Mobile Cellular Big Data: Linking Cyberspace and the Physical World with Social Ecology

Fengli Xu, Yong Li, Min Chen, and Sheng Chen

Abstract

Understanding mobile big data, inherent within large-scale cellular towers in the urban environment, is extremely valuable for service providers, mobile users, and government managers of the modern metropolis. By extracting and modeling the mobile cellular data associated with over 9600 cellular towers deployed in a metropolitan city of China, this article aims to link cyberspace and the physical world with social ecology via such big data. We first extract a human mobility and cellular traffic consumption trace from the dataset, and then investigate human behavior in cyberspace and the physical world. Our analysis reveals that human mobility and the consumed mobile traffic have strong correlations, and both have distinct periodical patterns in the time domain. In addition, both human mobility and mobile traffic consumption are linked with social ecology, which in turn helps us to better understand human behavior. We believe that the proposed big data processing and modeling methodology, combined with the empirical analysis on mobile traffic, human mobility, and social ecology, paves the way toward a deep understanding of human behaviors in a large-scale metropolis.

The past few years have seen a dramatic growth in mobile traffic, contributed by billions of mobile devices as the first-class citizens of the Internet. The global cellular network traffic from mobile devices is expected to surpass 24 exabytes (an exabyte is approximately equal to 10^{18} bytes) per month by 2019 [1], which is 9 times larger than the traffic served by the existing cellular network in 2014. Such a huge volume of mobile traffic forms large-scale mobile big data recording human's activities in the physical world, behaviors in the cyberspace, and interactions with the urban social ecology. Here, social ecology refers to the complex relationship between human behaviors and urban environments. More specifically, in this article, the study of social ecology is carried out through investigating urban functional regions, such as transport hub, business district, shopping mall, and residential area. Therefore, while we are embracing a world with ambient cellular connectivity, there is also a critical and challenging problem: how to understand the patterns of data traffic in cyberspace and human mobility in the physical world profoundly [2–6], especially their inherent relationship.

On a more practical note, understanding the hidden patterns of humans' activities and behaviors in cyberspace, the physical world, and social ecology in a large-scale urban environment is extremely valuable for service providers, mobile users, and government managers of modern cites [7, 8]. If the traffic patterns of a cellular network can be *identified and modeled*, the service provider can exploit the modeled traffic patterns and customize a strategy for its individual cellular tower for providing services, instead of using a uniform strategy, such as using the same load balancing and data pricing algorithms on each tower. Mobile users also benefit from the traffic modeling because they can then choose towers with predicted lower traffic and enjoy better services. More profoundly, management departments of government will benefit from such mobile big data analysis as well because they may infer the social ecology and human economy activities by interpreting these data recorded by mobile networks [9].

On the other hand, understanding humans' behaviors in cyberspace and the physical world as well as their interaction with social ecology by analyzing mobile big data is challenging for three reasons. First, the recorded data experienced by thousands of cellular towers deployed in large-scale modern cities is highly complicated and hard to analyze. For example, our measurement includes over 9600 cellular towers and 150,000 subscribers, where lots of redundant and conflicting logs are observed. To identify patterns and behaviors embedded in the data associated with thousands of cellular towers, designing a system that is able to clean and handle large-scale big data is needed. Second, we do not have a priori human behavior patterns in cyberspace and the physical world. Without these profiles of human behavior patterns, it is challenging to group the huge amount of data experienced by thousands of cellular towers into a small number of meaningful patterns, which are vital for further understanding human behaviors. Third, the traffic of a cellular tower is affected by many factors (time, location, etc.). These factors are often correlated with each other and further complicate the analysis task. For exam-

Min Chen is with Huazhong University of Science and Technology.

Fengli Xu and Yong Li are with Tsinghua University.

Sheng Chen is with University of Southampton and King Abdulaziz University.

This work is supported by the National Basic Research Program of China (973 Program) (No. 2013CB329105), National Nature Science Foundation of China (No. 61301080, 61572220), and Research Fund of Tsinghua University (No. 20161080099)

ple, significant traffic variation is observed at both fine-grained (hours) and coarse-grained (days) timescales, and across towers deployed in different locations [10]. By addressing these challenges, in this article, we investigate how to extract and model the user behaviors and patterns embedded in thousands of cellular towers in a large-scale urban environment via a credible dataset collected by one of the largest commercial mobile operators.

Our main contribution comprises three parts. First, we reveal people's behavior patterns in cyberspace and the physical world, in terms of traffic consumption and human mobility patterns, respectively. Specifically, we find out that cyberspace traffic consumption and physical world human mobility have temporal patterns and are tightly correlated with each other. Second, we link cyberspace and the physical world with social ecology by first detecting the key mobility patterns embedded in the dataset and then investigating their links with urban functional regions. Third, with the established link between cyberspace, the physical world, and social ecology, we find that the average traffic-consuming rate and human migration pattern are correlated with social ecology. More importantly, we further analyze the characteristics of human behavior in different urban functional regions, which deepens our understanding of human behaviors in large-scale urban environments.

The rest of this article is organized as follows. The introduction of the mobile big data investigated and the required preprocessing techniques employed is first presented, followed by an overall visualization of the temporal features of mobile big data. Then, how people behave in cyberspace and the physical world is investigated, and we further our understanding of people's behaviors by linking them with social ecology. Finally, the last section summarizes our study and discusses future studies.

