Article

Urban Analytics and City Science

EPB: Urban Analytics and City Science
2024, Vol. 51(4) 803–822

DOI: 10.1177/23998083231198721
journals.sagepub.com/home/epb
Sage

Understanding street-level urban vibrancy via spatial-temporal Wi-Fi data analytics: Case LivingLine Shanghai

Yan Zhang
Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

Chengliang Li
Graduate School of Architecture, Planning and Preservation, Columbia University, New York, NY, USA

Jiajie Li
Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

School of Software Engineering, Tongji University, Shanghai, China

Zhiyuan Gao
College of Civil Engineering, Tongji University, Shanghai, China

Tianyu Su
Graduate School of Design, Harvard University, Cambridge, MA, USA

Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

Can Wang
College of Architecture & Urban Planning, Tongji University, Shanghai, China

Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

Hexin Zhang
College of Design and Innovation, Tongji University, Shanghai, China

Teng Ma
State Key Lab of Disaster Reduction in Civil, Tongji University, Shanghai, China

Yang Liu
College of Design and Innovation, Tongji University, Shanghai, China

Weiting Xiong, Ronan Doorley and Luis Alonso
Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

Corresponding author:
Jiajie Li, MIT Media Lab¸ 7 Mass. Ave., E14/E15, Cambridge, MA 02139-4307, USA.
Email: jiajie@mit.edu

Yongqi Lou
College of Design and Innovation, Tongji University, Shanghai, China

Kent Larson
Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract
Urban vibrancy is a topic of great concern in the field of urban design and planning. However, the definition and measurement of urban vibrancy have not been consistently and clearly followed. With the development of technologies such as big data and machine learning, urban planners have adopted new methods that enable better quantitative evaluation of urban performance. This research attempts to quantify the impact on the urban vibrancy of the urban interventions introduced by the LivingLine project in a residential neighborhood renovation made in Siping Street, Shanghai. We use Wi-Fi probes to process collected mobile phone data and segment people into different categories according to commuting patterns analysis. We use a pre-trained random forest model to determine the specific locations of each person. Subsequently, we analyze the behavior patterns of people from stay points detection and trajectory analysis. Through statistical models, we apply multi-linear regression and find that urban intervention (well-curated and defined lab events deployed in the street) and people's behavior are positively correlated, which helps us to prove the impact of urban intervention on street dynamics. The research proposes a novel, evidence-based, low-cost methodology for studying granular behavior patterns on a street level without compromising users' data privacy.

Keywords
Street activities, Wi-Fi data, machine learning, urban vibrancy, multi-linear regression, behavioral analysis

Introduction

The core of urban vibrancy is real-time human activities (Barreca et al.,2020; Huang et al.,2020; Lang et al.,2020). To stimulate urban vibrancy, urban designers have developed a handful of urban intervention toolsets, from shaping the physical features of urban spaces as a vehicle for inviting human activities, to designing social events as a catalyst for social interactions (Griffiths, 1995). Thus, understanding and evaluating the impact of these interventions on human dynamics is urgent to guide strategic urban interventions (Zhong et al., 2020).

Among the various scales of activities that urban designers may look at, the street scale is the most human-centric and intriguing, but also challenging scale due to the requirement of fine data granularity (Resch and Szell, 2019). People's daily life happens on the street scale, so tracking their activities inside such a small temporal and spatial frame is a meaningful challenge. With the rapid development of big data technology, planners have a better understanding of crowd behavior and traffic patterns in urban spaces (Bellomo et al., 2016; Zhang et al., 2021). However, big data research mostly focuses on the macro-scale such as cities. For street-scale activities, the existing big data is often inaccurate enough to be more closely integrated with the space in detail. Moreover, most of urban big data is controlled by Internet companies and mobile operators, which is not affordable and accessible (Magalhaes, 2021).

Therefore, we need new methods to conduct more in-depth studies of the activities on the street scale. This study proposes a novel methodology to collect low-cost and easy-to-deploy street scale Wi-Fi data. This data is used for studying behavioral patterns of street visitors and to understand their correlation with social events held in the neighborhood. Do these social events lead to a more vibrant, diversified street space?

Related works

Street dynamics and urban vibrancy measurements

One of the pioneers in urban studies, Jane Jacobs believed vibrant city life could flourish only in neighborhoods that met several conditions (Jacobs, 1961). Streets with a clear pedestrian realm, narrower lanes, separation of local traffic, and more complex intersections were all identified as elements contributing to the vibrancy of the place (Jacobs, 1993). Based on the traditional definition, researchers developed different metrics to measure urban vibrancy. Ye et al. (2017) employed new urban data with analytical methods and found that both typology and density are significant for urban vibrancy. Barreca et al. (2020) tried to define and measure urban vibrancy by studying its relationship with neighborhood services and the real estate market.

With the development of sensing technologies, many researchers provided new insights for studying street dynamics. Empowered by newly available sources of spatial big data, Huang et al. (2019) proposed a comprehensive measure of urban vibrancy to investigate the associations between vibrancy and various urban built environment indicators. Williams et al. (2019) utilized image capture, location tracking, etc., and evaluated the performance according to the Gehl Method to measure urban public life. Long and Zhou (2016) defined the concepts of street vibrancy, developing the factors for quantitatively evaluating street vibrancy at the street level, and adopted linear regression for identifying the impact of each factor on the street vibrancy. Long and Huang (2019) investigates urban design's impact on economic vibrancy in 286 Chinese cities, revealing a significant positive relationship using various urban form indicators.

Application of Wi-Fi data

The application of Wi-Fi data for urban vibrancy analysis has drawn increased attention, albeit with a persistent gap concerning street-scale analysis. While diverse research has contributed to this domain, some noteworthy attempts merit special mention. For instance, endeavors have been made to identify activity patterns and population distribution in the context of urban vibrancy, such as the efforts by Calabrese et al. (2010) in categorizing wireless access points and Kontokosta and Johnson (2017) in extracting activity patterns in Lower Manhattan. Simultaneously, innovations in data integration have been pursued, exemplified by Shen et al. (2020) combining travel and smartphone usage behavior and Hu et al. (2020) coupling Wi-Fi probes with location data for population distribution study. Efforts have also been made to enhance the quality of Wi-Fi data; notable is Chilipirea (2019)'s attempt to refine Wi-Fi remote positioning data for crowd-movement analysis. Prentow et al. (2015) and Kim (2018) further expanded the scope by leveraging abundant Wi-Fi spatio-temporal data for large-scale building complexes and correlating Wi-Fi access points with urban vibrancy measures, respectively. Despite these significant contributions, a conspicuous gap persists in applying Wi-Fi data to street-scale urban vibrancy analysis—a critical area for understanding fine-grained urban dynamics. This paper addresses this lacuna, offering fresh insights into the potential of Wi-Fi data in studying street-level urban vibrancy Figure 1.

Research framework and methodology

Research goal

The research question of this study is: how is urban vibrancy associated with physical changes and events?

The objectives of this research include:

- To propose a novel methodology for studying finer-grained urban activities using Wi-Fi signal data and segmenting them into meaningful groups.
- To understand what factors are correlated with the density and diversity of urban activities and how significant the correlations are.

