

## MIT Open Access Articles

*Discovering urban spatial-temporal  
structure from human activity patterns*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Shan Jiang, Joseph Ferreira, Jr., and Marta C. Gonzalez. 2012. Discovering urban spatial-temporal structure from human activity patterns. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing (UrbComp '12). ACM, New York, NY, USA, 95-102. ACM New York, NY, USA ©2012

**As Published:** <http://dx.doi.org/10.1145/2346496.2346512>

**Publisher:** Association for Computing Machinery (ACM)

**Persistent URL:** <http://hdl.handle.net/1721.1/77151>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike 3.0



# Discovering Urban Spatial-Temporal Structure from Human Activity Patterns

Shan Jiang  
Massachusetts Institute of  
Technology 77 Mass. Ave. E55-  
19E Cambridge, MA 02142 USA  
+1 (857) 654-5066  
shanjiang@mit.edu

Joseph Ferreira, Jr.  
Massachusetts Institute of  
Technology 77 Mass. Ave. 9-532  
Cambridge, MA 02139 USA  
+1 (617) 253-7410  
jf@mit.edu

Marta C. Gonzalez  
Massachusetts Institute of  
Technology 77 Mass. Ave. 1-  
153 Cambridge, MA 02139 USA  
+1 (617) 715-4140  
martag@mit.edu

## ABSTRACT

Urban geographers, planners, and economists have long been studying urban spatial structure to understand the development of cities. Statistical and data mining techniques, as proposed in this paper, go a long way in improving our knowledge about human activities extracted from travel surveys. As of today, most urban simulators have not yet incorporated the various types of individuals by their daily activities. In this work, we detect clusters of individuals by daily activity patterns, integrated with their usage of space and time, and show that daily routines can be highly predictable, with clear differences depending on the group, e.g. students vs. part time workers. This analysis presents the basis to capture collective activities at large scales and expand our perception of urban structure from the spatial dimension to spatial-temporal dimension. It will be helpful for planners to understand how individuals utilize time and interact with urban space in metropolitan areas and crucial for the design of sustainable cities in the future.

## Categories and Subject Descriptors

H.2.8 [Database Applications]: data mining, spatial databases and GIS; I.5.3 [Clustering]: clustering algorithms and similarity measures.

## General Terms

Algorithms, Design, Measurement, Human Factors, Performance

## Keywords

Urban spatial-temporal structure, Human activity intensity, Kernel density estimation, Time-cumulative spatial activity density, Computational social science

## 1. INTRODUCTION

Cities, home to billions of people, are complex systems [1, 2]. For decades, urban spatial structure, measured by the degree of spatial concentration of population and employment, has been studied by urban scholars to describe the structure and organization of cities, and their function and role in people's life [3-5]. On the one hand,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UrbComp'12, August 12, 2012, Beijing, China.  
Copyright 2012 ACM 978-1-4503-1542-5/08/2012 ...\$15.00.

with the improvement of transportation systems, cities have evolved from monocentric to polycentric forms in their spatial configurations [3, 6-8]. On the other hand, with the advances in information and communication technologies (ICT), cities have been racing to be "smarter", in terms of their human, social, and environmental capital profiling [9, 10]. As a result, today swelling cities have become incidentally data repositories of human activities (gained from the emerging massive urban sensors such as GPS, mobile phone, and online user-generated social media). These facts, plus the spectacular ability of researchers to collect and analyze data, have helped us understand the nature of human mobility (e.g., high predictability in daily routine [11, 12]) and the dynamics of cities [13, 14], and provided a great potential for planners to optimize the value of existing infrastructures in the city [15].

While celebrating opportunities these massive urban sensing data bring to us, researchers have also realized challenges of adopting them, due to privacy and legal restrictions and economic constraints. With little or no information about either the social demographic characteristics of the individuals or the types of activities they are conducting, our understanding of the causes and underlying reasons of human behavior are still inadequate [16]. For example, in the study by Becker et al. [17], although it is promising to see plausible estimates of the spatial distribution of residence by users of different phone usage patterns (e.g., classified "workers" or "partiers") based on call detail records (CDR), we still cannot get a complete picture of human activities in non-home/work categories, as it is hard to infer or validate those types of activities from the CDR data. Yet since cities have been playing increasingly important roles as consumption centers [18], urban planners are pressed to know how cities are used by different types of people for different types of activities. Given the nature of these urban sensing data (such as CDR), similar challenges are faced by other studies (e.g., Eagle et al. [19]), where non-home/work activities are hard to be differentiated. Meanwhile, it is also unclear how knowledge gained from targeted groups of communities (e.g., universities [19]) may apply when the scale is enlarged to metropolitan area and beyond.

Activity-based travel surveys collected by planners to develop transportation and activity models for cities, on the other hand, have the potential to complement the new insights in human activities and mobility gained from massive urban sensing data as discussed above. If responded accurately, travel surveys can inform us about "who, what, when, where, why and how of travel for each person in a surveyed household" [20]. These questions have been in the center of urban planning, geography and

transportation fields for decades [21-26], due to their importance in helping us understand the complexity and dynamics of cities.

In this paper, we concentrate on numerical methods to mine the spatial-temporal activity patterns of individuals obtained from a recent travel diary survey in the Chicago Metropolitan Area. The advances of this study lie in three folds.

First, we expand the traditional understanding of urban structure from spatial dimension based on static density of population to spatial-temporal dimension, which we call urban spatial-temporal structure. We measure this structure by defining a spatial distribution of activity intensity with temporal information extracted from the travel survey data.