Dataset, Preprocessing, and Overall Visualization

This section provides the detailed information of the mobile big data investigated and the preprocessing needed. In addition, we visualize the temporal distribution of cellular traffic and subscribers, which benefits analysis.

Dataset Description

Our utilized mobile big data is an anonymized cellular trace collected by one of the largest mobile service providers in Shanghai during the whole month of August 2014. The trace contains the detailed mobile data usage record of 150,000 users, and each entry in the trace includes the identify ID of device (anonymized), start-end time of data connection, base station (BS) ID, address of BS, and the amount of third generation (3G) or Long Term Evolution (LTE) data consumed in each connection. The trace logs 1.96 billion tuples of the described information, contributed by over 9600 BSs, which contains the traffic logs of 2.8 petabytes (petabyte = 10^{15} bytes) cellular data traffic, 92 terabytes (terabyte = 10^{12} bytes) per day, and 7 GB per BS on average. This large-scale and fine-grained dataset ensures that our human behavior analysis and modeling is credible.

Preprocessing

The trace collected by the service provider needs to be preprocessed because of the existence of redundant and conflicting traffic logs as well as incomplete information of BSs' locations. The preprocessing includes three steps. First, redundant and conflicting logs are eliminated, such as the identical traffic logs caused by technical issues. Second, to solve the problem of incomplete information, we convert the addresses of BSs to their geographical longitudes and latitudes through application programming interfaces (APIs) provided by an online map service. This conversion gives us the precise geographical location of each BS, which is important for analyzing the ground truth of urban functional regions. The last step of preprocessing is segmenting the 31-day traffic trace of a tower into thousands of chunks, each of which contains a 10-minute traffic log. The 10-minute segmentation is chosen because it is the smallest time interval in which a cellular tower can experience non-zero traffic.

Data Visualization

Before diving into a deep analysis of mobile data traffic, the visualization is first displayed for the distributions of the temporal traffic and number of active users provided by the 9600 BSs, from which two interesting observations can be made.

First, the data embeds the fundamental temporal patterns of mobile data traffic. Figure 1 shows the aggregated and normalized traffic and the number of active users at different timescales. More specifically, Fig. 1a depicts the profile of normalized traffic and number of active users in one day (August 7, 2015, Thursday), where the aggregated network traffic looks similar to the profile of active users, and both are tightly coupled with the sleep patterns of humans; that is, high cellular traffic and a large number of active users are observed during the day, and low volumes are experienced overnight. Figure 1b shows the profile of normalized traffic and number of active users over one week (August 3–9, 2015). In addition to the repeated daily patterns of Fig. 1a, we observe from Fig. 1b that the peak traffic and the number of active users at the weekend are lower than those on a weekday. This suggests that mobile users are less active during the weekend and consume less cellular data traffic, which has also been found in [3]. Figure 1c illustrates the traffic distribution over the month (August 3–31, 2015), which shows that the traffic exhibits a periodical pattern on the order of a week, and weekend traffic is lower than weekday traffic. Figure 1d depicts the temporal patterns of the number of active mobile users over the month. On one hand, the profile of number of active users given in Fig. 1d exhibits similar patterns with the cellular traffic profile shown in Fig. 1c. On the other hand, the number of users is more stable during weekdays than the cellular traffic, which indicates that the number of active users in Shanghai does not vary significantly on different weekdays.

Human Behaviors in Cyberspace and the Physical World

In this section, human mobility and cellular traffic consumption patterns are investigated in order to understand human behaviors in cyberspace and the physical world. In addition, the relationships between human mobility in the physical world and traffic consumption patterns in cyberspace are further analyzed to provide insights of the link between cyberspace and the physical world.

Human Mobility in the Physical World

Human mobility is an important topic, which has been extensively studied in the past decade [11, 12]. With our mobile big data, mobile users can be located by checking the locations of the BSs to which they are connected. Therefore, it provides fine-grained location of large-scale mobile users, which is ideal for studying human mobility. In particular, human mobility can be studied through investigating dynamic distribution of mobile user population, which offers a new angle to aid our understanding of how humans move in an urban environment at a macro scale.

