of vehicles on a vessel tends to make it thicker, and higher speeds tend to contract its length (and vice versa). The latter behavior was chosen in order to transmit a global impression of the perceived distances within the city. This behavior shrinks the city when the traffic velocities are higher, and distends it in the rush hours when the city faces congestion problems. The vessels were also colored accordingly: lower average speeds imply the darkening of a vessel, expressing slower circulation and stagnant blood **(figure 3)**. The vessels' visualization results in an artifact with crude aesthetics that is innate to our visual metaphor, pulsating for each rush hour and stressing which roads are more congested. The compression and distention of each artery and the entire city caricature shows the perceived distances and deviations from the average speeds in traffic.

Caricaturing Singapore

Another classical technique that exhibits caricaturist features is the fish-eye view, introduced by George Furnas (1986) and later applied to maps (Keahey and Robertson 1996). This technique augments a geographical region of interest, while distorting its periphery in order to maintain other parts of the map on the canvas. The main difference from cartograms is in how the data caricature works: cartograms distort one dimension (geographic positions) to emphasize another (e.g., GDP), while in fish-eye views the same dimension (geographic positions) is distorted to emphasize a restricted set of the same dimension (region of interest). In addition, fish-eye views also carry a visual metaphor that makes a device to build data portraits: the magnifying glass.

Caricaturing Singapore's data set followed this approach. We used a data lens to solve the problem of displaying the details of each bus stop on a street level while keeping a global context. Nevertheless, the magnification provided by a classic fish-eye view is not sufficient for the magnification requirements of this visualization. To attain it, a new type of lens was developed, one that transforms the behavior of a magnifying glass (typically the

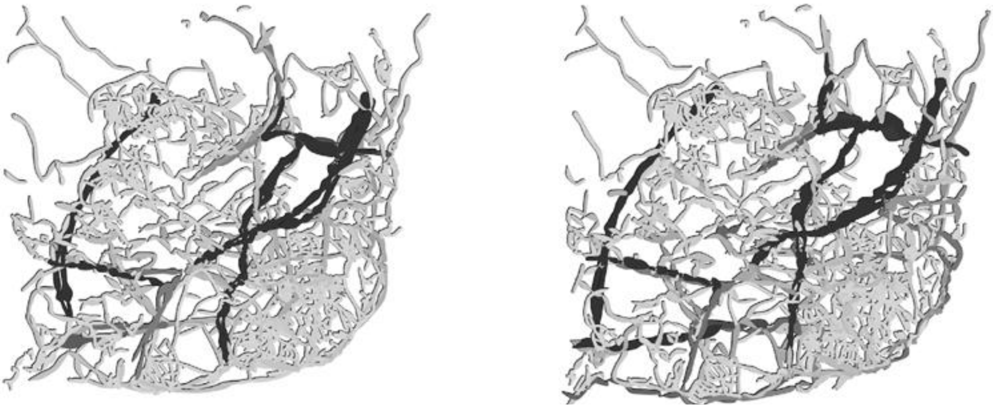


Fig. 3.
Blood vessels in Lisbon at 7:04 a.m. (*left*) and 8:44 a.m. (*right*), just before and after the morning rush hour. It can be observed that at the beginning of the rush hour, the main vessels of Lisbon carry a high number of vehicles, but without traffic congestion problems as the vessels are contracted. At 8:44 a.m. the average speed in the main vessels decreases, originating an expansion of vessels and of the entire city.

projection of the map on an hemisphere on top of the plane) into a projection on a very oblate ellipsoid on top of the plane (figure 4).

The implemented data lens is an interactive visualization tool that allows the user to uncover layers of information that reveal at greater detail the activity of Singapore's bus network. The lens can be dragged over the city's bus network or remain focused on a set of bus stops. The properties of the lens that a user can modify (the position, size and zoom level) help distinguish even very nearby bus stops. In addition, the lens provides access to every feature in the data set by filtering several types of information layers: the line numbers of buses currently stationary at each bus stop, the number of passengers on each bus, and the aggregate paid fares by passengers alighting at any one stop. Users can rapidly switch between these information layers to explore correlations between bus lines, stops, waiting times, passenger load, and paid fare prices.

Remarks

Data caricatures aim at pushing the boundaries in contemporary information visualization by balancing the distance from the dispassionate depictions of scientific visualization with the verge of artistic visualization. This balance is maintained by regarding data caricatures as instruments for design with a clarification intent – to communicate effectively and efficiently with an audience.

Data caricatures are framed in a figurative approach together with photographs, portraits, caricatures, and disfigurements for visualization. A photograph is mainly about direct mappings, or

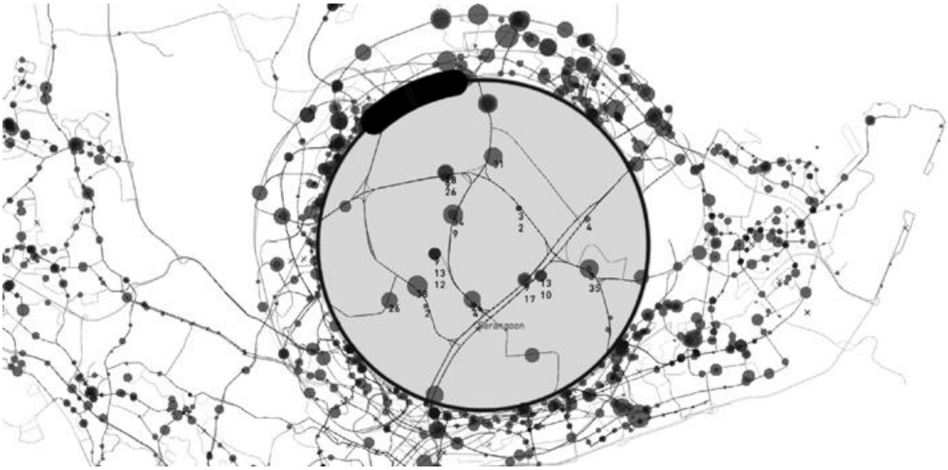


Fig. 4.
This implementation results in a much more diverting and interesting browsing of the space, where the user can direct the lens to a cluster of points, and fluidly unveil and understand each of its constituent parts.

visualization without reduction. Portraits differ from photographs by introducing strong visual metaphors. Visual metaphors have a semantic intent, constituting a figurative evidence of certain characteristics in data in addition to the ones directly mapped. Such evidence is elaborated on a graphical level and result in less abstract, but more expressive artifacts. A data caricature embraces this idea of semantic visual metaphors, extending it with exaggerations in order to emphasize aspects of data. Such exaggerations can culminate in pure distortions by introducing coarse inaccuracies – data disfigurements.

We implemented and applied the notion of caricatures to the visualization of cities. We believe that by using a caricatural approach to the portrayal of cities we can approximate audiences by using strong metaphorical semantics and delivering more concrete messages. For example, inhabitants of a city already have a mental image of certain aspects of it – if we compare such an image to the representation of an ideal model, then we have a distortion that is of interest to convey and where caricatures can adequately be used. Furthermore, such distortions can also be a natural consequence of visualization models that follow an approximation strategy to complicated mathematical problems. Such as in our work, agent- or physics-based models that smoothly adapt to data are much more generic than instantly reacting ones, enabling us to recycle them to other cities and even other types of data. Indeed, such models are ideal for a caricatural approach since they can be directly traduced to nature-based aesthetics, which, accordingly to Greg Judelman (2004), are often used in the context of visualization to decode complexity. Such as in nature, where complexity exists in various scales, the usage of nature-inspired systems seems to be a natural caricatural approach to cities, depicting information with different levels of granularity while maintaining the essence of the message. With this, we believe that caricatures are capable of playing a major role in information visualization, augmenting efficacy, efficiency, and memorability of communication, presenting, and approximating cities to people.

Bertin, Jacques. 1967. *Sémiologie graphique: les diagrammes, les réseaux, les cartes.* Paris: Mouton & Gauthier-Villars. Published in English as *Semiology of Graphics,* translated by William J. Berg. Madison: University of Wisconsin Press, 1983.

Cox, Donna. 2006. "Metaphoric Mappings: The Art of Visualization," in *Aesthetic Computing,* edited by Paul Fishwick. Cambridge, MA: The MIT Press, 89–114.

Dorling, Daniel. 1996. *Area Cartograms: Their Use and Creation.* Concepts and Techniques in Modern Geography 59. Norwich, England: School of Environmental Sciences, University of East Anglia.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, August 2–4, Portland, OR,* edited by Evangelos Simoudis, Jiawei Han, and Usama Fayyad. Menlo Park, CA: AAAI Press, 226–231.

Finkel, Raphael A., and Jon Louis Bentley. 1974. "Quad Trees a Data Structure for Retrieval on Composite Keys," *Acta Informatica* 4, no. 1: 1–9.

Furnas, George W. 1986. "Generalized Fish-eye Views," *ACM SIGCHI Bulletin* 17, no. 4 (April): 16–23.

Judelman, Greg. 2004. "Aesthetics and Inspiration for Visualization Design: Bridging the Gap Between Art and Science." *In Proceedings. Eighth International Conference on Information Visualisation, 2004, July 14–16, London, England.* Los Alamitos, CA: IEEE Computer Society Press, 245–250.

Keahey, T. Alan, and Edward L. Robertson. 1996. "Techniques for Non-linear Magnification Transformations," in *Proceedings IEEE Symposium on Information Visualization '96, October 28–29, San Francisco,* edited by Stuart Card, Stephen G. Eick, and Nahum Gershon. Los Alamitos, CA: IEEE Computer Society Press, 38–45.

Manovich, Lev. 2011. "What Is Visualisation?" *Visual Studies* 26, no. 1.: 36–49.

Playfair, William. 1801. *The Statistical Breviary; Shewing the Resources of Every State and Kingdom in Europe.* London: T. Bensley, for J. Wallis et al.

Ramachandran, Vilayanur S., and William Hirstein. 1999. "The Science of Art: A Neurological Theory of Aesthetic Experience," *Journal of Consciousness Studies* 6, nos. 6/7 (June/July): 15–51.

Rautek, Peter, Ivan Viola, and M. Eduard Groller. 2006. "Caricaturistic Visualization," *IEEE Transactions on Visualization and Computer Graphics,* 12, no. 5 (September): 1085–1092.

Redman, Lenn. 1984. *How to Draw Caricatures.* Chicago: Contemporary Books.

Tufte, Edward R. 1983. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

Viégas, Fernanda B., and Martin Wattenberg. 2007. "Artistic Data Visualization: Beyond Visual Analytics," in *Online Communities and Social Computing: Second International Conference, OCSC 2007, July 22–27, Beijing China,* edited by Douglas Schuler. Berlin: Springer, 182–191.

Philipp Hövel, Filippo Simini, Chaoming Song, Albert-László Barabási

Computational Models of Mobility: A Perspective from Mobile Phone Data

In recent years more and more data has been generated, contributing to the rise of network science (Albert and Barabási 2002; Lazer et al. 2009). The available networks encompass a plethora of fields and mechanisms, including protein-protein interaction, food webs, information flow, data transmission, transportation, social interaction, and many more. In particular, social networks can be studied in greater detail as an increasing number of data sets become readily available in the digital age. Among them are networks reflecting friendships, consumption behavior, and mobility.

The analysis of what is generally termed *big data* from social contexts allows for the inference of human behavior (Barabási 2005; Eagle and Pentland 2006; Onnela et al. 2007a; Eagle and Pentland 2009). To put it even stronger, empirical data is a crucial component to be able to understand, forecast, and model aspects of daily life such as human mobility. One type of data set that is especially suited to serve as a proxy for the study of mobility arises from the usage of mobile phones (González and Barabási 2007; González et al. 2008; Wang et al. 2011; Bagrow et al. 2011; Bagrow and Lin 2012). The empirical data set used in this chapter, for instance, consists of anonymized call data records (CDRs) of 10 million customers from a single phone company. A CDR is generated every time one of the users initiates a call or sends a text message. It contains an anonymized caller identifier, an anonymized callee identifier, and time and date of the call or text message as well as the location of the caller. The latter is approximated by the location of the nearest mobile phone tower that handles the call. Similar data sets have been used to investigate the geography of communication networks and space-dependent communities (Lambiotte et al. 2008; Blondel et al. 2010; Expert et al. 2011). An example can be found in the strong language-dependent divide between Dutch- and French-speaking regions of Belgium. This divide is identified by community detection of the mobile phone communication network, a method that sorts network nodes (users) into clusters (communities) that have many internal and few external, intercommunity links.

The aim of this chapter is to provide a perspective on human mobility based on the analysis of mobile phone data. We address the following questions: How predictable are we in our daily routines (Song et al. 2010b)? What are the mathematical laws and functional dependencies, also known as scaling laws, of human travel (Song et al. 2010a)? What are the ingredients for models that describe the observed patterns (Simini et al. 2012)?

In order to answer these questions, we consider not only the mobility aspect of the data providing information about the individual mobile phone user's whereabouts (where does a call originate from), but also the social information of the data set (who calls whom). Furthermore, our analysis spans distance-wise several orders of magnitude ranging from urban via regional to countrywide distances.

Predictability of Human Mobility

To get a flavor of the data included in this study, the left panel of **figure 1** displays the exemplary trajectory of a single user during a one-month-long observation period. This particular user visited N=31 different locations, which are given by the closest mobile phone tower and recorded every time the user placed a call (black dots). The set of tower positions is depicted by gray dots, and the gray lines mark the approximate reception areas derived from a Voronoi lattice. This lattice is given as the boundaries between locations that are closest to different towers. In other words, all that locations within the Voronoi cell of a specific tower have in common is that they have a smaller distance to this tower than to any other tower. This method is a standard procedure also applied by mobile phone network providers to roughly determine the reception area of a tower. The black lines indicate the trajectory by connecting the positions of the chronological sequence of calls placed by the user.

Among the 31 locations there is a large number of places that are only visited a few times, but only a small number to which the user frequently returned. This typical behavior can be abstractly visualized by the mobility network shown in the center panel. There, each network node corresponds to one location and two nodes are connected by a link if they are places of two subsequent calls. The size of the node reflects the percentage of communication events originating from the respective tower, that is, the percentage of calls and text messages via that tower. In this visualization one can easily identify a central hub that is connected to a large number of less-frequented nodes. Ranking the different locations according to their visitation frequency yields a power-law dependence, as depicted in the right panel of **figure 1**. This is also known as Zipf's law (Zipf 1946).

Further quantitative insight into the mobility is given by the *radius of gyration*. This quantity is calculated as the average distance of all recorded locations to the center of mass of the trajectory and thus provides a characteristic length scale of travel. Mathematically, the radius of gyration r_g for each user, who is recorded during L events at positions $\vec{r}_1, \ldots, \vec{r}_L$, is defined as

(1) $$r_g = \sqrt{\frac{1}{L} \sum_{i=1}^{L} \left(\vec{r}_i - \vec{r}_{\mathrm{cm}}\right)^2}$$

with the center of mass $$\vec{r}_{\mathrm{cm}} = L^{-1} \sum_{i=1}^{L} \vec{r}_i$$

Calculating r_g for all users on record leads to the normalized distribution shown in **figure 2**. The curve has the shape of a fat-tailed distribution. This indicates that most trips are local and cover only small distances. Similarly, only a few users display a characteristic travel distance of tens or hundreds of kilometers.

Fig. 1.
Top: **Trajectories of an anonymized mobile phone user who visited the vicinity of** ***N*****=31 different towers during a month-long observational period, in which the user placed a total of** ***L*****=119 calls. Each time a user makes a call, the closest tower that routes the call is recorded, pinpointing the user's approximate location (black dots). Gray lines represent the Voronoi lattice, approximating the area of reception of each tower (gray dots). The black lines correspond to the recorded movement of the user between the towers.** ***Center:*** **Mobility network of the same user. The size of the nodes corresponds to the percentage of calls the user made in the vicinity of the respective tower, and the widths of the edges are proportional to the frequency of the observed direct movement between the respective towers.** ***Bottom:*** **Rank of location vs. visitation frequency (Zipf plot).**

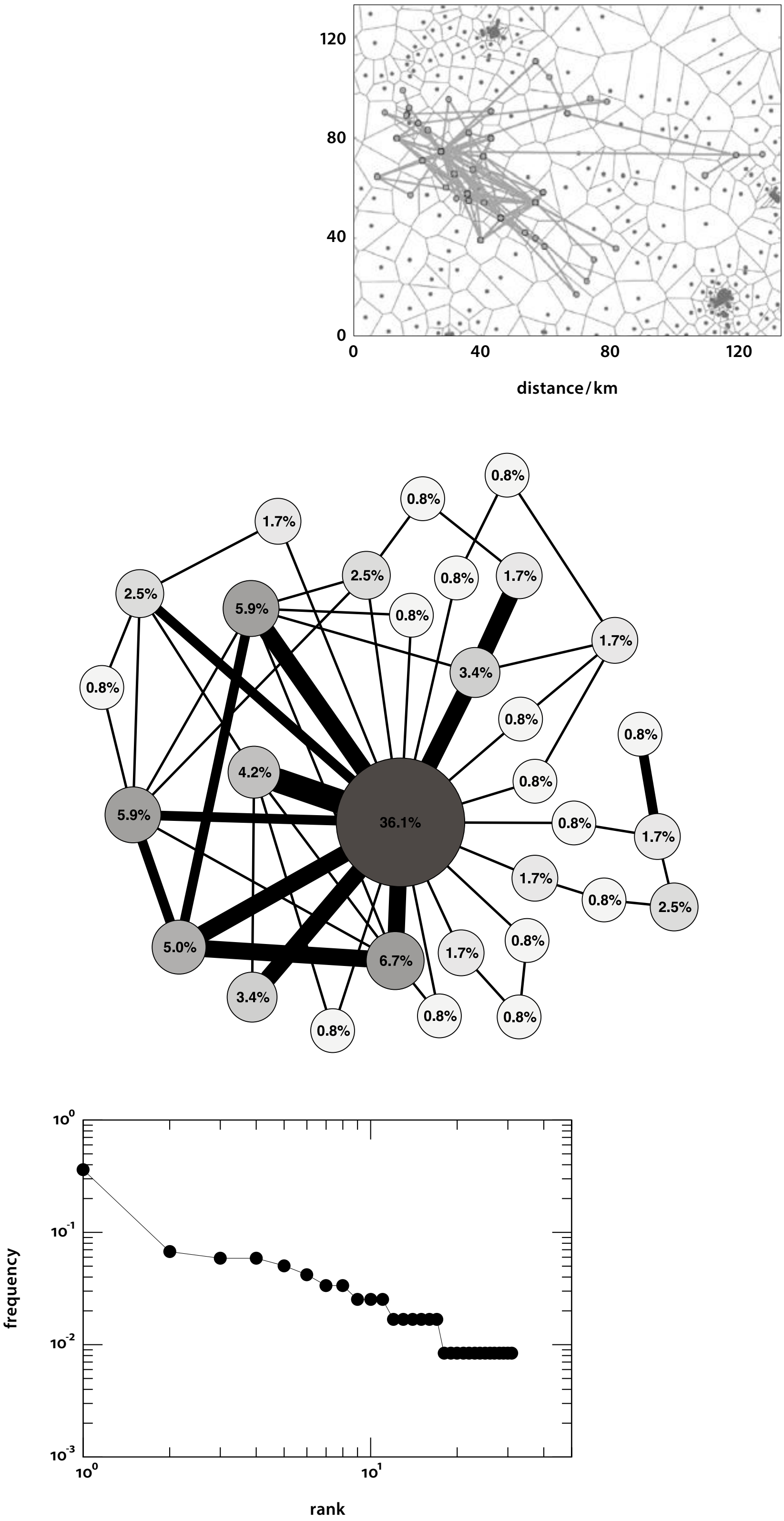
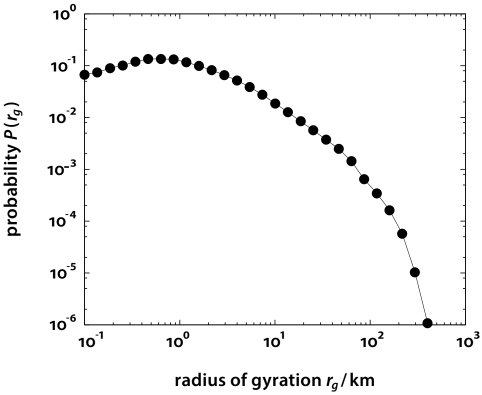


Fig. 2.
Normalized distribution of the radius of gyration r_g. Data: February 2009



In addition to the analysis of characteristic travel distances, we make use of the temporal information embedded in the data set (when a call is placed). This addresses the question of predictability. In short, we aggregate all calls recorded for each of the 168 hours of the week over a period of several weeks. That is, we extract for each user a list of places, where he/she called someone on Mondays between 12:00 a.m. and 1:00 a.m., 1:00 a.m. and 2:00 a.m., all the way to Sundays between 11:00 p.m. and midnight. For each of the 168 lists, we then determine the primary location, which is the tower that handled the most calls. The ratio R_i of the number of recordings at this tower to all calls,

$$(2) \qquad R_i = \frac{\text{number of appearances at primary location}}{\text{total of number of appearances}} ,$$

provides a measure of regularity for each hour of the week, i=1, …, 168. An example: If a user is found 8 times at one location and 2 times at another for one particular hour, then we obtain a value of regularity of $R_i = 8/(8+2) = 0.8$. **Figure 3** depicts both the average number of locations N (top panel) and the averaged regularity R (bottom panel) for each of the 168 hours of the week. Here the average is taken over a class of users with between 12 and 100 events per day. The curve confirms what one might have intuitively expected: People are most regular with the least number of locations during early-morning hours and least regular with the largest number of locations during commuting times between home and

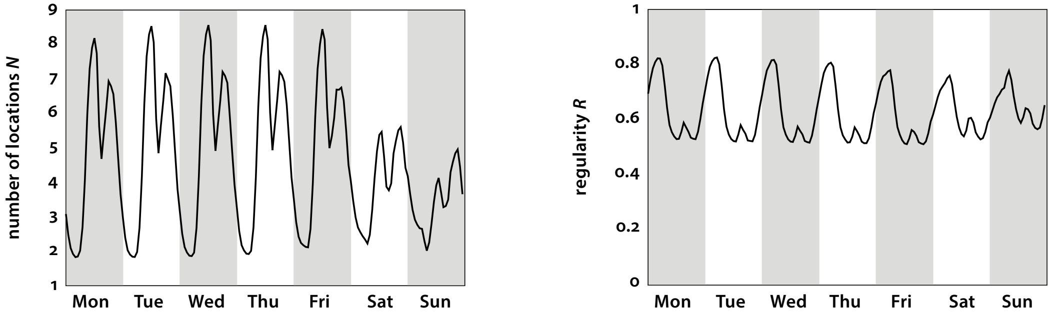


Fig. 3.
Regularity of mobility: Ratio *R* to find a user at his/her most frequent location and number of different locations *N* with an hourly resolution for each day of the week. Data: May 2009 for a class of users between 12 and 100 events on a daily average

workplace. This effect repeats for each weekday but is less pronounced on the weekends.

A question that links the radius of gyration introduced in the beginning and regularity is: Are users with a larger characteristic travel distance (larger r_g) more or less regular in their behavior? To account for the difference in the total number of locations, we consider the relative regularity R/R_{rand}, where $R_{\text{rand}} = 1/N$ serves as a reference case. This case assumes that the user visits the N recorded positions in a random but uniform manner. One might expect that users who do not travel far are more regular, but evaluation of the data exhibits the opposite trend. Figure 4 shows the weekly average R/R_{rand} in dependence upon the radius of gyration. Counterintuitively, we find that users with larger radius of gyration have a larger relative regularity. In other words, people become more regular if they travel longer distances.

Borrowing techniques from information theory and statistical physics, it is also possible to estimate the maximum predictability of a user's whereabouts. For details involving entropy and Fano's inequality see Song et al. (2010b) and references therein. The main result is twofold: (1) The predictability narrowly peaked at 0.93, meaning that the locations are picked at random not more than 7 percent of the time; (2) surprisingly, the predictability does not drastically decrease for users who have a large radius of gyration, but it saturates at 0.93. This means that, regardless of the distance a group of individuals travel, the set of places they visit remains predictable 93 percent of the time.

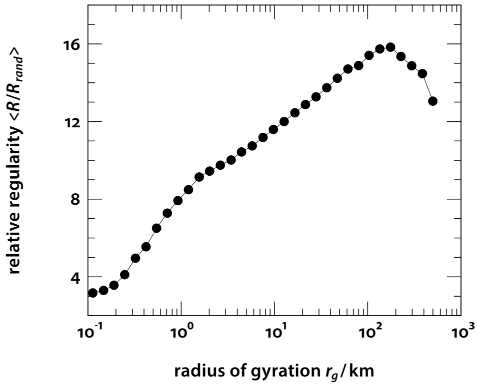


Fig. 4.
Averaged relative regularity R/R_{rand} vs. radius of gyration r_g indicating that users with large r_g have high relative regularity

Mathematical Functionalities of Human Mobility

So far we have discussed empirical evidence for regularity and the characteristic length scale of human travel. In the following, we present simple models that reproduce the empirical findings. At first, we focus on a stochastic model for individual mobility and its accuracy concerning the radius of gyration. Later we also introduce a model to account for migration patterns, which can be derived from first principles.