Second, we cluster individuals according to their daily activity patterns and spatial-temporal traces. By doing so, it enables us to go from the traditional classification of individuals as students, workers and other types to more diverse groups, and enriches our understanding about human activities in the city. Combining the clusters of individuals with the spatial dimension, we identify sub-regions of the metropolitan area with inherent connections to the performed activities.

Finally, we combine the above measures together to analyze and visualize the time-cumulative spatial densities of various activity types by sub-regions of Chicago for different types of individuals. By comparing the spatial distributions of the intensity of various activities, we demonstrate that enormous information about urban spatial-temporal structure can be quantitatively analyzed and vividly illustrated. It will be very useful for urban planners to understand how urban areas have been used in space and time by different types of individuals, and will be crucial for them to propose solutions for sustainable cities.

## 2. DATA SOURCE AND STUDY AREA

To understand the urban spatial-temporal structure, in this research, we study the Chicago Metropolitan Area as an example. We employ a publicly available large-scale “Travel Tracker Survey”<sup>1</sup> conducted by the Chicago Metropolitan Agency for Planning (CMAP) in 2008 [27]. As this survey is designed to estimate the regional travel demand, the carefully planned sampling strategy ensures a good representation of the total population in the region, which includes eight counties—Cook, Du Page, Grundy, Kane, Kendall, Lake, McHenry, and Will—from the Northeastern Illinois Region, and two counties—Porter and LaPorte—from the Northwestern Indiana Region.

In this study, since we focus on the urban spatial-temporal structure on weekday, we use the survey records from Monday through Thursday as a sample to represent an average weekday, which contains daily activities of 23,527 distinct individuals. For each distinct individual, the survey records every activity destination, arrival and departure time, location (the longitude and latitude pair), activity type (such as home, work, school, shopping, recreation, etc.), and duration at the destination, in 24-hours of the surveyed day. To facilitate the analysis of this study, we transform the individuals' survey records as described above to minute-by-minute records with information of latitude and longitude location, time and activity type<sup>2</sup>. This data transformation allows us to

<sup>1</sup> Source: <http://www.cmap.illinois.gov/travel-tracker-survey>. Retrieved on June 12, 2012.

<sup>2</sup> Since we do not know individuals' exact travel path between destinations, and the movements usually amounts to a small portion of people's daily life within 24-hours (around 5%), in order to fill in the gap between two consecutive destinations, we make a simplified

explore the spatial-temporal structure of the Chicago metropolitan area in this study.

## 3. URBAN SPATIAL-TEMPORAL STRUCTURE

Urban spatial-temporal structure (USTS) extends the traditional concept of urban spatial structure by incorporating the temporal information. Specifically, it contains information of time stamps of individuals' activities, activity locations, activity types, and optionally, personal attributes (which are not indispensable, but may provide rich socioeconomic context). In this section, we use kernel density estimator (KDE) to estimate several measurements of USTS and present empirical findings of the Chicago urban spatial-temporal structure.

### 3.1 Measurements of Urban Spatial-Temporal Structure

In order to capture the urban spatial-temporal structure, we propose to measure a “spatial-temporal activity density” at each time instant of the day and analyze it through KDE. For the purpose of visualization and demonstration, we integrate this density with respect to time, and get a “time-cumulative spatial activity density”, which measures *spatial distribution of activity intensity*—a normalized sum of individuals' activity duration in a study area during a time period of interest. Human activity intensity can be crucial for many purposes, such as the analysis and prediction of energy consumption, infrastructure usage and business opportunities, etc.

#### 3.1.1 Preliminary

We use a 3-dimensional space  $\mathcal{S}$  to describe time and 2-dimensional activity locations as follows:

$$\mathcal{S} \triangleq \{\mathbf{s} = (t, x, y) | t \in [t_0, t_1], (x, y) \in \mathcal{A}\} = [t_0, t_1] \times \mathcal{A} \quad (1)$$

where  $t$ ,  $x$ , and  $y$  are the time, longitude, and latitude, respectively,  $[t_0, t_1]$  defines the time period of interest (e.g., one or several minutes, hours, days, weeks, etc.), and  $\mathcal{A}$  is the set of (longitude, latitude)-pairs for the study area.

#### 3.1.2 Spatial-Temporal Activity Density

We use the distribution density  $\pi$  on  $\mathcal{S}$  of a particular activity in a study area  $\mathcal{A}$  during a time period of interest ( $[t_0, t_1]$ ) as the measurement of urban spatial-temporal structure. We call the measurement  $\pi$  the *spatial-temporal activity density*, which is defined as the probability density of individuals' presence at destination<sup>3</sup>  $(x, y)$  at time  $t$  for the corresponding activity purpose, and the density  $\pi$  satisfies

$$\int_{\mathcal{S}} \pi(\mathbf{s}) d\mathbf{s} = \int_{[t_0, t_1] \times \mathcal{A}} \pi(t, x, y) dt dx dy = 1. \quad (2)$$

---

assumption that people move in a straight line with constant speed when they travel. It is worth noting that this filled-in information is only used in subsection 4.2.2 *Clusters of Daily Traces*, and due to the small proportion of time spent in travel, the estimation error caused by this approximation is very limited. Readers should also note that since travel is not considered as activities at destinations, therefore this treatment of data transformation will have little effect on the estimation of urban spatial-temporal structure discussed in the paper.

<sup>3</sup> We define *destination* as a place where people go due to the need of committing a certain type of activity (e.g., work, school, shopping, recreation, etc.). Here, we do not consider individuals' movement en route while measuring activity density. Therefore our treatment of traces will not affect the accuracy of the estimation of spatial-temporal activity density and the time-cumulative spatial activity density.