Study area

The primary data used in this study were collected on NICE 2035 LivingLine Street (Supplementary Figure S3), a renovated street of a residential neighborhood located in Siping Road Residential District, Shanghai (Jiang et al., 2020).

We maintained a publicly accessible wireless network consisting of 50 access points (APs) distributed throughout LivingLine Street in Shanghai, in which there are 26 outdoor APs covering the whole street to capture population and mobility patterns, and 24 indoor APs covering labs and

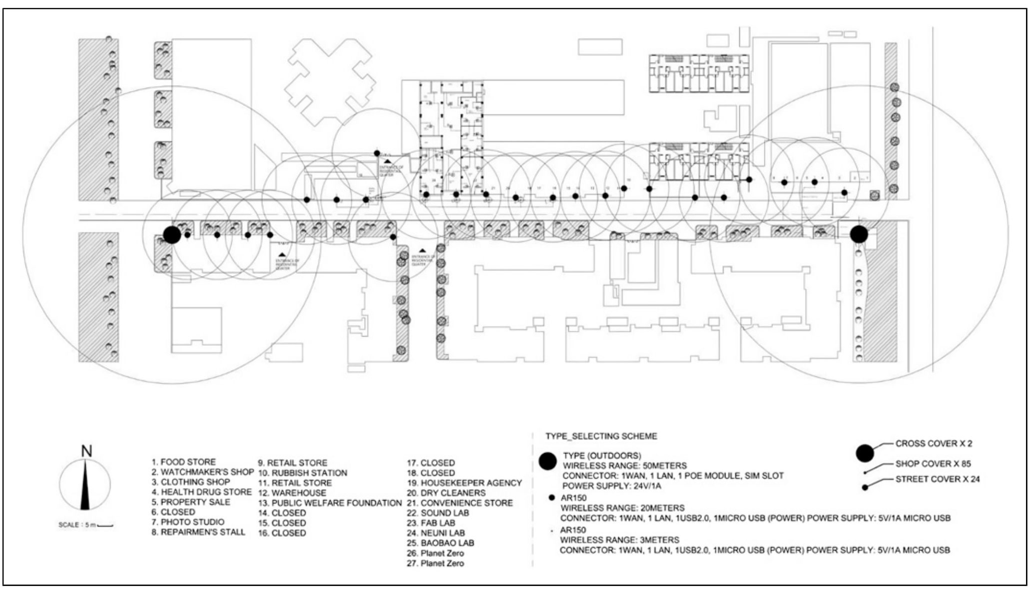

Figure 1. Study area and AP setup. In the NICE 2035 project, there are 24 indoor APs covering labs and public places to record human activities, and there are 26 outdoor APs spanning the entire street to record population and mobility patterns.

public spaces of NICE 2035 for capturing human activities. From October 1, 2018, to December 31, 2018, 10,687,173 observations and 32,211 unique macs were identified.

Overview of the approach

As previously described, the goal of this study is to propose a comprehensive methodology that helps to understand human dynamics on the street scale and to find correlations between urban interventions and street dynamics. As such, this methodology can be useful in the impact evaluation stage after the implementation of urban renewal projects or place-making activities and has the potential to support the decision-making stage by inferring outputs of different intervention plans.

As shown in Figure 2, this methodology covers four steps: Data Collection and Cleaning, Grouping and Positioning, Street Activities Analysis, and Regression Analysis.

Data collection and cleaning

Unlike cellular data, which only triggers observations when devices make phone calls or send text, Wi-Fi APs constantly collect request data from Wi-Fi-enabled devices in the surrounding area, regardless of whether the devices connect to the network.

Each AP has a detection range of 15 m and collects data at a frequency higher than 1 hertz. The APs are plugged into a SIM card to gain access to the internet and are configured to upload newly collected data to our server every 15 minutes. Once a pedestrian walks through the street, he or she will pass through detection areas of different APs. Therefore, we can collect the Received Signal Strength Indicator (RSSI) data between the pedestrians' mobile device and all the APs.

Each observation from the Wi-Fi AP includes the following variables: a unique mac address of the probe, a timestamp, the MAC (Media Access Control) address of devices connected to this probe and the device's received signal strength indicator (RSSI).

To prepare the raw data for analysis, we took three steps of data preprocessing: converting data format, filtering out data from non-mobile users, and physical cleaning. Privacy and confidentiality are non-trivial concerns in the use of location-based data. In this analysis, device MAC addresses are

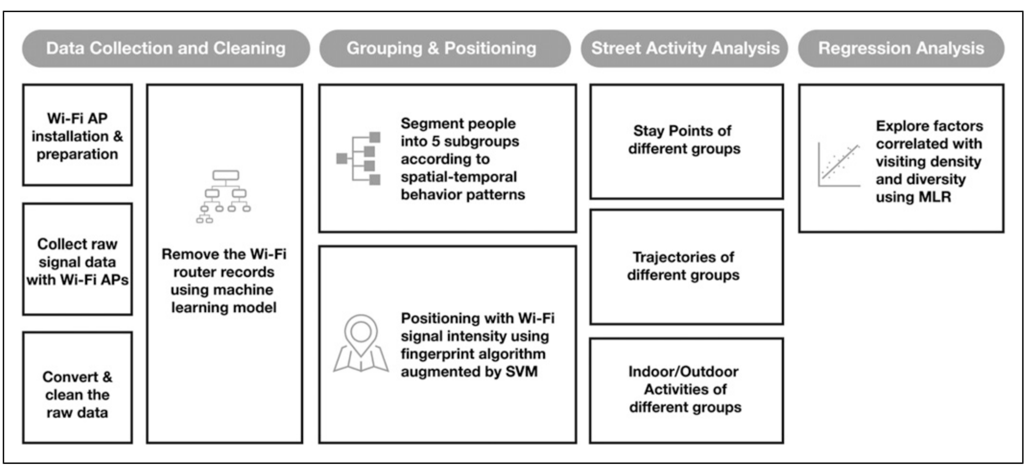

Figure 2. Overview of research methodology, which covers four steps: data collection and cleaning, people grouping and positioning, behavioral pattern analysis, and regression analysis.

anonymous and only aggregated counts are used for analysis. (Details are shown in supplementary material section 12).

Grouping and positioning

As the street is shared by people with different demographic and social attributes, it's important to segment them into meaningful categories, so urban designers can understand the composition and design interventions for each group specifically.

Besides, activities are inherently related to both temporal and spatial dimensions. To detect different activities such as staying, walking, and interacting with a certain urban space, we need to have relatively precise positioning of the devices on the street.

Grouping. Existing methods to study public activities at street scale usually focus on collecting demographic information. For instance, Liu et al. (2019) introduced a big data method using spatial-temporal mobility characteristics and interest preferences information in indoor Wi-Fi positioning data to infer people's gender and age. But it depends heavily on customer interest preferences, which are less distinct in a street than in a market. They also find it difficult to make inferences about group visitors because a group of people may have similar mobility patterns but vary in relation to one another. In terms of urban intervention on streets, we believe demographic information could be obtained from census data or surveys, while mobility patterns could be used to extract more knowledge about people's activity on the street. Compared with other data sources, Wi-Fi data's strength lies in its ability to track individual devices to find recurring patterns after accumulating a certain amount of data. The application of Wi-Fi data is being gradually introduced into the state-of-the-art in the context of spatial, urban, and behavioral analysis.

