



Discovering the influence of facility distribution on lifestyle patterns in urban populations

Chao Fan^{a,*}, Fangsheng Wu^b, Ali Mostafavi^{a,**}

^a Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, 77843-3136, USA

^b Department of Computer Science and Engineering, Texas A&M University, College Station, TX, 77843-3112, USA

ARTICLE INFO

Keywords:

Human mobility
Population lifestyles
Facility distribution
Urban computing

ABSTRACT

The spatial structures of cities defined by population distribution, distribution of facilities, and mobility have a significant impact on lifestyles of residents and their wellbeing. In this study, we analyze millions of mobile phone data points to infer significant sequences of visited facilities by individuals, cluster people with similar patterns of life activity sequences, and define lifestyles based on the patterns in each cluster. We find that lifestyles of a large number of people can be captured using a small set of activity sequences, while a small portion of populations have lifestyles with a variety of activity sequences. Facility proximity in spatial constraints is positively correlated with the volume of human movements, and is a significant factor in formation of the majority of lifestyle patterns. Differences in facility proportions between two neighborhoods contribute to cross-neighborhood travels for life activities, but its effect could be mediated by the geographical distances between neighborhoods. Our findings demonstrate that the widely studied scaling laws in these areas are not independent but rather connected through a deeper underlying reality, which has important implications for urban planning and city management policies to enhance equal accessibility.

1. Introduction

City dwellers worldwide suffer from unequal access to services (Hong et al., 2021), traffic congestion (Saber et al., 2020) and disease spread (Fan et al., 2021). These challenges primarily arise due to the distribution of urban facilities for people's basic necessities and transportation systems for human activities (Chang et al., 2021). Equitable planning of public services and infrastructure facilities may hold the key to support different needs of people and overcome challenges faced by residents (Jackson, 2003). Planning of urban facilities requires a deep understanding of spatial structures of cities at the nexus of populations, activities, and facilities (Yao et al., 2018; Yuan et al.,). For the past several years, human activity patterns have been extensively investigated (Wang et al., 2018), and important contributions have been made in statistical models, such as the universal scaling laws (Brockmann et al., 2006). Existing highly abstracted probability models (Song et al., 2010; Yan et al., 2017), however, are limited in explaining the complex phenomenon underlying the interactions between population and facilities in urban spatial networks. Given the heterogeneity of factors that influence population's activity patterns (Alexander et al., 2015), it is

crucial to advance our understanding of the spatial structure of cities and its impacts on human life activity patterns.

Many studies have explored the scope of human activities in a variety of aspects. A remarkable research discovery related to activity patterns is the spatial probability distribution of human trajectories characterized by jump sizes and waiting times (González et al., 2008). Deriving from the basic laws of human motion, a collection of studies (Gallotti et al., 2016; Hu et al., 2011; Zhao et al., 2014) has been simulated to model and predict human activities. For example, Yan et al. proposed a universal model combining the memory effect with population-induced competition to enable accurate prediction of individual and collective activity patterns, including scaling behaviors and trajectory motifs (Yan et al., 2017). Sun et al. showed scaling properties in the structuring of physical encounters from the spatial-temporal patterns in collective daily regularities of activities on public transportation (Sun et al., 2013). With the availability of detailed call records and location-based social platforms, the potential exists to exploit data-driven activity prediction models, such as gravity models (Zipf, 1946), intervening opportunity models (Noulas et al., 2012; Stouffer, 1940), radiation models (Ren et al., 2014) and social-based models (Wang et al., 2019). Leveraging

* Corresponding author.

** Corresponding author.

E-mail addresses: chfan@tamu.edu (C. Fan), amostafavi@civil.tamu.edu (A. Mostafavi).

<https://doi.org/10.1016/j.dibe.2024.100348>

Received 23 April 2023; Received in revised form 30 October 2023; Accepted 22 January 2024

Available online 1 February 2024

2666-1659/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

large sets of empirical data, these models provide important insights regarding movement distance (Yan et al., 2014), commuter activities in large transportation networks (Ren et al., 2014), and population flux across regions. Despite the prevalence of computational models, activity patterns are purely drawn from geographical data records of populations. Without sufficient evidence related to the interactions between human activities and other elements in cities (such as facilities), the understanding on tendencies and frequencies of human activities is still limited.

To obtain a comprehensive understanding of human activity patterns, recent studies (Barbosa et al., 2018) have explored the relationships between human activities and other factors, such as social networks (Wang et al.), job-imposed programs (Rivera et al., 2010) and activity memories (Liu and Yan, 2020). The interplay between social ties and human activities is could have profound outcomes and consequences in broad domains. For example, using the activity and communication records of mobile phone users, a prior study showed that the similarity between two individuals' activities strongly correlates with their proximity in the social network (Wang et al.). Further, Deville et al. derived a scaling relationship connecting human activity and communication patterns through the finding of spatial constraints in both behaviors (Deville et al., 2016). In addition to mobile phone data, empirical studies on location-based social media data also reveal the possibility of explaining human activities with social relationships (Gao et al., 2013). A typical example is Cho et al.'s work (Cho et al.,). The study finds that humans experience a combination of geographically limited periodic activities and seemingly random jumps correlated with their social networks. Although these studies have uncovered the influences of social factors on human mobility, they are only loosely related to specific movements. Hence, the findings obtained from these studies tend to be limited in guiding planning for public services and facilities in cities.

The spatial structures of cities defined by geography, facilities, and demographics have been emphasized in existing studies (Louail et al., 2015). Urban areas are composed of facilities, such as restaurants, financial institutions, and grocery stores, which contribute to their functionality. Population flows constitute the backbone of the spatial networks of cities by connecting facilities and locations (Louail et al., 2015). Human life activities occur between urban facilities to obtain necessities and desired activities, such as food, health care, and entertainment (Qin et al., 2012). Recent studies (Zeng et al., 2017) have attempted to characterize human activities between urban facilities. Examples of studies include the development of algorithms to predict the life activities in points of interest (POIs) (Lu et al., 2016), discovering regions of different functions using mobility entries between regions and POIs (Yuan et al.,), and modeling the demand of POIs by exploring the daily needs of people identified from their large-scale mobility data (Liu and Yan, 2020). These prior studies propose methods to explain the characteristics of human activities between POIs, fueling concomitant advances in areas of both human activities and urban facilities (Hanson and Schwab, 1987). Owing partly to progress enabled by this research, a number of recent studies (Wang et al., 2017) further leverage the movements among facilities to develop technologies for predicting the next facility visit in the activity sequence of an individual. For example, Yao et al. proposed a method by incorporating the degree of temporal matching between individuals and POIs into personalized POI recommendations (Yao et al., 2018). In addition, a variety of deep learning models, such as convolutional networks with variational attention mechanism (Gao et al.) and recurrent neural networks (Gao et al.), are employed to learn the historical movement behaviors for next-POI prediction. The increasing availability of large-scale location-based datasets enhances the resolution and predictability of human movements across facilities. But few existing studies have quantified the relationships between facility distribution and human sequential activities, using large-scale population activity data. Hence, a quantitative understanding of the extent to which the distribution of urban

facilities shapes human activity patterns in their regular lives is still missing.

To this end, building on rich and fundamental literature, we uncover diverse lifestyle patterns of collective activities extracted from mobile phone data based on the temporal order of visits to urban facilities (defined as human life activity sequences). The resulting lifestyle clusters of people and the patterns of life activity sequences are further associated with our proposed metrics of facility distribution. To explain the influence of facility distribution on human life activities, we developed and tested multivariate models in each cluster. The presented work provides insights regarding the extent to which urban facilities can shape and alter the way of human life activities, implying interpretable behavioral patterns of the population at large. To test the generalizability of the findings, we apply computational experiments on mobile phone data collected from 12 U.S. metropolitan counties, including Los Angeles, Harris (Houston), King (Seattle) and Cook (Chicago). Theoretical contributions related to the drivers of human mobility and practical applications regarding urban planning strategies drawn from the findings are identified and discussed.