The individual mobility model builds on the empirical observation that our daily travel experience is dominated by many short distances, short waiting times, and frequently visited locations, whereas large excursions and long periods between subsequent trips are rare (González et al. 2008). Mathematically, the probability distribution $P(\Delta t)$ of the waiting time Δt, for instance, is formulated as a fat-tailed distribution $P(\Delta t) \sim |\Delta t|^{-1-\beta}$ with $0 \leq \beta \leq 1$. This formula reflects in the form of a mathematical functionality that long waiting times are less likely than short periods between two trips. The jump-size distribution that describes the statistics of the length of a trip given by the distance between visited towers has a similar form with an exponent α. One could try to design a model simply on the assumption of these two findings leading to a continuous-time random-walk model (Brockmann et al. 2006), but as shown by Song et al. (2010 a) this does not capture all aspects of human mobility.

In addition to the scaling laws for the selection of the waiting time and jump size, two generic mechanisms should be included that are schematically depicted in figure 5. The first mechanism

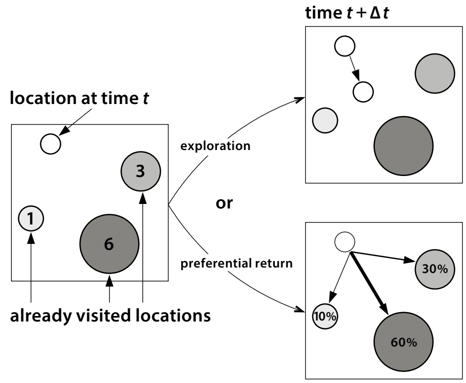


Fig. 5.
Schematic diagram of the individual mobility model: The configuration at time *t* is shown in the left box, indicating the user's current and previously visited locations (here: *N*=4, including the present location). The size of circles drawn at each location is proportional to the visitation frequency. At time *t*+Δ*t* (with Δ*t* chosen from a fat-tailed distribution) the user can either go to a new location from his/her present location with some probability, where that distance is again based on a fat-tailed distribution (exploration; *upper right*), or return to a previously visited location with the counterprobability, where the place of return will be chosen with a probability equal to the frequency of his/her earlier visits (preferential return; *lower right*).

is *exploration*. After a waiting time Δt the individual relocates to a previously unknown position. Both the waiting time and the jump size are chosen according to empirically confirmed fat-tailed distributions. This case occurs with a probability $P_{\text{exploration}} = \varrho N^{-\gamma}$, where N denotes again the number of previously visited locations. In other words, the likelihood of exploring a new location becomes smaller if one has already been to a large number of different locations. The parameters $\varrho \in (0,1]$ and $\gamma \geq 0$ can be determined from the data. With the counterprobability $1\text{-}P_{\text{exploration}}$, the user returns to an already visited position. As a consequence of the functional dependence of $P_{\text{exploration}}$, one is more inclined to return to a known place if one has already visited many locations. Furthermore, the return position is picked from the set of visited locations with a probability proportional to the visitation frequency, that is, more frequently visited places have a higher likelihood to be selected for a return visit. This coins the notion of *preferential return*.

The validity of this model can be tested considering the radius of gyration discussed earlier in this chapter (see equation 1 on page 112). The result is shown in figure 6. The left and right panels display the data and model predictions, respectively. The shapes of the curves and the scaling exponents are in good agreement supporting the assumption of the proposed model.

To demonstrate the good agreement of the individual mobility model with the data at hand, figure 7 shows the case of a single user as predicted by the model ($\alpha=\beta=0.6$, $\gamma=0.2$, $\varrho=0.4$). The comparison to figure 1 already suggests a very similar set of mobility characteristics

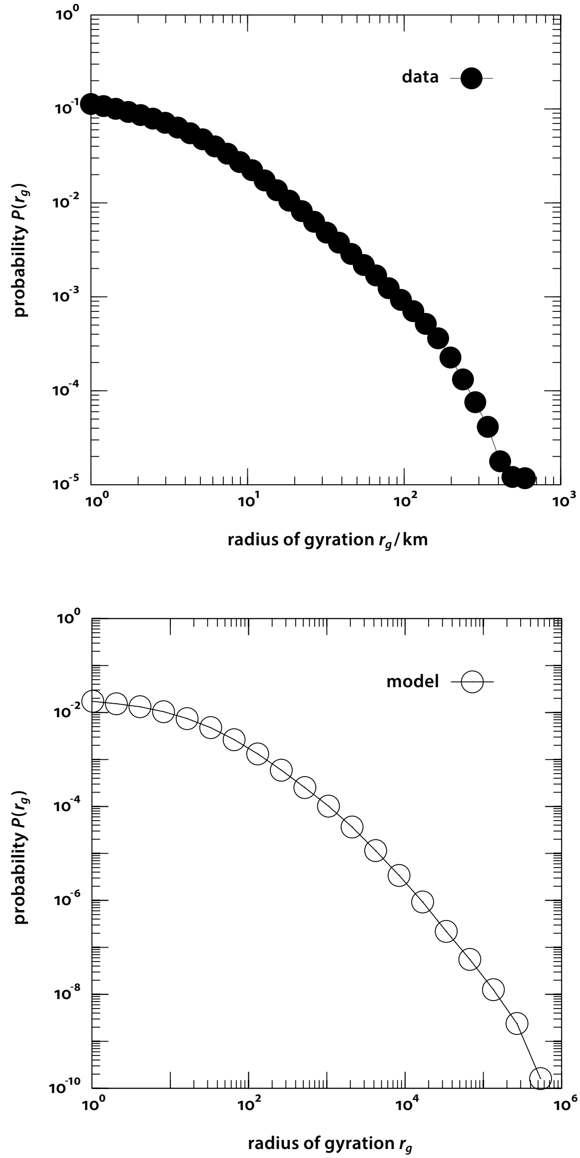
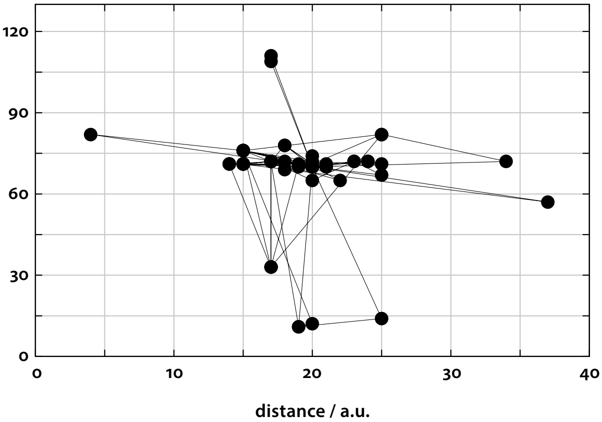


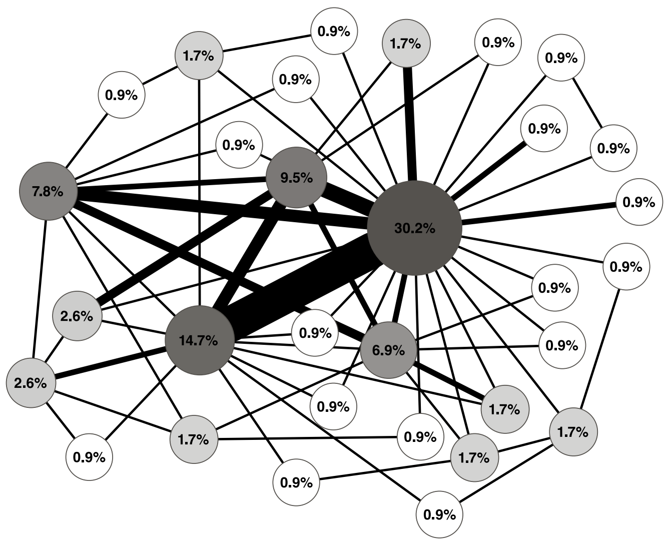
Fig. 6.
Normalized distribution *P*(r_g) of the radius of gyration r_g derived from the mobile phone data (*top,* observation period: 1000 hours) and the individual-mobility model (*bottom,* arbitrary spatial units). Reproduction of Figs. 4(a) and (c) in Song et al., 2010

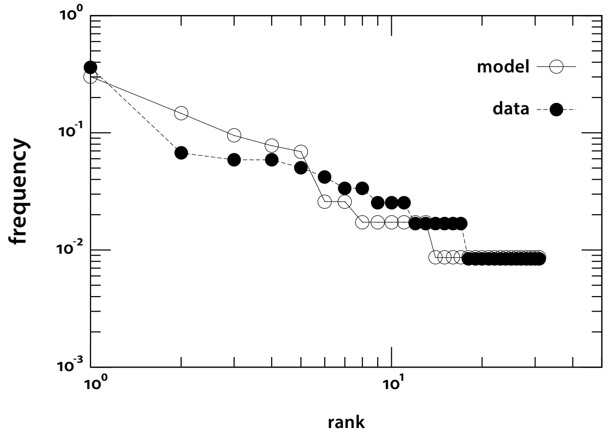
based on visual observation of the trajectories and the mobility network. The model also yields a quantitative agreement as illustrated by the Zipf plot (bottom panel of **figure 7**), which displays the result of both the empirical data and the model.

In addition to the results presented above, the individual mobility model also reproduces other aspects like the probability to find a user with a specific number of visited locations or an ultra-slow growth of the radius of gyration over time. Here the notion of

Fig. 7.
Top: **Trajectories as predicted by the individual mobility model (α=β=0.6, γ=0.2, ρ=0.4). The model user behaves similar to the one depicted in figure 1 and visits N=31 location during L=116 calls.**
Center: **Mobility network of the same user (layout as in figure 1).**
Bottom: **Rank of location vs. visitation frequency (Zipf plot), where the empirical data of figure 1 is added as filled dots.**







an ultra-slow process refers to a much slower growth than expected for the standard reference case of a continuous-time random walk model (Brockmann et al. 2006; details can be found in Song et al. 2010 a).

Universal Mobility Model

Next to the individual mobility model with its inherent fluctuations, which nicely reproduces the scaling properties of human mobility, one can also formulate a parameter-free model to describe mobility and migration patterns. It is only based on the population density. Hence, it does not require any a priori information such as waiting time or jump-size distributions. This model is called a *radiation model,* because it is derived using the following physical concepts (Simini et al. 2012): Particles (commuters) emitted at a source (location of origin) can be absorbed by the environment (terminate their commute at a surrounding location) with a certain probability. The traveled distance (length of commute) depends upon the thickness of the material (population density of the surrounding locations).

Thus, the radiation model aims to estimate the commuting fluxes, i.e., the average number of commuters traveling per unit of time between any two locations in a country. It starts from the population m_i and n_j at pairs of locations i and j, respectively. Furthermore, it takes into account the aggregated population between i and j by introducing the total population s_{ij} in a circle of radius r_{ij} centered at i. The model predicts the average flux $\langle T_{ij} \rangle$ from i to j by the following formula

(3) $$\langle T_{ij} \rangle = T_i \frac{m_i n_j}{\left(m_i + n_j\right)\left(m_i + n_j + s_{ij}\right)}$$

where $$T_i = \sum_{i \neq j} T_{ij}$$

denotes the total number of people starting from location i. Considering N_c commuters in total, this number is simply given by

$$T_i = m_i N_c \Big/ \sum_i m_i \quad \text{, where} \quad \sum_i m_i$$

is the population of the whole country under investigation.

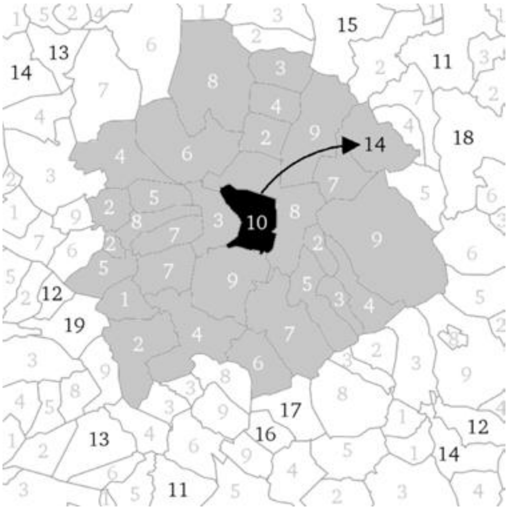


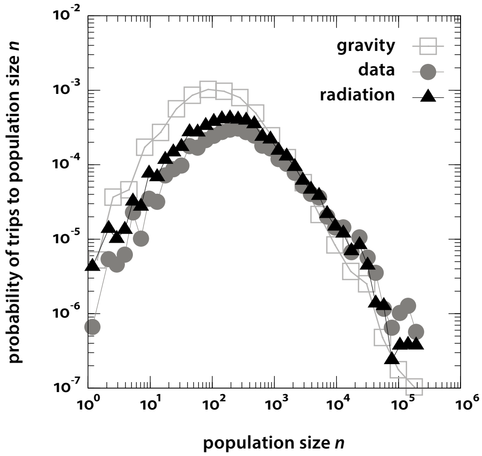
Fig. 8.
Schematics of the radiation model: The numbers in the municipalities denote the respective benefits, which are the highest offer received from that area. An individual at the center location (black) commutes to the nearest area with a higher benefit (arrow).

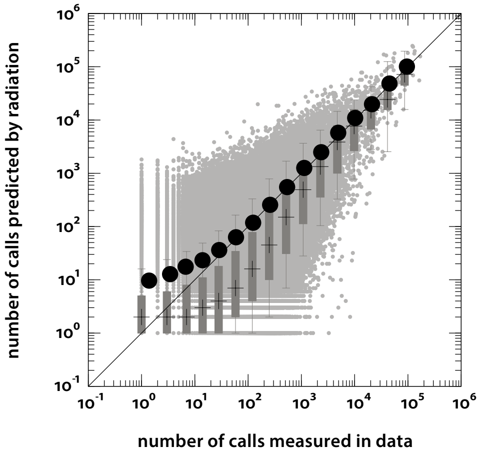
Figure 8 provides a schematic diagram of the radiation model. It is described as follows: An individual at a certain location will commute to the nearest location that offers a better benefit than his/her home location, where the benefit of each location (shown as numbers in figure 8) indicates the highest offer from the local companies, whose number is assumed to be proportional to the population of each location. In the depicted, exemplary case, a person at the center municipality (benefit 10, black area) is predicted to commute to the municipality with benefit 14 as indicated by the arrow, because it is the closest area with a higher benefit. The gray region marks the corresponding circle of radius r_{ij} that enters the model (3). All other locations within that circle offer a smaller benefit than the home location.

Simini et al. (2012) showed that the radiation model is superior to other models that predict migration patterns, such as the widely used gravity model. It determines the number of individuals T_{ij} traveling from location i to j as follows:

(4) $$T_{ij} = \frac{m_i^{\alpha} n_j^{\beta}}{f\left(r_{ij}\right)} \quad ,$$

where the deterrence function $f(r_{ij})$ is usually assumed to be a power-law or exponential and α and β are adjustable exponents. The analogy to Newton's gravity law provides the name. It was assumed that migration between two cities is like gravitation,





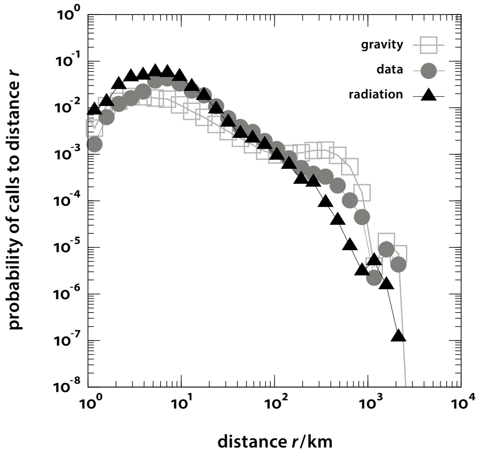


Fig. 9.
Top: **Probability distribution of a call over a distance *r*.**
Center: **Probability distribution of a call to a municipality with population size *n*.**
Bottom: **Comparison between measured and simulated fluxes for each pair of municipalities. Data: Number of phone calls between users living in different municipalities during a period of 4 weeks with a total number of 38,649,153 calls placed by 4,336,217 users. The data was aggregated to obtain the total number of calls between every pair of municipalities. Reproduction of figures 3(g) to (i) from Simini et al. (2012)**

with the masses of objects, which are subject to their mutual gravitational force, substituted for population densities at different locations. Accordingly there is more interaction, i.e., flux, between close and populous places. This model, however, suffers from several limitations such as a rigorous derivation, a well-funded choice of the deterrence function and parameters, a requirement to fit parameters to data, an unboundedness in the limit of large population sizes, and an intrinsic neglect of the terrain between locations *i* and *j*. All of these can be overcome by the radiation model.

In general the radiation model can be applied to various contexts such as long-term migration and transport-driven processes like freight transportation. Here we focus on its predictive power concerning mobile phone data. **Figure 9** illustrates the comparison between data, the gravity model, and the radiation model. The top panel depicts the distribution of the probability that a call is placed between two locations with a distance *r*. While the gravity model has discrepancies both for small and large distances, the radiation model matches the data better. The same holds for the probability of trips to locations with a population size *n*, which is shown in the center panel. Finally, the bottom panel displays the fluxes between all pairs of municipalities measured in the data compared to their predicted value (gray dots). The black dots correspond to the mean number of predicted calls for each bin. The candlestick bar indicates the 9th and 91st percentiles. This interval overlaps for all bins with the diagonal added as a thin black line.

Summary

We have demonstrated how big data – in our case mobile phone data – can be used to analyze human mobility over several orders of magnitude. We identified scaling laws as well as regularity of daily routines, and presented simple models that reproduced the empirical findings to a large extent. The universality of the discussed scaling laws and the applicability of the models suggest a relevance and importance for the scale of human behavior and infrastructure.

PH acknowledges support of the German Academic Exchange Service (DAAD) via a postdoctoral fellowship.

Albert, Réka, and Albert-László Barabási. 2002. "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics* 74, no. 1 (January): 47–97.

Bagrow, James P., and Yu-Ru Lin. 2012. "Mesoscopic Structure and Social Aspects of Human Mobility," *PLoS ONE* 7, no. 5: e37676.

Bagrow, James P., Dashun Wang, and Albert-László Barabási. 2011. "Collective Response of Human Populations to Large-Scale Emergencies," *PLoS ONE* 6, no. 3: e17680.

Barabási, Albert-László. 2005. "The Origin of Bursts and Heavy Tails in Human Dynamics," *Nature* 435, no. 7039: 207–211.

Blondel, Vincent, Gautier Krings, and Isabelle Thomas. 2010. "Regions and Borders of Mobile Telefony in Belgium and in the Brussels Metropolitan Zone," *Brussels Studies* no. 42 (October 4).

Brockmann, Dirk, Lars Hufnagel, and Theo Geisel. 2006. "The Scaling Laws of Human Travel," *Nature* 439, no. 7075: 462–465.

Eagle, Nathan, and Alex Pentland. 2006. "Reality Mining: Sensing Complex Social Systems," *Personal and Ubiquitous Computing* 10, no. 4 (May): 255–268.

Eagle, Nathan, and Alex Pentland. 2009. "Eigenbehaviors: Identifying Structure in Routine," *Behavioral Ecology and Sociobiology* 63 (May): 1057–1066.

Expert, Paul, Tim S. Evans, Vincent Blondel, and Renaud Lambiotte. 2011. "Uncovering Space-Independent Communities in Spatial Networks," *Proceedings of the National Academy of Sciences* 108, no. 19: 7663–7668.

González, Marta C., and Albert-László Barabási. 2007. "From Data to Models," *Nature Physics* 3, no. 4: 224–225.

González, Marta C., César A. Hidalgo, and Albert-László Barabási. 2008. "Understanding Individual Human Mobility Patterns," *Nature* 453, no. 7196: 779–782.

Lambiotte, Renaud, Vincent Blondel, Cristobald de Kerchove, Etienne Huens, Christophe Prieur, Zbigniew Smoreda, and Paul Van Dooren. 2008. "Geographical Dispersal of Mobile Communication Networks," *Physica A: Statistical Mechanics and Its Applications* 387, no. 21: 5317–5325.

Lazer, David M. J., Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science," *Science* 323, no. 5915: 721–723.

Onnela, Jukka-Pekka, Jari Saramäki, Jörkki Hyvönen, Gábor Szabó, David M. J. Lazer, Kimmo Kaski, János Kertész, and Albert-László Barabási. 2007b. "Structure and Tie Strengths in Mobile Communication Networks," *Proceedings of the National Academy of Sciences* 104, no. 18: 7332–7336.

Onnela, Jukka-Pekka, Jari Saramäki, Jörkki Hyvönen, Gábor Szabó, Marcio Argollo de Menezes, Kimmo Kaski, Albert-László Barabási, and János Kertész. 2007a. "Analysis of a Large-Scale Weighted Network of One-to-One Human Communication," *New Journal of Physics* 9, no. 6 (June): article 179.

Simini, Filippo, Marta C. Gonzalez, Amos Maritan, and Albert-László Barabási. 2012. "A Universal Model for Mobility and Migration Patterns," *Nature* 484, no. 7392: 96–100.

Song, Chaoming, Tai Koren, Pu Wang, and Albert-László Barabási. 2010a. "Modelling the Scaling Properties of Human Mobility," *Nature Physics* 6, no. 10: 818–823.

Song, Chaoming, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010b. "Limits of Predictability in Human Mobility," *Science* 327, no. 5968: 1018–1021.

Wang, Dashun, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-László Barabási. 2011. "Human Mobility, Social Ties, and Link Prediction," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 21–24, San Diego, CA*. New York: Association of Computing Machinery, 1100–1108.

Zipf, George K. 1946. "The P1 P2/D Hypothesis: On the Intercity Movement of Persons," *American Sociological Review* 11, no. 6 (December): 677–686.

Kael Greco

Seeing the City through Data / Seeing Data through the City

In early 2007 a group of Google Earth users made a curious discovery in San Diego. Panning across the newly available satellite imagery, the armchair Magellans noticed a set of structures that formed an inexplicable shape when viewed from above; a Nazi swastika (**figure 1**). The find went viral – well before the concept of *going viral* existed – and was picked up by major news outlets. It was quickly discovered that the complex, whose surrounding roadways are coincidentally named after WWII-related sites, was actually built in 1967 by the US Navy. Now visible to anyone with an Internet connection, the base's plan view caused great public outcry, and the resulting political pressure led to a $600,000 reconstruction project to unmake the abhorrent shape (Perry 2007). "We don't want to be associated with something as symbolic and hateful as a swastika," a spokesperson said. The Navy claims the exact form and orientation of the structure was wholly unintentional, and simply the consequence of a humiliating planning oversight. But whether deliberate or not, it's clear the project's planners, architects, and builders were not anticipating a god's-eye-view perspective on the finished project. As ridiculous as the whole episode was, it highlights a particularly powerful idea: New ways of seeing can

Fig. 1.
Aerial image of a US Navy building complex close to San Diego.
Source: US Geological Survey © 2013 Google

drastically refigure our understanding of, and relationship to, place. Google had reprocessed and reframed urban data in a fashion that literally enabled a new perspective on the city – in this case revealing the largest government-subsidized swastika on the planet.

Of course, not all urban data sets are as intrinsically visual as satellite imagery. Many require the context of urban spatiality to understand and operationalize them – i.e., they gain salience and power when seen *through the city*. The de facto example of this power – thanks in large part to Edward Tufte's proselytization (2001, 24) – is on display in John Snow's 1854 cholera outbreak map **(figure 2)**. In the middle of the nineteenth century the city was inextricably tied to health, as planners and theorists aimed to disentangle the growing health problems that came with increasingly denser agglomerations of people. Of particular concern during the mid-1850s was the cholera epidemic. Basic germ theory had not yet been accepted by the medical field, but by mapping cholera deaths across Soho, Snow was able to tangibly communicate the idea that cholera was transmitted not through infected air, but rather through contaminated water and food.