To understand the urban spatial-temporal structure, it would be helpful to visualize the spatial-temporal activity density. However, as it is defined on a 3-dimensional space  $\mathcal{S}$ , a direct static visualization becomes impossible. To circumvent the visualization barrier of spatial-temporal activity density, we explore two alternatives. One is to examine the distribution density  $\tilde{\pi}_{\bar{t}}$  on  $\mathbf{A}$  at a fixed time instant  $\bar{t}$ .<sup>4</sup> The other is to consider a time-cumulative spatial activity density.

### 3.1.3 Time-cumulative Spatial Activity Density

A time-cumulative spatial activity density  $\pi_{a,b}$  on  $\mathbf{A}$  is defined as follows

$$\pi_{a,b}(x, y) = c_{a,b} \int_{[a,b]} \pi(t, x, y) dt \quad (3)$$

where  $[a, b] \subseteq [t_0, t_1]$ , and  $c_{a,b}$  is a positive constant to make  $\pi_{a,b}$  satisfy the restriction that the integral of density  $\pi_{a,b}$  on  $\mathbf{A}$  equals 1, i.e.,  $c_{a,b}$  is a constant of normalization.<sup>5</sup> We use this measurement to understand the spatial distributions of intensity of various activity types and explore urban spatial-temporal structure quantitatively, which will be presented in section 3.2.

### 3.1.4 Kernel Density Estimation

In this study, we use the kernel density estimator to estimate the time-cumulative spatial density of a certain set of activities. Suppose that we have  $n$  observations of a set of activities  $\{\mathbf{s}_i = (t_i, x_i, y_i)\}_{i=1}^n \subseteq [a, b] \times \mathbf{A}$ , then the corresponding time-cumulative spatial density is estimated by

$$\hat{\pi}_{a,b}(x, y; H) = \frac{1}{n} \sum_{i=1}^n K_H \left( \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_i \\ y_i \end{pmatrix} \right),$$

where the *bandwidth matrix*  $H$  is a  $2 \times 2$  symmetric positive definite matrix,

$$K_H \left( \begin{pmatrix} x \\ y \end{pmatrix} \right) = |H|^{-1/2} K \left( H^{-1/2} \begin{pmatrix} x \\ y \end{pmatrix} \right),$$

and  $K$  is a 2-variate kernel function that satisfies

$$\int_{\mathbf{A}} K \left( \begin{pmatrix} x \\ y \end{pmatrix} \right) dx dy = 1.$$

Here we follow the popular practice to take  $K$  to be the standard 2-variate normal density, a spherically symmetric kernel,<sup>6</sup>

$$K \left( \begin{pmatrix} x \\ y \end{pmatrix} \right) = \frac{1}{2\pi} \exp \left( -\frac{1}{2} (x^2 + y^2) \right),$$

which implies that  $K_H \left( \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_i \\ y_i \end{pmatrix} \right)$  is the  $N \left( \begin{pmatrix} x_i \\ y_i \end{pmatrix}, H \right)$  density in vector  $\begin{pmatrix} x \\ y \end{pmatrix}$ . To simplify the selection of  $H$ , we assume that  $H = h^2 I$ , where  $h > 0$  and  $I$  is the  $2 \times 2$  identity matrix. It is clear that a larger bandwidth parameter  $h$  leads to a smoother estimation and a smaller  $h$  gives more fluctuations. In this study, the bandwidth  $h$  is chosen to minimize the mean integrated squared error (MISE). For detailed discussion about nonparametric kernel density estimation and the MISE criterion,

<sup>4</sup> The notation  $\tilde{\pi}$  means that it is obtained from  $\pi$  via normalization.

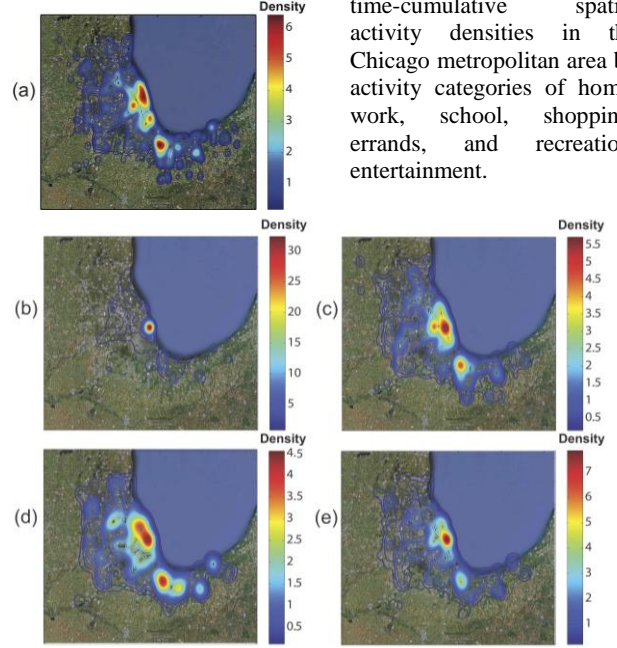
<sup>5</sup> Note that  $\tilde{\pi}_{\bar{t}}$  and  $\pi_{a,b}$  are closely related: If we assume that  $\pi$  is continuous in  $t$  (which is a very reasonable assumption), then we have  $\pi_{a,b} \rightarrow \tilde{\pi}_{\bar{t}}$ , as  $a, b \rightarrow \bar{t}$ .

<sup>6</sup> Normal density is in the class of admissible kernel functions, and people often choose it for the simplicity of interpretation. For detailed discussion about choice of kernel functions, please refer to [28] Wand, P. and Jones, C. *Kernel Smoothing*. Chapman & Hall, 1995.

and the applications of KDE on topics of interests in the urban realm, readers may refer to Wand and Jones [28], Kwan & Hong [29], and Yuan et al.[14], respectively.