2. Data and methods

2.1. Dataset collection

We analyzed mobile phone data provided by X-mode, Inc., a data company that collects anonymized location data from numerous applications from a great number of mobile phone devices (X-mode). X-Mode works with more than 70 developers of more than 300 applications, using their proprietary software development kit (SDK) technology. Globally, more than 50 million active people per month are sharing their location every 5–7 min. The dataset used in this study consists of anonymized location data in 12 metropolitan counties in the United States in February 2020, including King County (Seattle) in Washington, Los Angeles County (Los Angeles) in California, Dade County (Miami) in Florida, and Dallas County (Dallas) in Texas. There are three main reasons to consider in selecting the 12 metropolitan counties. First, population size is an important consideration as the study would like to reveal the patterns that can be representative for a large population. These metropolitan counties are in the areas of highly populated cities such as Houston, Los Angeles, and Seattle. Hence, the findings from these counties can be representative for most urban populations. Second, the selected counties could be geographically heterogeneous so that the results of this study can account for the spatial heterogeneity and improve its representativeness. The selected counties cover the east, north, west, south and middle of the U.S. Finally, the selected counties should allow the residents to have all defined lifestyle activities. Based on our experiments, we find that the lifestyle activities identified from our study are observed in all these counties. This criterion makes the findings generalizable across different counties in the U.S. These considerations in selecting the counties in our study could ensure the results and findings are representative and generalizable across the country. (See more details in Supplementary Information.) The data sample has a wide set of attributes, including anonymized device ID, latitude, longitude, speed and UTC (Coordinated Universal Time) time of observation. This dataset allows us to capture the interaction of populations with locations (places) based on when and where a device was and how long it spent in a specific place.

The data was shared under a strict contract with X-mode through their academic collaborative program in which they provide access to de-identified and privacy-enhanced mobility data for academic research. All researchers processed and analyzed the data under a non-disclosure agreement and were obligated to not share data further or to attempt to re-identify data. We acknowledge the issues such as biases and errors might be present in our mobile phone data. First, the representativeness of the mobile phone data has been verified by many existing studies which ensure the data could reflect the actual residents

in different neighborhoods, which ensure the validity of the data. Second, in preprocessing the data, we remove the devices that do not have consistent data points every day in the study period. These errors may be because the misdetection and location of the GPS signals. Removing these data points would allow us to have a high-quality data set that can well represent the trajectories of the people.

The place data was provided by SafeGraph, Inc., a location intelligence data company that builds and maintains accurate POI data and store locations within the United States (SafeGraph,). The dataset includes basic information, such as POI IDs, location name, address, category, and brand association for POIs where people spent time and money. SafeGraph, Inc. has collected data for about 6.5 million active POIs. In this study, we primarily adopted the coordinates of POIs and their categories. SafeGraph labels each POI with its NAICS (North American Industry Classification System) category code (United States Census Bureau), which is the standard used by Federal statistical agencies in classifying business establishments. In this study, we used a four-digit NAICS code which categorizes POIs at the industry group level. SafeGraph Data Consortium shared the data through their academic program, which provides data at no charge to the academic community. The program allows researchers to publish papers and uncover insights that will lead to better policy-making, new innovations, and business growth.

To further identify the geofence of each POI, we employed the U.S. building geometry data from Microsoft US Building Footprints (Microsoft,). This dataset contains 125,192,184 computer-generated building footprints (polygon geometries) in all 50 US states. This data is licensed by Microsoft under the Open Data Commons Open Database License. We correlated the coordinates of POIs from SafeGraph to the polygons of the buildings to identify the geofence of each POI. There might be a case, like malls and plazas, in which one building consists of multiple POIs, leading to challenges of specifying the geofences for each POI. In this

case, hence, we have to associate the building polygon with all these POIs. Since such cases are rare in this study, our findings will not be affected. The geofences of the POIs can further assist to capture the location of a mobile device.

The polygon and demographic data for census block groups (CBGs) were obtained through the American Community Survey (ACS) Census 5-year 2014–2018 dataset (United States Census Bureau). This dataset is the premier source for detailed population and geographical information about United States communities. The data is publicly available through the US Census Bureau.

2.2. Human lifestyle characterization

The first step of this study is to characterize the lifestyles of populations based on their visits to POIs. Algorithms to detect human lifestyle activities would take a large volume of computer memory. To balance the computational cost and validity of the results, this study focuses mainly on individuals who generated daily mobile phone records in February 2020 in each county. For all entries of each individual device, we located coordinates in a chronological sequence within the geofences of POIs. The records within the geofences are labeled by the POI IDs and their four-digit NAICS category codes. By doing so, we generated the sequences of visited places of each individual device during February. Fig. 1a shows the distribution of visits to a variety of POIs based on the records in our dataset. Restaurants and other eating places, accounting for more than 20 % of visits, are the most frequent places visited. Clothing stores and offices of physicians are also frequently visited, but not as often as restaurants. In this study, we selected the POIs in the top 12 categories, which account for 90 % of population activities.

Our main goal was to uncover significant human life activity sequences hidden in normal life activities based on the records of visited

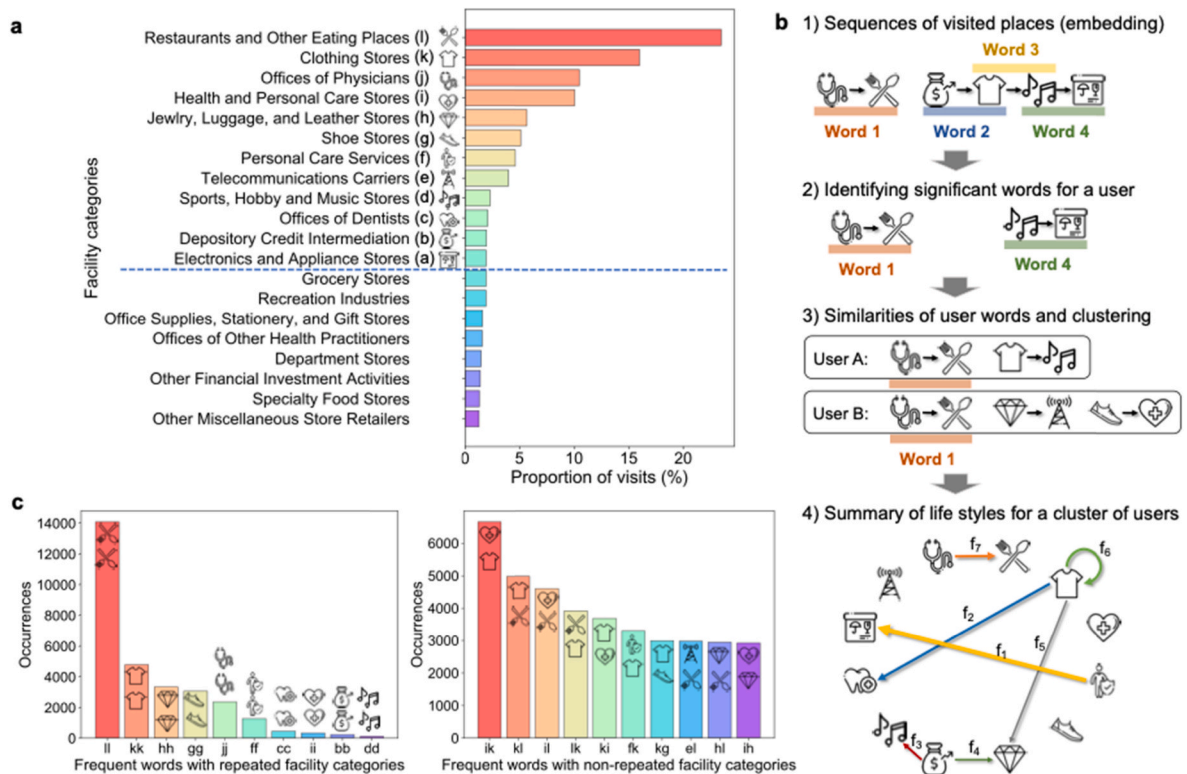


Fig. 1. Methods for identifying lifestyles and basic analyses. **a.** Proportion of visits in different categories of facilities during the study period. The types of facilities are defined by the North American Industry Classification System (NAICS). For simplicity, the categories of facilities are represented by characters. **b.** Schematic illustration of methods of identifying life activity sequences and lifestyle clusters. **c.** Examples of significant words that are frequently adopted by the populations. The left plot shows the top 10 most frequent words with repeated facilities, and the right one shows the top 10 most frequent words with non-repeated facilities.

places. Hence, the methods adopted in this study encode the sequences of visited places in a concise representation to allow processing that reveals hidden “grammar rules” and that identifies significant behavioral patterns (Fig. 1b). To this end, we first assigned a letter to each selected POI category (Fig. 1a). Then, we applied the Sequitur algorithm (Nevill-Manning and Witten, 1997) to infer a context-free grammar from a sequence of letters based on the life activity sequences of all devices in our dataset. Using the grammar rules, we generated “words” (combinations of letters) to compress the sequences of visited places for devices. The words were of letters representing a chronological sequence of continuous visited places visited in succession. Since the words are incorporated in the grammar which ensures that each word appears in the records of empirical data at least twice, the words are representations of human life activity sequences by individual devices.