Data-Overload and the Curse of Dimensionality

Like the march of technology, the practices of visualization and mapping are inextricably tied to the promise of the infinite, or as James Corner put it, "mapping *unfolds* potential; it re-makes territory over and over again, each time with new and diverse consequences" (1999, 213). This thought, coupled with the proliferation of the big data paradigm, has furthered the intoxicating belief that any and everything can be described, manipulated, modeled, and simulated through computation. In the era of big data, quantity

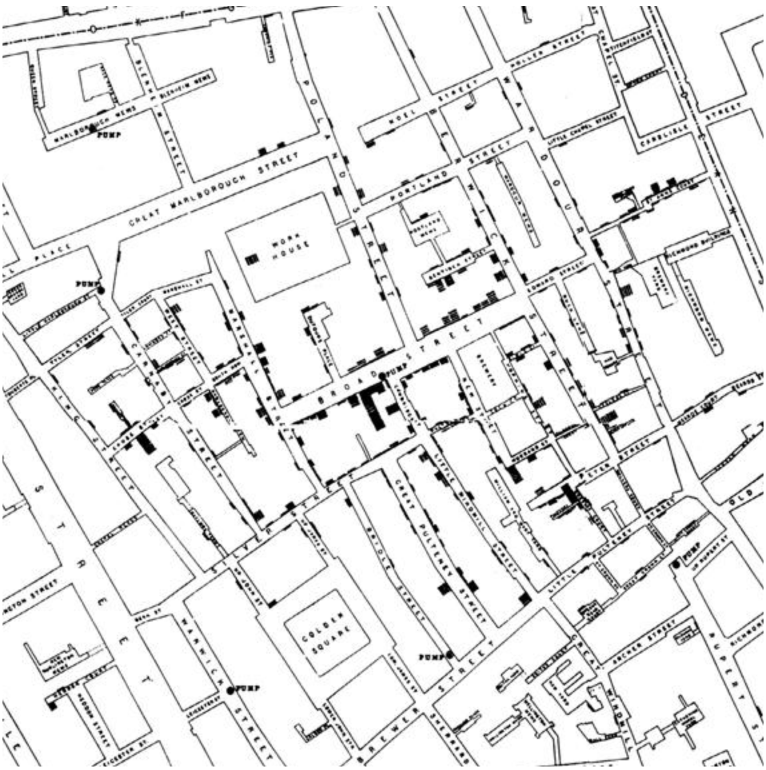


Fig. 2.
Dr. John Snow's spot map of death cases during the cholera epidemic of 1854 in London's Golden Square
Source: www.wikipedia.org

is king, and every additional dimension brings us closer to perfection in representation. Jorge Luis Borges's "On Exactitude in Science," a cautionary tale on the pursuit of the one-to-one ideal, has been all but forgotten. We no longer balk at the prospect of comprehensively describing and simulating our world – Google has assured us this is an inevitability, and the new existential dread sits on the crumbling line between the real and the simulated (Beane et al. 2012); now we're afraid of losing the ability to tell where the physical world ends and the datascape begins.

While an ontological exploration of reality in this context is admittedly a bit cheeky, underneath this goofy progression of logic rests the subtle implication that, again, more is more – every piece of datum carries, at the very least, a modicum of descriptive power. This prevailing notion is that important stories sit somewhere within all data, and consequently, the task of analysis and representation is to simply uncover these stories. And thus, the march toward data-absolutism continues, instilling a tendency to cast meaning where it simply doesn't exist – to identify or construct false patterns in the great static that is big data.

What does this all mean in the context of the city? On a fundamental level it underscores the fact that urban data are powerful and capricious; they encapsulate countless dimensions at myriad physical and temporal scales, and a wide gulf exists between possible relationships and actionable results. Our struggle is now in reclaiming a sense of legitimate, verifiable meaning from the morass, i.e., reorienting our processes of modeling, simulation, and representation to distill value while also keeping validity in check. The following case study illustrates our approach to achieving this delicate balance.

The City of Riyadh and the UTS Project Vision

Rapid economic and demographic changes throughout Saudi Arabia are posing new challenges and opportunities for the Kingdom. Of particular concern is the explosive growth of the nation's capital, Riyadh, where development is quickly outpacing transportation infrastructure. Further straining the city's already limited roadway capacity is an explosion in vehicular traffic; between 1987 and 1995, automobile trips increased at the rate of 9 percent per year.

The UTS project aims to develop an innovative, highly dynamic urban traffic system to address the mobility challenges specific to Riyadh. To this end, the project is based on creating an alternative to traditional intelligent transportation systems by taking advantage of the digital traces of our everyday lives to create models for mobility analysis, intervention, and planning, for policy makers, planners, and development professionals, as well as for the citizens of Riyadh themselves.

Conceptually, the project includes both short- and long-term components. Over the short term it hopes to analyze the performance of human mobility over existing infrastructure; identifying hidden inefficiencies and potential points of improvement across the urban transportation system. The long-term perspective is geared toward future city development, as it serves to address the region's higher-order trends in an effort to forecast scenarios for increased service allocation and growth.

The culmination of the project rests in the formulation of a mobility data browser, a platform to facilitate collection, analysis, modeling, and representation of a variety of urban data sets and a variety of timescales. Beyond its specific focus on mobility, the tool itself does not project any overarching agenda; rather, the hope is to aid in the creative combination of data streams that are typically perceived as incongruent, all in the hopes of revealing previously imperceptible or obscured connections.

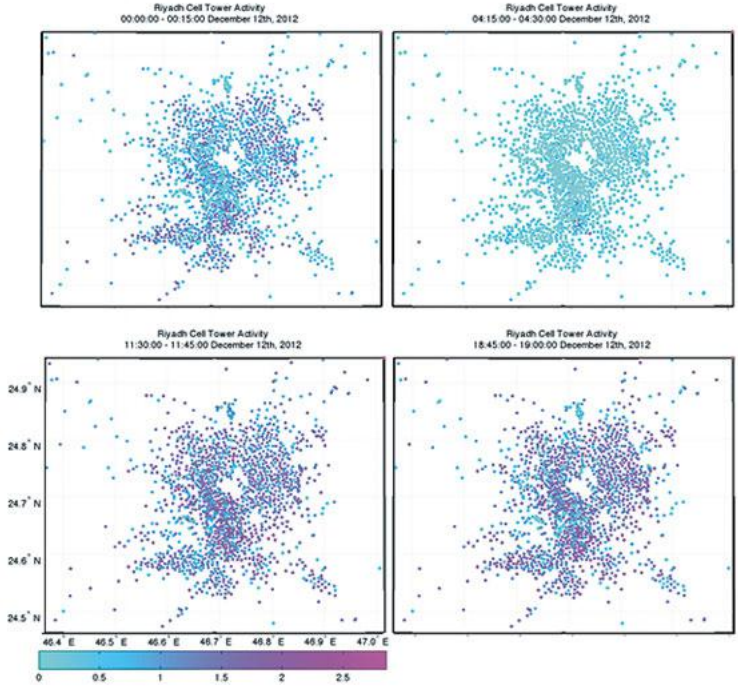
Before diving into various analysis methodologies and representational strategies, it is important to clarify the data foundation on which the project is based. We partnered with telecom companies in the region to collect roughly one month of total phone activity across the country. The cell phone is one of the most powerful real-time sensing mechanisms currently available to us; the ubiquity of digital devices allows us to capture extremely high-resolution traces of humanity across a variety of dimensions. Saudi Arabia's mobile phone penetration is above 198 percent – an astonishing figure suggesting that many across the Kingdom own more than one mobile device. We aggregated nearly 100 million daily network connections, assigned to more than 10,000 unique cell towers. Each anonymized call detail record (CDR) held a precise time and duration measure for the connection, the caller's location, the type of connection (phone call, SMS, Internet query, etc.), and the user's type of service (subscription, prepaid, etc.).

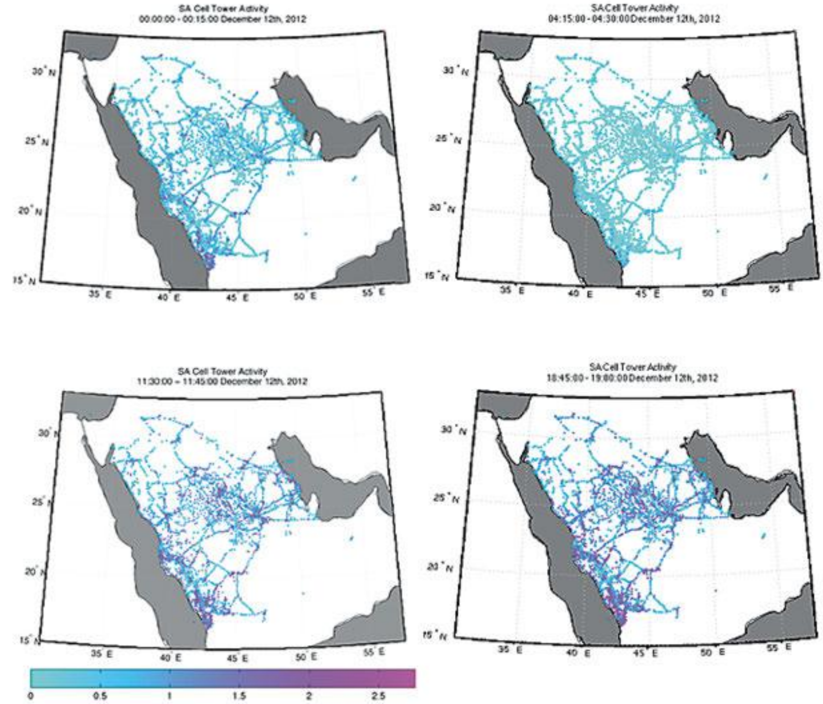
Analysis and Visualization: Seeing Riyadh through Data

We began by spatializing the data and generating a suite of activity profiles at different physical and temporal scales. **Figure 3** shows cellular usage patterns over one day across (1) the city of Riyadh, and (2) the entirety of Saudi Arabia. The aggregate cellular activity (number of calls, texts, and data queries) over an interval of fifteen minutes is plotted by color (dark to light on a logarithmic scale) for each cell tower. These static images present a powerful impression of the dynamics of social life across the region (alongside charting out the territory's telecommunications backbone), but ultimately they offer mere suggestions rather than answers, eschewing the rhythm or pace of life. Bringing more context to the fore – city imagery, infrastructural form, and temporal dimensionality – we arrive at a much richer representation of the rhythm and pace of life across the city.

Figure 4 shows cellular activity through color, transparency, and height (again in logarithmic scale) pixelated across the metropolitan expanse of Riyadh. As opposed to seeing the cell towers as discrete points in the city, we show network traffic interpolated over a 100 × 100 grid. In this sense, each grid cell is assigned an intensity based on its distance to surrounding antennas and their activity levels using a Gaussian smoothing function. The temporal activity is interpolated in a similar manner, and although impossible to demonstrate through print, the final visualization projects a much more natural view – a strong portrait of the social character of the city. With the inclusion of satellite imagery as a base map, we arrive at a unique view of how the social rhythm of the city is expressed over built form. As a basic sanity check for our data, we see – as one would anticipate – very low activity levels through the early morning hours before the downtown core becomes buried in an eruption of traffic that slowly spreads outward to hang over the region for the remaining hours of the day. We also see clear subcenters emerge that correlate with construction density, and these subcenters appear to be partitioned by the roadway network itself.

Fig. 3.
Activity snapshots across Riyadh (*top*) and Saudi Arabia (*bottom*)





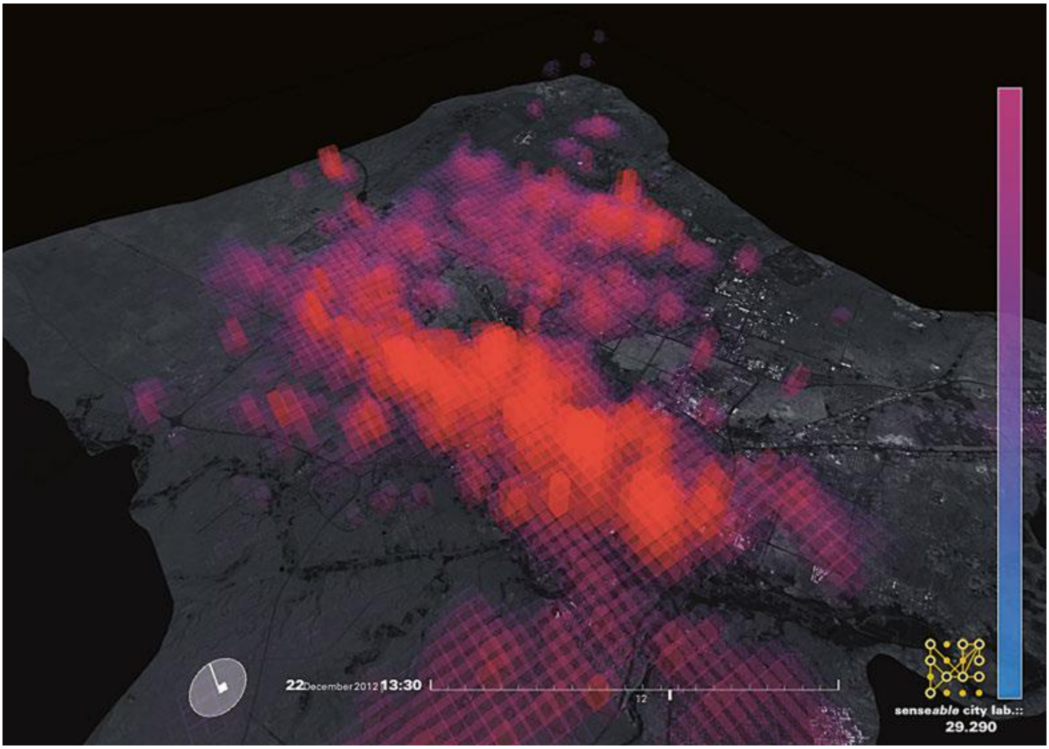


Fig. 4.
Cellular activity averaged across the geography of Riyadh. Height and color represent aggregate activity over a 15-minute interval. Visualization: Kael Greco, SENSEable City Lab 2013

The city's shifting activity profile also highlights a rich temporal signature of communication that is all Riyadh's own. Watching the oscillations of the activity landscape, a unique character emerges – we see that the city really doesn't come alive before noon, and peaks in aggregate activity around 6:15 p.m. With a careful eye, we can begin to pick out subtle regional delineation across the time slices: the residential neighborhoods to the southwest and northeast of the downtown core activate well before the rest of the city, and experience the strongest interhour fluctuations throughout the course of the day. Lastly, the temporal dynamism presents some peculiar discontinuities throughout the day – almost as if all phone traffic is suddenly halved at strange intervals. (We'll return to this phenomenon later.)

Inferring Home / Work Locations

Expanding our time intervals to capture broader day and night variation we can begin to differentiate home and work locations across the city – a methodological precursor to inferring urban land use. We define home locations as the most visited cell tower areas during weekday nighttime hours, and work locations as the most visited cell tower areas during weekday daytime hours. This essentially breaks down to filtering users who make the majority (60 percent) of nighttime calls in one place between the hours of 10:00 p.m. and 6:00 a.m., and the majority of daytime calls between 9:00 a.m. and 3:00 p.m. in another place.

The process left us with approximately 2 million weekday home-work pairs, but the question became: How can we put these to use? Can the combination of these pairs tell us something new about the operational structure of city? Of course, each home-work dyad by its very nature defines the start and endpoint of a commute, making a fundamental step toward understanding travel demand, but is it possible to see something unique in the data?

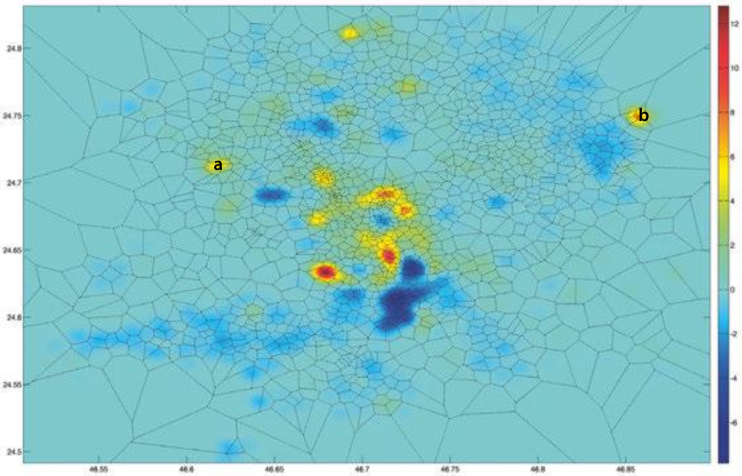
We began by accumulating each home and work location bounded by the expanse of the city and geographically smoothing the results in two separate maps. We then created an additional map to highlight the extremes, subtracting resident locations from work locations, shown in **figure 5**.

The map highlights the discrepancy between the entirely residential and entirely worksite locations, showing some monocentrically clustered worker hotspots that follow the overall spatial logic of the city. At the periphery we also see a number of universities show up strongly as work locations. Lastly, we see high agglomerations of residences to both the south and east of the city, with smaller pockets scattered throughout. Again, this is in line with a subjective survey of the land through Google maps and discussions with locals.

Detecting Mobility Communities

The home/work visualizations above point to an organizational logic of the city. Can empirical analysis confirm the strong regional clusters we see in the preceding maps? And if so, how can we visually explore the implications of the underlying community structures? If we conceptualize all of home/work commutes as a citywide mobility network, we can conceivably break this network into subcommunities by applying a regional delineation algorithm.

Fig. 5.
Map of work/home location differential
a) King Saud University
b) King Saud Bin Abdulaziz University for Health Sciences



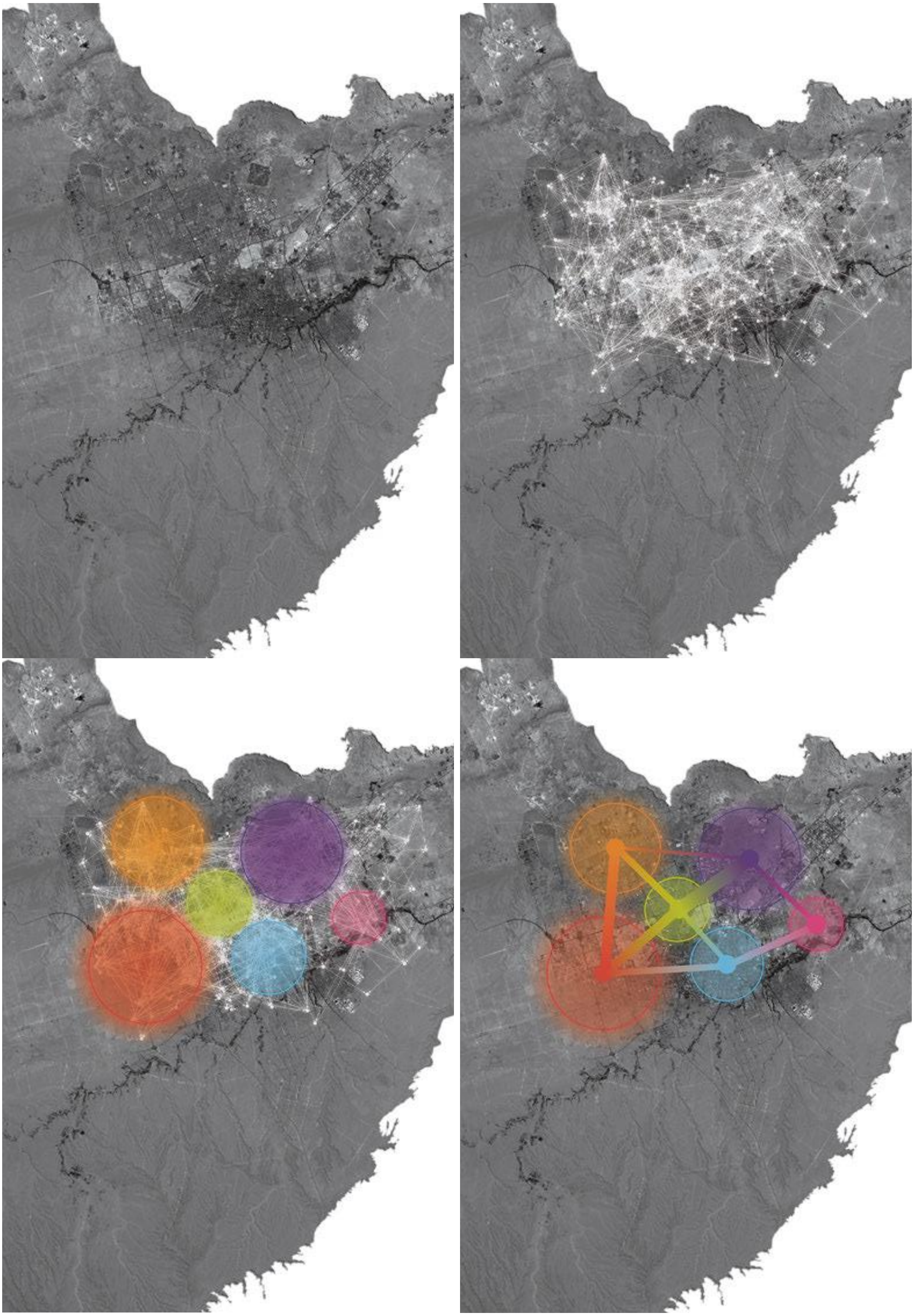
The process begins with the citywide network of connected cell tower locations, where a weighted directed edge between two nodes is defined as the cumulative trip flow between them. The algorithm then uses a modularity optimization scheme, such that subnetworks are clustered in a way that minimizes total internal arc disruption. Each resulting subcommunity represents an area where the majority of commuters live and work. In total, we identified seventeen distinct communities in Riyadh.

By overlaying our results on the geography of the city, a number of interesting relationships are revealed. Most strikingly, the clusters correlate very closely with the main arteries of the city. Mobility communities seem to be heavily reliant on the street network itself, underscoring the city's overall dependence on highway infrastructure. These results also support the commonly held belief that heavily trafficked streets, on many levels, are instruments of segregation and control, or perhaps more optimistically: Good streets make good neighbors.

From Social Rhythm to Directionality

Facing page:
Fig. 6.
Partitioning of the origin/destination network into mobility communities. Visualization: Kael Greco, SENSEable City Lab 2013

CDRs have the ability to tell the stories of urban inhabitants in near real time. From the planning perspective, one of the most meaningful stories we can glean is an individual's mobility pattern, which, at an aggregate level, describes one of the most vital components of urban analysis: origin-destination matrices. Constructing accurate



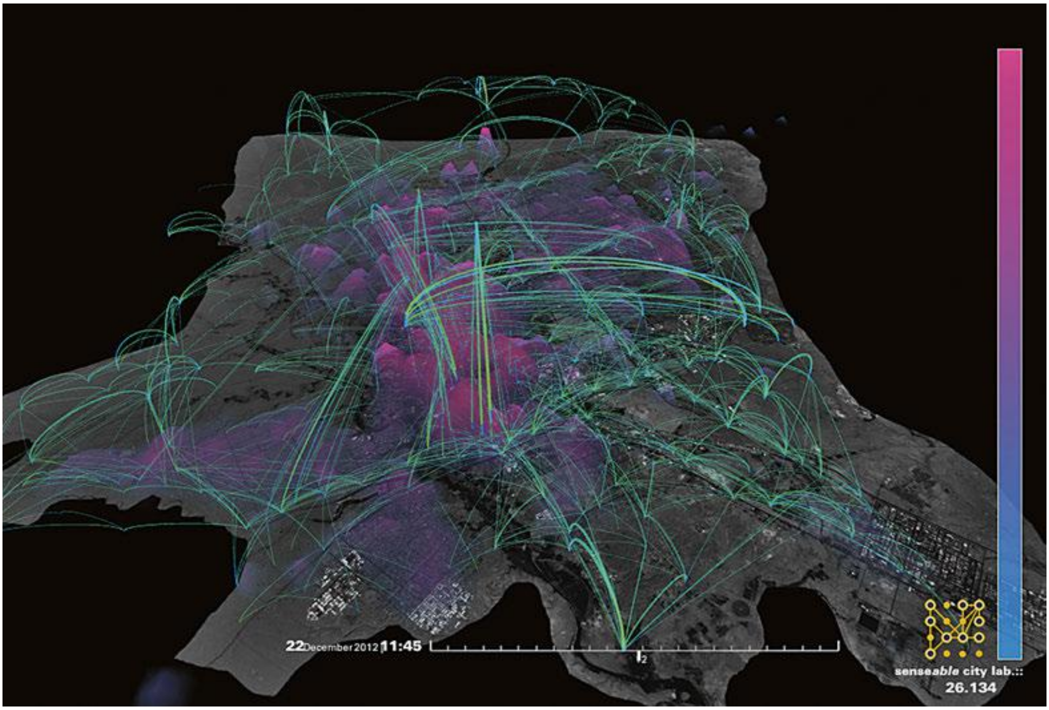


Fig. 7. Origin / destination traces. Visualization: Kael Greco, SENSEable City Lab 2013.

O-D matrices is a crucial component for transportation network optimization – not only for assessing moment-to-moment capacity constraints, but also for forecasting future needs. Traditionally, O-Ds are constructed through onerous census surveys that are conducted every five to ten years. The process is long and costly, and when completed, only provides a rudimentary snapshot of travel demand.

