2022 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress

2022 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics) | 978-1-6654-5417-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics55523.2022.00027

Mining Composite Spatio-Temporal Lifestyle Patterns from Geotagged Social Data

Suparna De
Dept. of Computer Science
University of Surrey
Guildford, UK
s.de@surrey.ac.uk

Usamah Jassat
Dept. of Electronic Engineering
University of Surrey
Guildford, UK
usamah.jassat@gmail.com

Alex Grace
Dept. of Electronic Engineering
University of Surrey
Guildford, UK
alex@adgrace.com

Wei Wang
School of Advanced Technology
Xian Jiaotong Liverpool University
Suzhou, China
Wei.Wang03@xjtlu.edu.cn

Klaus Moessner
Faculty of Electrical Engineering
Chemnitz University of Technology
Chemnitz, Germany
klaus.moessner@etit.tu-chemnitz.de

***Abstract*—As social networks become increasingly integrated with their users' daily lives, and users are willing to publicly share data about their offline activities on these networks, the resultant data offers a powerful tool to non-intrusively understand city dynamics as it captures human behaviour and interactions. In this paper, we derive lifestyle patterns from the Foursquare social network data, using matrix factorization and tensor decomposition as unsupervised methods to extract latent spatio-temporal behavior patterns. The extracted patterns offer precise definition of activity levels associated with specific lifestyles and showcase that users' behaviors are a combination of several lifestyles, in contrast to traditional circadian topology theory which classifies individuals to a specific temporal pattern. The obtained patterns can provide deeper insights into city dynamics, the people within them and how society functions.**

***Index Terms*—Foursquare, spatio-temporal lifestyle, tensor computing, non-negative matrix factorization**

I. Introduction

The increasing availability of smartphones and the growing popularity of location-based social networks' (LBSNs) use offers a new sensing paradigm for urban dynamics by observing the collective characteristics of urban residents [1]. LBSNs link a physical location to online data through acts such as check-ins to a city point-of-interest (POI), or geotagging photos or messages. The resultant cyber-physical-social system (CPSS) data [2], [3] can help to characterize spatio-temporal social and economic aspects of different urban areas [4]–[6].

By switching the perspective from a city to its citizens, these digital traces from LBSNs reflect how citizens experience their city [7], based on their work/social travel patterns and types of POIs visited at various temporal points. Understanding different lifestyles helps to gain insights into social groups' daily movements, cultural boundaries and latent social patterns (e.g. segregation of citizens into different regions) [7], [8], and offers possibilities for understanding better the interaction of citizens with public spaces. We define 'lifestyle' from existing literature, as "the way in which a person or group lives, including their interests, behaviors, and behavioral orientations" [8].

Prior works on understanding lifestyles of different demographics have used targeted one-off surveys [9], coupled with location-tracking smartphone apps for residents or tourists [10], [11], required developing custom mobile apps for crowd-sourced lifestyle profiling of the older population [12], or the use of proprietary data such as call data records for activity-based pattern mining [13]. A recent work by Hu *et al.* [8] used LBSN data for deriving spatial and temporal patterns analogously.

In this work, we leverage data from the Foursquare[1] LBSN that represents the "natural unconstrained behavior" [8] of citizens, to infer lifestyles of urban residents. Foursquare is one of the biggest LBSNs and is built around the premise of providing personalized recommendations of places to visit (venues) based on a user's previous location check-ins. By enabling its users to generate check-ins to venues listed on the platform, Foursquare provides a rich source of information of the activities of users (i.e. according to the type of 'venue' visited, such as restaurant, music venue, home etc.) at different times of the day/week [7].

With the resultant geo-tagged data combining both static spatial and social elements of the urban environment with the dynamic characteristics of human check-ins, appropriate techniques need to be designed to uncover the latent lifestyle patterns from this multi-dimensional data as well as being able to handle large matrices resulting from the large dataset. Non-negative matrix factorization (NMF) has been successfully employed in the literature to extract interpretable latent patterns

This work (Grant number: EP/W032473/1) has been supported through the UKRI Strategic Priority Fund as part of the wider Protecting Citizens Online programme and is associated with the National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online (REPHRAIN).