## 3.2 Chicago Metropolitan Area General Spatial-Temporal Structure

By using kernel density estimation method described above, we estimate the activity intensity in the Chicago Metropolitan Area over 24 hours to understand the Chicago Urban Spatial-temporal structure. Figure 1 shows the time-cumulative spatial activity densities in the Chicago metropolitan area by activity categories of home, work, school, shopping/errands, and recreation/entertainment.



**Figure 1 Chicago time-cumulative spatial densities of different activity types (a) home, (b) work, (c) school, (d) shopping and errands, and (e) recreation and entertainment, on an average weekday.**

We can see from the figure that areas with intensive person-hours for home activities are along the north lakeshore, south lakeshore, and around Oak Park. Downtown Chicago alone has extremely intensive work activities—very high total person-hours for working—compared with the rest of the region. School activities in the region are mainly concentrated in the city of Chicago (where major universities and schools are clustered), and the southeastern part of the region (where Purdue University is located). Several regions with intensive person-hours for shopping activities exist, including the downtown and northern part of the City of Chicago, the City of Evanston, Schaumburg, southeast area of the region (cities of Hammond and Schererville). Areas that have been visited and used intensively for entertainment and/or recreational activities is narrower than their counterparts for shopping activities in space, but with some overlaps, including the downtown and northern part of the City of Chicago, City of Evanston, Oak Park, and the southeastern part of the Chicago metropolitan region.

As discussed in the introduction section, these results are interesting and informative; however, we want to further explore how the urban spaces are utilized by different types of individuals with different activity patterns. In other words, we want to ask “do students, workers, or other types of individuals use urban spaces in similar or different ways?” “Are there differences in activity

intensities by similar types of individuals across different subareas of the region?” “In what ways are their activity intensity and usage of urban areas in space and time similar or different?”

## 4. CLUSTERING DAILY ACTIVITY PATTERNS AND TRACES

In order to answer these questions and to further understand urban spatial-temporal structure, clustering individuals both by their daily temporal-activity patterns and by their spatial-temporal trace becomes important and natural components of this study.

### 4.1 K-Means Clustering via PCA

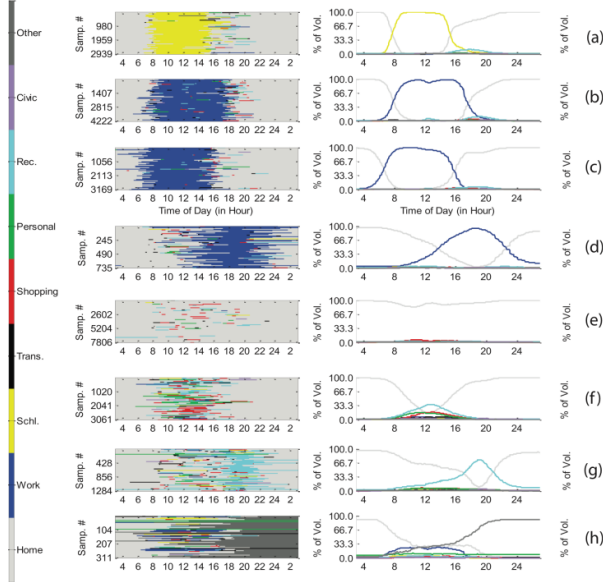
The  $K$ -means algorithm has been widely applied to partition datasets into a number of clusters [30]. It performs well for many problems. However, the computational cost can be very high, when the dimension of the data is large [31]. As the principle component analysis (PCA)/eigen decomposition method is widely employed for dimension reduction, it is a common practice to utilize the  $K$ -means algorithm via PCA and obtain successful applications [32, 33]. In an earlier study [34], we explain in detail the processes and validity of applying the  $K$ -means algorithm via PCA method to cluster individuals into different groups according to their activity patterns in the Chicago Metropolitan Area. For details about the  $K$ -Means clustering via PCA method, readers may refer to Section 4 of Jiang et al. [34].

### 4.2 Daily Activity Pattern and Trace Clusters in Chicago

In this section, we first review the clustering results of activity patterns in the Chicago Metropolitan Area on an average weekday [34], and then we cluster individuals’ daily spatial-temporal traces using a similar process.

#### 4.2.1 Clusters of Daily Activity Patterns

Figure 2 summaries the analysis results of individuals’ daily activity pattern clusters on an average weekday in the Chicago Metropolitan Area.



**Figure 2 Clustering of individuals' weekday activity patterns in Chicago, with clusters (a) students, (b) regular workers, (c) early-bird workers, (d) afternoon workers, (e) stay-at-home, (f) morning adventurers, (g) afternoon adventurers, and (h) overnight adventurers. Adapted from [34] Jiang et al.**

The first column of Figure 2 displays individuals’ daily activity sequences for each cluster. The second column shows the aggregated volume of different types of activities in the metropolitan area during a specific time interval over 24 hours. The results contains 8 types of personal activity patterns, including students (12.50%), regular workers (17.90%), early-bird workers (13.50%), afternoon workers (3.10%), the stay-at-home (33.20%), the morning adventurers (13.00%), the afternoon adventurers (5.50%) and the overnight adventurers (1.30%). Notice that this represents much richer information about urban groups than the traditional classification in which only 47% of the individuals can be categorized as students or workers based on their daily activity patterns.

#### 4.2.2 Clusters of Daily Traces

We cluster individuals’ spatial-temporal traces according to their space-time similarities by applying the  $K$ -Means algorithm via PCA in a very similar procedure to the study we conducted earlier [34].