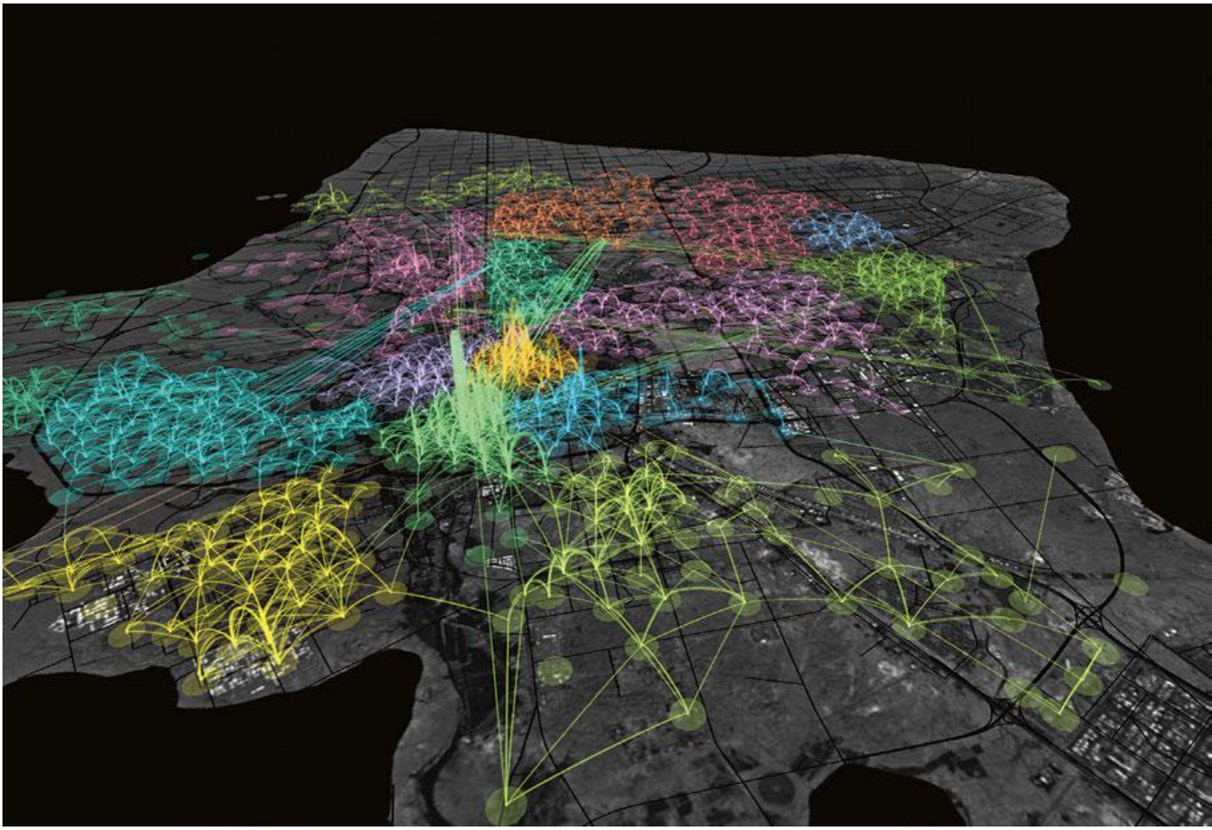
While some have proposed the installation of exhaustive sensor networks to bypass these inadequacies, our approach is to leverage the ubiquity of the sensing apparatus already in our environments, namely cell phones. By collecting and filtering each user's mobile activity as a sequence of cell tower locations, we are able to estimate a population's travel demand in terms of origins and destinations of individual trips. We've shown that these approximated O-D flows hold a strong correlation to census estimates (Calabrese et al. 2011); however, this approach includes the added benefit of capturing travel demand at highly dynamic time slices ranging from seasonal variations to hourly fluctuations. Such a high temporal resolution has the potential to reshape our understanding of urban mobility.

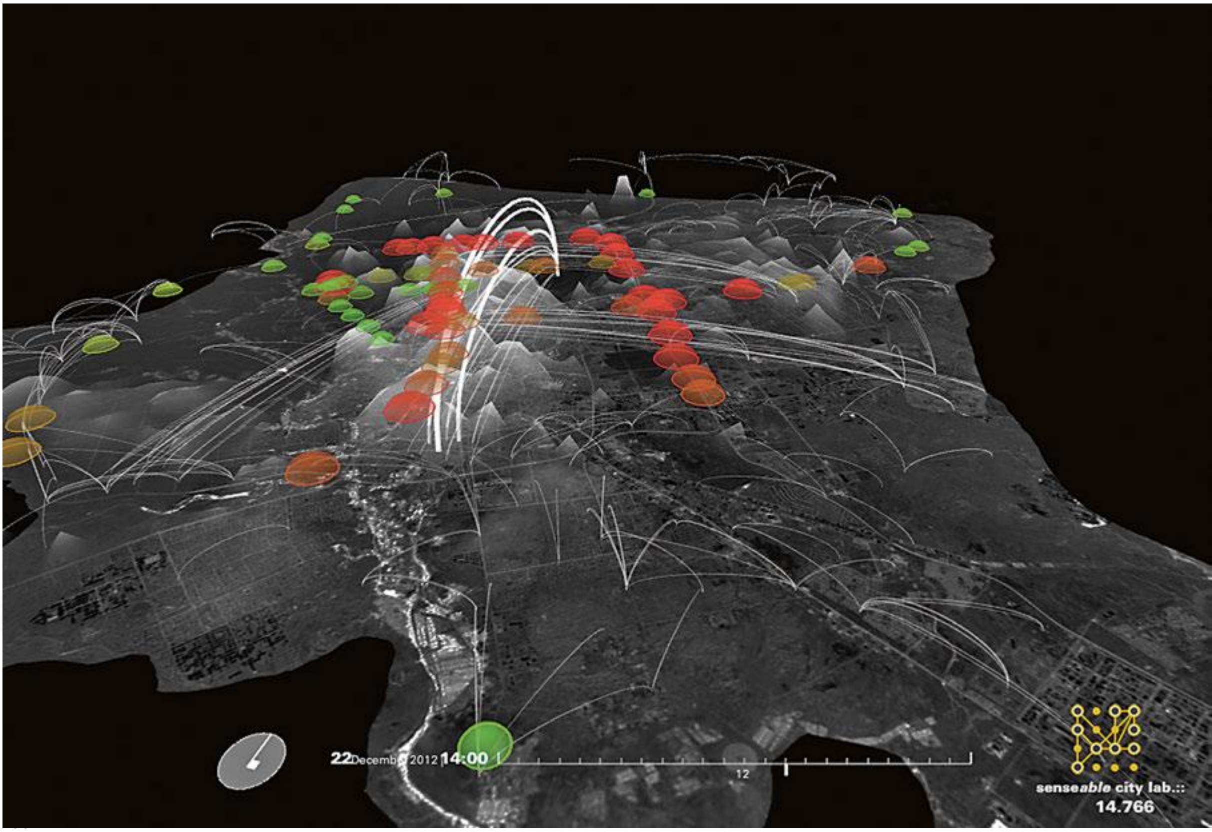
We first constructed O-D matrices on an hourly basis in an effort to show them alongside the network activity. In visualizing the result, we represent each "trip" as an arc that rose from originating to terminating cell tower. Each arc embodies a variable number of trips, and to illustrate this we altered its thickness and height in correspondence to the intensity of activity along that route (on a logarithmic scale). To further highlight directionality, a color scheme has been applied that shows origins in blue and destinations in green (figure 7). The O-D arcs are drawn over the same city base geography, on top of the social interaction mesh from above, in an effort to reveal unseen connections between the two data sets.

The resulting dynamic maps hold a striking similarity to the local intuition of vehicular flows across the city, and the overall O-D arcs correspond quite closely to the underlying street network. Most notably, the visualized results show intense activity along the city's main arteries, King Fahd Road and both the Northern and Eastern Ring roads. Additionally, after conducting a series of workshops and interviews with community leaders and planning professionals, these results were found to be in close alignment with the citizenry's subjective understanding of commuting patterns. To further validate our findings, we compared the estimated O-D flows with the best ground-truth measurements of roadway activity that were available: historical traffic counts. These volumes were collected by moving a small collection of pneumatic tube sensors across the city, intersection by intersection, at forty-eight-hour intervals.

These counts were built into the visualization as half-spheres placed at their respective intersections. Each sphere changes shape and color at an hourly rhythm in line with the measured volume. Again, the main sections of the city line up quite nicely, however, some intriguing O-D activity can be seen to the southeast of the city center that, unfortunately, has no corresponding car count figures to compare against. This remains an open area of exploration for the future.

The final step of this line of analysis will be transforming these mesoscale commuting flows to activity on the road network itself.

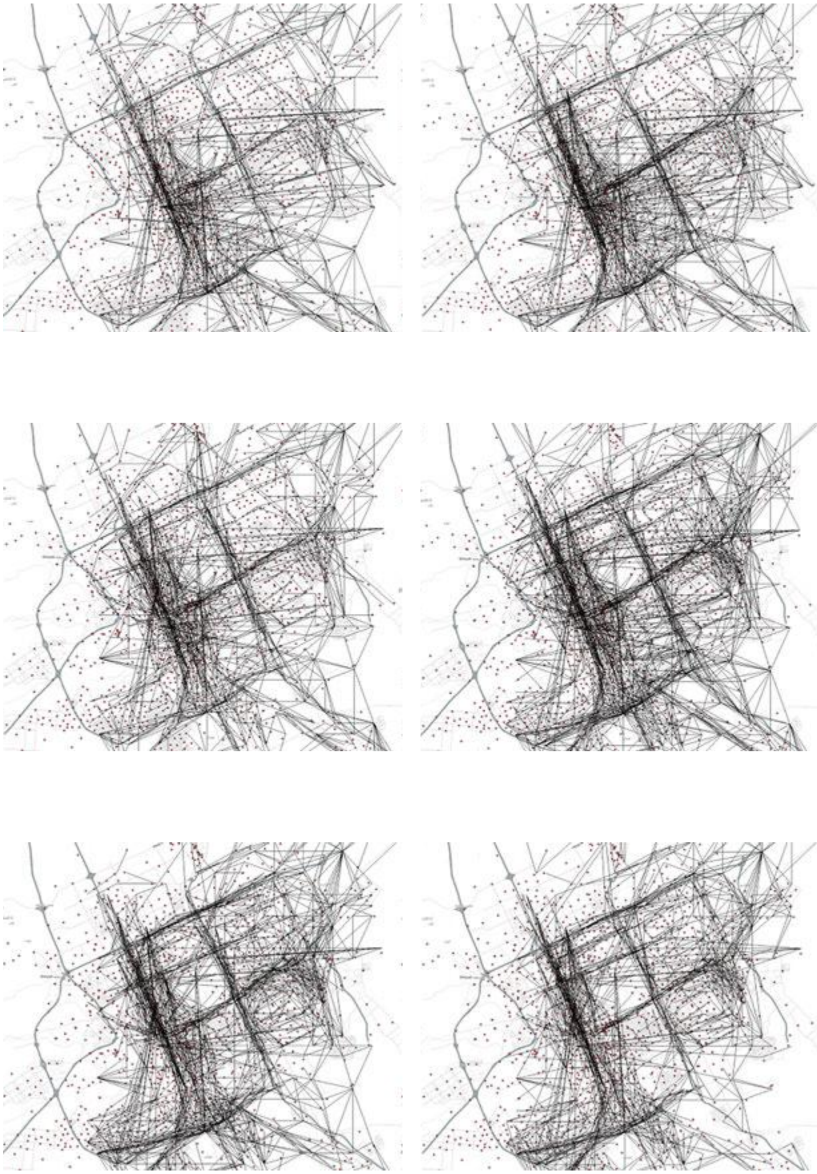



22
2012
14:00
12
senseable city lab.::
14.766

Facing page, top:
Fig. 8.
Mobility communities resulting from the partitioning process. Visualization: Kael Greco, SENSEable City Lab 2013

Facing page, bottom:
Fig. 9.
Combining O-D flows with historical traffic counts.

Right: **Fig. 10.**
O-D matrices across Riyadh from 6:00 a.m. to 9:00 a.m., 9:00 a.m. to 12:00 p.m., 12:00 p.m. to 3:00 p.m., 3:00 p.m. to 6:00 p.m., 6:00 p.m. to 9:00 p.m., and 9:00 p.m. to 12:00 a.m.



By probabilistically transposing our collected O-D trips to a detailed Geographic Information System (GIS) database of road segments, we are able to quantify the impact various mobility communities have on the aggregate transportation network – demonstrating the potential for an exhaustive real-time representation of commuting. The technique has the power to quantitatively identify both overburdened arteries and the specific congestion points that initiate the city's daily gridlock **(figure 6)**. Similar studies have found that just a small number of drivers from a small number of neighborhoods can be responsible for tying up key roads. Due to their reliance on already sensitive metropolitan roads, Wang et al. (2012) identified 15 census tracts (out of 750 total) at the heart of Boston's traffic problem. Whether or not this will be true in Riyadh remains to be seen.

The Speculative Frontier: Seeing Data through Riyadh

The preceding research utilized inherently "social" data to uncover patterns of large-scale human dynamics across the urban form, resulting in temporally precise representations of Riyadh's pulse and cyclically shifting directionality. With this introductory analysis under way we can shift our focus to the other end of the continuum that was put forth in the introduction – uncovering hidden social trends by filtering the network of human interaction through the city itself. There are a multitude of contextual factors that make Riyadh, and the Kingdom of Saudi Arabia as a whole, a fascinating subject for geospatial cultural analysis. This begs a number of more theoretical questions: What compelling sociocultural factors have we captured in our data that can be activated through the frame of the city? Returning to John Snow's cholera map, are we able to reveal hidden facets of social life by affixing our data to the structure of the city? And consequently, what can this transposition further teach us about the character and composition of urban form in Saudi Arabia? What follows is merely a trajectory for further study. Its primary objective is to highlight questions whose solutions could be found somewhere at the murky interface between social and physical geographies. These brief thought experiments are by no means exhaustive, but they are presented in the hopes of sketching a periphery for future inquiry.

Starting with the broad context of Saudi Arabia, one can't help but notice gender segregation is the most obvious place to begin any sort of comparative social analysis. The nation is known for its rigid gender norms. From the compulsory *niqabs* (veils) to the requisite accompaniment of a male guardian, Saudi women face a heavily protected – if not unconditionally restricted – existence. Can the backdrop of the city shed light on how this defines daily life? Using our anonymized CDRs in combination with demographic information will allow us to empirically explore how this segregation manifests itself across the country. In particular, Saudi Arabia holds the dubious distinction of being the only country in the world that prohibits women from driving cars. With few options for public transportation, women must instead depend on either male relatives or hired drivers to move about their environs. Is it possible to quantify this restriction spatially – can we reconstruct and represent the geography of exclusion? What would such a landscape look like? Which areas are associated with the most and least female accessibility and how are they distributed across various cities? Can we then disentangle the culturally imposed restrictions from those that are made physical through urban design? How does this segregation manifest itself through the human interaction network – does the spatial constraint hold here as well? And finally, what are the resulting implications for city policy makers, planners, and managers?

Another cultural phenomenon that is unique to the Arab world is the daily call to prayer. As intimated above, we found an intriguing pattern in mobile activity distributions that was unlike any other country or city we've analyzed before. At various points in the day activity would simply drop off for around thirty to forty minutes before picking back up to its typical trend. These inactivity "valleys" were actually the result of these daily prayer times. Millions of Muslims across the country put down their phones to turn and face the holy city of Mecca to give prayer five times a day. Shops and businesses essentially close down for roughly twenty to thirty minutes while the religious police – the Mutaween – surveil the streets in the hopes of sending all loiterers to the nearest mosques. To our surprise, our activity distributions very closely capture this behavior. The precise timing of these calls to prayer depend on the position of the sun in the sky, and thus, by differentiating the CDR distributions into western, central, and eastern regions, we are able to see the prayer times moving across the country as demonstrated in the figure above. This presents another series of intriguing questions. As we've already seen in the "social pulse" visualization **(figure 3)**, this sudden dip in cellular activity is identifiable when applied to the geography, but can we quantify and map the intensity of the disruption and show which areas are most affected by calls to prayer? What is the relationship of disruption to the spatial organization of urban environments – how would our measure of intensity correlate to patterns of land use? Does it follow the density distribution of mosques? Leading from this, can we detect and illustrate how prayer time disruptions are expressed through mobility? Do average trip lengths shorten during prayer windows as one would anticipate? Lastly, can the intensity of disruption serve as a proxy for regional religiosity? Can we find any correlation between religiosity and characteristics of the human interaction network?

We will continue forward with all of these questions in mind, using the urban context as the pivot point on which we can orient our gaze in the search for answers. And through this careful, back-and-forth negotiation into and away from the spatial frame, we hope to arrive at a collection of representations that capture new ways of seeing both the city and the social forces operating through the city.

Beane, Silas R., Zohreh Davoudi, and Martin J. Savage. 2012. "Constraints on the Universe as a Numerical Simulation," http://arxiv.org/abs/1210.1847 (accessed May 4, 2013).

Calabrese, Francesco, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. 2011. "Estimating Origin-Destination Flows Using Mobile Phone Location Data," *IEEE Pervasive Computing* 10, no. 4 (April): 36–44.

Corner, James. 1999. "The Agency of Mapping: Speculation, Critique and Invention," in *Mappings,* edited by Denis Cosgrove. London: Reaktion, 213–252.

Perry, Tony. 2007. "Navy to Mask Barracks Shaped Like a Swastika," *Los Angeles Times,* September 26. http://articles.latimes.com/2007/sep/26/local/me-swastika26 (accessed April 30, 2013).

Tufte, Edward. 2001. *The Visual Display of Quantitative Information,* 2nd ed. Cheshire, CT: Graphics Press.

Wang, Pu, Timothy Hunter, Alexandre M. Bayen, Katja Schechtner, and Marta C. González. 2012. "Understanding Road Usage Patterns in Urban Areas," *Scientific Reports* 2: article 1001.

Places – Implications for Design

Andres Sevtsuk

Networks of the Built Environment

“Cities happen to be problems in organized complexity,” commented Jane Jacobs on the ballet of daily life on Manhattan’s Hudson Street in the 1960s. “The variables are many, but they are not helter-skelter; they are ‘interrelated into an organic whole.’” (Jacobs 1969, 433) Jacobs was reacting to the prevailing urban design discourse of her day, which claimed to bring order to complex social life through formal interventions. The problem was, Jacobs argued, that architects and planners had a poor understanding of the social and economic interactions that take place in dense urban environments and their interventions were therefore disconnected from the real needs of a place. Further, it was not clear to what extent the form of the environment played a role in their success at all, since the activities of Hudson Street were shaped by numerous cultural, historical, and geographic factors beyond form.

“How much a park is used depends, in part, upon the park’s own design,” Jacobs remarked. “But even this partial influence of the park’s design upon the park’s use depends, in turn, on who is around to use the park, and when, and this in turn depends on the uses of the city outside the park itself. Furthermore, the influence of these uses on the park is only partly a matter of how each affects the park independently of the others; it is also partly a matter of how they affect the park in combination with one another, for certain combinations stimulate the degree of the influence from one another among their components. [...] No matter what you try to do to it, a city park *behaves* as a problem in organized complexity, and that is what it is." (Ibid., 433–434)

Fifty years forward, the challenge of describing and analyzing complex spatial interactions in the built environment still remains one of the central challenges for urban design (Batty 2005). It would be naive to suggest that urban designers lack the interest or the willpower to delve into the social organization and invisible forces that shape places in contemporary cities. On the contrary, there is ample evidence that investigating the workings of diverse and heterogeneous urban environments in detail is widely popular (for example, Belanger et al. 2001; Rienets, Siegler et al. 2009; Busquets 2006; Sorkin 2009).

Some critics have suggested that developing a better understanding of the interactions between social processes and urban form is also hampered by designers' limited education in social sciences. A number of urban sociologists have alerted urban designers to remain wary of what Webber has called "some deep-seated doctrine that seeks order in some simple mappable patterns, when it is really hiding in extremely complex social organization instead" (Webber 1963, 54). While it is true that most designers are not trained in qualitative and quantitative methods of social analysis, there is also a rich body of literature within as well as outside of the urban design field to offer rigorous examples of good social analysis that uncover the complex interaction between the physical configuration of space and its occupancy patterns (Gehl 2010; Whyte 1980; Peattie 1968; Gans 1962). The studies demonstrate that what might appear as complexity to an outsider typically conceals order that remains yet to be uncovered or, as Jacobs put it, organized complexity.

In this chapter we argue that managing spatial and social analysis of complex urban environments is not only challenged by research methods and analytic skills needed to describe and investigate the interactions between the form and function of a place, but also by the conventions of spatial representation in which the problems under study are depicted. We argue that the most pervasive medium for describing the built environment – the plan – comes with certain limitations that make it difficult to use for studying complex spatial interactions between different users of a neighborhood. Every built environment contains a spatial order, which determines relationships of proximity and adjacency between different buildings, public spaces, and routes that connect them. These relationships influence how different circulation routes are utilized, how visible or connected public spaces are, or how conveniently buildings are located with respect to one another. These spatial patterns, in turn, determine what places are better or worse for particular land uses, which public spaces different building tenants routinely encounter, and how the activities of one space might influence the others. We suggest that a network representation of the built environment offers an effective framework for capturing and operationalizing such relationships of urban form.

The plan – a two-dimensional depiction of the form, and sometimes of the functions of a built environment – remains the best known and the most utilized medium of spatial representation among designers and scholars of the city (Conzen 1960; Moudon 1986; Anderson 1993). Plans are powerful tools that convey spatial information in ways that are readily comprehensible to professionals across disciplines. Yet plans can be misinterpreted, and the rich variety of content and meaning they contain is easy to miss (Mandelbaum 1990; Hoch 2002; Ryan 2011). Perhaps most important for the study of complex urban environments, plans store a wealth of information on the built environment, but leave all interrelationships of proximity, adjacency, and interconnectivity between its various elements to be gauged and interpreted by the eyes of a reader. Plans do not embody explicit information about the connections between its elements (e.g., streets, buildings, institutions, etc.); these connections need to be estimated visually by inspecting what is connected to what, how, and why. Put alternatively, plans are rich in elements of the built environment, but poor in conveying the interrelationships between these elements; the quality of their analysis is consequently largely dependent of the quality of their analyst.

Reading spatial relationships from a plan is possible, but labor-intensive and far from trivial. While one-to-one relationships are generally easy to read from a plan – reading a route from a subway station to a particular building is relatively simple – doing the same for one-to-many relationships can be complicated. Gauging the relationships from a subway station to all possible buildings that are located within a ten-minute walk along the available pedestrian routes is not trivial and takes time. Yet this might be an important criterion for deciding the location for a new station. Add to that the constraints of street crossing (e.g., traffic lights, underpasses), a narrowed focus on only buildings of a particular use category (e.g., only residential buildings) and the different sizes of the buildings (e.g., the number of dwelling units in each building) and we quickly arrive at a complex problem that is hard to digest. Business owners choose locations according to access to their clients or suppliers, residents according to nearby amenities, and municipal infrastructure investments are more likely to be approved for more utilized sites. Such relationships are important to understand if planned environments are to attract their desired users and public spaces their desired activity patterns. Gauging how the built environment might impact such decisions from a plan is difficult and requires multiple sets of spatial relationships to be read simultaneously. Doing it fast enough to keep pace with an urban designers' thought process is even more challenging. Human brains tend to operate in a serial manner and are quite poor at processing multiple parallel computations simultaneously (Minsky 1988). The reader may try, for instance, to memorize two or three limited sets of numbers simultaneously. An analysis of spatial relationships in real urban environments may necessitate a processing of hundreds or thousands of such relationships in parallel.