[1]https://foursquare.com

978-1-6654-5417-9/22/$31.00 ©2022 IEEE

DOI 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics55523.2022.00027

from large spatio-temporal data [8], [14]. We employ NMF to extract temporal lifestyles guided by circadian topology (CT) [15], as evidenced through the work/rest behavioural patterns derived from the data, as well as spatial aspects corresponding to users' visits to types of venues. These are then expanded into lifestyle patterns covering both spatial and temporal dimensions by applying third-order tensor decomposition [16]. Tensors have been recognised as an efficient method for representation of heterogeneous CPSS data and for extracting complex patterns from high-dimensional data [17], [18]. We apply the developed methods to Foursquare check-in data collected in the city of London.

The contributions of our work are as follows:

- our work translates ubiquitous LBSN data into interpretable lifestyle (human mobility) patterns, reflecting the interactions of residents with city spaces,
- the unsupervised data mining pipeline is able to extract latent *composite*, i.e. joint spatio-temporal lifestyle patterns, rather than analogous spatial and temporal distributions,
- the uncovered lifestyle patterns can be interpreted as a combination of individuals' diurnal and spatial habits, supported with a precise description of different activity levels at various temporal points of the day for each lifestyle.

The rest of this paper is organized as follows: in Section II, we review the related work in extracting lifestyle and relevant latent patterns from CPSS data. Then, we present the dataset collection process and analyse its characteristics in Section III. Section IV introduces the NMF and tensor decomposition methods and how these are applied to the lifestyle derivation problem, with results presented in Section V. Section VI concludes the paper with a discussion of the obtained results and outlines the future research.

II. Related Work

Lifestyle research has previously been conducted through surveys, and focused on assigning individuals to their relevant CT work/rest behavior group [9]. In contrast to such targeted surveys that are reliant on user participation and self-reported observations, our work on capturing human lifestyles through cyber-physical-social data traces from their daily activities offers an unobtrusive way for data collection and analysis at large scales.

Existing crowdsensing approaches include research to determine the older population's lifestyle in Singapore [12], with user profiles derived from smartphone sensor data (location, noise and light) fused with travel patterns (and their frequencies) to POIs. The resulting profiles clustered users (into 3 groups) according to their age, gender, ratios comparing their 'stay' at their home location versus that at indoor/outdoor/public/private POI, as well as the type of POI (health/food/shopping). A comparable approach to extract individual spatial mobility networks in terms of transitions between home and stay locations (i.e. work, shop, lunch/dinner) from mobile call data records, validated with household travel surveys, was investigated in [13]. The approach taken in our work, in contrast, employs openly available social network data and results in a more fine-grained and precise description of the diurnal activity levels of the wider population across venues arranged over a wider set of intuitive categories. Similar to our work, Hu *et al.* [8] used Foursquare data to find multiple patterns in check-in data of individuals, including daily temporal lifestyle patterns, i.e. hours that individuals are most active and correlating to college or commute lifestyles, and the times that different types of POIs had the most activity. Our approach is motivated by this work; we further enhance the analysis to derive joint spatio-temporal patterns.

Geotagged social network data, specifically Foursquare data, has been employed to describe different regions of a city in terms of the visits, along different temporal ranges, to venues located in different regions [7] and to understand area boundaries by clustering venues according to the number of check-ins [19], [20]. Research by Silva *et al.* [21] analysed Foursquare data to better understand the dynamics of cities, through city-level check-in heatmaps, comparing check-in times with check-in transition (source to destination) Foursquare category, time and type (weekday/weekend) of day, to enable a quantitative comparison of several cities. Foursquare data has also been used to measure urban deprivation [22] by correlating venue categories to the neighborhood deprivation score as well as for correlating check-ins to diurnal pollutant levels and traffic volume [23]. This shows the applicability of Foursquare data as a useful and reliable indicator of several urban dynamics' features.