In our study, first we use occurrence frequency to classify users into frequent visitors, occasional visitors, and seldom visitors. The definitions of each group are in the following Table 1.

After applying the classification rules to all the data, we aggregated data in a single day and counted unique devices in each category to capture the trend of the overall population. To further segment the visitor population, we developed a 2-level rule-based classification model based on these hourly temporal patterns. (Detailed rules are shown in section 12 of supplementary material and Table S7).

Figure 3 visualizes the mean weekly population by hours, showing meaningful patterns in each user group. Almost every class has two constantly recurring peaks in a day. For commuters, these peaks typically appear at 8 a.m. and 6 p.m., which is the same as passersby, aligned with the rush hours. For residents, they go out a bit earlier on average, at 6 a.m. and come back home at 7 or 8 p.m. Resident workers don't need to commute so their population has less fluctuation. When it comes to visitors, we find that besides the morning and evening commuting time, there is another lower peak at around 7 or 8 p.m., which suggests nearby residents may take a walk after dinner. As expected, the populations of most groups all go down to zero, except for residents and resident workers. There are

Table 1. The occurrence frequency was used to classify users into frequent visitors, occasional visitors, and seldom visitors.

Class	Definition
Frequent	Have visited in at least 1/2 out of all the days
Occasional	Have visited less than 1/2, but more than 2/7 out of all the days
Seldom	Have visited in at least 2/7 out of all the days

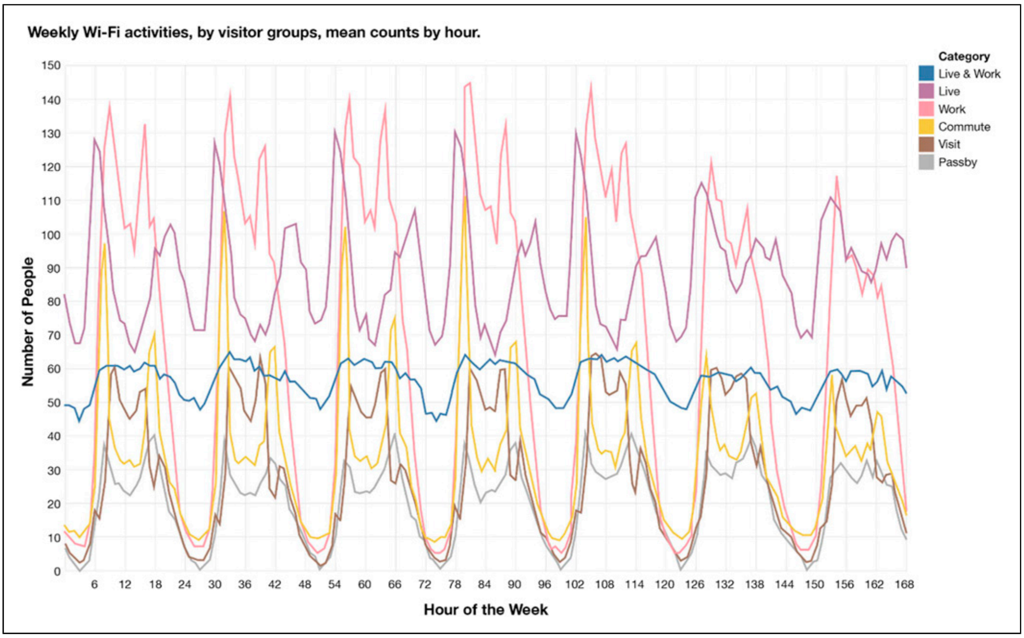

Figure 3. This figure shows weekly population statistics at hourly granularity, grouped by refined visitor patterns categories (work & live, live, work, commute, visitors, passersby).

also obvious decreases on weekends for commuters, workers, and residents. These patterns suggest that our classification model can capture the distinctive activity patterns from Wi-Fi data and has merit in segmenting people into meaningful groups.

Positioning. In general, there are three kinds of methods (Xia et al., 2017) that utilize the strength of the received signal to obtain a user's position: trilateration, approximate perception, and fingerprint positioning.

We tried both the trilateration method and path loss model at first, but the performances were insufficient. Then we utilized the Fingerprint Positioning method, which requires neither knowing the position of APs nor converting RSSI into spatial distances. Instead, it requires us to make a fingerprint library to conduct the positioning, which has higher accuracy than trilateration. There are two central phases in fingerprint positioning: offline training and online positioning.

In offline training, we first collected fingerprints in the positioning area in advance. We divided the street into a 10 m*10 m grid, which has y cells in total. At the center of each cell, we have a mobile device that remained there for more than 5 min, for which we knew its MAC address. Then we collect all the RSSI values captured by each AP about this device, and the device's current position as a fingerprint. The fingerprint database gradually covered the whole street during an iterative process.

Then the database was used as the training data set to train a machine learning algorithm. Using Scikit-learn, different ML algorithms were tested, including K Nearest Neighbor, Random Forest, Support Vector Machine, and Gradient Boosting.

In the online positioning stage, the collected signal strength information was uploaded to the server. The information would serve as the input of the regression model to conduct positioning precisely.

As the training/test set for the machine learning model, there is a dataset with 678 samples with 50 dimensions (features). By applying different machine learning algorithms to the data, the performances shown in Supplementary Table S5 reflect the different accuracies of each model. We chose the Random Forest model due to its high accuracy (82.52% accuracy on test set) and low variances (standard deviation equals 12.5389 on test set). A visualization of the positioning result is shown in Figure 4.

Street activities analysis

To build a more convincing and comprehensive model articulating the correlation between urban vibrancy and NICE2035 events (Supplementary Table S3), a behavior pattern analysis, including stay points and trajectories of people in the street, needs to be performed.

Stay points analysis. Stay points can reflect pedestrians' activity information and transform spatial trajectories into semantic trajectories. Through the number, density, location, and time interval of stay points, we can analyze the impact on pedestrians, and then determine whether the urban intervention changes people's activities and behavior patterns and urban vibrancy.

The detection of staying points can be used to help mine the user's GPS trajectory. These staying points are places and locations of interest to specific users, upon which businesses could provide personalized recommendations for users (Li et al., 2008). The same stay point detection algorithm is used to process geolocated telecom data to analyze and perform cluster analysis to obtain activity groups of different scales and forms of types and personalities (Noyman et al., 2019). In addition, a speed-based recognition algorithm is also used to extract the staying points in the trajectory (Cai et al. 2020).

There are two thresholds that need to be defined: time and space (timeThreh & distThreh). After analyzing the data according to the scale of the venue and the fine-grained basic cognition and control of the time of pedestrians' activity, we suppose 10 min and 20 min can be used as two thresholds to detect stay points.