We aggregated the words for each individual device to characterize people’s movements among facilities. The words are attributed to a specific individual only when they occur more frequently than corresponding words generated from a random process with the same lengths of life activity sequences. To detect significant words, we generated 1000 randomized life activity sequences for each individual device and applied the Sequitur algorithm to identify words (Di Clemente et al., 2018). The significance of a word is then calculated using Z-score:

$$Z_i = \frac{N_i^{real} - \bar{N}_i^{rand}}{\text{std}(N_i^{rand})} \quad (1)$$

where N_i^{real} is the number of words i in the life activity sequences from real-world data; N_i^{rand} is the number of words i in the life activity sequences generated from randomized processes. In calculating the Z-score Z_i for a specific word i in the life activity sequences of an individual device, we took the average and standard deviation of the N_i^{rand} from the randomized processes. Referring to an existing study (Di Clemente et al., 2018), the significant words for an individual device are the words with Z-score greater than 2. By doing so, we generated a list of significant words, $\{w_1, w_2, \dots, w_n\}$, as the representations of a device’s consecutive movement as a proxy of daily behavior.

The third step was to examine the generality of lifestyle patterns among all individuals and to identify ubiquitous characteristics of people with different lifestyle patterns. In this step, we calculated the similarity of the lifestyles for all individuals in a county. With the lists of significant words for all devices, we performed a Jaccard similarity (Niwtanukul et al., 2013):

$$J(a, b) = \frac{|a \cap b|}{|a \cup b|} \quad (2)$$

where $J(a, b)$ is the Jaccard similarity score between two individuals; a and b represent the lists of significant words for two individuals. As described by the formula, the values of the Jaccard similarity ranges from 0 to 1. Values close to 1 indicate strong similarities between the lifestyle characteristics of two individuals. Using the results from Jaccard similarity, we obtained a relation matrix which describes the similarities of life activity sequences among all individuals in a county.

Finally, we created a network of individuals using the relationship matrix. The edges between two individuals are created only if the Jaccard similarity between these two individuals is greater than a set threshold. Here, the selection of a threshold depends on the connectivity and sparsity of the network. The trade-off is made to identify appropriate number of clusters in the network. After several tests, we used a Jaccard similarity of 0.2 for the network in King County. Then, we applied the Louvain algorithm to identify communities (a.k.a., clusters) in the network (Waltman and Van Eck). The Louvain algorithm is an unsupervised algorithm that maximizes the modularity of the network and partitions the network into communities where the nodes within communities are more highly connected than the nodes between communities (Fan et al., 2020). As such, we identified multiple clusters by

grouping individuals with similar lifestyle patterns. In each cluster, we decomposed the words into pairs between two consecutive POI categories, indicating the chronological orders to visiting places in human life activities. We aggregated the pairs of POI categories for all individual in the same cluster and calculated the average frequencies of the pairs per individual (i.e., $f_1, f_2, \dots, f_7$ in Fig. 1b). Based on the average frequencies, we obtained a matrix of the patterns of life activity sequences for each cluster of individuals (Supplementary Information), considered as a representation of the human lifestyle. In the Sequitur algorithm where we aim to obtain the sequence of activities of the humans, we set the sequence of activities to be two. We split the higher order activities and the sequence of two activities to represent the behavioral patterns of the humans, which can enhance the generalizability of this study and findings. The parameters for Jaccard similarity and Louvain algorithms are determined by our experiments. Based on existing studies and the consideration of generalizability, we consider the number of life patterns in a city could be 3 to 5. Setting this parameter in our algorithm, we can then determine the threshold for Jaccard similarity and Louvain algorithm. We also tested different values for many parameters. However, as prior knowledge provides us with sufficient assumption for our experimental settings, the results are stable from our experiments. Hence, we use these parameters in our analyses.

3. Results

3.1. Characteristics of words and human lifestyles

We analyzed the sequences of human life activities for 12 metropolitan counties in the United States. In the main text of this paper, we show the results for King County (Seattle) in the state of Washington. (The results for the rest of the counties can be found in the Supplementary Information.) Fig. 1c displays the occurrences of significant words among all individuals in King County. We find that the occurrence of the word representing the activity sequences from restaurants to restaurants is particularly higher than other words. This type of life activity sequences could imply a pattern of some people who prefer dining out. The words representing the sequences from clothing stores to clothing stores and the sequences from jewelry stores to jewelry stores also indicate the shopping behaviors of the people. Regarding the non-repeated visited places, the words showing the sequences from health care facilities to clothing stores and the sequences from clothing stores to restaurants, are the most commonly adopted lifestyles among the people.

We detected, in total, more than 10,000 significant words for individuals in King County as presented in Fig. 2a. The distribution of the occurrences of the words, to some extent, are close to a power-law distribution. A large set of words unique to individuals rarely recurs. There is, however, still a certain proportion of words that have a high probability of frequently appearing in the lives of different people, which leads to this fat-tailed distribution of word frequencies. The distribution decays faster at around 3000 frequencies, and the tail diminishes at 10,000 frequencies. Such distribution of word frequencies has been observed in all 12 metropolitan counties investigated in this study (Fig. 2b). Despite differences in demographic characteristics across these counties, the word frequencies follow almost the same probability functions. The properties of this universal distribution indicate that, in a certain number of populations, human lifestyles are concentrated in a small set of words, along with a variety of other words, but the frequencies of the concentrated words are bounded by certain values. Multiple factors, such as the distribution and capacities of urban facilities, may influence and result in such a distribution of human life activities. This study will explore mainly the influence of facility distributions.

We also noticed that the time intervals for movements between two consecutive places in a word vary between groups based on their number of occurrences among the population (Fig. 2c). For example,