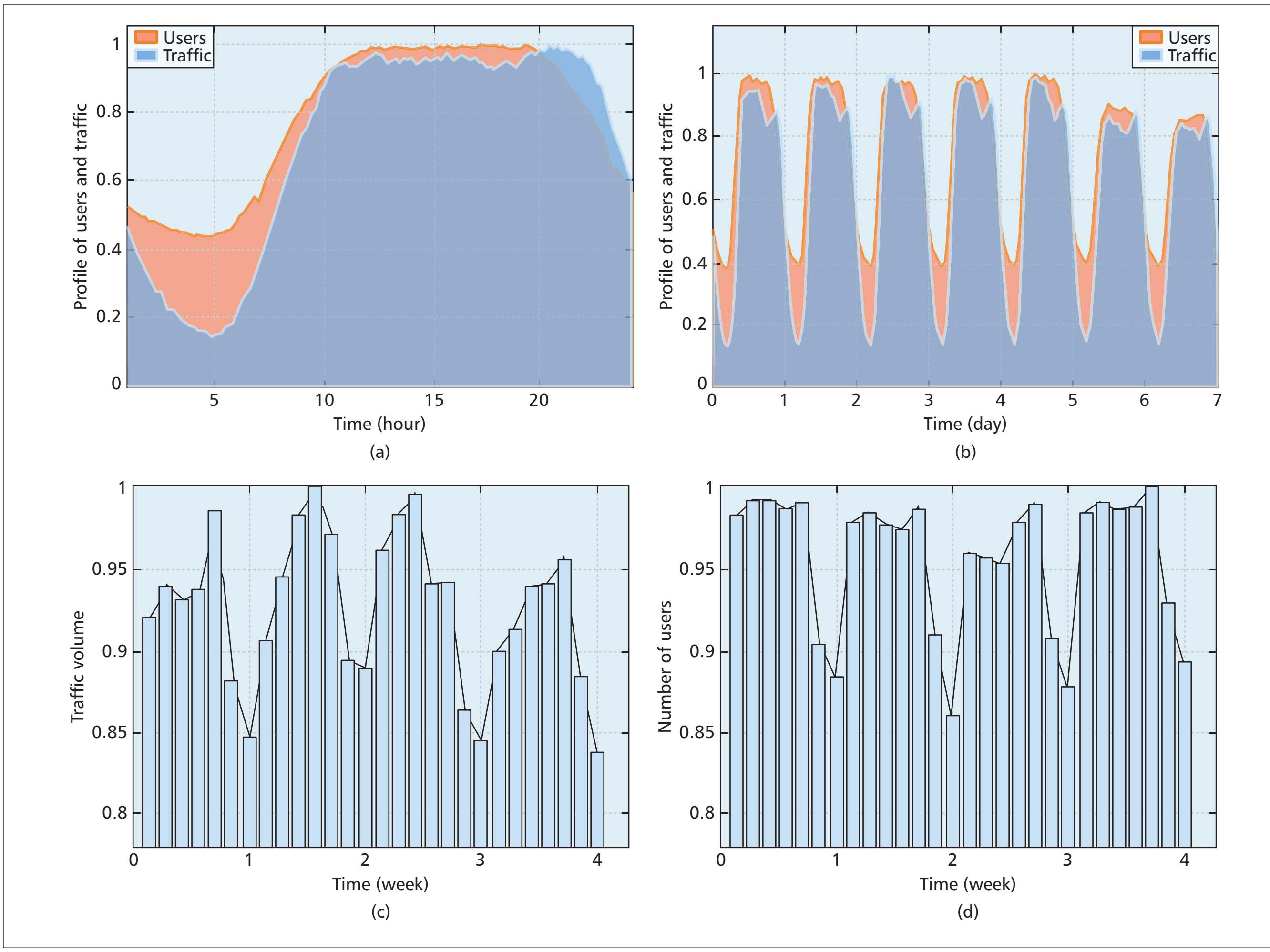

Figure 1. Variation of the normalized traffic and the number of active users at different timescales: a) profile of active users and traffic in one day; b) profile of active users and traffic in one week; c) traffic volume in one month; d) number of active users in one month.

The bottom two plots of Figs. 2a and 2b show the spatial distributions of the number of active mobile users at two different times of a day (August 7, 2015, Thursday). From these results, it can be seen that the city center possesses the highest density of mobile users. Also, we can see that the spatial distribution of mobile users varies with time during a day. In particular, the spatial distribution at 4 a.m. is different from that at 4 p.m. This suggests that human mobility patterns vary during a day in an urban environment, which is probably governed by activity patterns of humans.

Traffic Consumption Patterns in the Cyberspace

Understanding the traffic consumption patterns in an urban environment is of great importance for cellular network load balancing, green operation, and smart pricing. With the help of the traffic logs recorded in our big dataset, we are able to analyze such patterns in an urban environment up to the period of one month. Let us specifically study the spatial distribution of cellular traffic during a chosen day (August 7, 2015, Thursday).

The top two plots of Figs. 2a and 2b present the spatial distributions of normalized cellular traffic at two different times. From these two figures, we observe that the highest traffic consumption rate always occurs in the city center for different times, which is probably associated with the highest density of mobile users at the city center. Furthermore, from Fig. 2b, we can see that the spatial distribution of cellular traffic is similar to the spatial distribution of mobile users at 4 p.m. However, observe from Fig. 2a that the cellular traffic's spatial distribution is different from the distribution of mobile users at 4 a.m. This indicates that the traffic consumption is not only correlated with the number of users, but also affected by other factors, such as traffic demand.

Relationship Analysis

From Figs. 2a and 2b, it can be seen that the traffic consumption is correlated with the number of users. Understanding the correlations between them will help us better understand human behavior in the physical world and cyberspace. Therefore, we analyze and quantify the correlations between human mobility and cellular traffic patterns.

To understand the relationship between traffic consumption and number of users, the cumulative distribution functions (CDFs) of the spatial and temporal correlations, respectively, between them are analyzed and presented in Fig. 2c. The spatial correlation is derived by computing the Spearman correlation coefficient at each time slot, while the temporal correlation is computed on each BS. Observing the results of Fig. 2c, the number of users and the traffic consumption rate has a strong correlation in the spatial domain, with most time slots having a correlation coefficient larger than 0.9. This suggests that at every time slot an area with more users is very likely to have higher traffic consumption. In contrast, the number of users and the traffic consumption exhibit surprisingly low correlation in the temporal domain, with about 20 percent of the BSs having negative correlation and about 40 percent of the BSs having a correlation coefficient lower than 0.4. This implies that in 50 percent of the BSs the number of users has a weak correlation with the traffic consumption rate.