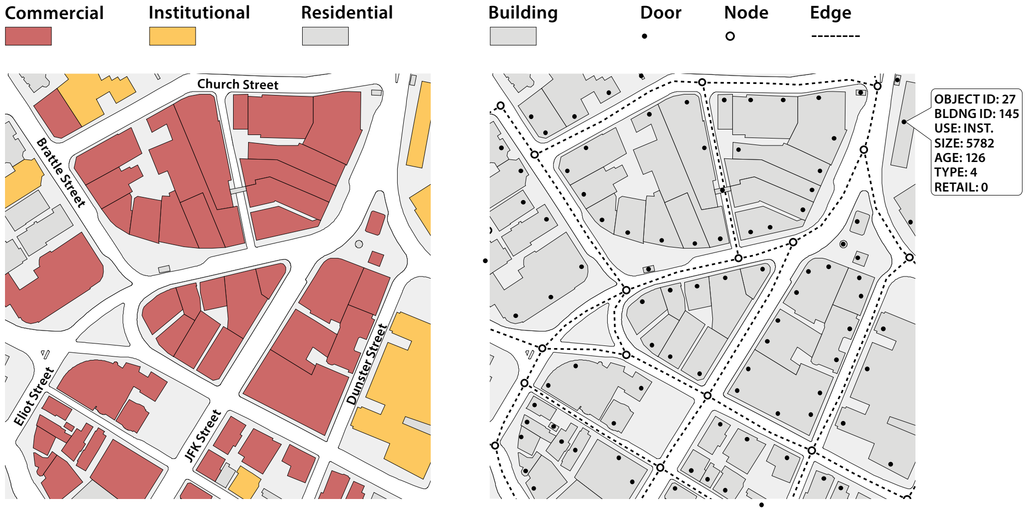
In order to represent and analyze such complex spatial relationships, urban designers and planners have started to use network-based models of the built environment. Unlike traditional plans, network-based representations of urban space encode explicit relationships between the elements of the network, documenting, for instance, how streets are connected to one another, how long the travel times between different districts, buildings, or rooms are, or how many people commute between them. Such linkage information is typically stored in one of two ways. First, it can be stored in a full origin-destination (O-D) matrix, where every element of a plan (e.g., zone, street segment, building, firm, etc.) is shown in a data column next to every destination, and a separate column is used to indicate the desired linkage information about each such connection. The linkage column may contain any kind of connectivity information, such as travel time, the amount of workers commuting between the origin and destination, the amount of economic inputs or outputs exchanged between them, and so on. This approach is relatively easy to analyze using database queries that can retrieve the desired spatial relationships between a set of origins and destinations. But this convenience comes at the cost of information storage – representing relationships between all individual location pairs in a separate table row requires very large tables, which grow as a square of the number of observations. For only 100 locations, the number of connections is 10,000. If all the relationships are symmetrical, that is, if connections from A to B have the same characteristics as those from B to A, then the table can be reduced to half the size. But with tens of thousands of locations, it may still be too large to analyze.

The second, more economical, approach is to represent all spatial relationships with an adjacency matrix. An adjacency matrix does not summarize the information about the full routes between each related location pair in the environment, but instead only stores the immediate neighbor adjacencies for each location. If the environment is modeled as a network of neighborhoods, then the adjacency matrix would capture each neighborhood's relationship to only its immediately adjacent neighborhoods. If the environment is modeled as a network of buildings and streets, then the adjacency matrix would capture each building's relationship to only its immediately adjacent buildings along the street network. Useful network analysis algorithms can then query this information and infer the full spatial relationships between all elements of the network from this shorter table. Querying the adjacency matrix requires more advanced algorithms than querying a full O-D table, but a lot less storage space. Contemporary algorithms for processing such tables allow vast spatial interactions to be analyzed in seconds (Vanegas et al. 2009).

There are a number of different ways of representing such information in networks and tables. What is important, however, is not so much the precise form of the network representation used – centered on land-use or urban form (Bhat et al. 2000), using actual network routes or as-a-crow-flies connections (Anselin 1988), primal or dual network representations (Hillier 1996; Porta et al. 2005), two-element or three-element networks (Sevtsuk 2010) – but the fact that spatial relationships in a given environment are depicted numerically, such that all desired linkages between places are explicitly encoded in a relationship table. These spatial relationships may depict connectivity in terms of traffic, material, information, or financial exchange. This is a major departure from traditional plans that has occurred quietly for most urban designers and physical planners during the past decade. Instead of requiring the reader of a plan to infer complex spatial relationships embedded in the environment visually and intuitively, network-based representations encode such information explicitly and allow the user to access large combinatorial summaries of spatial connections on the fly. Network models automate the analysis of numerous parallel relationships in urban space and allow the analyst to use that information in urban design decision making almost instantaneously. This is profoundly changing how we describe and analyze complex urban environments, paving a way for more informed decision making in real-world planning problems.

In the following we describe one of such models – the Urban Network Analysis Toolbox – developed at the City Form Lab (Sevtsuk and Mekonnen 2012). There are many other network-based approaches to describing built environments; we use the one we have developed to illustrate the more general functionality of network representations of urban space (Levin 1964; Casalaina and Rittel 1967; Rittel 1970; Tabor 1970; March and Steadman 1971; Hillier 1996; Porta et al. 2005; Xie and Levinson 2007; Okabe and Sugihara, 2012; Miller and Wu, 2000; Jiang and Claramunt, 2002 ; Peponis and Bafna, 2008; Vanegas et al. 2012).

The Urban Network Analysis Toolbox – an open-source and free plug-in for ArcGIS – models the built environment using three basic elements: edges, representing paths along which travelers can navigate; nodes, representing the intersections where two or more edges intersect; and buildings, representing the locations where traffic from streets enters into indoor environments or vice versa. Buildings can be replaced by any other point locations on



Origin	Destination	Distance	Transactions
1	2	242	3
1	3	11	0
1	4	95	0
1	5	141	1
2	1	56	2
2	6	136	8
2	7	25	0
3	9	12	0
3	11	85	0

Fig. 1.
Left: **Plan representation of Harvard Square**
Right: **Network representation of the same area, with an adjacency matrix below**

the network: public spaces, transit stations, utility facilities, etc. Our unit of analysis thus becomes a building (or other location identifier on the network), enabling the interrelationships to be computed separately for each building.

Each building, street, and intersection carries an additional set of attributes describing its real-life properties. These attributes, stored in another table, can describe any measurable properties of these elements: for buildings, their size, height, establishment mix, demographic occupancy, etc.; for streets, their directionality, traffic capacity, sidewalk characteristics, etc. The weighted representation of interconnected elements opens up a range of possibilities for studying different kinds of spatial relationships between buildings in a network of city streets. This network representation framework is illustrated in figure 1. The left side of the figure presents a fragment of Harvard Square in Cambridge, Massachusetts, in a plan drawing, with color-coded land uses. The same plan drawing is shown as a network on the right. Each building in the network is connected to its nearest circulation path at a discrete location – its entrance doors in this case. Note, however, that a building can have several doors and connections to the network.

A network representation of both form and function of an area provides a basis for complex spatial analysis. The three elements of urban form describe the physical pattern of urban infrastructure – the two- and three-dimensional geometry of built form and its circulation routes, the shape of public space, and paths that connect them. Using attributes within these categories allows us to further differentiate the parameters of these elements – building volumes, the spacing or placement of buildings with respect to circulation spines, the capacity or direction of routes, etc. Table attributes also allow us to describe the functions of each element – which activities are located where, how many people they accommodate, and how the activities connect with one another. Activities are typically categorized into loose groupings, such as living, working, or playing spaces, but they can also change from one activity to another or intensify in use depending on time of day or day of week. Together, such indicators aggregate into a complex description of a place, where everything can be related to everything else around it (Tobler 1970). The relationships are not helter-shelter, they are explicitly encoded into the adjacency matrix and attribute table, organized by the analyst. Let us now look at the Bugis area in Singapore to apply this type of a representation on a real, complex urban environment.

Bugis is located in downtown Singapore, encompassing an area of roughly a square kilometer. It is a historical area that was developed as part of the Raffles Plan and covered through the nineteenth and early decades of the twentieth centuries with traditional shop houses. Since the 1960s, the area has been gradually redeveloped with multistory deep-floor-plate commercial structures that accommodate a vast, heterogeneous mix of activities.

Figure 2 illustrates an interior view of Bugis Street, a multistory bazaar of hundreds of small retail and food businesses located at the center of the area. There are a total of more than 4,000 individual businesses including 1,769 retailers, 559 service providers, 519 eating and drinking establishments, 130 offices, 38 hotels, 24 educational institutions, and 19 entertainment facilities within an area of roughly 0.8 square kilometers around the Bugis Mass Rapid Transit (MRT) station. Bugis is one of the busiest, and indeed most complex urban environments in Singapore.

Fig. 2.
Interior alley of Bugis Street in Singapore

The area was surveyed by researchers of the City Form Lab in fall 2012, who recorded every door, building, and business, along with their size, use category, and a few other economic characteristics. The survey covered all publicly accessible floors in the area, with roughly half of the businesses on the ground floor and the other half on the upper floors or underground. The researchers also documented the entire pedestrian path network in the area, both indoors and outdoors, on grade, above and below grade – observing over 32 linear kilometers of walking paths within less than a square kilometer of land: 35 percent of these paths were outdoors, 26 percent outdoors but covered (e.g., arcades), and 37 percent were indoors on various levels. **Figure 3** shows this information encoded in a network. The red dots indicate individual businesses, the gray lines the pedestrian paths, and the black lines the ground-floor structural building walls.

We demonstrate network analysis of this area using two types of spatial connectivity indices: Betweenness and Reach (Sevtsuk and Mekonnen 2012). The first involves foot-traffic prediction in different parts of the site; the second models accessibility to food establishments.

In order to estimate where and how people might be walking in different parts of this area, we looked at walking routes from the MRT station to retail destinations in the whole area. According to interviews on site, a large part of the crowd in Bugis comes there to shop by MRT. We based the analysis on the assumption that

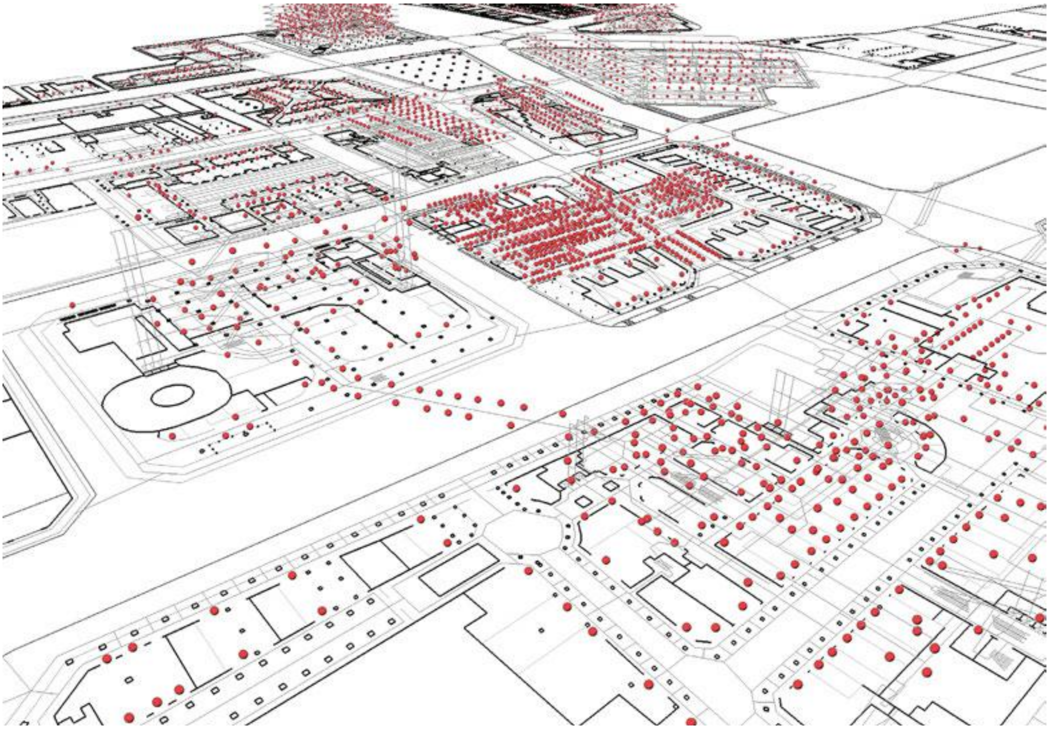


Fig. 3.
Three-dimensional network representation of the Bugis area in Singapore

a walking trip is made from the Bugis MRT station to each of the 1,769 retail establishments in different parts of the site along the shortest available path. We modeled all these paths using Betweenness analysis in the UNA Toolbox and kept track of which network segments are most trafficked in the process. The Betweenness metric thus captured the number of estimated passersby at each network segment who walk from the MRT station to a retail destination along the shortest paths.

Figure 4 shows the results, color-coding the footfall from green to red as the traffic increases. We find the highest expected pedestrian activity in Albert Mall, Bugis Street, and Bugis Junction – all major shopping destinations in the area. There is also a peak of activity near Arab Street and Haji Lane, both historical streets, lined almost continuously with stores in old shop houses. Each of these places is indeed crowded in reality (figure 2). Perhaps more important, network analysis allows us to predict not only general areas of activity concentration, but even particular

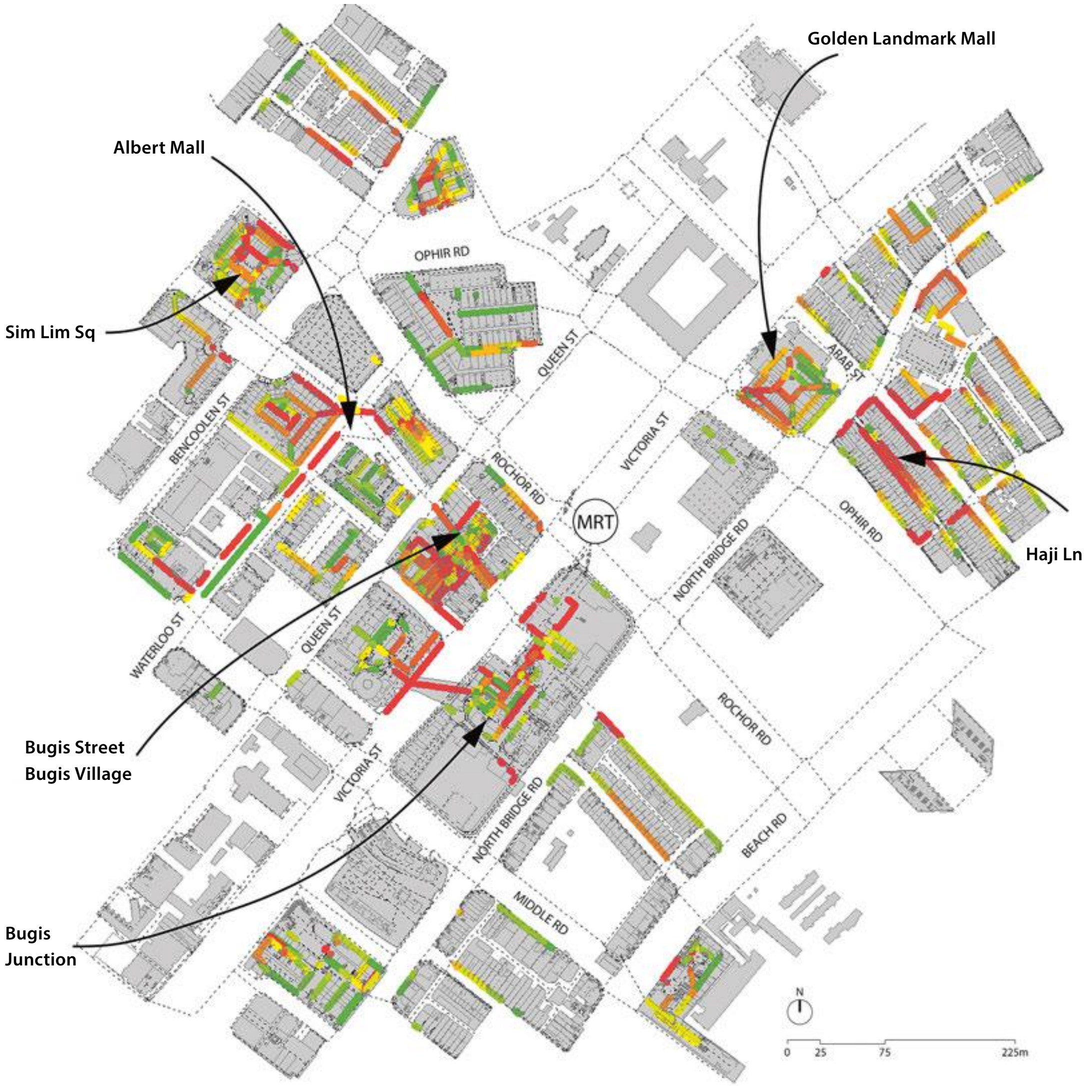


Fig. 4.
Betweenness analysis, indicating expected pedestrian traffic from the Bugis MRT station to all individual retail destinations in the area

Fig. 5.
Reach analysis, indicating how many eating and drinking places can be reached on foot within a 200-meter radius from each door

street segments or indoor corridors throughout a whole district that are likely to attract a lot of traffic. The variation in movement flow on different segments can be an important factor for explaining their business mix and use patterns. Eating establishments, for instance, often prefer to locate on paths with high foot traffic between other origins and destinations. Businesses that do not depend entirely on passersby but value customer spillovers from other businesses, may occupy the second-best locations that remain near pedestrian currents, but on side streets, where rents are cheaper (Sevtsuk 2010).

When we analyzed which locations in this area have best access to food establishments, we see that restaurants and drinking places do locate quite close to retail destinations and the pedestrian routes that lead to them. **Figure 5** illustrates a network-based Reach accessibility metric specified to eating and drinking establishments within a 200-meter walking radius. The Reach index estimates how many particular types of destinations – food establishments, in our case – are available from each origin within a given walking radius (e.g., 200 meters). The more eating and drinking destination available, the higher the index.

The results in **figure 5** suggest that restaurants, hawker stands, and bars are typically clustered near retailers and the pedestrian paths that lead to them.[1] The highest concentration is found between Albert Mall and Bugis Street, where numerous food stalls cluster at the Albert Market and Food Center. There is also a concentration of food vendors in Bugis Street and Bugis Junction and on Arab Street and Muscat Street, which branches off to its right. In front of Albert Center, a pedestrian can reach up to 146 different food establishments within a 3-minute walk. These four areas are the centers of gravity for food. Overall, there are three times as many retailers as food establishments in the area, but even the lighter green locations on the map reach 30 to 50 eating and drinking venues in a 200-meter walking radius, suggesting that the area is not poor in culinary options.

If we zoom in on one of the walking paths between the Bugis MRT station and a retail store – a computer hardware shop on the fourth floor of Sim Lim Square electronics mall – then we can further qualify the characteristics of a particular route in our analysis

1 We do not assess the statistical significance of these location choices here, but an interested reader may find such an analysis in Sevtsuk (2010).

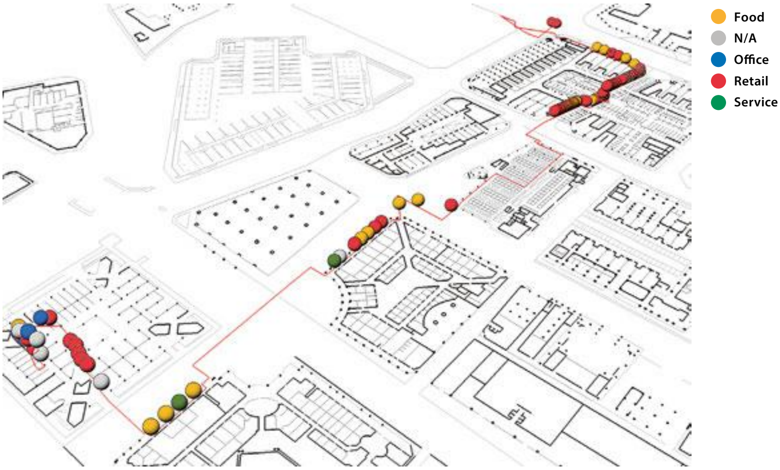


Fig. 6.
A typical walking route from a store on the fourth floor of Sim Lim Square electronics mall to the Bugis MRT station

(figure 6). This path, which depicts a typical shopper's visit to the area, passes 86 businesses before arriving at its destination, 58 of which are retailers, 20 eating places, 8 offices, and 3 service establishments. A comparison of such paths leading to different socio-economic destinations can be valuable for a number of applications – explaining, for instance, the attraction of different paths to pedestrians, how different groups of people experience the city, or for studying microeconomic clustering between establishments.

A networked representation of the built environment presents a powerful framework for describing and analyzing complex urban environments. It is already being used in numerous digital urban models, and its applications are likely to grow quickly in the coming years. Unlike traditional plans, network models of urban space explicitly encode information about the connectivity between different actors and places they represent, making complex spatial analysis between the different elements of the environment possible within seconds on a computer. They overcome the slow and challenging process of reading spatial relationships, typical of traditional plans. But the analysis of spatial networks therefore also depends on the relationships that have been encoded into their tables. Documenting such relationships is an important first step in making use of such methods.

Networked representations of city environments are not, however, alternatives to traditional plans, but rather complements. As urban designers, we know that visual readings of plans are more nuanced, sensitive, and powerful than anyone might be able to explain. Plan reading will remain vital to urbanists. Underlying network connections embedded in these plans can simply augment the static representations with powerful spatial interconnections that are hard to gauge otherwise. They help automate labor-intensive counting and measuring tasks that a reader of the plan may not be able to perform mentally, and allow her to utilize such information instantaneously for studying or manipulating the plan.

The graphic plan interfaces of network models will also allow the analyst to overcome the shortcomings of an overparameterized model. The interrelationships of form and function that network models embody are of course not comprehensive and can miss a number of important dimensions of a place – its history or its broader social, cultural, or environmental context. But the graphic interface of network models allows them to be interpreted in the same way as traditional plans, with supplements. A holistic approach to urban spatial analysis still requires "a pragmatic outlook [that] embraces context and seeks continuity among diverse viewpoints" (Hoch 2000, 54). Digital interconnectivity between the elements of a plan improves rather than hampers holistic thinking.

Finally, we should remember that a novel representation of a place does not necessarily lead to a better understanding of its underlying complexity. But by providing a clear framework for describing a multitude of simultaneous spatial relationships embedded in its structure, network models of the built environment eliminate a major burden of reading such relationships visually and allow the designer to focus on the analysis rather than the description of the problem.

Anderson, Stanford. 1993. "Savannah and the Issue of Precedent: City Plan as Resource," in *Settlements in the Americas: Cross-Cultural Perspectives,* edited by Ralph Bennett. Newark, DE: University of Delaware Press, 110–144.

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer Academic Publishers.

Batty, Michael. 2005. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals.* Cambridge, MA.: The MIT Press.

Bélanger, Pierre, Chuihua Judy Chung, Joshua Comaroff, Michael Cosmas, Sonal Gandhi, A. David Hamilton, Lan-Ying Ip, Jeannie Kim, Gullivar Shepard, Reshma Singh, Nathaniel Slayton, James Stone, and Sameh Wahba. 2001. "Lagos: Harvard Project on the City," in *Mutations,* edited by Rem Koolhaas et al. Barcelona: ACTAR.

Bhat, Chandra, Susan Handy, Kara Kockelman, Hani Mahmassani, Anand Gopal, Issam Srour, and Lisa Weston. 2000. "Development of an Urban Accessibility Index: Literature Review." Center for Transportation Research, The University of Texas at Austin. Report completed for the Texas Department of Transportation.

Busquets, Joan. ed. 2006. *Aleppo. Rehabilitation of the Old City.* Cambridge, MA: Harvard University Graduate School of Design.

Casalaina, Vince, and Horst W. J. Rittel. 1967. "Generating Floor Plans from Adjacency Matrices."

Conzen, M. R. G. 1960. *Alnwick, Northumberland: A Study in Town Plan Analysis,* 2nd rev. ed. (1969). London: Institute of British Geographers.

Gans, Herbert J. 1962. *The Urban Villagers: Group and Class in the Life of Italian-Americans.* New York: Free Press of Glencoe.

Gehl, Jan. 2010. *Cities for People.* Washington, DC: Island Press.

Hillier, Bill. 1996. *Space Is the Machine: A Configurational Theory of Architecture.* Cambridge: Cambridge University Press.

Hoch, Charles J. 2002. "Evaluating Plans Pragmatically," *Planning Theory* 1, no. 1 (March): 53–75.

Jacobs, Jane. 1961. *The Death and Life of Great American Cities.* New York: Random House.

Jiang, Bin, Christophe Claramunt, and Michael Batty. 1999. "Geometric Accessibility and Geographic Information: Extending Desktop GIS to Space Syntax," *Computers, Environment and Urban Systems* 23, no. 2: 127–146.

Levin, Peter Hirsch. 1964. "The Use of Graphs to Decide the Optimum Layout of Buildings," *Architects' Journal* 140, no. 15: 809–815.

Mandelbaum, Seymour. 1990. "Reading Plans," *Journal of the American Planning Association* 56, no. 3 (Summer): 350–356.

March, Lionel, and Philip Steadman. 1971. *The Geometry of Environment: An Introduction to Spatial Organization in Design.* London: RIBA Publications.

Miller, Harvey J., and Yi-hwa Wu. 2000. "GIS Software for Measuring Space-Time Accessibility in Transportation Planning and Analysis," *GeoInformatica* 4, no. 2: 141–159.