First, we sample the individuals’ locations and activities every five minutes, using the transformed data described in Section 2. Let  $n_T$  denote the total number of individuals in the sample, then for each individual  $i = 1, \dots, n_T$ , we have an 576 (=288×2) dimensional vector  $\mathbf{T}_i = (x_{i,1}, \dots, x_{i,288}, y_{i,1}, \dots, y_{i,288})'$  to describe his/her trace, where  $x_{i,j}$  and  $y_{i,j}$  is individual  $i$ 's longitude and latitude in the  $j$ -th time instant being sampled.

Second, we deal with  $\mathbf{T}_i$  in the same way as we deal with  $\mathbf{a}_i$  in Section 4 of Jiang et al [34]. We arrange the thus obtained 576 eigenvalues in descending order, i.e.,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{576} \geq 0$ , and the eigenvector  $\mathbf{v}_k$  that corresponds to the  $k$ -th eigenvalue is called the  $k$ -th *eigen trace*.

Third, we reconstruct the individual  $i$ 's spatial-temporal trace  $\mathbf{T}_i$ , by using a subset of eigen traces as follows. Let  $\mathbf{d}_i = \mathbf{T}_i - \bar{\mathbf{T}}$ , where  $\bar{\mathbf{T}} = \frac{1}{n_T} \sum_{i=1}^{n_T} \mathbf{T}_i$  is the sample mean, and suppose the projection of  $\mathbf{d}_i$  onto the first  $h$  eigen traces  $\{\mathbf{v}_1, \dots, \mathbf{v}_h\}$  are  $(z_1, \dots, z_h)'$ . According to formula

$$\hat{\mathbf{T}}_i = \bar{\mathbf{T}} + (\mathbf{v}_1, \dots, \mathbf{v}_h)(z_1, \dots, z_h)' \quad (4)$$

we obtain a vector  $\hat{\mathbf{T}}_i = (\hat{x}_{i,1}, \dots, \hat{x}_{i,288}, \hat{y}_{i,1}, \dots, \hat{y}_{i,288})' \in \mathbb{R}^{576}$ . We define the *reconstruction error*  $e(\mathbf{T}_i)$  for  $\mathbf{T}_i$  as the *average reconstruction deviation*, namely,

$$e(\mathbf{T}_i) = \frac{1}{288} \sum_{j=1}^{288} \text{dist}((x_{i,j}, y_{i,j}), (\hat{x}_{i,j}, \hat{y}_{i,j})) \quad (5)$$

where  $\text{dist}((x_{i,j}, y_{i,j}), (\hat{x}_{i,j}, \hat{y}_{i,j}))$  is the distance (in kilometers) between two locations  $(x_{i,j}, y_{i,j})$  and  $(\hat{x}_{i,j}, \hat{y}_{i,j})$ . Given any  $\varepsilon > 0$ , we can find some  $h > 0$ , so that the *average reconstruction error*,  $\frac{\sum_{i=1}^{n_T} e(\mathbf{T}_i)}{n_T}$ , caused by neglecting the projections onto the ignored eigen traces  $\{\mathbf{v}_{h+1}, \dots, \mathbf{v}_{576}\}$  is no greater than  $\varepsilon$ . Let  $\varepsilon_0 > 0$  be the acceptable error level, and define  $h(\varepsilon_0)$  to be the smallest  $h$  such that the average reconstruction error induced by using the first  $h$  eigen traces is no greater than  $\varepsilon_0$ . We then call  $h(\varepsilon_0)$  the *appropriate number of eigen traces*. In this study, we take  $\varepsilon_0 = 0.5$  kilometers, and we get  $h(\varepsilon_0) = 33$ . Thus, eigen traces  $\{\mathbf{v}_1, \dots, \mathbf{v}_{33}\}$  are used in the reconstruction of  $\mathbf{T}_i$  and in the  $K$ -means clustering of individuals’ spatial-temporal traces.

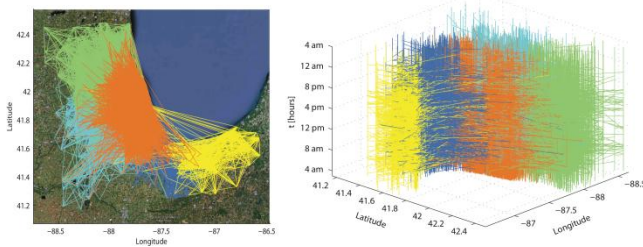
Lastly, according to Dunn's Index (DI), which maximizes inter-cluster distances while minimizing the intra-cluster distances, our analysis shows that with a clustering number of 2, it provides the most stable partition of individuals (i.e., the higher the index, the

better the clustering results) (see Table 1). However, since we are more interested in a sophisticated clustering of individuals than the dichotomy grouping, we find that when the clustering number is 5, it provides a second-best alternative yet much richer partition of individuals' daily traces.

**Table 1. Cluster numbers ( $k$ ) and the Dunn's Index (DI)**

$k$	2	3	4	5	6
DI	2.431	1.501	1.853	2.017	1.253
$k$	7	8	9	10	11
DI	1.288	1.365	1.38	1.387	1.230
$k$	12	13	14	15	16
DI	1.370	1.059	1.063	1.088	1.092

Figure 3 shows the clustering results of individuals' spatial-temporal traces with  $k=5$  in space. The left panel of the figure shows a 2-D view of the clustering results with each color representing a cluster, and  $x$ -,  $y$ -axes representing longitude and latitude; while the right panel of the figure shows a 3-D view of the same clustering results, with an additional vertical  $z$ -axis representing the time dimension. This type of 3-D view of individuals' traces is called spatiotemporal prism by geographers [35], and is visually helpful to understand people's movement in space and time.



**Figure 3 Clustering of individuals' daily traces on an average weekday in Chicago (cluster number=5).**

As we know, in people's daily life, their travel from one destination to another in space and time is not usually confined by administratively defined boundaries such as county boundaries. Therefore clustering a region into subdivision by using individuals' spatial-temporal traces can reveal underlying inherent connections in the region better than using administratively defined areas such as counties.