Linking with Social Ecology

Based on the above analysis, we have some basic understanding of human behavior in the cyberspace and physical world in urban environment. A natural question to ask is does human behavior relate to the social ecology? To answer this question, the links between the physical world and social ecology are established by detecting the key patterns of human mobility. Then, with knowledge of social ecology, we deepen our understanding of human behavior in cyberspace and the physical world.

Discovering the Links with Social Ecology

Discovering the links between the physical world, cyberspace, and social ecology is nontrivial, because we have little knowledge of the relationships among them. However, inspired by a key observation that the human mobility patterns of the same geographical context tend to be similar, we implement and evaluate a system to discover the links by detecting the key mobility patterns.

Detecting Key Mobility Patterns: Our system is composed of three key elements: data cleaner, pattern identifier, and metric tuner.

Data Cleaner: The data cleaner is a distributed traffic analysis system implemented in Hadoop, which is able to tackle large-scale unstructured mobile big data. The key to designing the data cleaner is a parallel transformer, which takes the time-domain logs of thousands of cellular towers as its input and converts each cellular tower's logs into a vector. A vector is constructed in two phases: aggregation and folding. In the first phase, each BS's number of users is aggregated in each 10-minute time slot to generate a vector representing its user number pattern. Then cellular towers' user number patterns of a month are converted into the patterns of a week (seven days) by averaging. The purpose of averaging is smoothing burst events experienced by cellular towers, such as parades.

Pattern Identifier: The pattern identifier takes the vectorized data from the cleaner and runs an unsupervised machine learning algorithm to identify the key patterns of human mobility. The pattern identifier addresses one key challenge of the mining process, unknown patterns, by exploiting hierarchical clustering [13]. The basic idea of hierarchical clustering is iteratively merging the nearest two clusters. It first considers each input point as a cluster and then bottom-up iteratively merges the nearest two clusters until the stop condition is met. In our application, correlation distance is utilized as the distance metric, and the distance between clusters is defined as average-linkage distance. In addition, a threshold value is set as the stop condition, which stops clustering when the distance between every pair of clusters is above the threshold value. To be more specific, the pattern identifier operates in the following three steps. First, it receives the predefined threshold value, takes the vectorized data as input, and considers each cell tower's data as a cluster. Second, it calculates the distances for all pairs of clusters. Third, it finds the minimum distance from the set of all the distances and compares it to the threshold value. If the minimum distance is above the threshold, the clustering is stopped, and the number of clusters gives the number of patterns identified, while the average pattern of every cluster is output as the identified pattern for each cluster. Otherwise, it merges the nearest two clusters and returns to the second step.

Metric Tuner: As the patterns of user variation are unknown, a key question is when should the identifier stop its clustering. In our system, the Davies-Bouldin index is utilized [14] to explicitly inform the identifier that the optimum number of patterns has been identified. The Davies-Bouldin index is utilized because it measures both the separation of clusters