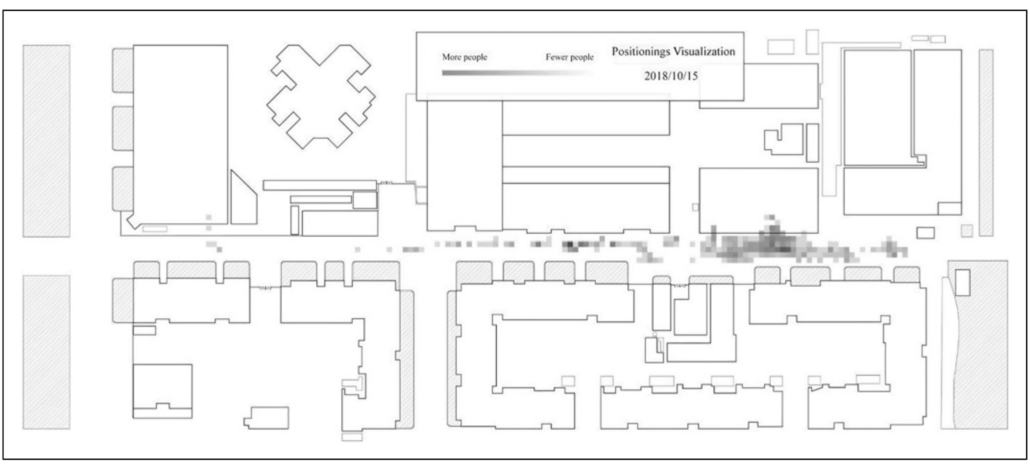

Figure 4. Visualization of positioning: daily people density heatmap (sample of Oct. 15, 2018). Each cell represents some records of people's appearances and the darker it is, the more records were captured.

Figure 5 shows the visualizations of positions and stay points. 6 different colored dots represent position points with circles representing stay points, of which the diameters shows stay time. These two graphs compare two Thursdays, on one of which there was activity during the chosen time (2: 40-2:50 p.m., Oct. 6, 2018), while on the other day, there was no event in NICE2035 Lab at all. From Figure 5 we can infer that there are more stay points during the time with activities. Also, a, b, c, e groups of people have more population as well as stay points than groups d and f, which include commute and pass-by people.

Trajectory analysis. We use trajectories to represent the movement activities of people in the street. From trajectory data, we can gain rich information about not only spatial but also social processes in urban environments.

We divided the block into eleven areas, shown in Supp Figure S9, according to the function of each part, classified the preprocessed data according to each person and hour, and regard each person's activity in 1 hour as a trajectory. From 8:00 a.m. to 20:00 p.m., we make statistics on the number and distance of the areas that the pedestrian trajectory passes through, and the results are shown in Supp Figure S1. It can be found that the trajectories for staying in identical place (max distance equals 0) and crossing all areas (max distance equals 10) are significant. So, we divide all people into three categories: single-region, whole-region, and partial-region. For each category, we also show the frequency analysis of people in each region respectively (Supp Figure S1). For the partial-region people dataset, which has both data sufficiency and diversity, we used cluster analysis to further mine the impact of the urban intervention on move events.

To perform the clustering analysis, we first performed the calculation of the affinity matrix based on the pre-processed data and obtained the similarity between every two trajectories. The similarity between trajectories is measured in two dimensions, geometric similarity, and semantic similarity. (Details are shown in supplementary material section 2).

As shown in Figure 6, we obtained four representative trajectory clusters and plotted them on a street section diagram (refer to Supp Figure S9) using edge bundling (Gansner et al., 2011). We must highlight that the C1 and C4 trajectories are clustered in the right half of the block, that is, Warehouse, Rubbish Station (Right), Residential Area (Right), and Market. This is the area with the highest frequency of move activities mentioned previously and combined with the semantics of the

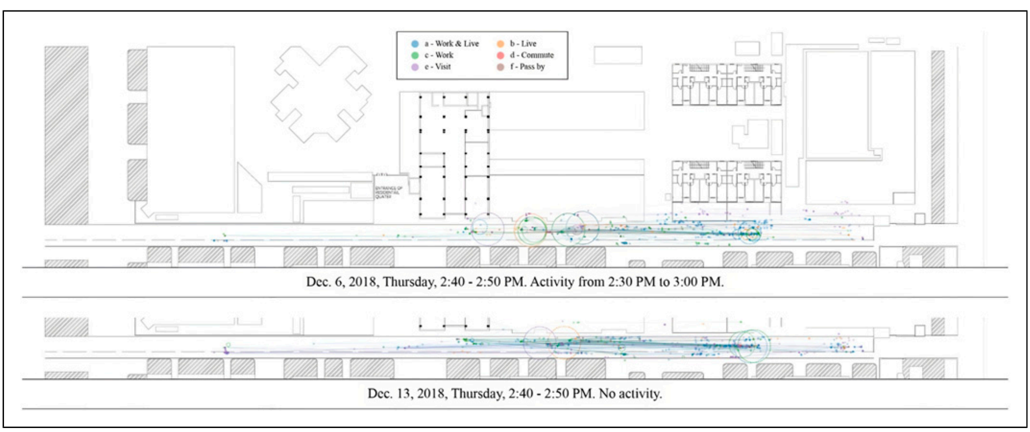

Figure 5. Location points and stay points, which infer that there are more stay points during the time with NICE2035 events. Also, a, b, c, e groups of people have more population as well as stay points than groups d and f, which include commute and pass-by people.

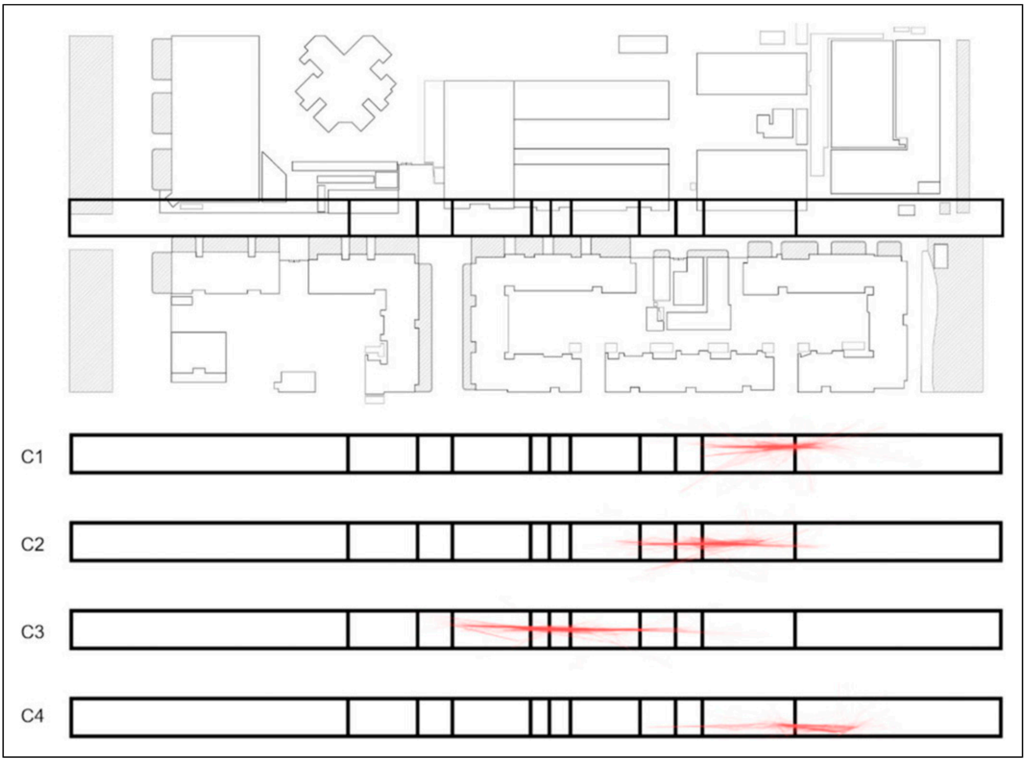

Figure 6. Four representative clusters generated by K-Medoids clustering, which shows that four different clusterings of trajectories might have associations with various types of activities in different areas.