III. Dataset Collection and Characteristics

A. Data and Model

We collected 138,140 Foursquare check-ins from 20,000 users across 10,853 London venues between March - May 2017. Foursquare does not provide a streaming API or a way to query for public location data. Developers can look up public check-ins if they have the check-in ID but they cannot search for public check-ins, making Foursquare data collection a non-trivial task. Since mid-2014, user check-ins have transitioned from Foursquare to an app created specifically for this purpose, called Swarm. Swarm is now the social network where the users check-in and share it with their friends, but the API is still served by Foursquare, which is now a review, services and area knowledge app fed from the Swarm data. The Foursquare API when given the public Swarm check-in ID returns the Swarm check-in data. This data contains:

- id : unique check-in id
- createdAt : seconds since epoch when check-in was created
- timeZoneOffset : minutes offset between when check-in occurred and same time in UTC
- type : checkin / shout / venueless
- user : user details
- venue : venue details containing name, location, category and other information.

When Foursquare users opt to link their Foursquare account with their Twitter account, a geo-tagged tweet is generated

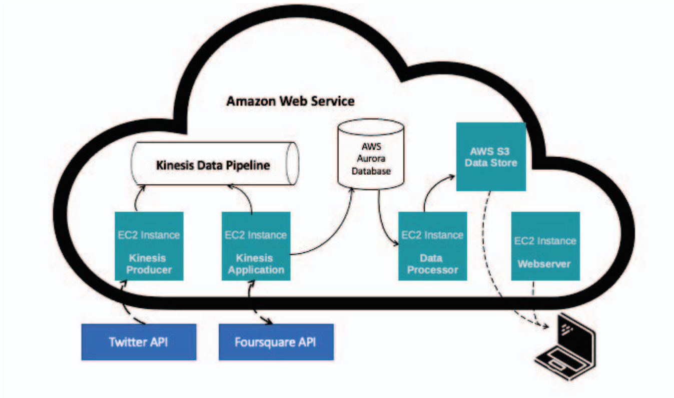

Fig. 1. AWS data collection and preprocessing pipeline

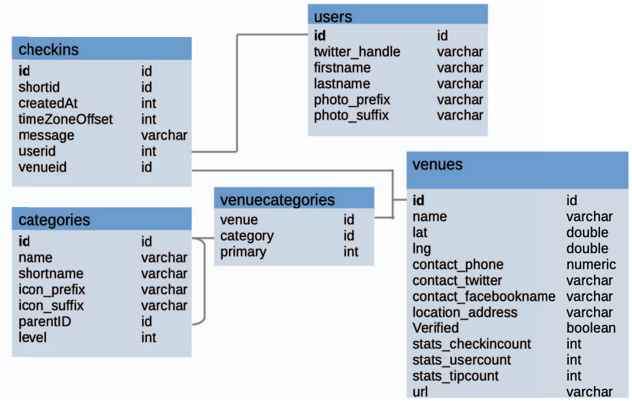

Fig. 2. Foursquare data schema

automatically when a Foursquare check-in is performed. In line with existing works [21], we employ a hybrid solution by using the Twitter streaming API and filtering for tweets with a Foursquare link containing the users' public Foursquare check-in. We set up a data collection pipeline on Amazon Web Services (AWS) cloud, as shown in Fig. 1, utilizing a mixture of EC2 instances (servers) and services offered by AWS such as the Kinesis Data Pipeline and AWS Aurora Database.

A Kinesis Producer connects to the Twitter streaming API, filtering for tweets containing a Foursquare URL, which are then pushed straight into the Kinesis Data Pipeline as a JSON string. The Kinesis Application then takes groups of tweets from the Kinesis Pipeline, extracts the Foursquare URL and a regex is applied to determine the check-in id. An example Foursquare URL extracted from a tweet is of the form: https://www.swarmapp.com/c/1fZ8p9S0jZA, where 1fZ8p9S0jZA is the Foursquare check-in ID. The Foursquare API is then queried with the check-in ID to retrieve the check-in details. The query returns a JSON string with the check-in data detailed above (in bullet points). The venue is the users' current location and is tagged with details such as the coordinates, name and venue category such as cinema, restaurant, home. The data collection code is made publicly available on github[2]. The JSON string is processed and mapped to a defined Foursquare data model and then added to the Amazon Aurora Database. The developed model and corresponding database schema is shown in Fig. 2.