Figure 4 shows the geographic location of counties that define the Chicago Metropolitan. From the geographical coverage of the clustered traces in Figure 3, we find that Cluster #1, depicted in dark blue, covers the Lake County (south to Cook) and the southeast corner of Cook County. Cluster #2, depicted in cyan, covers most area of Du Page, the north part of Will and Grundy, and eastern side of Kane and Kendall counties, and the northern part of Grundy County. Cluster #3, depicted in green, covers the north part of Cook County, the Lake County (north to Cook), and the McHenry County. Cluster # 4, depicted in yellow, covers the Porter County and LaPorte County (in the State of Indiana). Cluster # 5, depicted in orange, covers the center of the Cook



**Figure 4 Counties in the Northeastern Illinois Region and Northwestern Indiana Region**

County, or the City of Chicago. It is interesting to analyze further the intrinsic reasons of this regional clusters defined by the user traces. It may be possible to find some demographic or economic reasons inherent in these aggregations.

## 5. URBAN SPATIAL-TEMPORAL STRUCTURE BY REGION, ACTIVITY PATTERN AND TYPE

With the previous analysis we obtained clusters of individuals by their daily activity patterns and their spatial-temporal traces (in Section 4). Together with the proposed measurement (i.e., time-cumulative spatial activity density) to estimate urban spatial-temporal structure (in Section 3), we are able to further explore the detailed urban spatial-temporal structure for individuals with different daily activity types defined as students, regular workers, early workers, afternoon workers, stay-at-home, morning adventurers, afternoon adventurers and overnight adventurers. We can observe now how these types of individuals distribute across different clustered sub-regions based on their time-cumulative spatial activity density.

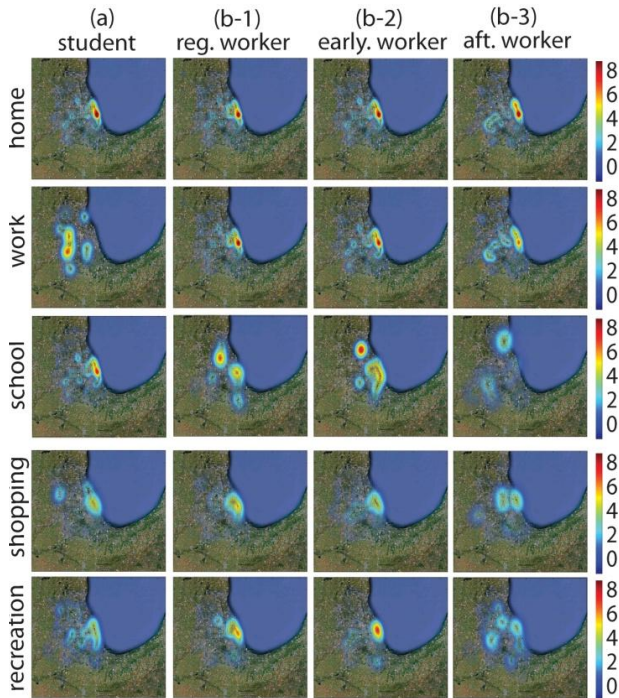
Due to limited space of the paper, and the importance of the City of Chicago, in terms of intensity of various activities (shown in subsection 3.2), as a demonstration, in this section we further explore the spatial temporal structure for individuals whose daily activities are heavily concentrated in this sub-division of the region (i.e., Cluster # 5 obtained from the trace clustering result in subsection 4.2.2), and focus on five types of activities for the eight types of individuals (obtained from the activity pattern clustering results in subsection 4.2.1). We first discuss the spatial distribution patterns of activity intensities on home, work and school activities for individuals engaged in a fixed daily activity, namely: students and workers (including early workers, early-bird workers and afternoon workers) in subsection 5.1. Then discuss the spatial distribution patterns of intensity of the same five types of activities for stay-at-homes and the three types of adventurers (i.e., morning, afternoon and overnight adventurers). The latter group includes individuals that are not traditionally classified by activity patterns and are much harder to model in urban simulators.

### 5.1 The Students and Workers

As depicted in Figure 5, we find that the spatial distribution of home activities for students and workers are quite similar—heavily concentrated in the downtown area and northern part of the city of Chicago (Figure 5, row 1). It also shows similar patterns in the work activity intensity for the three types of workers, but quite different for the students (with multiple centers in the west part of the city) (Figure 5, row 2).

The spatial distribution of school activity intensity for students matches in a very similar way to that of their home activity intensity. While for regular workers and early-bird workers, there are two centers—one in the north part of the city, and one in the downtown area of the city. For the afternoon workers, the school activity intensity is very weak, but mostly concentrated in the far northern and western edge of the city (Figure 5, row 3).

Shopping activity intensity for the regular and early-bird workers are similar—concentrated in downtown Chicago; and with two additional sub-centers for the students—one in the north, and one in the northwest (around the O'Hare airport). And there are two centers hand-by-hand in the center of Chicago for the afternoon workers.