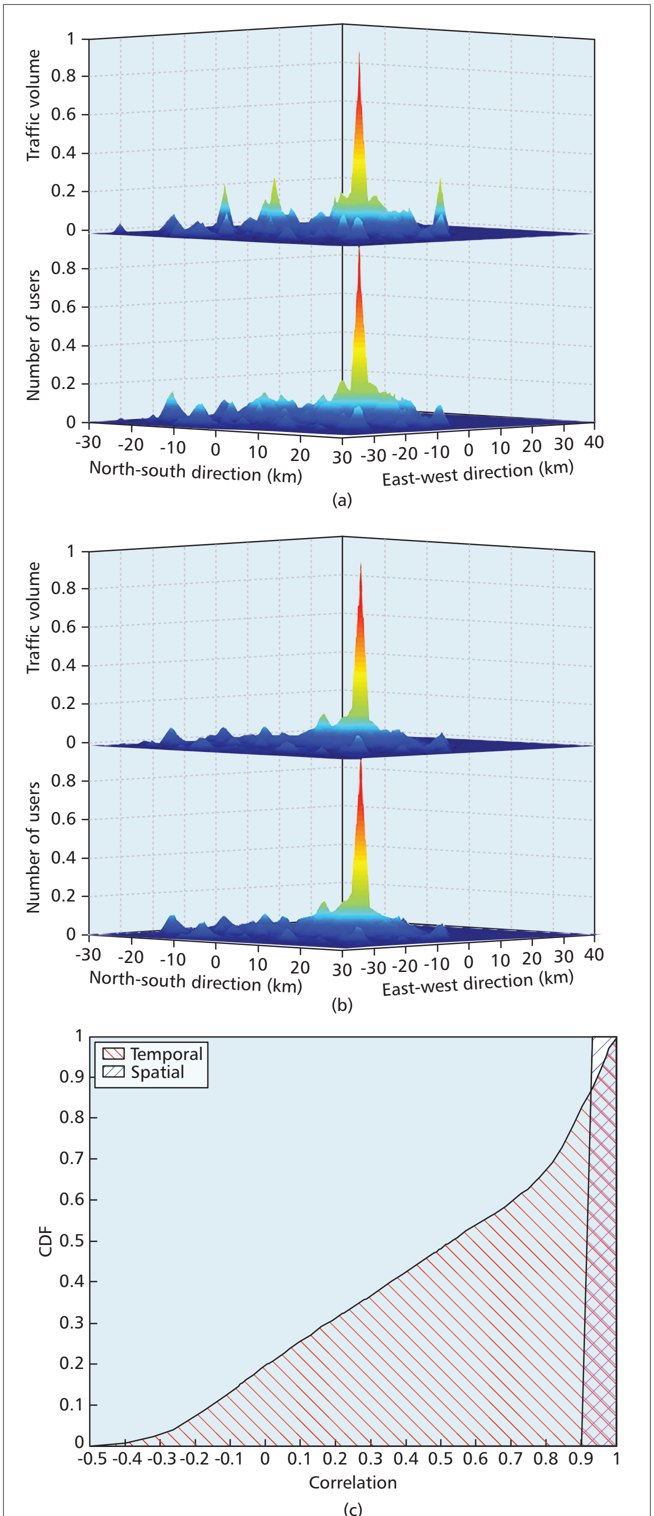

Figure 2. The spatial distribution of normalized traffic consumption and number of active users, and the CDF of their correlation: a) spatial distribution at 4AM; b) spatial distribution at 4PM; c) CDF of correlation between number of active users and traffic.

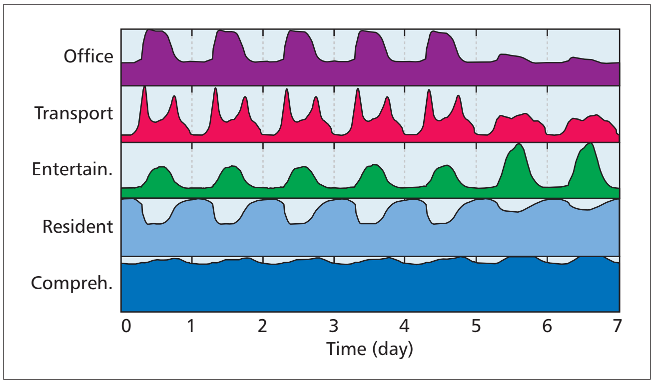

Figure 3. Patterns of number of active users for the five identified clusters.

and cohesion within clusters, which mathematically guarantees a good clustering result. When a minimum Davies-Bouldin index is obtained, the optimum number of patterns is identified.

Figure 3 shows the five time-domain patterns identified by our system from over 9600 cellular towers. The five clusters differ in terms of the time where the peak of user number appears as well as the number of users experienced during weekdays and weekend. The percentage of each cluster's cellular towers is shown in the upper right column of Table 1, which indicates that the first cluster has the most cellular towers and the second cluster has the least.

Linking Mobility Patterns with Social Ecology: After obtaining the clustered BSs, the next question to ask is how to link these clusters to the social ecology of the city (i.e., the urban functional regions). We build the linker via investigating the ground truth of urban functions in different regions.

To start with, the distribution of points of interest (POIs) is investigated in each cluster to establish the links between human mobility patterns and urban functional regions. A POI is a specific point location of a certain function such as a restaurant or shopping mall. An area's POI distribution reflects its urban function and can be considered as ground truth [15]. Therefore, studying the POI distribution of an area can help us to accurately find out the urban function of that area. To calculate the POI distribution, we measure the numbers of four types of POIs, which are resident, transport, office, and entertainment, within 200 m of each cellular tower. Then different regions' POI distributions are summarized in the lower part of Table 1. The maximum value of each column is shaded dark blue, which shows the dominant urban function in the corresponding row (i.e., cluster). According to Table 1, cluster 1 corresponds to an office area, cluster 2 corresponds to a transport area, cluster 3 corresponds to an entertainment area, cluster 4 corresponds to a residential area, and cluster 5 corresponds to a comprehensive area. Therefore, regions are classified into their dominant urban functions. If a region does not have an obvious dominant urban function, it is classified as comprehensive. With the help of POI data, we manage to establish the links between the human mobility of the physical world with urban social ecology.

Understanding Human Behaviors with Social Ecology

After discovering the links between the cyberspace, physical world and social ecology, we are able to further our analysis to better understand human behavior in both the cyberspace and physical world by focusing on human behaviors in different functional regions of social ecology. To characterize the features of cellular traffic patterns in different urban functional regions, the normalized traffic patterns for both weekday and weekend are presented in Fig. 4.

Functional regions	Cluster index	Percentage
Office	1	45.72%
Transport	2	2.58%
Entertainment	3	9.35%
Resident	4	17.55%
Comprehensive	5	24.81%

Cluster	Points of interest			
	Office	Transport	Entertain	Resident
#1	0.1034	0.0813	0.0515	0.0439
#2	0.1012	0.2000	0.1020	0.0473
#3	0.0976	0.1201	0.1674	0.0474
#4	0.0232	0.0285	0.0269	0.0528
#5	0.0453	0.0373	0.04030	0.0508

Table 1. Percentage and averaged normalized points of interest of cellular towers classified in each cluster.