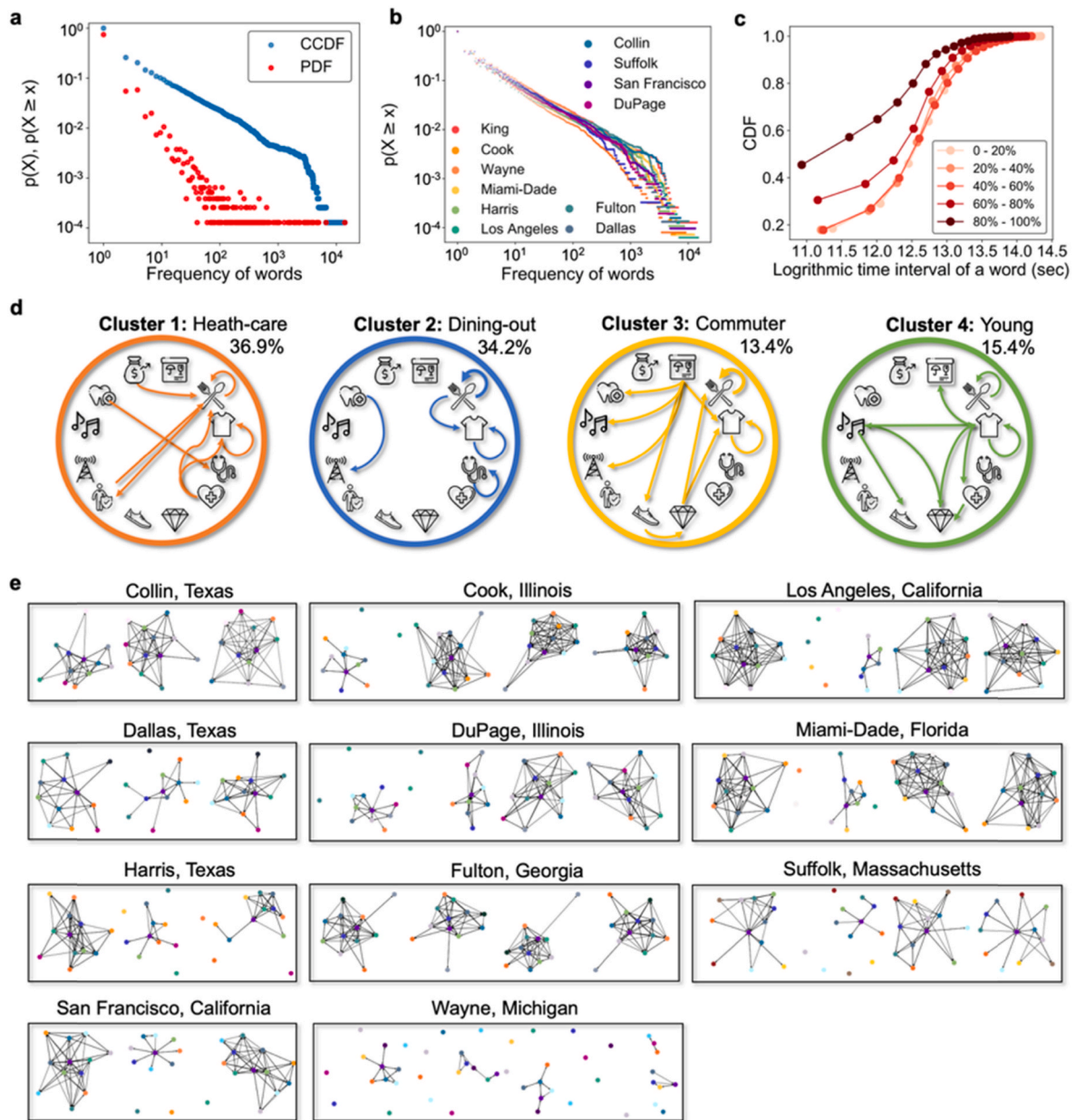


Fig. 2. Semantic analysis of embedded words and lifestyle clusters. **a.** Probability density function plot of the frequencies of words and its complementary cumulative density function of the frequencies of words. The y-axis scale is exponents of 10. **b.** Complementary cumulative density function of the frequencies of words in 12 metropolitan counties. As shown in the figure, the frequencies of word occurrences follow a similar probability function. **c.** Cumulative density function of the time interval between two places in a word. The words are categorized into five groups based on their frequencies in human lives. The words in the group of 0–20 % represents the words that are the rarest words, while the words in the group of 80–100 % represent the most commonly occurring words. The x-axis is a logarithmic scale with base e . **d.** Lifestyle patterns for four clusters of the population based on their visiting behaviors. Here, we show some unique and most frequent life activity sequences in each cluster to highlight people's unique characteristics. **e.** Illustration of networks of lifestyle patterns indicating frequent life activity sequences across different types of facilities in other 11 metropolitan counties in the United States. The nodes represent different types of facilities, and the directed links represent the life activity sequences. (More details are provided in the Appendix).

about 45 % of the words within the top 80–100 % most frequent words were the most frequently occurring words, with less than 16-h intervals. In other groups of words, however, the proportion of words with less than 16-h intervals are much smaller (less than 30 %). Hence, in general, the words that more commonly reflect individuals' lives tend to have shorter time intervals between the two consecutive facilities, while the words that are rarely in human lives tend to have longer time intervals. This may be because the facilities in the rarely adopted words are far from each other, forming geographical barriers preventing the individuals to access them in a consecutive manner. This result highlights the influence of geographical proximity on human lifestyles and the

tendencies of the population moving across a long geographical distance for their life needs. In the following sections, we explore the influences of facility distribution on human lifestyles and population flow across CBGs.

With the representations of human life activities, we further identified four clusters of people who share similarities in their life activity sequences (Fig. 2d). As displayed in the figure, some sequences frequently occurred across all lifestyle clusters, while some are unique in specific clusters. People in the same cluster, however, tend to have similar significant and frequent life activity sequences, while those who are in different clusters have evident variations of lifestyles. The activity

sequences across different POI categories are not equal, indicating the tendencies of people to have specific lifestyles. Nevertheless, based on the uniqueness and frequencies of the life activity sequences, we named the four clusters of people: (1) health-care, (2) dining-out, (3) commuter, and (4) young. People in each cluster have distinct characteristics. Specifically, people in cluster 1, health care, have particularly more life activity sequences between different health care facilities and more sequences between health care and restaurants, compared to those in other clusters. The second cluster, dining-out, mainly contains people who prefer dining out evidenced by the most commonly adopted word is from restaurants to restaurants. Meanwhile, people in cluster 2 have very few life activity sequences, compared to people in other clusters. The first two clusters account for about 70 % of the individuals in our dataset. The sizes of the rest two clusters are similar. People in cluster 3, commuters, seem to be households with significant life activity sequences from electronics and appliance stores to other stores, such as

health care, music, jewelry, telecommunication and shoe stores. These people take care of many errands related to daily lives. The last cluster, the young, shows unique life activity sequences, such as the movement from music stores to shoe and clothing stores. The lives of these people mimic conventional behaviors of young individuals. Based on the discussion of people with various lifestyles, we observed that the focuses of life activity sequences vary across different clusters of people in a county, and the frequencies of life activity sequences also vary among people within a cluster.

The variations of life activity sequences among the population in a county were also examined in different counties. Fig. 2e uses surrogate network models to represent the lifestyle patterns in 12 metropolitan counties using directed networks of POIs. Each county may have three or four lifestyle clusters. Unique and frequent life activity sequences form different lifestyle patterns in clusters. It is obvious that some categories of facilities are closely connected or at the core of the network, while

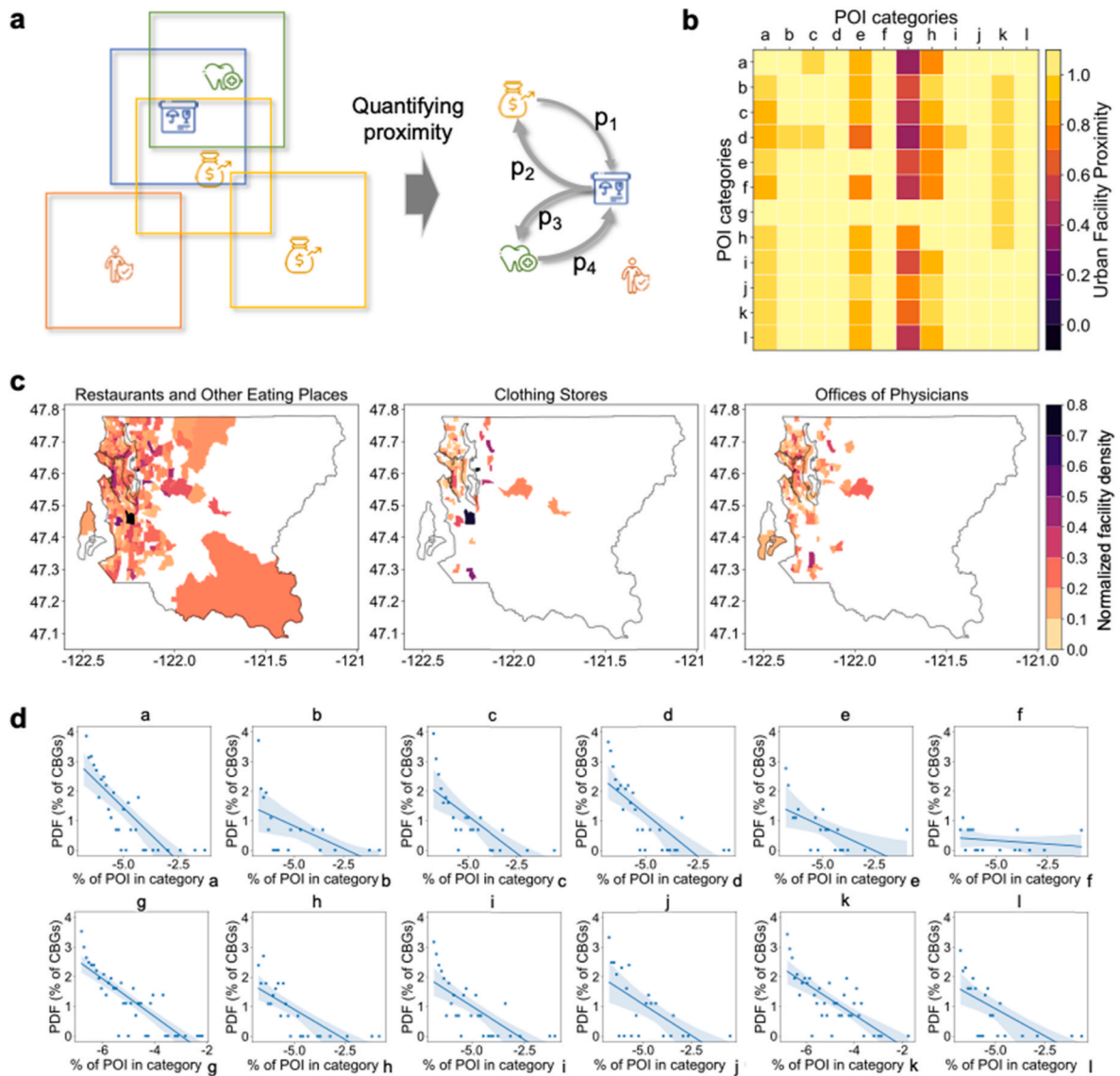


Fig. 3. Distribution patterns of urban facilities. **a.** A schematic illustration for quantifying facility proximity. The bounding box is set for each facility; the probability of a type of facility having another type of facility within its bounding box. **b.** The proximity matrix of 12 types of facilities in King County. The light color indicates that two types of facilities tend to locate closely, while dark color indicates that two types of facilities tend to be far away from each other. **c.** The distribution of three types of facilities in King County. The density of facilities is the proportion of a specific facility in a neighborhood among all the same facilities in the county. To better visualize the distribution, the density values on the maps are transformed using gamma function. **d.** The probability density functions of 12 facilities that describe the distributions of the facilities at the CBG level in the King County. The x and y axis are in logarithmic scale with base e. The plots are titled using the representative characters of POI categories.