Minsky, Marvin. 1988. *The Society of Mind.* New York: Simon & Schuster.

Moudon, Anne Vernez. 1986. *Built for Change: Neighborhood Architecture in San Francisco.* Cambridge, MA: The MIT Press.

Okabe, Atsuyuki, and Kokichi Sugihara. 2012. *Spatial Analysis Along Networks: Statistical and Computational Methods.* Statistics in Practice. Hoboken, NJ: Wiley.

Peattie, Lisa Redfield. 1968. *The View from the Barrio.* Ann Arbor: University of Michigan Press.

Peponis, John, Sonit Bafna, and Zongyu Zhang. 2008. "Connectivity of Streets: Reach and Directional Distance," *Environment and Planning B: Planning and Design,* 35, no. 5: 881–901.

Porta, Sergio, Paolo Crucitti, and Vito Latora. 2005. "The Network Analysis of Urban Streets: A Primal Approach," *Environment and Planning B: Planning and Design,* 33, no. 5: 705–725.

Reilly, William J. 1929. *Methods for the Study of Retail Relationship.* Austin: Texas University Bureau of Business Research monograph no. 4.

Rieniets, Tim, Jennifer Sigler, and Kees Christiaanse, eds. 2009. *Open City: Designing Coexistence.* Amsterdam: SUN.

Rittel, Horst W. J. 1970. "Theories of Cell Configuration," in *Emerging Methods in Environmental Design and Planning,* edited by Gary T. Moore. Cambridge, MA: The MIT Press.

Ryan, Brent D. 2011. "Reading through a Plan." *Journal of the American Planning Association* 77, no. 4: 309–327.

Sevtsuk, Andres. 2010. "Path and Place: A Study of Urban Geometry and Retail Activity in Cambridge and Somerville, MA," Ph.D. diss., MIT, Cambridge, MA.

Sevtsuk, Andres, and Michael Mekonnen. 2012. "Urban Network Analysis: A New Toolbox for ArcGIS," *Revue international de géomatique* 22, no. 2: 287–305.

Sorkin, Michael. 2009. *Twenty Minutes in Manhattan.* London: Reaktion Books.

Tabor, Philip. 1970. *Traffic in Buildings 4. Evaluation of Routes.* Land Use and Built Form Studies. Working paper 20. Cambridge: University of Cambridge.

Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography* 46, no. 2: 234–240.

Vanegas, Carlos A., Daniel G. Aliaga, Bedřich Benes, and Paul Waddell. 2009. "Interactive Design of Urban Spaces Using Geometrical and Behavioral Modeling," *ACM Transactions on Graphics* 28, no. 5 (December): article 111.

Vanegas, Carlos A., Ignacio Garcia-Dorado, Daniel G. Aliaga, Bedřich Benes, and Paul Waddell. 2012. "Inverse Design of Urban Procedural Models," *ACM Transactions on Graphics* 31, no. 6 (November): article 168.

Webber, M., 1963. "Order in Diversity: Community without Propinquity," in Cities and Space: *The Future Use of Urban Land,* edited by Lowden Wingo. Baltimore: The Johns Hopkins University Press, 23–54.

Whyte, William Hollingsworth. 1980. *The Social Life of Small Urban Spaces.* Washington, DC: Conservation Foundation.

Xie, Feng, and David Levinson. 2007. "Measuring the Structure of Road Networks," *Geographical Analysis* 39 (July): 336–356.

Markus Schläpfer

How Polycentric Are Our Cities?

Today’s large metropolitan areas, from New York and London to Singapore and Tokyo, all comprise a multitude of interconnected centers: places where our daily activities for employment, leisure, and travel are highly concentrated. A deep and quantitative understanding of this urban spatial structure is of considerable importance for our contemporary sprawling cities, where it is ever less obvious what constitutes a center. As an example, being able to predict how a new business hub interacts with already existing centers, and eventually changes the citywide mobility patterns, is a prerequisite for developing efficient measures against traffic congestion and air pollution. The recent availability of many new, large-scale data from human activities, such as those automatically collected from mobile phone networks, opens unprecedented possibilities to shed light on the increasing coexistence of multiple centers, offering numerous new applications for urban planning, infrastructural investments, and environmental impact mitigation.

The spatial structure of urban landscapes in terms of centers and subcenters is constantly evolving over time (Batty 2013; Bertaud 2004). From a historical perspective, it was mainly the explosive rise of private car ownership together with the introduction of mass public transportation which has led to a qualitative transformation of almost all major cities during the last century: from a rather simple monocentric structure with a clear central business district (CBD) and a surrounding suburban hinterland toward a more complex polycentric structure which pictures the coexistence of multiple activity centers (Anas et al. 1998). This trend is expected to continue in the near future, giving rise to new challenges for improving the performance of cities at large. It is therefore hardly surprising that the characterization of polycentric urban structures has gained much attention in the urban planning and economics community during recent years (Ibid.). Nevertheless, due to its multifaceted nature, polycentricity (or polycentrism) has remained a rather fuzzy concept, still lacking a commonly accepted definition. Generally, as is illustrated in **figure 1**, the two most important aspects of polycentricity are (1) its morphological dimension, which denotes the size and spatial distribution of centers; and (2) its functional dimension, which additionally addresses the linkages between different centers such as the daily flow of commuting people or the strength of business and social network connections (Green 2007; Burger and Meijers 2012). Morphological polycentricity has been traditionally assessed in

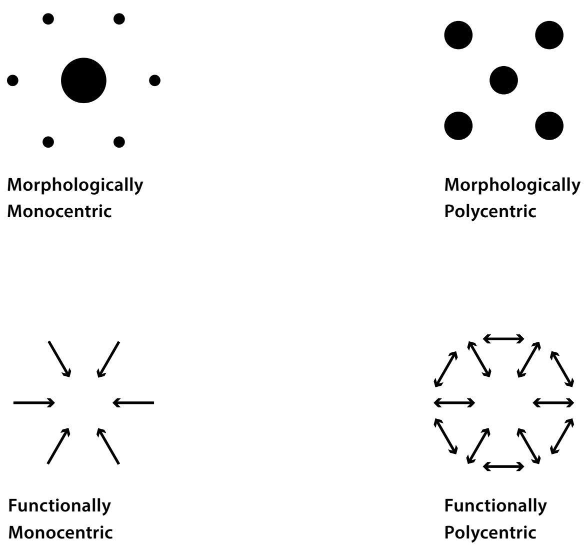


Fig. 1.
Morphological and functional dimensions of polycentricity, adopted from Burger and Meijers (2012)

a rather straightforward manner by applying simple thresholds for population, employment, or business density. It has been difficult, however, to appropriately quantify functional polycentricity, mainly due to the scarcity of large-scale, individual-based-movement or social-interaction data. Consequently, going beyond fundamental conceptualizations, little is known so far about the spatiotemporal interactions of people with different centers and subcenters.

On these premises, the new wave of user-generated data sets from mobile phones, credit cards, online social networks, and numerous other sources, is providing us with an unprecedented potential to scrutinize and eventually quantify these intrinsic urban interactions. Recent research has already taken a first step in this direction by analyzing millions of subway journeys in London based on the Oyster card electronic-ticketing system, showing a hierarchical nature of different centers in terms of stations and commuter inflows (Roth et al. 2011). In order to substantiate the value-added of harnessing the electronic footprint of urban dwellers, the following section provides an exemplary application of analyzing massive mobile phone CDRs (call detail records) to quantify the polycentricity of large metropolitan areas. CDRs provide us with an accurate statistical picture of individual-based people movements in cities (Isaacman 2010), while capturing a large fraction of the population due to the usually high penetration rate of mobile phones. It becomes immediately obvious that these practical strengths outperform traditional, questionnaire-based survey or census data, whose collection is substantially more difficult and cost-intensive, which are limited to a small subset of the population, and which provide a merely static system snapshot.

In the course of an ongoing research project initiated and led by the SENSEable City Lab, we are developing novel algorithms and data-mining techniques, allowing us to translate CDRs into the spatiotemporal distribution of people and to subsequently unfold the complex patterns of polycentricity (Schläpfer et al. 2014). Within the framework of the Singapore-MIT Alliance for Research and Technology (SMART), one of the largest mobile phone operators of the city-state of Singapore has provided us with massive anonymized CDRs for academic research purposes. This data has recently been collected for billing purposes and covers more than half of the population during several consecutive months with approximately 10 million phone call events per day. Each record corresponds to a single call event. It consists of (1) the anonymized (i.e., surrogate) numbers of the two connected mobile phone users, (2) the call duration, (3) the time stamp of the call initiation, and (4) the two cell towers routing the call, together with their geographic location.

This georeferenced call data is utilized to reveal the polycentric structure of Singapore, by identifying the most important locations at which phone users tend to converge during their daily intraurban movements. The first step of our analysis procedure is the calculation of so-called origin-destination (O-D) matrices, a concept borrowed from the field of transportation engineering. Simply put, for each mobile phone user our algorithm estimates the home location (or *origin*) by the cell tower area where most of the calls are made during evening and night hours. We subsequently partition the total urban area of Singapore into a regular 0.5 km × 0.5 km grid (implying a total of 1,423 cells including at least one cell tower), and determine for each user the visited grid cells (or *destination*). As a second step of our analysis, we count for each grid cell the number of users who are living within a given distance and who are visiting the considered location with a given frequency. The result is a comprehensive picture of the relative importance, or *attractiveness,* of each pixel in the grid, telling us how many people are attracted, how often, and from how far away. This, in turn, allows us now to systematically study the spatiotemporal emergence of centers from a citywide perspective. For instance, if we consider only small radii of attraction by limiting our focus to those visits which are made in close vicinity of the home location, we find that the cells which attract a large number of people are highly scattered over the entire city area **(figure 2a)**. This pattern makes intuitive sense, as it may well highlight very local attraction points such as local shopping facilities and other amenities. However, if we gradually expand the radius of attraction by increasingly considering visitors who live further away, the grid cells which attract a large number of people tend to become more and more clustered in space, giving rise to the emergence of distinct centers, which indeed correspond well to our subjective perception of central places in a city. As an example, widening the considered radius of attraction around each grid cell to 15 kilometers, we can visually identify well-known centers in Singapore as coherent areas of cells with a high number of visitors, such as the downtown core in the southern part of the city, Jurong in the western part, or Woodlands in the northern part (figure 2b). If we further widen the catchment area of each grid cell toward including visitors from the entire urban region, we increasingly observe a rather monocentric structure with a pronounced downtown area, as is illustrated in figure 2c. Thus, Singapore's downtown core attracts a large fraction of dwellers from all over the metropolitan area, while other centers such as Jurong seem to have a smaller catchment area from which the majority of visitors is attracted. In other terms, the degree of polycentricity is highly dependent on the spatial scale of the assessment.

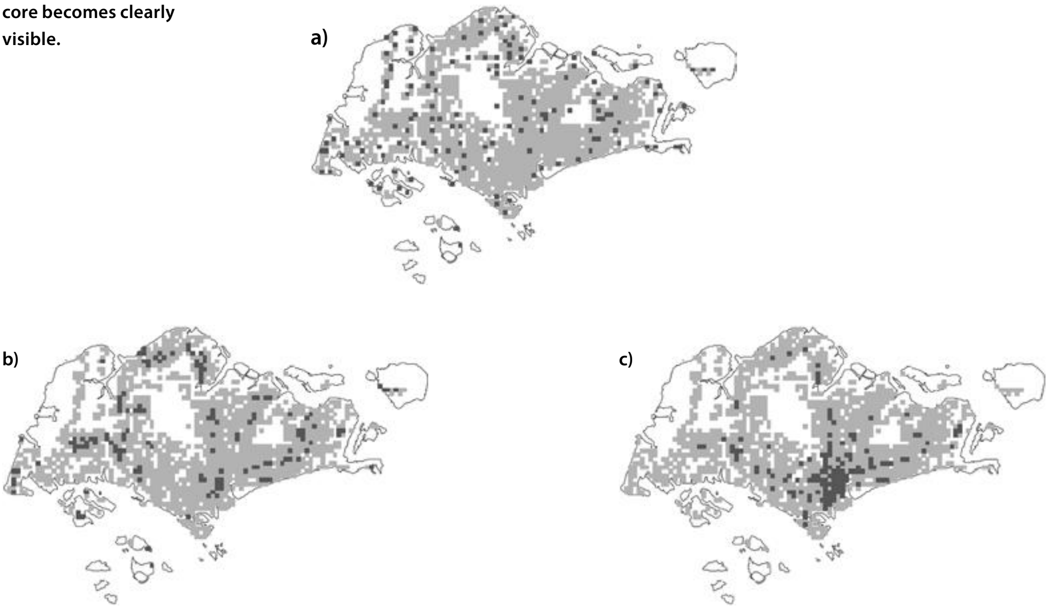


Fig. 2.
Revealing the spatiotemporal polycentricity in Singapore. The marked (dark) cells correspond to the top 10 percent of locations in terms of the relative people inflow.
a) For low values of the radius of the considered catchment area (*r* = 1 km) highly localized and dispersed centers can be observed.
b) For medium distances (*r* = 5 km) centers of regional importance emerge.
c) For a large radius of the catchment area (*r* = 15 km) the predominance of the downtown core becomes clearly visible.

Going beyond visual inspection, we can quantify the change of the relative importance of each center with increasing spatial scale by measuring the size of each center as the number of connected grid cells (sharing a common border) with a given people inflow. As shown in figure 3 for the examples of the downtown core and Jurong, the size of both centers first increases with the growing

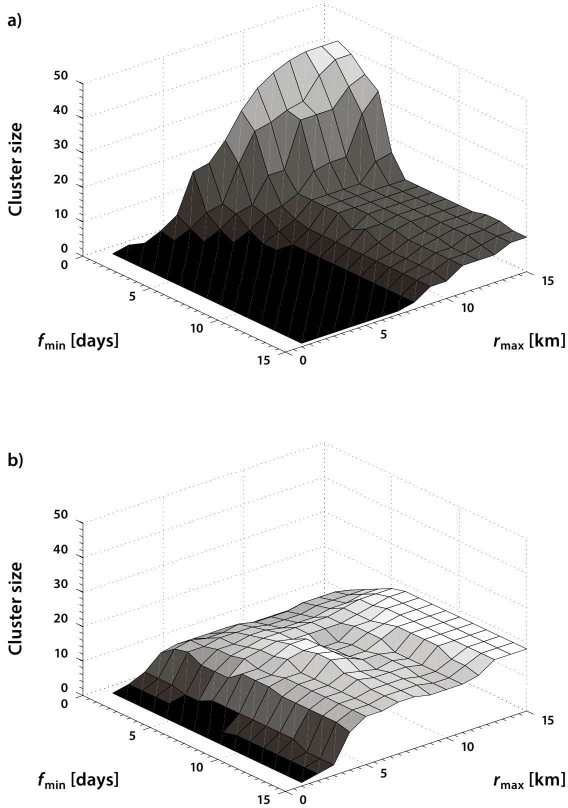


Fig. 3.
Relative importance of urban centers as measured by the number of connected cells (cluster) that belong to the top 10 percent of all cells in Singapore in terms of the relative people inflow, as a function of the radius of the catchment area r_{max} and the minimum visiting frequency f_{min}.
a) Downtown core;
b) Jurong

radius of the considered catchment area. However, after exceeding a certain radius the size, and thus the relative importance, of Jurong stabilizes, while the downtown core is gradually increasing in terms of people inflow. This changing importance of different centers with increasing geographic scale, as we have observed for the case of Singapore, suggests an intrinsic hierarchical organization of different centers and subcenters.

Besides clarifying the role of the spatial scale in the characterization of polycentricity, CDRs also provide rich insights into the role of the temporal scale. To that end, we assessed the impact of the frequency with which each grid cell in Singapore is visited. Similar to increasing the distance of the home location, we considered only those users who visited a given cell with a minimum frequency, which we define based on the minimum number of days a user connects to the network while being in this cell. Figure 3 shows the relative importance of the two exemplary centers, again expressed in terms of their size, as a function of the minimum visiting frequency. The relative importance of the downtown core decays rapidly when restricting the analysis to regular visitors only. Hence, merely irregular visitors contribute to the overarching importance of the downtown core. Conversely, the relative importance of the more local center Jurong shows even a slight increase with increasing visiting frequency (figure 3), thus attracting mainly visitors on a regular basis. This shows that the temporal aspect of the attractiveness of urban centers has to be considered as important as the spatial counterpart. It is interesting to note that our findings based on the intraurban flow of people seems to be in line with longstanding principles of economic geography, explaining the central location of those specialized goods and services which attract a larger pool of people but on a less regular basis (Krugman 1996). A prominent example is the Central Place theory of Christaller, which attempts to explain the spatial distribution of cities (Christaller 1933).

The present study of Singapore shall mainly serve as an illustrative example for the promising potential of analyzing heretofore unavailable, user-generated data, in light of developing a deeper understanding of the complex organization of large metropolitan areas. In particular, we leveraged the information provided by mobile phone CDRs so as to uncover the trajectories of urban dwellers and to subsequently identify the spatial and temporal scale as the two main ingredients of polycentricity. These basic insights may be taken as a starting point toward a formal, data-driven definition of functional polycentricity. Moreover, it will be interesting to enrich the study by the quantification of people flows between different centers, as well as to extend it to other major cities across different geographies, cultures, and economies, so as to compare their overall performance and efficiency in terms of travel behavior or environmental impacts such as emissions. Thereby, the necessary data collection needs to be done in a privacy-respecting way, which requires novel data mining, storage, and anonymization approaches (Trantopoulos et al. 2011). Nevertheless, such insights gained may hopefully help city planners and decision makers to navigate toward a sustainable urban future while maintaining a high quality of life.

Anas, Alex, Richard Arnott, and Kenneth A. Small. 1998. "Urban Spatial Structure," *Journal of Economic Literature* 36: 1426–1464.

Batty, Michael. 2013. *The New Science of Cities*. Cambridge, MA: The MIT Press.

Bertaud, Alain. 2004."The Spatial Organization of Cities: Deliberate Outcome or Unforeseen Consequence?" Working Paper 2004-01. University of California, Berkeley.

Burger, Martijn, and Evert Meijers. 2012. "Form Follows Function? Linking Morphological and Functional Polycentricity," *Urban Studies* 49: 1127–1149.

Christaller, Walter. 1933. *Central Places in Southern Germany,* translated by C.W. Baskin. London: Prentice-Hall, 1966.

Green, Nick. 2007. "Functional Polycentricity: A Formal Definition in Terms of Social Network Analysis," *Urban Studies* 44: 2077–2103.

Isaacman, Sibren, Richard Becker, Ramón Cáceres, Stephen Kobourov, James Rowland, and Alexander Varshavsky. 2010. "A Tale of Two Cities." In Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications, 19–24. *HotMobile* 10. New York, NY, USA: ACM.

Krugman, Paul R. 1996. *The Self-Organizing Economy.* Cambridge, MA: Blackwell Publishers.

Roth, Camille, Soong Moon Kang, Michael Batty, and Marc Barthélemy. 2011. "Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows," *PLoS ONE* 6, no. 1: e15923.

Schläpfer, Markus, Michael Szell, and Carlo Ratti. 2014; in preparation. "Human Spatial Organization and the Emergence of Urban Centers."

Trantopoulos, Konstantinos, Markus Schläpfer, and Dirk Helbing. 2011. "Toward Sustainability of Complex Urban Systems through Techno-Social Reality Mining," *Environmental Science & Technology* 45, no. 15: 6231–6232.

Luís M. A. Bettencourt

The Kind of Problem a City Is: New Perspectives on the Nature of Cities from Complex Systems Theory

Cities are not organisms, any more than they are machines, and perhaps even less so. They do not grow or change of themselves, or reproduce or repair themselves. They are not autonomous entities, nor do they run through life cycles, or become infected. [...] But it is more difficult, and more important, to see the fundamental ineptness of the metaphor and how it leads us unthinkingly to cut out slums to prevent their "infectious" spread, to search for an optimum size, to block continuous growth, to separate uses, to struggle to maintain greenbelts, to suppress competing centers, to prevent "shapeless sprawl," and so on.
Kevin A. Lynch, Good City Form *(1984)*

•

Jane Jacobs's Challenge

The title of this essay is taken from the challenge posed by Jane Jacobs in her influential book the *Death and Life of Great American Cities* (1961). In the last chapter of her book she provides a conceptual basis for her arguments about urban form by defining cities as problems in organized complexity (Weaver 1958). Ideas about complexity in science were brand-new at the time, and Jacobs readily realized their importance for understanding cities through her serendipitous association with Warren Weaver. Weaver, in turn, was fresh from explaining the significance of Claude Elwood Shannon's new theory of information to the world (Shannon and Weaver 1949). These were exciting times, when a series of new integrative perspectives and concepts were sweeping like wildfire through many disciplines. But it was perhaps too early then to define in more precise terms what kind of complex systems cities really are. Clearer formalizations of social networks took another fifty years to emerge. The spectacular diversity and scope of new urban forms and growth patterns that we observe in cities today was also still unknowable then.

In my view, the challenge of defining the kind of problem a city is goes well beyond a principled rejection of the urban renewal planning practices of Jane Jacobs's day. It also asks for more than the consideration and adoption of her observations about the nature of urban spaces. Today some of these observations may even appear a little parochial, too close to what she observed in New York's West Village of the late 1950s. What lives on as a challenge is the creation of new and better reconceptualizations of cities as complex adaptive systems (to use more modern language) and the development of a more scientifically grounded practice of urban planning guided by such principles.

My main objective here is to address this challenge by discussing the essential character of cities with the benefit of more than fifty years of research and of new insights from the emerging science of cities as complex systems.

I will show that taking a perspective of cities as integrated social networks imbedded in space and time, and requiring general properties for their open-ended land use and infrastructure development, provides us with a new unified model of urbanization. Such a perspective is supported quantitatively by an enormous body of empirical research, characterizing thousands of cities across the globe at very different levels of development, at present and in the past (Bettencourt 2013). In terms of urban design, this conceptualization of cities emphasizes the importance of generative models, where local structures remain to be developed by agents possessing particular goals and information, but must also be constrained by the function of the city as a whole, as an open-ended "social reactor."

A New View of Cities as Complex Systems

There is no longer much of an excuse to ignore many of the measurable properties of cities. Cities across the globe and through time are now knowable like never before, across many of their dimensions: social, economic, infrastructural, and spatial.

The main empirical findings from a new body of work in geography and complex systems (Bettencourt 2013; Koolhaas 1995; Batty 2005; Bettencourt et al. 2010; Nordbeck 1971; Bettencourt and West 2010) can be summarized succinctly in the observation that all cities realize both certain spatial economies of scale as they grow and, simultaneously, attain general socioeconomic productivity gains (Congress of New Urbanism 2008; Alexander et al. 1977). For example, when we compare two cities of different population sizes in the same urban system (usually a nation) we tend to find that the larger is a little bit denser, and that its volume of infrastructure networks per capita (roads, pipes, cables) is smaller. We also find that the larger city is usually wealthier (as measured by its per capita GDP or wages), more expensive, and more productive culturally and technologically, e.g., in terms of numbers of creative endeavors or patents filed. Thus, these measurements confirm general familiar expectations that larger cities are not only more congested and expensive, but also more productive and exciting culturally.

These effects are universal, in the sense that they are observed across cities in many urban systems, from the US to China, and from Brazil to Germany (Bettencourt 2013). They also apply over time, in contemporary urban systems, as well as in the spatial patterns of settlement of the pre-Hispanic basin of Mexico measured over a 2,000-year period, which must have evolved independently from their Old World counterparts (Ortman 2013).

More specifically, one finds that these properties of cities change gradually with city size and do not point to any special city size. With each doubling of city population, on average, a city contains about 10–20 percent less infrastructure volume per capita. It also displays a 10–20 percent increase in rates of wealth production, innovation, and (if left unchecked) of less benign products of human socioeconomic interactions, such as violent crime. While definitions of cities vary, these findings apply to the city as a whole, conceived as mixing populations or unified labor markets.

These definitions result in a view of functional cities as metropolitan areas (Bettencourt 2013), rather than their central cores or other particular administrative divisions, which have no real socioeconomic meaning. For small cities this distinction is less important.

These findings and especially the generality of their applicability and magnitude, from developed to developing cities, are certainly a starting point but they still do not provide a new model of what cities are. To that end I have recently proposed a new theoretical framework (Bettencourt 2013) that derives these properties quantitatively, as well as many others, from a simple conceptualization of cities as urban social and infrastructural networks colocated in space and time.

The central idea is that cities are first and foremost large social networks. In this sense cities are not just large collections of people, they are agglomerations of social links. Space, time, and infrastructure play a fundamental role in enabling social interactions to form and persist, and in allowing them to become open-ended in terms of increased connectivity, and sustainable from the point of view of energy use and human effort.