sections, these two types of move activities are related to people's routines. As for C3 movement events, they have a very significant east-west displacement, so most of the trajectories may be related to traversing the LivingLine, such as commuting. Regarding C2, these types of movement activities pass through the Warehouse(H) and frequently reach the convenience store and housekeepers sections (E,F,G). Compared to C1 and C4, the visualization shows that the trajectories of cluster C2 are attracted to the left side of the street, and geometrically these trajectories connect the Lab section(D) and the right end of the street, which shows that C2 is the most relevant cluster for trajectories related to lab-related events. We then used statistics to confirm these correlations.

As shown in Table 2 and Figure 7, we calculated the frequency and percentage of move activities for days with NICE2035 events and days without NICE2035 events separately, according to the four representative clusters. In Figure 7, it is observed that the frequency of C1 is the highest in the routine situation. In contrast, there is a significant increase in the C2 category move activities on NICE2035 events days, overtaking C1 as the highest frequency category, which confirms the correlation between C2 and NICE2035 events.

Regression analysis

The ultimate objective of this research is to provide a methodology for understanding the impact of urban interventions on street dynamics, especially on the mobility patterns of different groups of people. In our case, we retain a special focus on whether the NICE2035 events have made a significant impact. We use a panel data regression model, a model that can capture the linear

Table 2. Comparison of the composition of move activities with and without NICE2035 events separately, according to the four representative clusters.

	Dates w/NICE2035 events		Dates w/o NICE2035 events	
Move activities type	Frequency	Percentage (%)	Frequency	Percentage (%)
C1(Routine N)	5880	32.68	27,555	33.63
C2(Lab)	6290	34.95	24,613	30.04
C3(Commuting)	4608	25.61	21,156	25.82
C4(Routine S)	1217	6.76	8602	10.50
Total	17,995	100	81,926	100

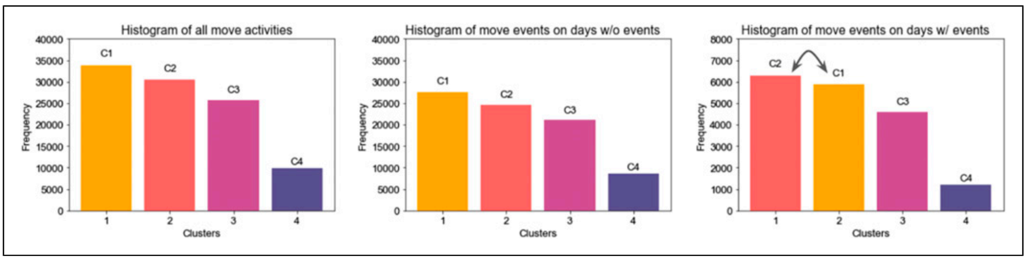

Figure 7. Histogram of move activities frequency by move activities type, which infers that the frequency of C1 is the highest in the routine situation and confirms the correlation between C2 and NICE2035 events.

relationship between multiple variables and features in panel data, to describe how different dimensions of human activities depend on several predictor variables.

Panel data, also known as longitudinal or cross-sectional time-series data, is a dataset in which the behavior of entities is observed across time. In our study, human dynamic patterns in a certain period of the day are considered unique subjects—for example, the number of devices belonging to the residents' group from 8 a.m. to 9 a.m., representing the cross-sectional dimension. Each subject is observed for multiple days over a 3-month span, representing the time series dimension. Panel data allows us to control for variables we are not interested in, like differences across hours of the day.

Equation (1) shows the equation for panel data regression in our study:

$$Yit = \beta_1 Events_{it} + \beta_2 Rain_{it} + \beta_3 Temp_{it} + \beta_4 Weekday_{it} + \beta_5 Month_{it} + \alpha_i + \mu_{it} \quad (1)$$

where $\alpha_i (i = 1...n)$ is the unknown entity-specific intercept for each entity. Yit is the dependent variable where i represents entity and t represents time. X_{it} represents an independent variable, and β_1 to β_4 are the coefficient for that independent variable. And μ_{it} is the error term.

Human dynamics may not only be influenced by urban intervention, but also by environmental factors such as rain and temperature. There are several time-variant dummy variables included as regressors: $Events_{it}$ denotes whether an NICE2035 event was happening at time t for subject i, and $Rain_{it}$ and $Temp$ represents the weather condition at the time. And because the data over a 3-month span has an obvious time-series characteristic, the human dynamic will naturally fluctuate between weekdays and weekends and vary between different months, which is not the focus of this study. A month dummy $Month_{it}$ and weekday dummy $Weekday_{it}$ are introduced to absorb the effects related to months and weekdays.

The model is fitted by finding the values of the parameters β_1 to β_5 so that the sum of squared estimation errors is minimized. Panel data regression can provide intuitive insights into the relationship between variables, including p-values, which indicate the level of statistical significance of a relationship. And R-Square is a measure of how well a linear regression model fits the observed data.

Correlation with populations. From the regression result, we find that NICE2035 events may have a certain positive impact on the number of residents and workers, but not significantly (Table 3). This means it still does not retain sufficient attraction for people who have already been working or living in this street. However, we found that NICE2035 events cause a significant increase in commuters, visitors, and passersby. The most obvious impact is on visitors, where the average number of visitors per hour will increase by 6.5 during the NICE2035 events period. This number was 2.26 for commuters and 2.95 for passersby. This shows that the NICE2035 events may have a major activating effect for attracting people outside of the street. They are the main target groups of the activities, and the result shows that the NICE2035 events have been an important reason for their visit.

In addition, we also found that rainy days have a strong negative correlation with workers, visitors, and passersby. This is also instinctive because the old alleys are not easy to traverse on rainy days. And compared to weekends, there will be more resident workers, workers, and commuters every hour on weekdays, while the number of residents is significantly reduced because they will leave the community to work outside, which also aligns with common sense.

Correlation with stay points. We detected the stay points from 8:00 to 20:00 for all 93 days of observation data, and analyzed the data as follows in Table 4:

The NICE2035 events have a positive impact on pedestrian activities and are directly reflected in the stay points. During the event period, the number of stay points for all groups of people has an increasing tendency. What is more significant is that the average number of stay points for residents per hour will increase by 4.040, and the number of visitors will increase by 2.973. The result shows that the NICE2035 events may have attracted these two groups of people to stay.