some categories of facilities are isolated due to the patterns of life activity sequences. Such variations of life activity sequences raise an important question regarding underlying factors, such as the distribution and proximity of urban facilities. In the following sections, we quantify urban facility proximity and quantitatively examine its relationship with the lifestyle patterns.

3.2. Quantifying urban facility distribution

To examine the influence of urban facility distribution, we created two metrics to measure the characteristics of facility distribution in cities. The first metric is the proximity of urban facilities, the importance of which has been highlighted in existing studies (Chen et al., 2008; Zhang et al., 2016). Considering the facilities across the entire county, the distances between individual facilities of two POI categories may not provide sufficient information to understand the facility proximity for all other facilities within the same category. In addition, the relationship between population mobility and the levels of movement scales (characteristic spatial scales in how people travel) is non-linear (Alessandretti et al., 2020): Daily human mobility does contain meaningful scales corresponding to spatial “containers” that restrict their movement behaviors. That is, movement frequencies and facility preferences of the population within a certain moving scale are stable. Beyond the scale, the movement frequencies and tendencies of the population significantly decrease.

To this end, we propose a scale-based proximity measure that indicates the overall location proximity of the facilities in the entire county. According to an existing study (Alessandretti et al., 2020), we assume that the smallest spatial moving scale of the population is approximately 2–3 km (km). We draw a bounding box of about 2 km for each facility (Fig. 3a). We measure the probability of one type of facility having another type of facility within its bounding box. For example, as shown in Fig. 3a—a proportion, p_1 , of depository intermediations having electronics stores within its bounding box, while there is p_2 of electronics stores having depository intermediations within its bounding box. In the illustration, since the personal care services are not in any of the bounding boxes of other types of POIs, the probability is zero. By calculating all the probabilities of other types of facilities being located within their bounding boxes, we generate a proximity matrix that describes the extent to which one type of facility in a county has other types of facilities within the spatial moving scale of the population (Fig. 3b). The values in the proximity matrix range from 0 to 100 %. As shown in the proximity matrix of the top 12 facilities in King County, we find that a great number of facilities are located in proximity to each other. For example, most of the clothing stores have depository intermediations, restaurants, and offices of physicians within their bounding boxes. In addition, we notice that the proximity matrix is asymmetric, indicating that all bounding boxes of one type of facilities could have another type of facilities, but not vice versa. For example, the shoe stores have the majority of other types of facilities within their bounding boxes, while the shoe stores are not often located in the boxes of other facilities. The asymmetry of the proximity matrix is due mainly to the variation in the numbers of facilities in different categories. This property of the metric also allows us to explore the relationship of the asymmetry to life activity sequences.

In addition, urban facilities are distributed heterogeneously in different CBGs (i.e., census block groups). The CBGs with a greater number of facilities may attract people from other CBGs, leading to higher population in-flow from the outside CBGs. Here, we capture the distribution of facilities by a simple metric, the portion of a specific type of facilities in the CBG among all facilities in the same category in the county. In Fig. 3c, we plot the distributions of three types of facilities in CBGs in King County. The distributions are extremely unequal, approximately following a fat-tailed distribution, and indicating that facilities tend to concentrate in a few CBGs, and but are scarce in other areas (Fig. 3d). The types of concentrated facilities in a neighborhood

also vary greatly in different CBGs. For example, restaurants distribute widely in the majority of King County CBGs, while clothing stores are mainly operated on the west side of the county. Such an unequal spatial distribution may relate to the volume and direction of the cross-CBG population flow due to the needs and incentives of human activity governing their mobility.

3.3. Influences of facility distribution

To uncover the influence of facility distributions, we coupled the human lifestyle matrix with the facility proximity matrix (Fig. 4a). The frequencies of life activity sequences are plotted against the physical proximity of urban facilities. We observed that the frequencies of life activity sequences also follow a fat-tailed distribution (Fig. 4a, the vertical distributions on the right side of the subplots), indicating that the majority of lifestyle sequences occur when one type of facility is often located in the bounding boxes of the other. There are only a few life activity sequences occurring when the two facilities are not closely located. This result demonstrates the presence of spatial movement scale related to mobility behaviors of the population across different POIs. People are most likely to move from an origin POI to a destination POI which is located within the spatial proximity (called “container” (Alessandretti et al., 2020)) of the origin POI. In addition, people with different lifestyles have a varying distribution of frequencies on life activity sequences at the proximity scale. For example, the life activity sequences of people in cluster 1 (health-care) are concentrated in high-frequency activity sequences, while people in cluster 4 (young) have a broader spectrum of sequence frequencies. Despite the existence of variations in the joint distributions of sequence frequencies and facility proximity among different clusters of people, the finding related to the existence of spatial movement scaling as an influence on lifestyle patterns of people is universal across four clusters of lifestyles.

To further explore variations in the joint distribution of life activity sequence frequencies and facility proximity, we examined the popularity of the facilities using a multivariate model, analogous to standard gravity models (Mazzoli et al., 2019). Here, we defined the popularity of a type of facility as the number of life activity sequences that include this type of facility. In the gravity model, the more popular the destination facility, the higher the population flow directed to the destination facility. Hence, to better quantify the influence of facility proximity on human lifestyle sequences and to explain variations in joint distribution, we used a multivariate model that takes the frequency of lifestyle sequences as the dependent variable, and the popularity of the facilities and the facility proximity as the independent variables. The model can be formulated as follows:

$$\log L_{ij} \sim \alpha_1 \log m_i + \alpha_2 \log m_j + \alpha_3 \log p_{ij} + \mu + \epsilon \quad (3)$$

where L_{ij} represents the frequencies of the lifestyle sequence from facility type i to facility type j , m_i is the popularity of the facility type i , m_j is the popularity of the facility type j , p_{ij} is the facility proximity factor which represents the proportion of facility type i having facility type j in its bounding boxes, α_1 , α_2 , and α_3 are coefficients, μ is the intercept, and ϵ is the regression error. It should be noted that all values for the variables are normalized in order to capture both the importance and significance of the variables.

Table 1 shows the results obtained from ordinary linear regression on the model for the four clusters of lifestyles. The popularity of facilities is always significant across four lifestyle clusters. Specifically, the coefficients for the popularity of the facilities in both POIs are positive and close to each other, meaning that both factors have a positive impact on the occurrences of the life activity sequences. Based on the results for clusters 2 (dining-out) and 3 (commuter), we find that the most influential factor is the facility proximity: the closer the two types of facilities in the county, the higher frequencies the life activity sequences would be. This finding has also been found in lifestyle patterns of other counties