From this perspective the observed scaling properties of cities with population size are obtained from four simple assumptions:

(1) Cities are mixing social networks, where every urbanite on a minimal budget afforded through his/her social connections can in principle connect with any other person in the city.[1] This determines the cost of movement of people, goods, and information. A city may be denser because it is larger, everything else being equal, and/or because the relative cost of movement is high.

(2) Networks of infrastructure are decentralized, meaning that they are built gradually as the city grows, with amounts of new infrastructure added proportional to typical urban spatial densities.[2] This means that there is no single center (no heart to the city) where all traffic must be relayed through (Samaniego and Moses 2008), and also that the infrastructure network is evolvable gradually as population size increases or technologies change.

(3) The individual effort to stay socially connected is independent of city size. This is made possible because (1) and (2) imply that as cities grow they become spatially denser, so that even

[1] This derives a relationship between total land area, A, and population N, of the form $A(N)=aN^{a}$, where a depends on transportation technology and the elasticity exponent $a=D/(D+H)$, where D is the dimension of space ($D=2$) and H is a fractal dimension measuring how accessible the city is to each individual. In practice, $a;2/3$

[2] This means mathematically that $A_n : d^{-1/2} N ; A_0 N^{5/6}$ where A_n is the volume of infrastructure (road surface) and $d=N/A$ is the overall population density.

if distances traveled by individuals stay approximately the same (constant effort) more people and institutions become available for contact. In other words, the city comes to us as it grows. This removes an important objection about larger cities from sociology and social psychology (Samaniego and Moses 2008; Simmel 1903; Wirth 1938; Milgram 1970) and allows cities, in principle, to be open-ended in terms of their size.

(4) The socioeconomic outputs of cities, from economic production to innovation but also crime, are proportional to the number of social interactions realized per unit of time. These in turn are determined on the average by the density of people in public networks.[3]

[3] This means that social outputs Y (total GDP, number of violent crimes) increase with population size according to a scaling relation of the form $Y : N/A_n \sim N^{7/6}$.

The mathematically inclined reader can find the translation of these statements into more formal language in Bettencourt 2013, as well as the derivation of several other properties discussed below.

What are the general implications of this conceptualization of cities? The first is that the view of cities in terms of social networks emphasizes the primary role of expanding connectivity per person and of social inclusion in order for cities to realize their full socioeconomic potential. In fact, cities that for a variety of reasons (violence, segregation, lack of adequate transportation) remain only incipiently connected will typically underperform economically compared to better-mixing cities. It has been argued that expanding social connectivity (in cities) is indeed the key to development and civic participation (Lynch 1960), for example by increasing opportunities for the division and coordination of labor, and the creation of more complex and sophisticated social and economic organizations (Holston 2008). It has also been noted that planning that manages to increase connectivity (e.g., by coordinating transportation and housing policies; see Bettencourt et al. 2012) is typically more successful than initiatives that address these issues separately. Certainly, on a small scale this is a central tenet of models for smart growth (Congress of New Urbanism 2008), but what these results emphasize is the need for social integration in huge metropolitan areas over their largest scales, not only at the local level, such as neighborhoods.

There are several important consequences for general land use in cities. First, the price of land rises faster with population size than average incomes. This is the result of per capita increases in both density and economic productivity, so that money spent per unit area and unit time, i.e., land rents, increases on average by 50 percent with every doubling of city population size! It is this rise in the price of land that mediates, indirectly, many of the spontaneous solutions that reduce per capita energy use and carbon emissions in larger cities. Cars become expensive to park, and taller buildings become necessary to keep the price of floor space in pace with incomes, thus leading to smaller surface area to volume. This, in turn, presents an opportunity for reducing heating and cooling costs per person. These effects may also create the conditions for public transportation to be a viable alternative to automobiles, even when the price of time is high and fast point-to-point transportation becomes increasingly more desirable. Thus, larger cities may be greener paradoxically as an unintended consequence of their more intensive land use and their higher economic productivity. Policies that increase the supply of land per capita or reduce transportation costs (such as urban renewal), while addressing other problems, will tend to create cities that are less dense and that require higher rates of energy consumption in buildings and transportation to perform the same social functions.

This raises the issue of population density. When we compare cities at the same time but across different sizes, we tend to find that larger cities in the same nation are denser. Nevertheless, in the United States and increasingly all over the world, we also find many modern urban forms, and especially many low-density large cities, such as Atlanta or Dallas. Are these lesser cities than the West Village that Jane Jacobs knew, or the walkable towns that smart growth planers advocate? The perspective of cities as interaction networks tells us how these urban forms can coexist as the result of variations in transportation technology and levels of economic development. The overall spatial extent of the city is determined by the interplay between interactivity and the relative cost of mobility. Personal preferences for certain types of neighborhood may also play a role, of course. In general, when it is possible to move fast across space, cities tend to become much more diaphanous and are able to spread out spatially while preserving their social connectivity. It is, in fact, the diffusion of fast transportation technologies (modern mass transit, but especially cars and road infrastructure) in developing world cities that is promoting rapid increases in their land area, sometimes faster than their population growth (Angel et al. 2011). This, of course, creates possible vulnerabilities: For example, if the cost of transportation relative to incomes suddenly rises (e.g., because it is tied to oil prices) then cities may not be able to stay socially connected, leading to expected decreases in their socioeconomic production. Ideas for shrinking cities that have lost population and economic activity, such as Detroit and other former industrial cities, apply the same ideas in reverse: By reducing the spatial extent of the city, social connectivity can be realized more pervasively and at a smaller cost, with the expectation that such places should "heal" their social fabrics by becoming more reconnected and revitalized.

Another interesting consequence of the theory is that total volume of infrastructure is expected to increase with city size faster than land area. This is a subtler mathematical prediction (see footnotes) that follows from the requirements that infrastructure networks are decentralized and are built gradually. In practice, this means that in larger cities infrastructure network volumes become a larger and larger part of space. This is intuitively clear in large and/or developed cities where cables, roads, pipes, etc., become a ubiquitous feature of urban landscapes, before they eventually are moved into the third dimension, either above or below ground in order not to crowd out other more fundamental uses. Thus, planning for incommensurate changes in land and infrastructure footprints with city size is necessary to enable growth, especially of large, dense cities.

Another aspect of this problem is that we can predict the amount of energy spent in transportation processes (not only of people, but also of goods, information, services, etc.) necessary to maintain the city connected. Perhaps surprisingly, given spatial and material economies of scale, energy dissipated in transportation increases faster than population and behaves, in fact, just like incomes or measures of innovation. This is also the expected cost of operating (but not, typically, of building) urban infrastructure: it has been measured, e.g., in terms of resistive losses in the power grids of German cities (Bettencourt 2013). It may also appear in the form of higher congestion in larger cities, which is always worse in the larger arteries (highways) than in small roads (Downs 1962). This suggests some original solutions for the problem of managing urban infrastructure costs and promoting sustainability in cities. Investing in efficiency improvements that reduce energy dissipation in transportation processes, especially in the wider tracts of urban infrastructure, should be done strategically, as a (nonlinear) function of city size.

In addition, the theory gives us some specific clues about how energy use in transportation may be minimized, as well as some of the limits to this effort. In particular it tells us that the ratio of social outputs to energy transportation losses, a measure of urban efficiency, is independent of city size. This is probably one of the most important reasons why cities appear across a huge diversity of scales.[4] The socioeconomic agglomeration advantages of larger cities are proportional to their dissipative costs. Increases in congestion costs mirror socioeconomic advantages of agglomeration. Cities that work well maximize the difference between net social benefits and infrastructural costs. Such measures provide an overall set of metrics that can guide adaptive urban planning and policy (Bettencourt 2013).

Thus, I have shown how the form and function of cities can be determined, at least overall, by a set of network principles that more fundamentally captures the kind of problem a city is. These ideas will continue to require empirical testing and further development, but the recent direct detailed measurement of human social interactions through mobile phone technologies (Schläpfer et al. 2013) and the observation of spatial scaling in urban systems of pre-Hispanic Mexico analogous to that observed in modern cities (Ortman 2013) lend further support to their generality.

Two further problems remain in our quest to understand the importance of cities in people's lives and in the development of nations. The first is to understand more directly how the structure of human interactions produces effects well known to economists and sociologists, who study, for example, crime or innovation within cities. Some advances were recently made, through the direct measurement of individual connectivity statistics via mobile communications

[4] Cities are scale-invariant, meaning that the distribution of city sizes does not contain any specific population number at which the properties of cities change dramatically. The statistics of city sizes in an urban system have been intensely studied since the 1930s. They follow approximately Zipf's law, which states that if we order cities by size N, according to their rank, r, we obtain a statistical rank-size rule of the form $N(r) = N_0/r$, where N_0 is the size of the largest city, at r=1.

(Schläpfer et al. 2013), but much more needs to be done. Better knowledge of this kind should also help plan for more successful public spaces (Whyte 1980). The second is the interaction between cities in an urban system. Measures of growth over time, from population to economic development, are properties shared by all cities in an urban system,[5] but how they are established and what determines their magnitude is perhaps one of the most important open problems in economics and urbanism. These questions seem to be now open to new theoretical efforts as data across scales, nations, and time establishes a clearer view of these problems.

[5] This is known as Gibrat's law: (population) growth rates are constant, on average, independent of city size, *N*.

Planning for the Future

The entire world is now urbanizing at breakneck speed and, as city growth takes place spontaneously in India, Africa, or Latin America, it is following no architectural master plan. Perhaps as much as a billion people worldwide today (United Nations Human Settlements Programme 2003) live in slums and more than that build their own shelter and organize their neighborhoods without the services of architects or urbanists or knowledge of any science of cities. The ongoing great urban transition, which entails the building of more urban infrastructure in the next few decades than over the entire history of our species, is largely an unplanned and spontaneous process. Against this chaotic but vital urban backdrop, what is the role of urbanism and of a science of cities?

For those involved in planning practice it may look like the complex systems perspective on the spatial city leaves too much unspecified. In particular, it says nothing about some of the elementary choices in planning such as the shapes of streets or neighborhoods, houses and buildings, specific uses of space, zoning, etc.

That planning should leave many of these choices unspecified, to be developed locally by individuals, organizations, and communities, is an altogether more radical statement. However, both urban history and fundamental scientific concepts about how complex systems are created and how they evolve suggests just that (Hayek 1945; Anderson 1972). This is sometimes known as the "planner's problem." The problem rests on the fundamental properties of information in complex systems made up of an immense number of heterogeneous agents: The planner cannot possibly know in practice all the myriads of ways in which people would like to develop urban spaces over the history of a city. Better choices are usually made by agents with more specific information, adequate to their goals and aspirations, so far as these are constrained not to limit similar choices made by others and their integration across urban scales. Thus, ensuring general constraints (such as those discussed in the previous section), together with basic rules at the local level (such as those inspired by vernacular architecture over many centuries now or by some forms of new urbanism or generative design), may provide a practical model for planning, especially in cities that are largely being built informally anyway. By helping to create the city from the general to the particular, through designs that are generative of the whole but not prescriptive of the small parts, planning for the city as a complex system may finally fulfill Patrick Geddes's aspiration "to undo as little as possible, while planning to increase the well-being of the people at all levels, from the humblest to the highest."

I hope to have demonstrated that cities are natural systems that evolve spontaneously in human societies under very general circumstances, whenever there are open-ended advantages to human sociality across scales. In this sense they are as natural as beehives or coral reefs, and should not be thought of as arbitrary human artifacts to be redesigned at will. At the same time, they are a different kind of complex system from other more limited forms of social organization in nature and can be, in fact, immensely more complex in the many forms of information they can embody and generate. Cities reveal at once the best and the worst aspects of humanity in terms of our creativity and imagination but also our tendencies for violence or discrimination. Because of this enormous potential for human development, cities should not be seen as systems to be controlled or resisted, but encouraged to evolve spontaneously in the direction of achieving the best open-ended expressions of our collective nature.

That then is our challenge. We are living the last few decades of the great urban transition and finally fulfilling our global potential as the most social of all species to create something altogether new in Earth's history. We have within sight age-old human aspirations, such as to eliminate extreme poverty, to end most injustice, to gain access to good health for all, and to do all that sustainably, in balance with the Earth's biosphere. All this will have to happen in cities, and it can now happen very quickly. Bigger data and a more scientific approach to cities will certainly help. But the ultimate challenge for all of us involved in influencing and practicing urban planning is to translate, apply, and further develop these new ideas to promote types of urban environments that can encourage and nurture the full potential of our social creativity, targeted at sustainable and open-ended human development.

I am especially indebted to Nicholas de Monchaux for several discussions on the history of planning and urbanism and their relation to complex systems concepts that helped me frame the ideas developed here. I thank Michael Mehaffy and Dietmar Offenhuber for comments and suggestions on earlier versions of the manuscript. This research was partially supported by the Rockefeller Foundation, the James S. McDonnell Foundation (grant no. 220020195), the National Science Foundation (grant no. 103522), the John Templeton Foundation (grant no. 15705), and by a gift from the Bryan J. and June B. Zwan Foundation.

Alexander, Christopher, Sara Ishikawa, and Murray Silverstein. 1977. *A Pattern Language: Towns, Buildings, Construction.* New York: Oxford University Press.

Anderson, Philip W. 1972. "More Is Different," *Science* 177, no. 4047: 393–396.

Angel, Shlomo, Jason Parent, Daniel L. Civco, Alexander Blei, and David Potere. 2011. "The Dimensions of Global Urban Expansion: Estimates and Projections for All Countries, 2000–2050." *Progress in Planning* 75, no. 2: 53–107.

Batty, Michael. 2005. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals.* Cambridge, MA: The MIT Press.

Bettencourt, Luís M. A. 2013. "The Origins of Scaling in Cities," *Science* 340, no. 6139: 1438–1441.

Bettencourt, Luís M. A., and Geoffrey B. West. 2010. "A Unified Theory of Urban Living," *Nature* 467, no. 7318: 912–913.

Bettencourt Luís M. A., Geoffrey B. West, José Lobo, and Deborah Strumsky. 2010. "Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities," *PLoS ONE* 5, no. 11: e13541.

Bettencourt, Luís M. A., Horacio Samaniego, and HyeJin Youn. 2012. "Professional Diversity and the Productivity of Cities," http://arxiv.org/abs/1210.7335 (accessed April 21, 2014).

Bettencourt, Luís M. A., José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. 2007. "Growth, Innovation, Scaling, and the Pace of Life in Cities," *Proceedings of the National Academy of Sciences* 104, no. 17: 7301–7306.

Congress of New Urbanism. 2008. "Canons of Sustainable Architecture and Urbanism: A Companion to the *Charter of New Urbanism,*" http://www.cnu.org/canons (accessed April 21, 2014).

Downs, Anthony. 1962. "The Law of Peak-hour Expressway Congestion," *Traffic Quarterly* 16, no. 3: 393–409.

Duranton, Gilles, and Matthew A. Turner. 2011. "The Fundamental Law of Road Congestion: Evidence from US Cities," *American Economic Review* 101, no. 6: 2616–2652.

Hayek, Friedrich A. 1945. "The Use of Knowledge in Society," *American Economic Review* 35, no. 4: 519–530.
Holston, James. 2008. *Insurgent Citizenship: Disjunctions of Democracy and Modernity in Brazil.* Princeton, NJ: Princeton University Press.
Jacobs, Jane. 1961. *The Death and Life of Great American Cities.* New York: Random House.
Koolhaas, Rem. 1995. "What Ever Happened to Urbanism?" in *S,M,L,XL,* edited by Rem Koolhaas and Bruce Mau. New York: The Monicelli Press, 959–971.
Lynch, Kevin. 1960. *The Image of the City.* Cambridge, MA: The MIT Press.
Milgram, Stanley. 1970. "The Experience of Living in Cities," *Science* 13, no. 3924: 1461–1468.
Nordbeck, Stig 1971. "Urban Allometric Growth," *Geografiska Annaler. Series B, Human Geography* 53, no. 1: 54–67.
Ortman, Scott G., Andrew Cabaniss, and Luís M. A. Bettencourt. 2013. "Urban Scaling in Prehispanic Central Mexico." Santa Fe Institute Working Paper 13-01-001, http://www.santafe.edu/media/workingpapers/13-01-001.pdf (accessed April 21, 2014).
Samaniego, Horacio, and Melanie E. Moses. 2008. "Cities as Organisms: Allometric Scaling of Urban Road Networks," *Journal of Transport and Land Use* 1, no. 1: 21–39.
Schläpfer, Markus, Luís M. A. Bettencourt, Sébastian Grauwin, Mathias Raschke, Rob Claxton, Zbigniew Smoreda, Geoffrey B. West, and Carlo Ratti. 2012. "The Scaling of Human Interactions with City Size," http://arxiv.org/abs/1210.5215 (accessed April 21, 2014).
Shannon, Claude Elwood, and Warren Weaver. 1949. *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.
Simmel, Georg. 1903. "Die Grosstädte und das Geistesleben." Published in English as "The Metropolis and Mental Life," in *The Sociology of Georg Simmel,* translated and edited by Kurt H. Wolff. New York: The Free Press of Glencoe, 1976, 409–424.
United Nations Human Settlements Programme. 2003. *The Challenge of Slums: Global Report on Human Settlements.* London: Earthscan Publications.
Weaver, Warren. 1958. "A Quarter Century in the Natural Sciences," *The Rockefeller Foundation Annual Report 1958.* New York: The Rockefeller Foundation, 1–122.
Whyte, William Hollingsworth. 1980. *The Social Life of Small Urban Spaces.* Washington, DC: Conservation Foundation.
Wirth, Louis. 1938. "Urbanism as a Way of Life," *The American Journal of Sociology* 44, no. 1: 1–24
World Bank. 2013. *Planning, Connecting, and Financing Cities – Now: Priorities for City Leaders.* Washington, DC: World Bank.

Stanislav Sobolevsky

Digital Approach to Regional Delineation

Globalization and information technology have accelerated social and economic processes, transforming cities, regions, and whole countries at an increasing rate. Urbanization is still increasing: at the beginning of the nineteenth century, only about 3 percent of the world population lived in cities. By the beginning of twentieth century, this portion had increased to 14 percent. In 2008, finally, the urban population of the world passed the mark of 50 percent,[1] reaching 60–80 percent today in highly developed countries. Rural and suburban populations play a significant role in the dynamics of urbanization and also feel its effect. The daily flux of ex-urban commuters extends the city beyond its administrative boundaries, constituting what has been called the functional city, covering larger metropolitan areas or urban agglomerations.

The increasing spatial and informational dynamics of contemporary cities, regions, and countries create new challenges for urban and regional planners. Regional designations become obsolete within years of their definition. Maintaining an understanding of the actual spatial interaction and its boundaries is important for governments setting appropriate regional social and economic policies and deciding on infrastructural questions, for

1 "Population Reference Bureau," http://www.prb.org

example, planning transportation systems that respond to the needs of their users in an optimal way. Detecting changes in the regional structure can provide important clues for processes and shifts occurring in a community or society.

In the past, these kinds of decisions exclusively relied on the estimation of human mobility fluxes obtained from manual surveys, which are expensive and slow to undertake and are usually conducted every few years. This approach has become less practical in our current culture, which requires swift response to the social and economic challenges of large and fast-paced cities.

However, with these new challenges come new possibilities for addressing them. The increased prevalence of electronic and digital information in various spheres of human interaction means that more and more human actions leave a digital trace. Time and duration of the call, as well as the approximate locations of caller and recipient are stored by the network operator. Paying for coffee or groceries using a credit card generates a similar trace, recorded by both the card service provider and the bank. Even taxi trips generate detailed digital information, thanks to the vehicles' built-in GPS systems. Taken together, these pieces of information form huge data sets of human activity, often discussed under the term *big data*. Today, such data is available from governmental institutions and private companies, anonymized to protect the identity of the individuals. For research, these data sets open unprecedented possibilities for investigating the laws and patterns of human behavior: mobility, temporal activity, and in particular human interactions. The latter could be considered as the underlying force connecting different places and creating the regions as they appear.

Measuring human interactions using big data created by human activity, we developed a computational approach for identifying regional delineation. We will show how human interaction in space manifests itself in communities that are much larger than the social network of an individual person, group, or community; nevertheless, they can be spatially described. The "digital revolution" has often led to the idea that mediated forms of human interaction would decrease or even eliminate the role of physical space. Nicholas Negroponte famously predicted: "the post-information age will remove the limitations of geography. Digital living will include less and less dependence upon being in a specific place at a specific time" (Negroponte 1995, 165). However, we will show that space still remains the key factor defining our life and interactions.

In our study we use countrywide data sets of records of human communication activity, namely cell or landline phone calls, collected over a defined period of time by phone service providers. The geography of the data sets include countries such as the United States, the United Kingdom, France, Portugal, Italy, Belgium, the Ivory Coast, and Saudi Arabia.

Great Britain

In a recent paper we focused on data from phone networks in Great Britain (Ratti et al. 2010). The data set covered more than 95 percent of Great Britain's residential and business landlines and contained 12 billion calls made during a one-month period. To safeguard personal privacy, individual phone numbers were anonymized by the operator before we received the data. Each caller's geographic location was aggregated by subregional switching facility groups, covering on average an area of 49 square kilometers, too coarse to recover any individual customer's address.

From these data we created a network of roughly 20.8 million nodes and 85.8 million undirected links connecting users with reciprocal connection, i.e., who made phone calls in both directions. We assume that the above network is a measure of human interactions at an individual level over all of Great Britain and aggregated it into a grid of 3,042 square pixels, each with dimensions 9.5 km × 9.5 km. We treated each pixel as a spatial node and measured its connection strength to every other pixel, thereby deriving a matrix of the total bidirectional traffic between each pair of spatial nodes in the geographic network **(figure 1)**. This way a weighted directed edge between two nodes is defined as the cumulative communication flow created by reciprocal calls between them.