Comparing Fig. 4a with Fig. 4b, the traffic patterns in different urban functional regions have distinct features on a weekday and on the weekend. For a weekday, the traffic patterns of the office area and entertainment area reach their peaks around noon, while the traffic pattern of the transportation area has two peaks in the morning and afternoon, and the traffic pattern of the residential area experiences a high value at night. In contrast, on the weekend, the traffic in the transportation area has only one peak, and the traffic patterns in the residential and office areas are different from those experienced during the week, while the entertainment area's pattern does not vary much.

To characterize the patterns of traffic consumption, the average traffic consumption for different urban functional regions are presented in Fig. 4c. As seen from Fig. 4c, the residential area possesses the highest traffic consumption, while the transportation area has the least traffic consumption. In addition, in the office and transportation areas, the traffic consumption is lower on the weekend than on a weekday, while in the residential and entertainment areas, the traffic consumption is higher on the weekend, which is consistent with human activity patterns.

The information of social ecology can also benefit our understanding of human behaviors in the physical world. A human's migration between different urban functional regions is an important aspect of understanding human mobility in an urban environment. Therefore, the migration probabilities (from other areas) to office and residential areas are presented in Figs. 5a and 5b, respectively. The migration probability from one region A to another region B is calculated by dividing the number of users migrating from A to B with the total number of users moving out of region A. In addition, a positive value represents people actually migrating from A to B, while a negative value represents people actually migrating from B to A. Therefore, the migration probability to each area sums up to 1 or –1 in each time slot, with a negative value indicating people migrating out of this area and a positive value suggesting that people are migrating into this area. As seen from Fig. 5b, most people migrate into an office area from a residential area from 5 a.m. to 9 a.m., and most people migrate out of an office area

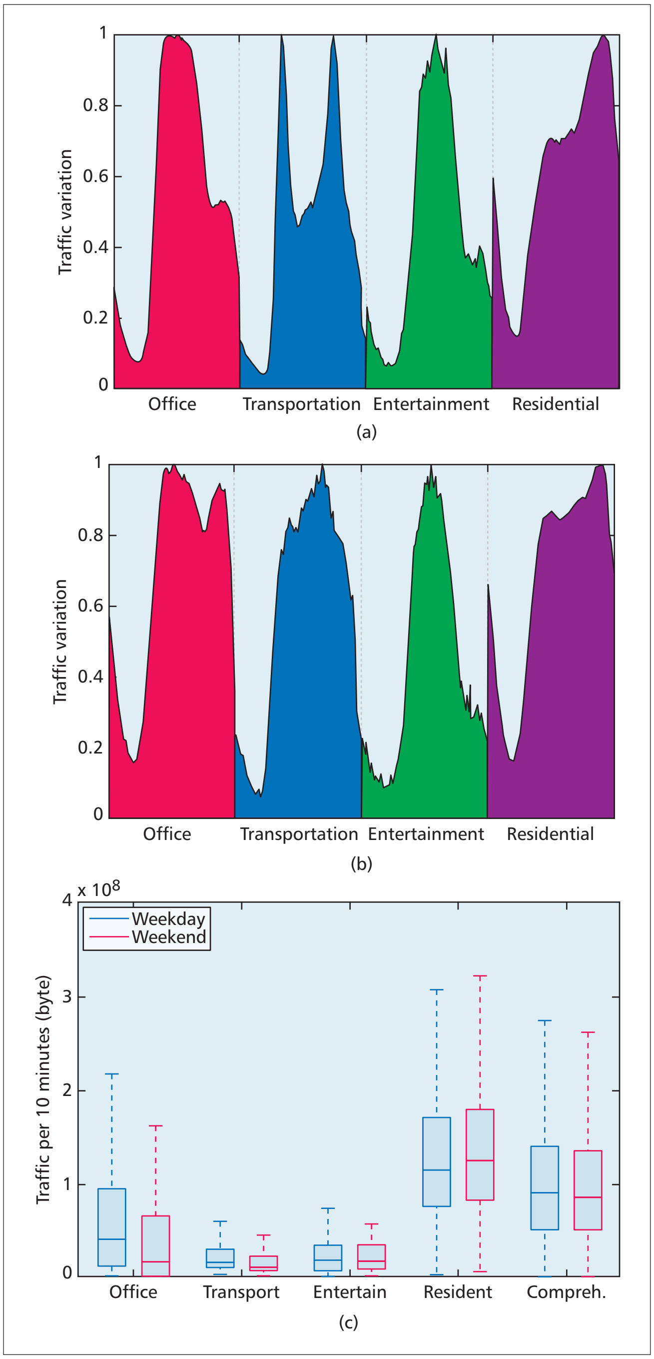

Figure 4. Characteristics of cellular traffic patterns in different functional regions: a) normalized traffic variations of different regions on weekdays; b) normalized traffic variations of different regions on the weekend; c) average traffic consumption rates in different regions.

to a residential area from 12 p.m. to 9 p.m. In addition, people begin to migrate from a transportation area to an office area from 8 a.m. to 10 a.m., while people begin to migrate from an office area to a transportation area from 1 p.m. to 5 p.m., as can be seen from Fig. 5a. Furthermore, the migration probability to a residential area is the opposite of that to an office area, as can be clearly seen by comparing Fig. 5a with Fig. 5b. This simply confirms that there is a strong connection between the office area and residential area, with most migration happening between these two areas. The above results clearly suggest that going to work is the main purpose of human migration between urban functional regions.