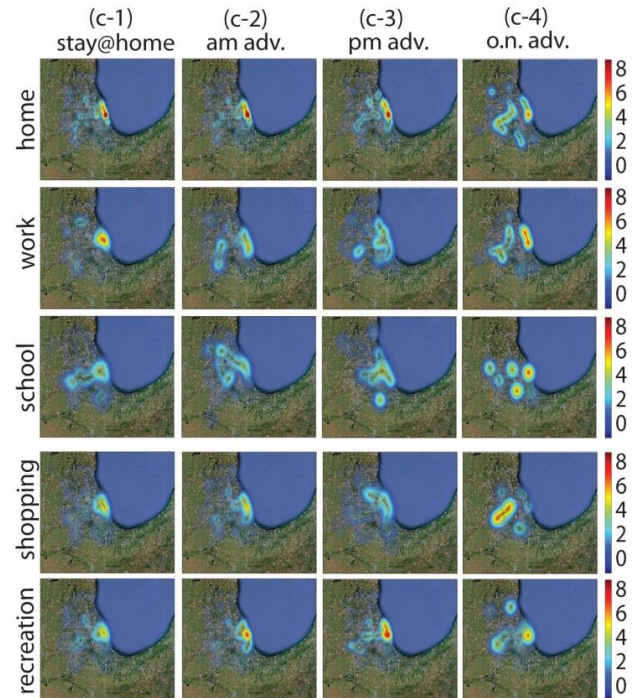
**Figure 5** Time-cumulative spatial densities of home, work, and school activities in the Chicago sub-region by (a) students, (b-1) regular workers, (b-2) early-bird workers, and (b-3) afternoon workers on an average weekday.

Interestingly, we see that the intensity of recreational activity for regular and early-bird workers are similar in geographical location (concentrated in downtown Chicago) compared to their shopping activity, but with stronger intensity— meaning that for these two types of workers, there are either more of them doing recreational (e.g., eating out with friends)/entertainment activities and/or they spend longer time in these activities than in shopping on weekday in the downtown area (Figure 5, row 4). The spatial distribution of recreation/entertainment activity intensity for afternoon workers are more dispersed in space and time than their counterparts of regular and early-bird workers. For students, their recreational activity centers are distributed along the lake shore from north to downtown Chicago, and also in the western part of the city.

To summarize, similar spatial distribution patterns of intensity exist for students' home and school activity, as well as for workers' home and work activity. In terms of the spatial distribution of shopping and recreational activity intensities, they are quite similar for regular and early-bird workers (concentrated in the city center), but exhibit diverse and multiple centers for the afternoon workers (which can be explained by their space-time flexibility during the day compared with regular and early-bird workers). For students, their spatial distribution of recreational activity intensity is similar to that of their school activity, but with lower density, and their shopping activity are more concentrated in the center and northern part of the city, as well as in the northeastern corner of the city near the O'Hare airport.

## 5.2 The Stay-at-Homes and Adventurers

Different from the students and workers, who spend most of their day on (spatially and temporally constrained) activities of school and work, adventurers have more time flexibility; and the stay-at-homes are more limited in terms of their spatial flexibility.



**Figure 6** Time-cumulative spatial densities of shopping/errands & recreation/entertainment activities in the Chicago sub-region by (c-1) stay-at-home, (c-2) morning adventurers, (c-3) afternoon adventurers, and (c-4) overnight adventurers.

We can see from Figure 6 that the spatial distributions of activity intensities of various types (except for home activity) are very diverse in space for the adventurers, but very uniformly concentrated for the stay-at homes. For the morning and afternoon adventurers, the second most intense activity right after the home activity is recreation/entertainment, and they are heavily concentrated in the city center, and moderately concentrated in both the north and south parts of the city for morning adventures, and only in the north for the afternoon adventurers.

Shopping activities for the overnight adventures are heavily concentrated along the southwestern corridor, and their recreational activities are also distributed in the far north part of the city.

The school and work activity intensity distribution in space for morning and afternoon adventurers are even more diverse and dispersed compared with their shopping and recreational activities, which means that very few adventurers conduct work/school activities, and their spatial distribution is more spread in the city than concentrated in the downtown area.

Interestingly, the spatial distributions of home and work activity intensities for the overnight adventures are very similar. But that of their school activity intensity is very diverse and spreads over different parts of the city.

## 6. CONCLUSIONS

This paper introduces a new concept of urban spatial-temporal structure, which expands the traditional theories of urban structure, and offers a framework to study and analyze it using activity-based travel survey data. We discuss the advantages of travel survey data in terms of their richness in revealing individuals' activity types, space and time presence, and we provide a new

perspective on how to mine survey data as complements to the recently emerging massive urban sensing data.

With the introduction of spatial distribution of activity intensity, measured by time-cumulative spatial activity density, and the employment of the kernel density estimator, we provide an approach to analyze the proposed urban spatial-temporal structure.

In order to discover similarities and differences in human activity patterns and spatial-temporal traces, we apply *K*-means algorithm via PCA method to cluster individuals into groups. By estimating and visualizing the time-cumulative spatial densities of various activities by person types (obtained from the clustering of daily activity patterns) in one of the sub-regions of the Chicago Metropolitan Area (identified from the clustering of individuals' traces), we are able to explore the diverse spatial-temporal structure of Chicago. Due to space limit, we do not demonstrate the results for the rest of the four sub-regions in the Chicago metropolitan area in this paper. However, abundant information obtained from other sub-regions will help planners gain solid understandings on "how different sub-regions of the metropolitan area have been utilized by different types of individuals for different activity types".

Answering these questions will be essential for urban planners and scholars to understand the dynamics and complexity of the polycentric city, and this paper offers new insights for urban planning through the estimates of spatial distributions of various activity intensities for different types of individuals. As activity intensity is very closely related to energy consumption, business opportunity and infrastructure usage, the measurements and approaches proposed here will be helpful for urban planners to design sustainable cities in the future.

## 7. ACKNOWLEDGMENTS

This research was funded in part by the MIT Department of Urban Studies and Planning, by the US Department of Transportation Region One University Transportation Center, and by the Singapore National Research Foundation (NRF) through the Singapore-MIT Alliance for Research and Technology (SMART) Center for Future Mobility (FM).