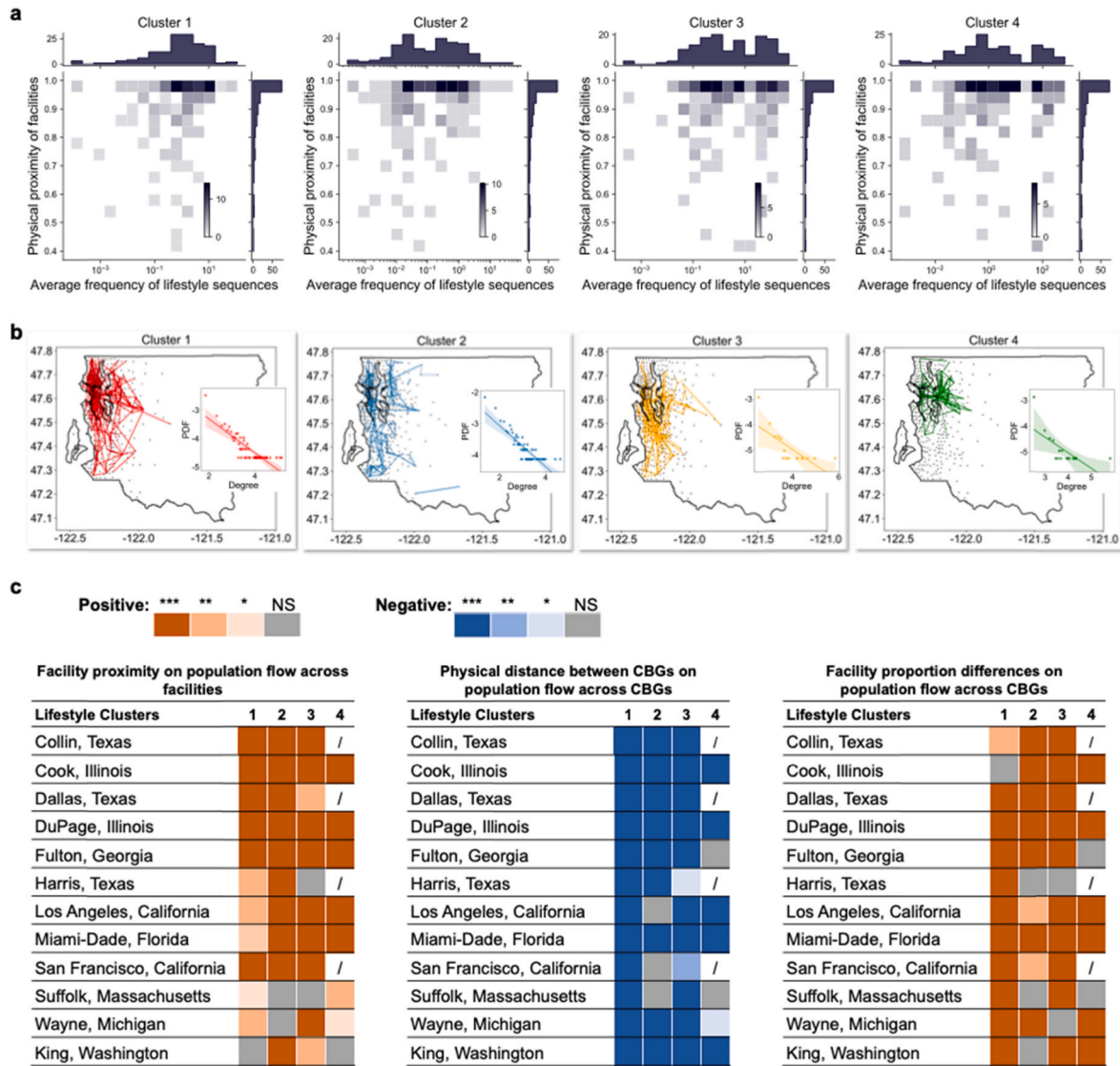


Fig. 4. The influences of facility distribution in King County, Washington State. **a.** Relationships between urban facility proximity and population lifestyles. **b.** Population flows between different neighborhoods for the individuals in four lifestyle clusters. The subplot on the top right corner shows the weighted in-degree distribution d_p of the CBGs in the King County. $d_p = \sum_{q \in \text{Neigh}(p)} T_{pq}$, where the T_{pq} is the population flow from CBG q to CBG p , and CBG q is one of connected neighbors of CBG p . The x and y axis are in logarithmic scale with base 10. **c.** The effects of facility and CBG characteristics on human lifestyle patterns indicated by population flow for 12 metropolitan counties in the United States. Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$; and NS (not significant) $p \geq 0.1$.

(Fig. 4c). Additionally, we noticed that in clusters 1 and 4, since the majority of the life activity sequences occur between proximate facilities, the facility proximity may not have a high influence on the presence of life activity sequences. In this context, instead, the popularity of the facilities plays a more important role to explain the population's life activity sequences among the proximate facilities.

In addition to the life activity sequences between proximate facilities, people also travel across neighborhoods to satisfy their needs due to the lack of specific facilities in some neighborhoods. As a result, life activity sequences may lead to cross-CBG movements. Due to the influence of facility discrepancy among different CBGs, human lifestyles stimulate the incentives of cross-CBG mobility, which may be one reason for cross-CBG movement. To examine the extent to which the interplay between human lifestyles and CBG-level facility distribution shapes the human movements between CBGs (i.e., frequencies of life activity sequences across CBGs), we plotted the cross-CBG life activity sequences on the geographical map of King County (Fig. 4b).

The cross-CBG movements in four clusters of people based on their life activity sequences form different networks of CBGs connected by

activity sequences. The in-degree distributions of the CBGs in the networks are fat-tailed (Fig. 4b). Observed from the geographic map, we found that most of the movements occur between the center of the county and other peripheral CBGs. The involvement of peripheral CBGs, however, distinctly varies across the four clusters of lifestyles. For example, the movements for life activity sequences in cluster 1 connect a large number of CBGs, including the CBGs distant from the center CBG. The movements in cluster 4, however, only connect the CBGs in the Northwest corner of the county. The CBGs where the life activity sequences are concentrated are not always the CBGs at the center of the county. For example, the movements in clusters 1 and 2 are concentrated in the CBGs in the northwest corner of the county, while the movements in cluster 3 are concentrated in the CBGs in the west part of the county.

Based on these results, we find that the difference in facility proportions among various CBGs could be an important factor influencing cross-CBG movements. Also highlighted by existing studies (Kadar and Pletikosa, 2018), the popularity of the specific facilities in a CBG and the physical distance between two CBGs also have an impact on the

Table 1

Regression results for the relationship between population flow across facilities and facility proximity.

	Dependent variable (L_{ij})			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Model Information				
Number of observations	144	144	144	144
F-statistic	20.53***	13.41***	19.06***	13.05***
Coefficients				
Popularity of facility category i (m_i)	1.047*** (0.193)	0.772*** (0.202)	0.931*** (0.190)	0.635*** (0.183)
Popularity of facility category j (m_j)	0.966*** (0.210)	0.727*** (0.203)	1.073*** (0.194)	0.967*** (0.190)
Facility proximity between i and j (p_{ij})	1.602 (1.399)	3.997*** (1.492)	3.350** (1.599)	2.766 (1.777)
Intercept	-16.594*** (3.676)	-9.580*** (3.209)	-18.387*** (3.730)	-13.176*** (3.558)
Model performance metrics				
R-squared	0.306	0.223	0.290	0.218
AIC	688.9	728.4	746.2	770.6

Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

cross-CBG travels. To further quantify the influences of facility distribution in conjunction with other factors, we used a multivariate model to incorporate related factors and compared their importance and significance on the cross-CBG life activity frequencies.

$$\log T_{pq} \sim \beta_1 \log m_q + \beta_2 \log k_{pq} + \beta_3 \log d_{pq} + \mu + \epsilon \quad (4)$$

where T_{pq} is the frequencies of life activity sequences from facilities in CBG p to facilities in CBG q , m_q is the popularity of the specific type of facilities in CBG q (frequencies of life activity sequences directing to the specific type of facilities in CBG q), k_{pq} is the difference of facility proportions between CBG p and CBG q for a specific type of facility (zero for negative values), d_{pq} is the physical distance between CBG p and CBG q (measured based on the centroid of the CBG), β_1 , β_2 and β_3 are coefficients, μ is the intercept, and ϵ is the regression error.

Table 2

Regression results for the relationship between population flow across neighborhoods and facility distributions.