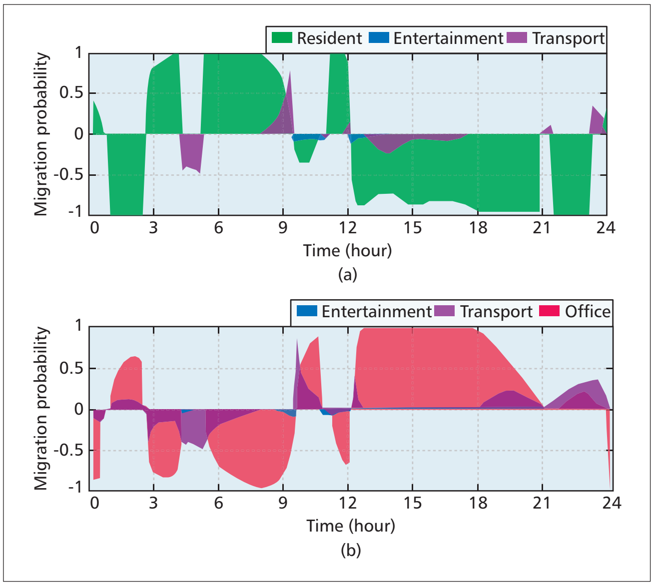

Figure 5. Migration probabilities to office area and residential area: a) to office area; b) to residential area.

Prospects and Discussion

With the rapid growth of mobile devices and ubiquitous cellular access, the cellular mobile network has become a gigantic sensing platform, which captures human behaviors in the physical world and cyberspace. For example, a cellular network records human access of all kinds of applications as well as human mobility and locations in the physical world. This detailed information enables us not only to analyze human mobility and traffic consumption patterns but also to study the links between human behaviors in the physical world, cyberspace, and social ecology. In our future studies, we plan to further investigate the links from the following aspects.

Spatial domain: In our current study, we have found out that the number of active mobile users has a strong correlation with cellular data traffic in the spatial domain. With the penetration rate of mobile devices reaching up to 96 percent over the world, mobile devices have become the best agent to monitor traces of human mobility. Therefore, based on cellular big data, future studies can be carried out to model the dynamic distribution of population, which not only is an important topic in human mobility, but also plays an important role in disease control, transportation scheduling, and other urban planning applications.

Temporal domain: In our current study, we have shown that human behaviors have strong temporal periodicity in the physical world as well as in cyberspace. Moreover, the patterns of human behaviors differ significantly in different urban functional regions. Therefore, based on the links with social ecology, we can better characterize and model human behavior in the physical world and cyberspace. Future studies can be carried out to model the temporal patterns of human behaviors with social ecology in mind and to develop applications based on it, such as a cellular network's dynamic load balancing schemes.

Events oriented: Detecting anomaly events in the physical world is an interesting topic that is of great importance in public safety. In our current study, we have observed that human behaviors in the physical world are tightly coupled with those in cyberspace. For example, a parade may cause spikes in cellular data traffic in particular regions. Therefore, by investigating the links between human behaviors in the physical world and cyberspace, we plan to develop an effective system to detect anomalous events in the urban environment.

Conclusion and Discussion

In this article, we carry out, to the best of our knowledge, the first study of human behaviors in cyberspace and the physical world embedded in large-scale 3G and LTE cellular towers deployed in an urban environment. Through investigating human mobility patterns and traffic consumption patterns, we characterize the features of human behaviors in the physical world, cyberspace, and social ecology. Our analysis reveals that human mobility and traffic consumption have strong correlation, and both have distinct periodical patterns in the time domain. Moreover, they are both linked with social ecology, which helps us better understand human behaviors. We believe that our analysis provides a systematic and comprehensive understanding of human behavior in social-physical-cyber space, and opens a new set of research directions.

References

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019, White Paper, Feb. 3, 2015.

[2] A. K. Das *et al.*, "Contextual Localization through Network Traffic Analysis," *Proc. INFOCOM 2014*, Toronto, Ontario, Canada, Apr. 27–May 2, 2014, pp. 925–33.

[3] B. Cici *et al.*, "On the Decomposition of Cell Phone Activity Patterns and Their Connection with Urban Ecology," *Proc. ACM MobiHoc 2015*, Hangzhou, China, June 22–25, 2015, pp. 317–26.

[4] Z. Su, Q. Xu, and Q. Qi, "Big Data in Mobile Social Networks: A QoE-Oriented Framework," *IEEE Network*, vol. 30, no. 1, Jan.–Feb. 2016, pp. 52–57.

[5] Z. Su, P. Ren, and Y. Chen, "Consistency Control to Manage Dynamic Contents over Vehicular Communication Networks," *Proc. GLOBECOM 2011*, Houston, TX, Dec. 5–9, 2011, pp. 1–5.

[6] K. Zheng *et al.*, "Big Data Driven Optimization for Mobile Networks toward 5G," *IEEE Network*, vol. 30, no. 1, Jan. 2016, pp. 44–51.