Correlation with trajectories. We counted the number of trajectories from 8:00 to 20:00 in 93 days and divided the trajectories into two parts: laboratory-related trajectories and laboratory-unrelated trajectories.

For the laboratory-unrelated trajectories (Table 5), we can see that resident workers, commuters and visitors are greatly affected by NICE2035 events, and the number of trajectories has increased significantly.

For laboratory-related trajectories (Table 6), the data shows that the impact of NICE2035 events on commuters is not significant. Resident workers and visitors are still greatly affected by NICE2035 events.

Interpretation on analysis result. Our regression analysis indicates that NICE2035 events have a limited impact on the number of residents and workers, but significantly increase the number of commuters, visitors, and passersby. Notably, the effect on the number of visitors is particularly significant, with an average hourly increase of 6.5 people, and there is an increasing trend in the duration of stay for both residents and visitors. Additionally, there is a significant increase in trajectories for resident workers, commuters, and visitors.

These findings suggest that NICE2035 events have the ability to attract more external populations to visit and stay. As Noyman et al. (2019) suggested, this staying behavior is an indication

Table 3. Regression results of population which reflect that NICE2035 events may have a certain positive impact on the number of residents and workers, but not significantly.

	Dependent variable					
	Resident workers	Residents	Workers	Commuters	Visitors	Passersby
Factor(events)1	1.292 (1.348)	2.685 (2.434)	0.685 (3.172)	2.262 (1.156)*	6.543 (2.582)**	2.951 (1.535)*
Factor(rain)1	0.458 (0.503)	2.873 (0.909)***	−6.228 (1.184)***	−0.729 (0.432)*	−8.847 (0.964)**	−4.184 (0.573)***
Temp	0.066 (0.074)	0.099 (0.133)	−0.414 (0.174)**	−0.100 (0.063)	−0.422 (0.141)***	−0.595 (0.084)***
Factor(weekday)1	3.343 (0.499)***	−13.415 (0.901)***	17.024 (1.174)***	4.137 (0.428)***	1.622 (0.956)*	−0.865 (0.568)
Factor(month) November	−0.655 (0.764)	12.061 (1.379)***	14.837 (1.797)***	2.498 (0.655)***	7.780 (1.463)***	2.535 (0.870)***
Factor(month) October	−6.769 (1.002)***	6.857 (1.809)***	10.350 (2.358)*	1.980 (0.859)**	15.034 (1.920)***	1.782 (1.141)
Observations	1157	1157	1157	1157	1157	1157
R^2	0.163	0.493	0.560	0.748	0.431	0.269
Adjusted R^2	0.149	0.485	0.553	0.744	0.422	0.258
F statistic (df = 6;1138)	12.285***	61.416***	80.430***	187.763***	47.985***	23.279***

Note. *$p < .1$; **$p < .05$; ***$p < .01$.

Table 4. Regression results of stay points showing that the NICE2035 events have a positive impact on pedestrian activities and are directly reflected in the stay points.

	Dependent variable											
	StayPoint_a	StayPoint_b	StayPoint_c	StayPoint_d	StayPoint_e	StayPoins_f	AveStay Time_a	AveStay Time_b	AveStay Time_c	AveStay Time_d	AveStay Time_e	AveStay Time_f
Factor (events)1	1.451 (1.002)	4.040 (1.318)***	1.851 (1.281)	0.662 (0.188)***	2.973 (0.768)***	0.281 (0.153)*	−0.053 (0.017)***	–0.004 (0.014)	0.006 (0.016)	0.152 (0.075)**	−0.024 (0.022)	0.200 (0.071)***
Factor (rain)1	1.065 (0.374)***	0.962 (0.492)*	−3.231 (0.478)***	−0.439 (0.070)*	−2.602 (0.287)**	−0.235 (0.057)***	0.003 (0.006)	−0.010 (0.005)*	−0.007 (0.006)	−0.118 (0.028)***	−0.025 (0.008)***	−0.108 (0.026)***
Temp	−0.138 (0.055)**	−0.183 (0.072)**	−0.197 (0.070)***	−0.003 (0.010)	0.004 (0.042)	−0.021 (0.008)**	0.00005 (0.001)	−0.002 (0.001)**	0.001 (0.001)	0.0003 (0.004)	0.002 (0.001)	−0.006 (0.004)
Factor (weekday)1	2.621 (0.371)***	−3.577 (0.488)***	6.070 (0.474)***	0.285 (0.070)***	−0.181 (0.284)	0.047 (0.057)	−0.003 (0.006)	0.024 (0.005)***	0.010 (0.006)*	0.115 (0.028)***	−0.009 (0.008)	−0.017 (0.026)
factor(month) November	1.818 (0.568)***	5.367 (0.747)***	3.634 (0.726)***	−0.084 (0.107)	−0.354 (0.435)	0.042 (0.087)	0.011 (0.010)	0.005 (0.008)	−0.024 (0.009)**	−0.028 (0.042)	−0.018 (0.013)	0.015 (0.040)
factor(month) October	−1.580 (0.745)**	–0.610 (0.980)	0.074 (0.952)	0.012 (0.140)	2.170 (0.571)***	−0.091 (0.114)	0.045 (0.013)***	0.001 (0.010)	−0.022 (0.012)*	−0.051 (0.056)	−0.019 (0.017)	−0.011 (0.052)
Observations	1157	1157	1157	1157	1157	1157	1157	1157	1157	1157	1157	1157
R^2	0.133	0.167	0.196	0.060	0.158	0.040	0.045	0.032	0.011	0.033	0.011	0.028
Adjusted R^2	0.119	0.154	0.184	0.045	0.145	0.025	0.030	0.016	-0.004	0.017	-0.004	0.013
F statistic (df = 6;1138)	29.051***	37.992***	46.302***	12.066***	35.706***	7.956***	8.881***	6.221***	2.186**	6.426***	2.162**	5.445***

a - resident workers; b - Residents; c -Workers; d - Commuters; e - Visitors; f - Passersby.
Note *p < .1; **p < .05; ***p < .01.

Table 5. Regression results of laboratory-unrelated trajectories, which shows that resident workers, commuters, and visitors are greatly affected by NICE2035 events, and the number of trajectories has increased significantly.

	Dependent variable					
	Resident workers	Residents	Workers	Commuters	Visitors	Passersby
Factor(events)1	3.843 (1.748)**	1.604 (1.192)	0.762 (0.899)	0.261 (0.104)**	2.603 (0.667)***	0.037 (0.101)
Factor(rain)1	−0.731 (0.652)	0.395 (0.445)	−1.215 (0.336)***	−0.103 (0.039)***	−1.644 (0.249)***	−0.054 (0.038)
Temp	−0.248 (0.096)***	−0.116 (0.065)*	−0.109 (0.049)**	0.002 (0.006)	0.059 (0.037)	−0.005 (0.006)
Factor(weekday)1	2.129 (0.647)***	−2.399 (0.441)***	2.927 (0.333)***	0.098 (0.038)**	−0.147 (0.187)	0.025 (0.037)
Factor(month) November	−1.261 (0.990)	1.070 (0.676)	−0.138 (0.509)***	0.003 (0.059)	−1.362 (0.378)***	−0.014 (0.057)
Factor(month) October	−12.363 (1.299)***	−3.554 (0.886)***	−3.172 (0.668)***	−0.027 (0.077)	−1.879 (0.496)***	−0.131 (0.075)*
Observations	1157	1157	1157	1157	1157	1157
R^2	0.314	0.284	0.251	0.322	0.192	0.305
Adjusted R^2	0.303	0.273	0.239	0.312	0.179	0.294
F statistic (df = 6;1138)	28.982***	25.088***	21.159***	30.089***	14.979***	27.764***

Note. *$p < .1$; **$p < .05$; ***$p < .01$.