	Dependent variable (T_{pq})			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Model Information				
Number of observations	6372	2346	2430	1792
F-statistic	326***	113***	2177***	169.5***
Coefficients				
Popularity of a type of facilities in CBG q (m_q)	0.124*** (0.004)	0.142*** (0.008)	0.121*** (0.006)	0.166*** (0.008)
Facility proportion differences between CBG p and CBG q (k_{pq})	0.017*** (0.002)	0.003 (0.004)	0.012*** (0.003)	0.026*** (0.004)
Physical distance between CBG p and CBG q (d_{pq})	-0.093*** (0.006)	-0.057*** (0.009)	-0.033*** (0.009)	-0.085*** (0.011)
Intercept	-0.198*** (0.026)	-0.166*** (0.042)	-0.026 (0.035)	-0.105** (0.045)
Model performance metrics				
R-squared	0.133	0.126	0.146	0.221
AIC	6818	2069	2177	1998

Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 2 shows the results obtained from ordinary linear regression on the models for four clusters of individuals. In all lifestyle clusters, the frequencies of lifestyle sequences are significantly proportional to the popularity of specific facilities in the destination CBG. The difference in facility proportions is a significant positive factor influencing the movements of people in clusters 1, 3, and 4, but not as important for cluster 4. Based on the observations of the lifestyle patterns in all counties (Fig. 4c), facility proportion differences tend to have a significant impact on population movements across CBGs. In addition, the physical distance between two CBGs tends to restrict the movement, even for the needs of lives (Fig. 4c). The farther apart the two CBGs, the lesser the movements for life activities between them. It is realistic that people would be attracted by the popularity of the facilities and their life needs, but may not move frequently due to mobility expenses and time constraints. Hence, although the differences in facility proportions could enhance human mobility between neighborhoods, physical distance creates barriers for people to move and diminishes the effects of facility popularity and proportion differences.

4. Discussion and concluding remarks

This study characterizes human lifestyles using large-scale aggregated mobile phone data and examines the influences of facility distribution on life activity sequences of urban population in metropolitan counties in the United States. Human lifestyles capture the ubiquitous life activity sequences and population flows across facilities. Quantitative investigation of lifestyles remains challenging in existing research. With high-resolution data, our study first extracted facilities visited by individuals in chronological order and then applied the Sequitur algorithm to identify significant life activity sequences for each individual. Furthermore, Jaccard similarity analysis was implemented to measure the number of shared life activity sequences among populations. The results reveal categorized clusters of people in based on their unique and frequent life activity sequences within counties. The frequencies of human life activity sequences displayed a fat-tailed distribution, showing variations of life activity sequences within and between lifestyle clusters. This pattern was observed in all 12 counties in this study, indicating a universal patterns of life activities in urban populations. To explain the variations and the universal distributions of human life activity sequences, we propose two metrics—scale-based facility proximity matrix and proportion of facilities in a neighborhood—to characterize the distribution of facilities. We observed that some facilities are often located within the moving scale of a specific type of facility, but not vice versa. We further explore the influence of the asymmetric property of facility proximity and fat-tailed distribution of facilities in CBGs on the frequencies of human life activity sequences. The results indicate the significant positive influence of facility proximity and facility proportion differences among CBGs on the frequencies of life activity sequences. But the positive effects of these factors may be mediated by the physical distances of two CBGs, which leads to insignificance of the factors in some lifestyle clusters.

The findings obtained in this study have multiple theoretical contributions and implications. In particular, understanding human lifestyle patterns has been a fundamental problem in urban science and city planning. The past few decades have witnessed a great amount of efforts in understanding dynamic human life activities in urban populations (Widhalm et al., 2015). While the universal laws and predictability of the activity patterns have been unveiled, the interactions among population, activities and locations, are still not fully understood. This study exhibits the fat-tailed distribution of life activity sequences, describing human movements motivated by the needs of individual lives. People not only share a few general life activity sequences, but also have their unique styles of movements. The distribution of life activity sequences and the clusters of lifestyles describe the statistical and scaling patterns of human behaviors at the county level, which expand our capabilities of capturing human activity patterns and exploring other phenomena

associated with human activities. A typical example is the equity of human access to facilities in disasters. People usually lose access to urban facilities due to disruptions of infrastructure. To prioritize disaster response and mitigation in different urban areas, assessing the needs of access to facilities in urban populations is important. Assessments based on human lifestyle patterns could provide additional information of life activities at different facilities. With the strategic planning of facility distribution and response actions, we could enhance the equity of human access to life services in disasters.

In addition, our work provides insights and opportunities for urban planning and development. Human activity is a complex phenomenon emerging as a result of confounding effects of a great number of factors (Wilson, 2008). The fat-tailed distribution of urban facilities and their proximity do significantly influence the frequencies of life activity sequences. Our observations from this study could advance the understanding of how people may interact with each other at urban facilities and to what extent facility planning can make a difference in the dynamic interaction among people. This implication gives us the confidence to propose urban plans as an effective tool to moderate mobility demand and to alleviate congestion at urban facilities in the long run. Examples of urban plans include redistributing existing facilities and developing new facilities to fulfill the needs of human lives. By doing so, human life activity sequences could be fulfilled within smaller spatial scales. Furthermore, human interactions, especially contact at facilities, would also be changed. People can have lower probabilities of contacting individuals from a variety of CBGs, with an optimal distribution of facilities in a county. This planning strategy and subsequent outcomes are typically important for mitigating disease propagation in health emergencies such as COVID-19, as it could particularly benefit the development of location-specific policies and reduction of contact probabilities.

Despite the significance of the findings and their contributions, this work has some limitations that could be addressed in future research with other datasets. First, daily movements and facility-visiting behaviors are the most common activities, but our dataset may fail to capture movement activities when people did not carry their mobile phones. Hence, it is important to also adopt other fine-grained datasets, such as credit card transaction data and banknotes of anonymized individuals to capture a comprehensive picture of the place-visiting and shopping behaviors. Second, the lifestyle patterns may vary among people with different social-demographic attributes, such as race and gender. Quantifying the relationship between social-demographic characteristics and lifestyle patterns contributes to the understanding of social variations of human lives. The individuals in our dataset, however, are de-identified for privacy protection purposes. We cannot obtain accurate social-demographic information about the individuals. Future studies could find ways to integrate anonymized location-based data with other data sets, such as surveys and social network data, to examine the influence of social-demographic features on human lifestyles. Finally, this study looks mainly into the similarities and consistency of results and findings across counties. It should be noted that different counties have different urban typologies, road networks, and service qualities, which may lead to variations of human lifestyles. Future studies are suggested to incorporate datasets about road networks and quality of services to unravel the disparities of human activities in different counties. Mobile phone data is only one type of the data that can indicate the lifestyle patterns of the populations. Future studies should consider different data sets and methods to test the stability of the findings from our study.

Code availability

The code that supports the findings of this study is available from the corresponding author upon request.

Author contributions

C.F., F.W., and A.M. designed the study. F.W. and C.F. implemented the method and empirical case study. A.M. support with data acquisition. C.F. and A.M. wrote the main manuscript. All authors reviewed the manuscript.

Declaration of competing interest

The authors declare that they have no competing interests.

Data availability

The authors do not have permission to share data.

Acknowledgements

This material is based in part upon work supported by the National Science Foundation under Grant CMMI-1846069 (CAREER), SES-2026814 (RAPID), the National Academies' Gulf Research Program Early-Career Research Fellowship, and the Microsoft Azure AI for a Public Health Grant. The authors also would like to acknowledge the data support from X-mode, Inc. and SafeGraph, Inc. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, Microsoft Azure, SafeGraph, or X-mode, Inc.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dibe.2024.100348>.

References

- Alessandretti, L., Aslak, U., Lehmann, S., 2020. The scales of human mobility. *Nature* 587 (7834), 402–407.
- Alexander, L., Jiang, S., Murga, M., González, M.C., 2015. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transport. Res. C Emerg. Technol.* 58, 240–250.
- Barbosa, H., Barthelemy, M., Ghoshal, G., James, C.R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J.J., Simini, F., Tomasini, M., 2018. Human mobility: models and applications. *Phys. Rep.* 734, 1–74.
- Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. *Nature* 439 (7075), 462–465.
- Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., Leskovec, J., 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589, 82–87.
- Chen, H., Jia, B., Lau, S.S.Y., 2008. Sustainable urban form for Chinese compact cities: challenges of a rapid urbanized economy. *Habitat Int.* 32 (1), 28–40.
- E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1082–1090.
- Deville, P., Song, C., Eagle, N., Blondel, V.D., Barabási, A.-L., Wang, D., 2016. Scaling identity connects human mobility and social interactions. *Proc. Natl. Acad. Sci. USA* 113 (26), 7047–7052.
- Di Clemente, R., Luengo-Oroz, M., Travizano, M., Xu, S., Vaitla, B., González, M.C., 2018. Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nat. Commun.* 9 (1), 3330.
- Fan, C., Jiang, Y., Mostafavi, A., 2020. Emergent social cohesion for coping with community disruptions in disasters. *J. R. Soc. Interface* 17 (164), 20190778.
- Fan, C., Lee, S., Yang, Y., Oztekin, B., Li, Q., Mostafavi, A., 2021. Effects of population co-location reduction on cross-county transmission risk of COVID-19 in the United States. *Applied Network Science* 6 (1), 14.
- Gallotti, R., Bazzani, A., Rambaldi, S., Barthelemy, M., 2016. A stochastic model of randomly accelerated walkers for human mobility. *Nat. Commun.* 7 (1), 12600.
- Gao, H., Tang, J., Hu, X., Liu, H., 2013. Exploring temporal effects for location recommendation on location-based social networks. *RecSys* 93–100. Proceedings of the 7th ACM Conference on Recommender Systems.
- Q. Gao, G. Trajcevski, F. Zhou, K. Zhang, T. Zhong, and F. Zhang, "DeepTrip: adversarially understanding human mobility for trip recommendation," Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 444–447..
- Q. Gao, F. Zhou, G. Trajcevski, K. Zhang, T. Zhong, and F. Zhang, "Predicting human mobility via variational attention," The World Wide Web Conference, 2750–2756..
- González, M.C., Hidalgo, C.A., Barabási, A.-L., 2008. Understanding individual human mobility patterns. *Nature* 453 (7196), 779–782.