## 8. REFERENCES

[1] Batty, M. *Cities as Complex Systems: Scaling, Interactions, Networks, Dynamics and Urban Morphologies* Springer, Berlin, DE, 2009.

[2] Portugali, J., Meyer, H., Stolk, E. and Tan, E. *Complexity Theories of Cities Have Come of Age: An Overview With Implications to Urban Planning and Design*. Springer, Heidelberg, 2012.

[3] Anas, A., Arnott, R. and Small, K. A. Urban Spatial Structure. *Journal of Economic Literature*, 36, 3 1998, 1426-1464.

[4] Horton, F. E. and Reynolds, D. R. Effects of Urban Spatial Structure on Individual Behavior. *Economic Geography*, 47, 1 1971, 36-48.

[5] Florida, R., Gulden, T. and Mellander, C. The rise of the mega-region. *Cambridge Journal of Regions, Economy and Society*, 1, 3, November 1, 2008, 459-476.

[6] Greene, D. L. Recent Trends in Urban Spatial Structure. *Growth and Change*, 11, 1, 1980, 29-40.

[7] McMillen, D. P. and McDonald, J. F. A Nonparametric Analysis of Employment Density in a Polycentric City. *Journal of Regional Science*, 37, 4, 1997, 591-612.

[8] Gordon, P., Richardson, H. W. and Wong, H. L. The distribution of population and employment in a polycentric city:

the case of Los Angeles. *Environment and Planning A*, 18, 1986, 161-173.

[9] Mahizhnan, A. Smart cities: The Singapore case. *Cities*, 16, 1 1999, 13-18.

[10] Ratti, C. and Townsend, A. The Social Nexus. *Scientific American* 305, 16 August 2011, 6.

[11] Song, C., Qu, Z., Blumm, N. and Barabási, A.-L. Limits of Predictability in Human Mobility. *Science*, 327, 5968, February 19 2010, 1018-1021.

[12] Gonzalez, M. C., Hidalgo, C. A. and Barabasi, A.-L. Understanding individual human mobility patterns. *Nature*, 453, 7196, 2008, 779-782.

[13] Calabrese, F., Colonna, M., Lovisolo, P., Parata, D. and Ratti, C. Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. *IEEE Transactions on Intelligent Transportation Systems*, 12, 1 2011, 141-151.

[14] Yuan, J., Zheng, Y. and Xie, X. *Discovering Regions of Different Functions in a City Using Human Mobility and POIs*, Beijing, 2012.

[15] Zheng Y., Zhang L., Xie X. and Y., M. W. *Mining interesting locations and travel sequences from gps trajectories*. Madrid, 2009.

[16] Nature Editorial A flood of hard data. *Nature*, 453, 7196 2008, 698-698.

[17] Becker, R. A., Caceres, R., Hanson, K., Loh, J. M., Urbanek, S., Varshavsky, A. and Volinsky, C. A Tale of One City: Using Cellular Network Data for Urban Planning. *Pervasive Computing, IEEE*, 10, 4, 2011, 18-26.

[18] Glaeser, E. L., Kolko, J. and Saiz, A. Consumer city. *Journal of Economic Geography*, 1, 1, January 1, 2001, 27-50.

[19] Eagle, N. and Pentland, A. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63, 7 2009, 1057-1066.

[20] Gandia, R. *City outlines travel diary plan to determine future transportation needs* Sun Media, Calgary, Alberta, Canada, 2012.

[21] Hägerstrand, T. Reflections on "what about people in regional science?". *Papers in Regional Science*, 66, 1, 1989, 1-6.

[22] Yu, H. and Shaw, S.-L. Exploring potential human activities in physical and virtual spaces: a spatio-temporal GIS approach. *International Journal of Geographical Information Science*, 22, 4, 2008, 409 - 430.

[23] Harvey, A. and Taylor, M. Activity settings and travel behaviour: A social contact perspective. *Transportation*, 27, 1, 2000, 53-73.

[24] Hanson, S. and Hanson, P. Gender and Urban Activity Patterns in Uppsala, Sweden. *Geographical Review*, 70, 3, 1980, 291-299.

[25] Hanson, S. and Kwan, M.-P. *Transport: Critical Essays in Human Geography*, Ashgate, Aldershot, 2008.

[26] Kwan, M.-p. and Lee, J. *Geovisualization of Human Activity Patterns Using 3D GIS: A Time-Geographic Approach*. Oxford University Press, Oxford, 2003.

[27] Chicago Metropolitan Agency for Planning *Chicago Travel Tracker Household Travel Inventory*, 2008.

[28] Wand, P. and Jones, C. *Kernel Smoothing*. Chapman & Hall, 1995.

[29] Kwan, M.-P. and Hong, X.-D. Network-based constraints-oriented choice set formation using GIS. *Geographical Systems*, 5, 1998, 139-162.

[30] Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G., Ng, A., Liu, B., Yu, P., Zhou, Z.-H., Steinbach, M., Hand, D. and Steinberg, D. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14, 1, 2008, 1-37.



- [31] Huang, Z. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, 2, 3, 1998, 283-304.
- [32] Ding, C. and He, X. K-means clustering via principal component analysis. In *Proceedings of the Proceedings of the twenty-first international conference on Machine learning*, ACM, 2004. Banff, Alberta, Canada.
- [33] Zha, H., Ding, C., Gu, M., He, X. and Simon, H. Spectral relaxation for K-means clustering. *Advances in Neural Information Processing Systems*, 14 (NIPS'01) 2001, 1057–1064.
- [34] Jiang, S., Ferreira, J. and González, M. Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, April 23, 2012, 1-33.
- [35] Miller, H. J. A Measurement Theory for Time Geography. *Geographical Analysis*, 37, 1, 2005, 17-45.