[7] Y. Dong *et al.*, "Inferring User Demographics and Social Strategies in Mobile Social Networks," *Proc. ACM SIGKDD 2014*, New York, NY, Aug. 24–27, 2014, pp. 15–24.

[8] M. Z. Shafiq *et al.*, "Characterizing Geospatial Dynamics of Application Usage in A 3G cellular Data Network," *Proc. INFOCOM 2012*, Orlando, FL, Mar. 25-30, 2012, pp. 1341–49.

[9] M. Kosinski, D. Stillwell, and T. Graepel, "Private Traits and Attributes are Predictable from Digital Records of Human Behavior," *Proc. Nat'l. Acad. Sci. USA*, vol. 110, no. 15, 2013, pp. 5802–05.

[10] H. Wang *et al.*, "Characterizing the Spatio-Temporal Inhomogeneity of Mobile Traffic in Large-Scale Cellular Data Networks," *Proc. 7th ACM Int'l. Wksp. HotPost*, Hangzhou, China June 22, 2015, pp. 19–24.

[11] C. Song *et al.*, "Limits of Predictability in Human Mobility," *Science*, vol. 327, no. 5968, 2010, pp. 1018–21.

[12] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding Individual Human Mobility Patterns," *Nature*, vol. 453, no. 7196, 2008, pp. 779–82.

[13] F. Corpet, "Multiple Sequence Alignment with Hierarchical Clustering," *Nucleic Acids Research*, vol. 16, no. 22, 1988, pp. 10881–90.

[14] U. Maulik and S. Bandyopadhyay, "Performance Evaluation of Some Clustering Algorithms and Validity Indices," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, Dec. 2002, pp. 1650–54.

[15] J. Yuan, Y. Zheng, and X. Xie, "Discovering Regions of Different Functions in a City Using Human Mobility and POIs," *Proc. ACM SIGKDD 2012*, Beijing, China, Aug. 12–16, 2012, pp. 186–94.

Biographies

FENGLI XU is currently pursuing his Ph.D. degree in the Department of Electronic Engineering, Tsinghua University. He received his B.S. degree from the School of Electronic Information and Communications at Huazhong University of Science and Technology in 2015. His research interests include human mobility, mobile big data mining, and mobile user behavior modeling.

YONG LI [M'2009] (liyong07@tsinghua.edu.cn) received his B.S. and Ph.D degrees from Huazhong University of Science and Technology (HUST)and Tsinghua University in 2007 and 2012, respectively. During 2012 and 2013, he was a visiting research associate at Telekom Innovation Laboratories and Hong Kong University of Science and Technology, respectively. During 2013 to 2014, he was a visiting scientist at the University of Miami. He is currently a faculty member in the Department of Electronic Engineering, Tsinghua University. His research interests are in the areas of mobile computing and social networks, urban computing and vehicular networks, and network science and the future Internet. He has served as General Chair, Technical Program Committee (TPC) Chair, and TPC member for several international workshops and conferences. He is currently an Associate Editor of the *Journal of Communications and Networking* and *EURASIP Journal of Wireless Communications and Networking*.

MIN CHEN [M'2008, SM'2009] (minchen2012@hust.edu.cn) is a professor in the School of Computer Science and Technology at HUST. He is Chair of the IEEE Computer Society (CS) Special Technical Community (STC) on Big Data. He was an assistant professor in the School of Computer Science and Engineering at Seoul National University (SNU) from September 2009 to February 2012. He received his Ph.D. in communication and information systems from the South China University of Technology, Guangzhou, China. He performed post-doctoral research at the University of British Columbia, Canada, and served on the faculty of SNU beginning in 2004. He has published more than 180 papers, including 90 SCI papers in the areas of the Internet of Things, mobile cloud, body area networks, healthcare big data, emotion-aware computing, robotics, and cyber physical systems. His Google Scholar citations have reached 6000+ with an h-index of 38. His top paper was cited 710 times, while his top book was cited 420 times as of April 2015. He received Best Paper Awards from IEEE QShine 2008 and ICC 2012. He has published *OPNET IoT Simulation* (HUST Press, 2015) and *Big Data Related Technologies* (Springer Computer Science Series, 2014). He has served as Associate or Guest Editor for seven IEEE/ACM international journals and magazines.

SHENG CHEN [M'1990, SM'1997, F'2008] (sqc@ecs.soton.ac.uk) received his B.Eng. degree from the East China Petroleum Institute in January 1982, and his Ph.D. degree from City University London in September1986, both in control engineering. In 2005, he was awarded the higher doctorate degree, D.Sc., from the University of Southampton, United Kingdom. From 1986 to 1999, he held research and academic appointments at the Universities of Sheffield, Edinburgh, and Portsmouth, all in the United Kingdom. Since 1999, he has been with the Department of Electronics and Computer Science, University of Southampton, United Kingdom, where he holds the post of professor in intelligent systems and signal processing. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural networks and machine learning, intelligent control system design, evolutionary computation methods, and optimization. He has published over 550 research papers. He is a Fellow of IET. He is a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia. He is an ISI highly cited researcher in the engineering category (March 2004). He was elected to a Fellow of the United Kingdom Royal Academy of Engineering in 2014.