Up to this point, the approach ignores geographic space; initially it was not even intended to be used for identifying regions. The original intent was to discover the structures and communities embedded in the network, based on what is known as

Fig. 1.
Telephone interactions in Great Britain. Visualization Mauro Martino, SENSEable City Lab

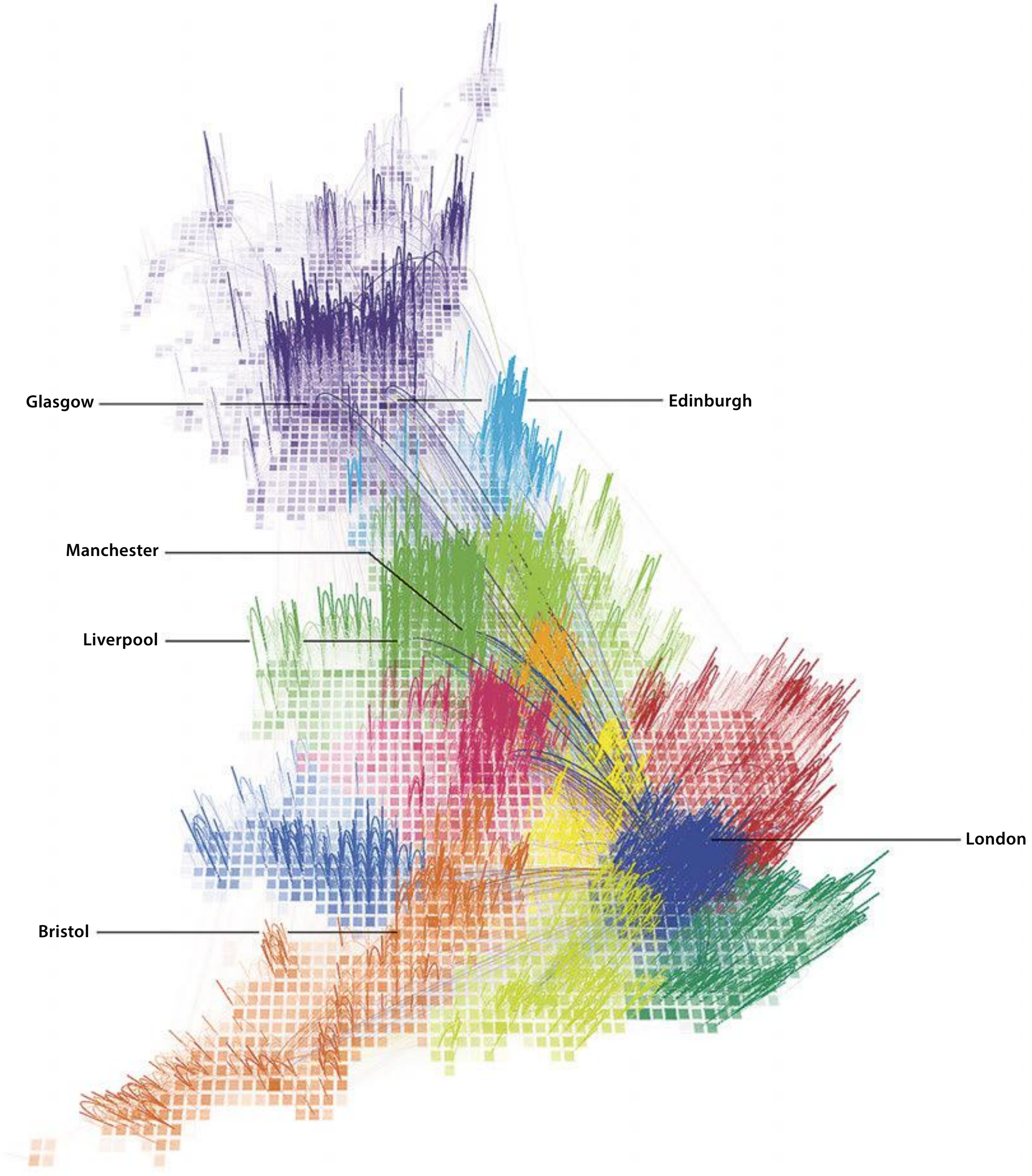
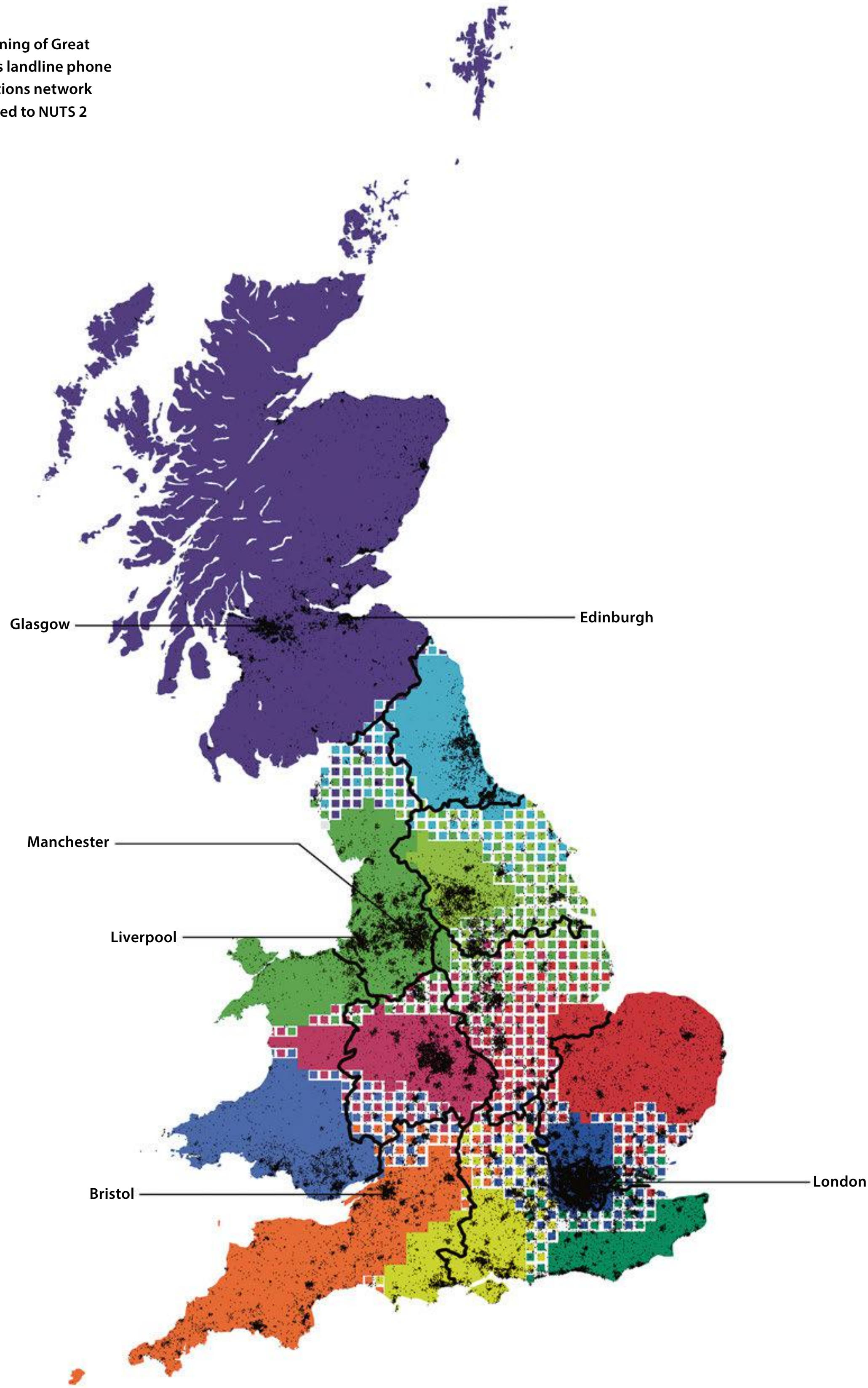


Fig. 2.
Partitioning of Great Britain's landline phone interactions network compared to NUTS 2 regions



the modularity optimization technique (Newman 2006). The modularity measure is introduced as a normalized quality function for the suggested community structure, which quantifies how strong the internal edges inside communities are and how weak the intercommunity edges are. More specifically the modularity measure scores every edge within the considered network using the difference between its actual strength and the average value expected from the network with the same total node strength but homogenously distributed edges. After that for each suggested community structure the overall modularity score is computed as the cumulative score of all internal community edges.

In terms of telephone interaction networks, this means that we compare the actual communication volume between each two locations with the average expectations' built-in assumption that all individual caller connections are equal to the network average. This way, the stronger-than-average links are preferably included as part of the detected community while the weaker links are avoided.

Our community detection algorithm then iteratively changes the initial community structure in order to improve the detection, for example by merging two existing communities into a larger one, splitting the existing community, or shifting a part of one community to another one and testing the result. These steps are repeated until no further improvement is possible. Again, the approach so far does not acknowledge geography, as it works with the network topology only. The number of resulting communities is not predetermined but is selected by the algorithm during the optimization process.

The resulting partitioning for Great Britain is shown in **figure 2**, where different communities are plotted using different colors while the official borders of 11 level 1 Nomenclature of Territorial Units for Statistics (NUTS)[2] are shown by solid black lines.

The first striking point is that the spatial projections of the resulting communities appear to represent geographically cohesive regions although geography played no role in the calculation. Another striking observation is that the number of the communities obtained from the modularity optimization process appears to be fairly consistent with the number of official regions in the country.

Comparing the particular shapes of the communities to the official administrative boundaries of the country, we observe

[2] "Eurostat. The NUTS classification," http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction

Fig. 3.
Partitioning of Great Britain's landline phone interactions network compared to NUTS 2 regions

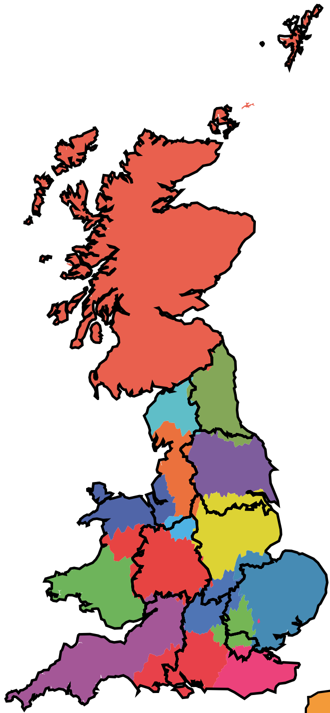


Fig. 6.
Bipartitioning of Belgium's cell phone interactions network compared to NUTS 2 regions

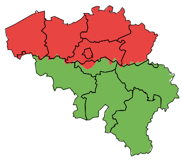


Fig. 5.
Partitioning of Belgium's cell phone interactions network compared to NUTS 2 regions

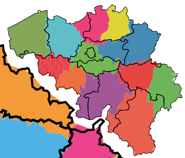


Fig. 4.
Partitioning of France's cell phone interactions network compared to NUTS 2 regions

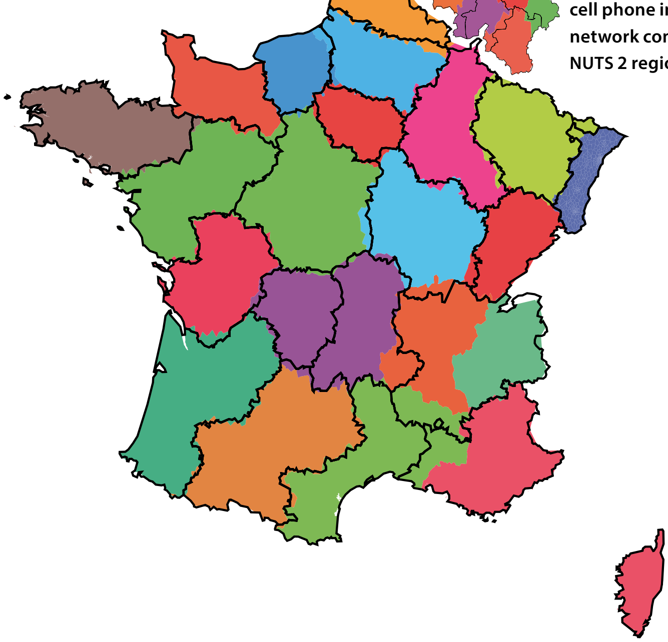
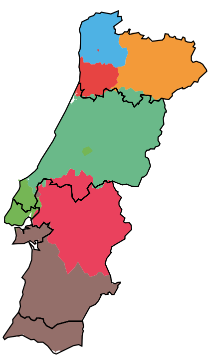


Fig. 7.
Partitioning of Portugal's cell phone interactions network compared to (from left to right) NUTS 2 regions, historical regional borders, and new borders proposed in the 1998 referendum



areas with a very close correspondence – for instance two of the nine official regions of England, South West England and East of England. Other regions resulting from our method, such as Scotland, appear identical to the official borders. However, deviations are also recognizable. It might not come as a surprise that the London region appears to be bigger than the official version – metropolitan areas around major cities are growing quickly, superseding existing boundaries. Wales, on the other hand, is split into Northern and Southern Wales, while the central part appears to be attached to West Midlands. This supports hypotheses from the transport and regional studies literature. Commuting data from the 2001 census confirms that Wales, in spite of its unique cultural and linguistic heritage, is well integrated with its English neighbors to the east (Nielsen and Hovgesen 2008).

The identification of new regions just west of London corroborates an earlier study of a "Western Crescent" of high-tech activity (Hall et al. 1987): a cohesive area that generally scores extremely well in measures of economic activity and low levels of deprivation, as measured by Gross Value Added (GVA) and qualifications (NVQ) for Berkshire, Buckinghamshire, and Oxfordshire.[3] While a more detailed and rigorous study is needed for drawing final conclusions, all the observations above give insight that the method is suitable for correctly identifying coherent regions based on the clusters found in the network of countrywide phone calls. Even more, this clustering approach seems to capture some of the aspects of human interaction more accurately than the official NUTS regional borders do.

Let us now see whether this approach works for countries other than Great Britain and also whether using cell phone data instead of landline calls only could make the analysis more accurate.

3 "UK statistics," http://www.neighbourhood.statistics.gov and http://www.statistics.gov.uk (accessed November 20, 2010).

France

Crossing the channel to France, we had the opportunity to study such a cell phone network. The locations of 14 million anonymized customers who made approximately 120 million calls during a period of 45 days are represented at the finer level of the cell tower, showing not only human mobility but also human interactions.

And the result (figure 4) – this and further results for Belgium and Portugal can be found in a recent paper (Sobolevsky et al. 2013) – demonstrates an even a more clear correspondence with the official borders apart from local deviations in the southeast along other minor local deviations. The slightly different partitioning algorithm used in this case was described by Vincent Blondel and others (Blondel et al. 2011).

Belgium

Unlike the French example, the locations of the customers for the Belgian cell phone network were defined based on their formal residence instead of the actual location during the call. Nevertheless, we see a clear correspondence between the official regional structure and the spatial projection of the network partitioning (figure 5) with the optimal number of regions given by the algorithm (Blondel et al. 2010). An interesting observation can be made if we constrain the number of resulting communities to two (the algorithm allows us to do so by simply limiting the iterative improvement steps) in order to get the optimal split of the country. We see (figure 6) the clear delineation of two major parts of Belgium – Flanders and Wallonia, splitting the country exactly along the Dutch-French language barrier. Strikingly, 97 percent of all communications in the network happen inside these two regions while only 3 percent happen between them. In a homogeneous distribution of the links in the network this value would be 50 percent, emphasizing the strong separation between these two parts of the country. Furthermore, nearly half of these 3 percent are the connections between Brussels attached to the northern part – Flanders – and Wallonia on the south. So the two parts of the country speaking different languages appear to be decisively separated, with Brussels as the major bridge between them.

Portugal

Finally, Portugal provides a remarkable example. Recently the country saw a lot of discussions on how Portugal should be administratively divided. In 1998 the regionalization question was the subject of a national referendum. Comparing our partitioning result for Portugal's cell phone network to the official territorial division of NUTS that currently exist, it is revealed that its second hierarchical level NUTS 2 (basic regions of the application of regional policies) is more coarse-grained (5 regions in continental Portugal) while NUTS 3 (small regions for specific diagnoses) is already much more fine-grained (28 regions in continental Portugal). Our partitioning is in between (7 regions) and matches historical regions better, showing the lasting impact of historical boundaries on human behavior, which can override modern categorization. The regional structure proposed on the referendum seems to demonstrate a much lower match with the regions identified from human interaction networks. This may explain partly why the referendum failed.

Conclusion

Similar findings – geographical cohesiveness of the communities of human interaction networks and their similarity to the official country regions in number and shape – were also confirmed for other different-scale countries from various continents: Italy, Ivory Coast, Singapore, and the United States. The community structure derived from the US cell phone network demonstrates numerous deviations from official state borders and could be an example of how real human interactions in a big, rapidly developing country go beyond an imposed and potentially outdated administrative division.

While a more precise analysis is needed to support planning decisions, the approach demonstrates the huge potential of human activity data for regional studies. Despite these limits, the general correlation between the community structure of telephone interaction networks and the regional structure of a country appears to be a robust, widely applicable finding. Therefore, it may be that this community structure should be considered in the face of changes or decisions affecting the regional structure of the country before more robust methodology is developed. And unlike the expensive, time-consuming surveys that have historically been the most common tool in regional studies, this new digital approach is able to give the outcome instantly, at literally no cost, provided that an up-to-date data set is available.

Blondel, Vincent, Gautier Krings, and Isabelle Thomas. 2010. "Regions and Borders of Mobile Telephony in Belgium and in the Brussels Metropolitan Zone," *Brussels Studies* no. 42 (October 4).

Blondel, Vincent, Pierre Deville, Frédéric Morlot, Zbigniew Smoreda, Paul Van Dooren, and Cezary Ziemlicki. 2011. "Voice on the Border: Do Cellphones Redraw the Maps?" *ParisTech Review* (November 15).

Calabrese, Francesco, Dominik Dahlem, Alexandre Gerber, DeDe Paul, Xiaoji Chen, James Rowland, Christopher Rath, Carlo Ratti. 2011. "The Connected States of America: Quantifying Social Radii of Influence," in *Proceedings, 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing (PASSAT/SocialCom), October 9–11, Boston, MA*. Los Alamitos, CA: IEEE Computer Society Press, 223–231.

Gallagher, Tom. 1999. "Unconvinced by Europe of the Regions: The 1998 Regionalization Referendum in Portugal," *South European Society and Politics* 4, no. 1: 132–148.

Hall, Peter, Michael Breheny, Ronald McQuaid, and Douglas Hart. 1987. *Western Sunrise: The Genesis and Growth of Britain's Major High Tech Corridor.* London: Allen & Unwin.

Negroponte, Nicholas. 1995. *Being Digital.* New York: Knopf.

Newman, M. E. J. 2006. "Modality and Community Structure in Networks," *Proceedings of the National Academy of Sciences* 103, no. 23: 8577–8582.

Nielsen, Thomas, and Henrik Harder Hovgesen. 2008. "Exploratory Mapping of Commuter Flows in England and Wales," *Journal of Transport Geography* 16, no. 2 (March): 90–99.

Ratti, Carlo, Stanislav Sobolevsky, Francesco Calabrese, Clio Andris, Jonathan Reades, Mauro Martino, Rob Claxton, and Steven H. Strogatz. 2010. "Redrawing the Map of Great Britain from a Network of Human Interactions," *PLoS ONE* 5, no. 12: 1–6.

Sobolevsky, Stanislav, Michael Szell, Riccardo Campari, Thomas Couronné, Zbigniew Smoreda, and Carlo Ratti. 2013. "Delineating Geographical Regions with Networks of Human Interactions in an Extensive Set of Countries," *PLoS ONE* 8, no. 12: e381707.

Albert-László Barabási is the Robert Gray Dodge Professor of network science and a distinguished university professor at northeastern university. His work on the limits of predictability and modeling the scaling properties of human mobility played a key role in raising the awareness of the research community to issues related to network-based studies of human mobility. This research, featured on the cover of *Nature,* also extends to the collective response of human populations to large-scale emergencies and link prediction of social ties.

Luís M. A. Bettencourt, a theoretical physicist by training, is a professor of complex systems at the Santa Fe Institute. His research deals with creating new theoretical frameworks to understand growth, change, and information processing in evolving complex systems, from cities and urbanization to economic and technological change to visual cognition. His work on cities has provided a new empirical basis and theoretical framework to understand urbanization around the world, through time, and at different levels of development. This work has been published in leading interdisciplinary journals and has been featured in international media, such as the *New York Times,* the *Economist, National Geographic, Smithsonian,* and *Scientific American.*

Pedro Cruz has a M.Sc. in informatics engineering from the University of Coimbra and has been a long-term collaborator with graphic design studio FBA. He was a visiting Ph.D. student at MIT's SENSEable City Lab and currently he is a researcher at the Cognitive and Media Systems Group/CISUC. His main research interests are information visualization, computational design, and computational art.

Fabien Girardin is a cofounder of the Near Future Laboratory, a research and design practice that investigates the near future possibilities for digital worlds. In his work he mixes qualitative observations with quantitative data analysis to gain insights on the integration and appropriation of technologies. Subsequently, he exploits the gained knowledge with engineering techniques to prototype and evaluate concepts and solutions.

Kael Greco is a researcher at the SENSEable City Lab and a graduate student in the department of urban studies and planning and the Technology Policy Program at MIT. His current research investigates the implications of emerging information and communications technologies in the realms of urban governance, policy, and planning. His work aims to redefine data science by creatively coupling quantitative methods and visualization strategies to better understand complex social phenomena.

Benedikt Groß is a speculative and interaction designer working across disciplines. He is fascinated by the relationships between people, their data, and their environments. Currently he studies design interactions at the Royal College of Art, UK. He is coauthor of *Generative Design,* one of the standard design books for computational design. He is an alumnus of MIT's SENSEable City Lab.

Philipp Hövel leads a junior research group at Technische Universität Berlin and the Bernstein Center for Computational Neuroscience. His research interests include nonlinear dynamics, control theory, and network science as well as dynamics on networks and network motifs. Between 2011 and 2013 he worked in the Barabási lab as a visiting scholar on a postdoctoral fellowship by the German Academic Exchange Service (DAAD).

David Lee is a Ph.D. candidate in urban studies and planning at MIT, and conducts research at the SENSEable City Lab. His work focuses on urban innovation districts, transforming infrastructure using real-time information systems, and elevating public discourse on critical urban issues with digital technology.

Penousal Machado is a teacher and leader of the Computational Design and Visualization Lab of the University of Coimbra. He is the author of more than seventy peer-reviewed journal and conference papers and the recipient of several scientific awards. Recently, his work was featured in *Wired* magazine UK and included in the *Talk to Me* exhibition at the Museum of Modern Art (MoMA) in New York.

Dietmar Offenhuber is an assistant professor at Northeastern University in the departments of art and design and public policy. His research focuses on the role of visualization and representation in urban governance and civic discourse; he led a number of research projects on infrastructure legibility and waste systems. Dietmar holds a Ph.D. from MIT in urban studies and planning and Master degrees from the Technical University Vienna and the MIT Media Lab. He has won several awards, including the NSF Visualization Challenge and the Paper of the Year Award from the *Journal of the American Planning Association.*

Carlo Ratti, an architect and engineer by training, teaches at the Massachusetts Institute of Technology, where he directs the SENSEable City Lab. He is also a founding member of international design office Carlo Ratti Associati. He is currently serving as a member of the World Economic Forum Global Agenda Council for Urban Management and a curator for the Future Food Pavilion for World Expo 2015. Carlo was named one of the "50 Most Influential Designers in America" by *Fast Company* and highlighted in *Wired* magazine's list of "50 People Who Will Change the World."

Francisca M. Rojas is a housing and urban development specialist at the Inter-American Development Bank. Previously, Francisca was research director with the Transparency Policy Project at Harvard's Kennedy School of Government, where she was also a postdoctoral fellow. She holds a Master in city planning and a Ph.D. from MIT in urban studies and planning. While at MIT, she was a researcher with the SENSEable City Lab.

Markus Schläpfer studied environmental engineering at ETH Zurich, where he also obtained his Ph.D. in 2010 in the field of energy systems modeling. Since 2011, he is a postdoctoral fellow at the SENSEable City Lab. The focus of his research is the understanding of cities by using digital traces of collective human behavior, eventually helping to design sustainable urban infrastructures.

Andres Sevtsuk **is an assistant professor in architecture and planning at the Singapore University of Technology and Design. Andres leads the City Form Lab (http://cityform.mit.edu) at SUTD, which investigates the influence of urban form on the social, economic, and environmental performance of cities using state-of-the-art spatial analysis tools. Before joining SUTD in 2011, Andres taught as a lecturer in architecture and urban studies and planning at MIT. He studied at L'École d'Architecture de la Ville & des Territoires (BArch) and MIT (SMArchS, Ph.D.) and has worked as an architect, urban designer, consultant, and researcher in Estonia, France, and the United States. He has published a number of articles and book chapters on urban design, urban technology, and spatial analysis.**

Filippo Simini **is a lecturer in transport and mobility modeling at the University of Bristol. Previously he has been a postdoctoral associate at the Center for Complex Network Research at Northeastern University in Boston, and at the department of physics at Budapest University of Technology and Economics. He received his Ph.D. in physics from the University of Padua.**

Stanislav Sobolevsky **is a researcher and coleader of the network and society group at the SENSEable City Lab at MIT since 2012. His background is in mathematics. Sobolevsky holds a Ph.D. (1999), Doctor of Science (2008) and the title of an associate professor (2009) from Belarus. His current research interests cover network science and big data analysis, the modeling of complex systems, and the theory of differential equations.**

Chaoming Song **is an assistant professor in the department of physics at the University of Miami. Prior to his current position, he worked as a research assistant professor in physics at Northeastern University together with Prof. Albert-László Barabási on the limits of predictability and modeling the scaling properties of human mobility, as well as on understanding the interplay between social network and human movements.**

Michael Szell **is a postdoctoral fellow at SENSEable City Lab. With a multidisciplinary background (mathematics, physics, computer science), his research interests cover the analysis and modeling of complex socioeconomic systems, social dynamics and mobility, and social networks. Michael is also developer of the Massive Multiplayer Online Game "Pardus" (www.pardus.at). At SENSEable City Lab he is involved in projects quantifying large-scale social networks and human behavior in cities.**

Anthony Vanky **is a researcher and partner strategist at MIT's SENSEable City Lab, and a Ph.D. student in urban studies and planning at the Massachusetts Institute of Technology. Trained as an urban designer and architect, he has worked on projects across the United States, and his design work has been exhibited widely, including at the Venice Architectural Biennale, the Dutch Design Week, and New Orleans DesCours. He has also been published in both academic and popular publications, including as a contributor to *Metropolis* magazine and the American Institute of Architects.**

Thank you to the MIT SMART program, the Center for Complex Engineering Systems at KACST and MIT, BBVA, The Coca Cola Company, Ericsson, Expo 2015, Ferrovial, the Regional Municipality of Wood Buffalo, Volkswagen Electronics Research Lab, and all the members of the MIT SENSEable City Lab Consortium for supporting the research.

Dietmar Offenhuber, Carlo Ratti (Eds)
DECODING THE CITY

Copy editing: Keonaona Peterson
Project management: Katharina Kulke
Layout, cover design and typography: Res Eichenberger, Zurich

Library of Congress Cataloging-in-Publication data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the German National Library
The German National Library lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

This publication is also available as an e-book (ISBN 978-3-03821-392-5) and in a German language edition (ISBN 978-3-03821-590-5).

P.O. Box 44, 4009 Basel, Switzerland
Part of De Gruyter GmbH Berlin/Boston

Printed on acid-free paper produced from chlorine-free pulp. TCF ∞

Printed in Germany

ISBN 978-3-03821-597-4

9 8 7 6 5 4 3 2 1

www.birkhauser.com