Table 6. Regression results of laboratory-related trajectories, which show the impact of NICE2035 events on commuters is not significant, and resident workers and visitors are still greatly affected by activities.

	Dependent variable					
	Resident workers	Residents	Workers	Commuters	Visitors	Passersby
factor(events)1	2.856 (0.881)***	0.966 (0.622)	0.390 (0.582)	0.048 (0.083)	1.661 (0.512)***	0.073 (0.079)
Factor(rain)1	−1.026 (0.329)***	0.024 (0.232)	−0.718 (0.217)***	0.015 (0.031)	−0.932 (0.191)***	−0.029 (0.030)
Temp	−0.219 (0.048)***	−0.151 (0.034)***	−0.080 (0.032)**	0.001 (0.005)c	−0.090 (0.028)***	0.001 (0.004)
Factor(weekday)1	2.057 (0.326)***	–0.827 (0.230)***	2.343 (0.215)***	0.074 (0.031)**	0.582 (0.189)***	0.021 (0.029)
Factor(month) November	−0.849 (0.499)*	0.813 (0.352)**	0.452 (0.330)	0.020 (0.047)	0.391 (0.290)	0.023 (0.045)
Factor(month) October	−7.680 (0.655)***	−2.259 (0.462)***	−2.143 (0.433)***	0.024 (0.061)	1.290 (0.380)***	0.115 (0.059)*
Observations	1157	1157	1157	1157	1157	1157
R^2	0.436	0.297	0.319	0.201	0.189	0.251
Adjusted R^2	0.427	0.286	0.308	0.189	0.176	0.239
F statistic (df = 6;1138)	48.844***	26.751***	29.562***	15.953***	14.743***	21.166***

Note. *$p < .1$; **$p < .05$; ***$p < .01$.

of higher social activity and interaction, which means NICE2035 events promote social engagement and interactions, thereby enhancing urban vibrancy.

Conclusion

This study proposes a comprehensive, extensible evidence-based methodology to study street-scale human dynamics, which are crucial for successful urban intervention design. Although the correlations we found between physical changes/events and urban vibrancy are not always significant, and the data accuracy of the model needs to be improved, this study has demonstrated the capability of understanding real-time human dynamics on the street, identifying correlations between urban intervention and street dynamics in both temporal and spatial dimensions. The data used in the case study of LivingLine Shanghai were purely based on Wi-Fi probe data, which is low-cost and easy to collect. The low cost of this approach enabled it to be implemented on a large scale. Additionally, the sound accuracy of Wi-Fi sensing makes it a powerful tool to study fine-grained real-time human behavior. Finally, there are two findings that we want to highlight as a validation of the methodology used: First, as explained, our proposed methodology successfully captured fine-grained temporal (by minutes) and spatial patterns of urban vibrancy that happened on Siping Street, Shanghai. Second, as expected, the density and diversity of laboratory-unrelated users are significantly correlated with the weather, weekday, holiday, and urban interventions.

This study has several limitations that should be considered. First, there might be a bias that most data is from younger generations, since the elders might don't use cell phones or smartphones; thus, the Wi-Fi probe method doesn't work for them. Besides, the smartphone operation system might change MAC address to keep user's privacy, which might create bias in the data collected in this method. Lastly, there is room for further improvement in the positioning accuracy of Wi-Fi, and it would be valuable to explore some cutting-edge approaches (Shen et al., 2016) to achieve this.

In a conclusion, with minor limitations, the methodology we proposed is yet one of the most promising and insightful methods for conducting urban measuring street-level urban vibrancy and understand its correlation with urban interventions without compromising users' data privacy. Our method has potential applications in analyzing the effects of fine-grained urban interventions on urban vibrancy, as well as providing quantitative evidence for space usage evaluation.

As the next steps, we plan to elevate our research by utilizing LivingLine as an urban living lab. We will gain insights from the daily urban dynamics of a traditional street, where we can implement our indoor–outdoor collaborative method to study people's travel between different types of spaces and their interactions with various objects. This will allow us to further explore our findings and demonstrate the potential applications of our approach in real-world settings.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

References

Barreca A, Curto R and Rolando D (2020) Urban vibrancy: an emerging factor that spatially influences the real estate market. *Sustainability* 12(1): 346.

Bellomo N, Clarke D, Gibelli L, et al. (2016) Human behaviours in evacuation crowd dynamics: from modelling to "big data" toward crisis management. *Physics of Life Reviews* 18: 1–21.

Cai X, Cao Y and Dong P (2020) Velocity-based trajectory stop point recognition algorithm. *Computer Systems and Applications* 29(4): 214–219.

Calabrese F, Reades J and Ratti C (2010) Eigenplaces: segmenting space through digital signatures. *IEEE Pervasive Computing* 9(1): 78–84.

Chilipirea C (2019) *Crowd Data Analytics as Seen from Wifi: A Critical Review*. Doctor of Philosophy. University of Twente.

Gansner ER, Hu Y, North S, et al. (2011, March) Multilevel agglomerative edge bundling for visualizing large graphs. In: *2011 IEEE Pacific Visualization Symposium*. IEEE, pp. 187–194.

Griffiths R (1995) Cultural strategies and new modes of urban intervention. *Cities* 12(4): 253–265.

Hu X, Shen P, Shi Y, et al. (2020) Using Wi-Fi probe and location data to analyze the human distribution characteristics of green spaces: a case study of the Yanfu Greenland Park, China. *Urban Forestry & Urban Greening* 54: 126733.

Huang B, Zhou Y, Li Z, et al. (2019) Evaluating and characterizing urban vibrancy using spatial big data: Shanghai as a case study. *Environment and Planning B: Urban Analytics and City Science* 47(9): 1543–1559.

Huang B, Zhou Y, Li Z, et al. (2020) Evaluating and characterizing urban vibrancy using spatial big data: Shanghai as a case study. *Environment and Planning B: Urban Analytics and City Science* 47(9): 1543–1559.

Jacobs J (1961) *Death and Life of Great American Cities*. New York: Random House.

Jacobs A (1993) *Great Streets*. Cambridge, Mass: MIT Press.

Jiang C, Xiao Y and Cao H (2020) Co-creating for locality and sustainability: design-driven community regeneration strategy in shanghai's old residential context. *Sustainability* 12(7): 2997.

Kim Y (2018) Seoul's Wi-Fi hotspots: Wi-Fi access points as an indicator of urban vitality. *Computers, Environment and Urban Systems* 72: 13–24.