- Hanson, S., Schwab, M., 1987. Accessibility and intraurban travel. *Environ. Plann.: Econ. Space* 19 (6), 735–748.
- Hong, B., Bonczak, B.J., Gupta, A., Kontokosta, C.E., 2021. Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nat. Commun.* 12 (1), 1870.
- Hu, Y., Zhang, J., Huan, D., Di, Z., 2011. Toward a general understanding of the scaling laws in human and animal mobility. *EPL* 96 (3), 38006.
- Jackson, L.E., 2003. The relationship of urban design to human health and condition. *Landsc. Urban Plann.* 64 (4), 191–200.
- Kadar, C., Pletikosa, I., 2018. Mining large-scale human mobility data for long-term crime prediction. *EPJ Data Science* 7, 1–27.
- Liu, E.-J., Yan, X.-Y., 2020. A universal opportunity model for human mobility. *Sci. Rep.* 10 (1), 4657.
- Louail, T., Lenormand, M., Picornell, M., García Cantú, O., Herranz, R., Frias-Martinez, E., Ramasco, J.J., Barthelemy, M., 2015. Uncovering the spatial structure of mobility networks. *Nat. Commun.* 6 (1), 6007.
- Lu, X., Yu, Z., Sun, L., Liu, C., Xiong, H., Guan, C., 2016. Characterizing the life cycle of point of interests using human mobility patterns. *Proceedings of the 1052–1063. ACM International Joint Conference on Pervasive and Ubiquitous Computing.*
- Mazzoli, M., Molas, A., Bassolas, A., Lenormand, M., Colet, P., Ramasco, J.J., 2019. Field theory for recurrent mobility. *Nat. Commun.* 10 (1), 3895.
- Microsoft, Microsoft/USBuildingFootprints: Computer generated building footprints for the United States GitHub. <https://github.com/microsoft/USBuildingFootprints..>
- Nevill-Manning, C.G., Witten, L.H., 1997. Identifying hierarchical structure in sequences: a linear-time algorithm. *J. Artif. Intell. Res.* 7 (1), 67–82.
- Niwattanakul, S., Singthongchai, J., Naenudorn, E., Wanapu, S., 2013. Using of Jaccard coefficient for keywords similarity. *Proceedings of the International Multiconference of Engineers and Computer Scientists* 1 (6), 380–384.
- Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C., 2012. A tale of many cities: universal patterns in human urban mobility. *PLoS One* 7 (5), e37027.
- Qin, S.-M., Verkasalo, H., Mohtaschemi, M., Hartonen, T., Alava, M., 2012. Patterns, entropy, and predictability of human mobility and life. *PLoS One* 7 (12), e51353.
- Ren, Y., Ercsey-Ravasz, M., Wang, P., González, M.C., Toroczkai, Z., 2014. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nat. Commun.* 5 (1), 5347.
- Rivera, M.T., Soderstrom, S.B., Uzzi, B., 2010. Dynamics of dyads in social networks: assortative, relational, and proximity mechanisms. *Annu. Rev. Sociol.* 36 (1), 91–115.
- Saberi, M., Hamedmoghadam, H., Ashfaq, M., Hosseini, S.A., Gu, Z., Shafiei, S., Nair, D. J., Dixit, V., Gardner, L., Waller, S.T., others, 2020. A simple contagion process describes spreading of traffic jams in urban networks. *Nat. Commun.* 11 (1), 1–9.
- SafeGraph, SafeGraph: Places Data & Foot-Traffic Insights SafeGraph. [https://www.safegraph.com/..](https://www.safegraph.com/)
- Song, C., Qu, Z., Blumm, N., Barabási, A.-L., 2010. Limits of predictability in human mobility. *Science* 327 (5968), 1018. LP – 1021.
- Stouffer, S.A., 1940. Intervening opportunities: a theory relating mobility and distance. *Am. Socio. Rev.* 5 (6), 845–867.
- Sun, L., Axhausen, K.W., Lee, D.-H., Huang, X., 2013. Understanding metropolitan patterns of daily encounters. *Proc. Natl. Acad. Sci. USA* 110 (34), 13774–13779.
- United States Census Bureau, North American Industry Classification System (NAICS) United States Census Bureau. <https://www.census.gov/naics/>.
- United States Census Bureau. (, december 19) *American community survey 2014–2018 5-year Estimates* United states census Bureau. <https://www.census.gov/newsroom/press-releases/2019/acs-5-year.html>.
- L. Waltman, and N. J. Van Eck, "A smart local moving algorithm for large-scale modularity-based community detection," *Eur. Phys. J. B.*
- D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "Human mobility, social ties, and link prediction," *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1100–1108..
- Wang, J., Kong, X., Xia, F., Sun, L., 2019. Urban human mobility: data-driven modeling and prediction. *SIGKDD Explor. Newsl.* 21 (1), 1–19.
- Wang, M., Yang, S., Sun, Y., Gao, J., 2017. Human mobility prediction from region functions with taxi trajectories. *PLoS One* 12 (11), e0188735.
- Wang, Q., Phillips, N.E., Small, M.L., Sampson, R.J., 2018. Urban Mobility and Neighborhood Isolation in America's 50 Largest Cities, vol. 0. *Proceedings of the National Academy of Sciences of the United States of America*, 201802537.
- Widhalm, P., Yang, Y., Ulm, M., Athavale, S., González, M.C., 2015. Discovering urban activity patterns in cell phone data. *Transportation* 42 (4), 597–623.
- Wilson, C., 2008. Activity patterns in space and time: calculating representative Hagerstrand trajectories. *Transportation* 35 (4), 485–499.
- X-mode. *X-mode | empowering Innovation with quality LocationX-mode.* <https://xmode.io/>.
- Yan, X.-Y., Zhao, C., Fan, Y., Di, Z., Wang, W.-X., 2014. Universal predictability of mobility patterns in cities. *J. R. Soc. Interface* 11 (100), 20140834.
- Yan, X.-Y., Wang, W.-X., Gao, Z.-Y., Lai, Y.-C., 2017. Universal model of individual and population mobility on diverse spatial scales. *Nat. Commun.* 8 (1), 1–9.
- Yao, Y., Hong, Y., Wu, D., Zhang, Y., Guan, Q., 2018. Estimating the effects of "community opening" policy on alleviating traffic congestion in large Chinese cities by integrating ant colony optimization and complex network analyses. *Comput. Environ. Urban Syst.* 70, 163–174.
- Z. Yao, "Exploiting human mobility patterns for point-of-interest recommendation," *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 757–758..
- J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 186–194..
- Zeng, W., Fu, C., Arisona, S.M., Schubiger, S., Burkhard, R., Ma, K., 2017. Visualizing the relationship between human mobility and points of interest. *IEEE Trans. Intell. Transport. Syst.* 18 (8), 2271–2284.
- Zhang, W., Cao, K., Liu, S., Huang, B., 2016. A multi-objective optimization approach for health-care facility location-allocation problems in highly developed cities such as Hong Kong. *Comput. Environ. Urban Syst.* 59, 220–230.
- Zhao, Z.-D., Huang, Z.-G., Huang, L., Liu, H., Lai, Y.-C., 2014. Scaling and correlation of human movements in cyberspace and physical space. *Phys. Rev. E* 90 (5), 50802.
- Zipf, G.K., 1946. The $P_1 P_2/D$ hypothesis: on the intercity movement of persons. *Am. Socio. Rev.* 11 (6), 677–686.