Kontokosta C and Johnson N (2017) Urban phenology: toward a real-time census of the city using Wi-Fi data. *Computers, Environment and Urban Systems* 64: 144–153.

Lang W, Hui EC, Chen T, et al. (2020) Understanding livable dense urban form for social activities in transit-oriented development through human-scale measurements. *Habitat International* 104: 102238.

Li Q, Zheng Y, Xie X, et al. (2008) Mining user similarity based on location history. In: GIS '08: 16th International Symposium on Advances in Geographic Information Systems.

Liu Y, Cheng D, Pei T, et al. (2019) Inferring gender and age of customers in shopping malls via indoor positioning data. *Environment and Planning B: Urban Analytics and City Science* 47(9): 1672–1689.

Long Y and Huang C (2019) Does block size matter? The impact of urban design on economic vitality for Chinese cities. *Environment and Planning B: Urban Analytics and City Science* 46(3): 406–422.

Long Y and Zhou Y (2016) Quantitative evaluation on street vibrancy and its impact factors: a case study of Chengdu. *New Architecture* 164: 52–57.

Magalhaes L (2021, December) The challenges of urban data ethics: big data and intelligent decision support systems. In *International Forum on Digital and Democracy. Towards a Sustainable Evolution*.

Noyman A, Doorley R, Xiong Z, et al. (2019) Reversed urbanism: inferring urban performance through behavioral patterns in temporal telecom data. *Environment and Planning B: Urban Analytics and City Science* 46(8): 1480–1498.

Prentow T, Ruiz-Ruiz A, Blunck H, et al. (2015) Spatio-temporal facility utilization analysis from exhaustive WiFi monitoring. *Pervasive and Mobile Computing* 16: 305–316.

Resch B and Szell M (2019) Human-centric data science for urban studies. *ISPRS International Journal of Geo-Information* 8(12): 584.

Shen J, Cao J, Liu X, et al. (2016) Feature-based room-level localization of unmodified smartphones. In: Smart City 360°: First EAI International Summit, Smart City 360°, Bratislava, Slovakia and Toronto, Canada, October 13–16, 2015. Revised Selected Papers 1. Springer International Publishing, pp. 125–136.

Shen J, Cao J and Liu X (2020) BaG: Behavior-aware group detection in crowded urban spaces using WiFi probes. *The World Wide Web Conference* May: 1669–1678.

Williams S, Ahn C, Gunc H, et al. (2019) Evaluating sensors for the measurement of public life: a future in image processing. *Environment and Planning B: Urban Analytics and City Science* 46(8): 1534–1548.

Xia S, Liu Y, Yuan G, et al. (2017) Indoor fingerprint positioning based on Wi-Fi: an Overview. *ISPRS International Journal of Geo-Information* 6(5): 135.

Ye Y, Li D and Liu X (2017) How block density and typology affect urban vitality: an exploratory analysis in Shenzhen, China. *Urban Geography* 39(4): 631–652.

Zhang X, Shen Y and Zhao J (2021) The mobility pattern of dockless bike sharing: a four-month study in Singapore. *Transportation Research Part D: Transport and Environment* 98: 102961.

Zhong J, Li Z, Sun Z, et al. (2020) The spatial equilibrium analysis of urban green space and human activity in Chengdu, China. *Journal of Cleaner Production* 259: 120754.

Yan Zhang (Ryan) is a Ph.D. student and research assistant in the City Science research group at the MIT Media Lab. His current research includes 1) urban big data analytics using cutting edge AI technologies such as deep learning and computer vision; 2) developing and deploying CityScope – an urban decision support system augmented by AI, AR, and TUI (tangible user interface) that democratizes the urban decision-making process.

Chengliang Li is an interdisciplinary designer and researcher. His interests include city science, computational design, data visualization, artificial intelligence, blockchain.

Jiajie(Chance) Li is a graduate student and research assistant in the City Science research group at the MIT Media Lab. He focuses on using computing and sensing technologies to gain insight into cities. Jiajie is currently working with MIT City Science Group on using Decentralized Autonomous Organization(DAO) to promote pro-social, pro-sustainable behaviors in cities.

Zhiyuan Gao is a computer science master's student at the University of Southern California, specializing in computer graphics, physics-based simulation, and deep learning. He has previously interned at the Tongji-MIT City Science Lab, where he worked on pedestrian pattern mining, and at the State Key Laboratory of Disaster Reduction in Civil Engineering, focusing on structural health monitoring.

Tianyu Su is an urban researcher and entrepreneur at the intersection of urban, human, and technology, with a hope to shape better cities by a deeper understanding of society and human behaviors.

Can Wang currently is a postdoc of Tongji University (China) and MIT (US), his research is focusing on spatial behavior analysis and pedestrian simulation.

Hexin Zhang is a Master in Interaction Design student at Tongji University and a student researcher in Tongji-MIT City Science Lab. In the LivingLine Shanghai project, he mainly works on Wi-Fi data collection, analysis, and visualization.

Teng Ma is a PhD. student in State Key Lab of Disaster Reduction in Civil. His research focus on using data-driven method to prevent disasters.

Yang Liu is a visiting scholar at MIT Media Lab, with research interests in urban data analysis, simulation, and artificial intelligence-aided design decision systems (interdisciplinary research in architecture/urban design/computer science/touchable human-computer interaction).

Weiting Xiong is a visiting student at MIT City Science and a Ph.D candidate from the School of Architecture, Southeast University, China.

Ronan Doorley is an engineer and data scientist, specializing in spatial analytics and high-resolution modeling of human behavior in cities. As a Research Scientist with the MIT City Science Group, he builds models and simulations to understand and predict urban phenomena such as traffic congestion, accessibility, spatial mismatches, and environmental impacts. Ronan earned his Bachelor in Mechanical Engineering and PhD in Transportation Engineering from Trinity College, Dublin. He is the recipient of awards such as the Fulbright Scholarship, the Trinity College Gold Medal, and the ICE Reed and Mallik Medal.

Luis Alonso is a research scientist in the City Science group and Principal Investigator of the Andorra Living Lab Project. He has a PhD in architecture and coordinates the Andorra Living Lab project. He oversees the integration of the group's diverse research topics (energy consumption, city simulation, urban mobility, innovation district, and smart housing) in City Science Network of Collaborative Cities, to provide comprehensive solutions for urban and country challenges, in order to transform cities into more diverse and vibrant "shuman scale ecosystems."

Lou Yongqi, PhD, is Dean of the College of Design and Innovation at Tongji University in Shanghai, Fellow of Royal Swedish Academy of Engineering Sciences (IVA), and Vice President of China Industrial Design Association(CIDA).

Kent Larson directs the City Science (formerly Changing Places) group at the MIT Media Lab. His research focuses on developing urban interventions that enable more entrepreneurial, livable, high-performance districts in cities. To that end, his projects include advanced simulation and augmented reality for urban design, transformable micro-housing for millennials, mobility-on-demand systems that create alternatives to private automobiles, and Urban Living Lab deployments in Hamburg, Andorra, Taipei, and Boston. Larson and researchers from his group received the "10-Year Impact Award" from UbiComp 2014. This is a "test of time" award for work that, with the benefit of hindsight, has had the greatest impact over the previous decade.