B. Dataset Exploration

To visualize the dataset characteristics, the Data Processor (in Fig.1) is used to extract data from the Aurora database and process it in accordance with each of the data models, resulting in a group of JSONs and GeoJSONs to represent the venues and model polygons and JSON object details; this data is persisted to the AWS S3 datastore which can be publicly accessed by a URL. The AWS S3 hosts the webserver. To visualize the venues and check-ins on the website, in the

[2]github.com/alexgrace95/Kinesis-Application-FoursquarePublicDataRecorder

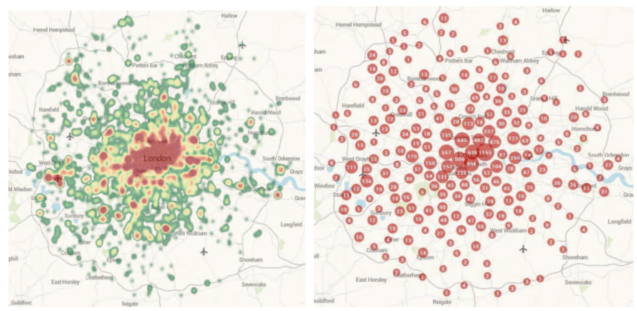

Fig. 3. Heatmap for London showing (a)venue density, (b)number of venues

form of heatmaps, the relevant data (JSON, GeoJSON and corresponding object details) are loaded by querying the AWS S3 API. The visualization uses the Java Topology Suite (JTS) for handling the GeoJSON files, polygons and point objects.

Figs. 3 and 4 show heatmaps depicting different views of the Foursquare venue density, number of venues across London and the number of check-ins, which all correlate with each other. As expected, the areas of higher density and more check-ins are both central London and Heathrow Airport, with smaller spikes around the suburbs.

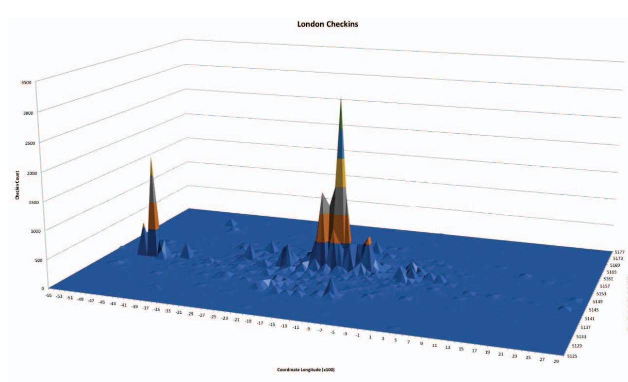

Fig. 4. Heatmap for London showing number of check-ins at each coordinate

IV. METHODOLOGY

In this section, we introduce the NMF and tensor decomposition processes for mining spatio-temporal lifestyles from the collected data. Since both NMF and tensor decomposition are matrix manipulation methods, the Foursquare data for the check-ins, venues and categories is first retrieved from the Aurora database into CSV files and then arranged into matrices according to the analysis methods.

A. Matrix Decomposition

NMF is an effective method to find latent hidden factors within large amounts of data and we apply it to find temporal and spatial patterns within Foursquare check-in data. NMF is one of several matrix factorisation techniques, the difference between NMF and other techniques, such as vector quantization and principal component analysis (PCA), is that NMF only works on non-negative matrices and works to find basis elements for this matrix that are also non-negative. The non-negative nature of the decomposition is useful as it allows us to model the vector as a positive combination of multiple parts and is intuitively applicable to datasets with all positive values, i.e. there is no such thing as negative user check-ins. The other alternatives for matrix decomposition such as PCA and factor analysis either rely on information from the correlation matrix, or have insufficient information derived from the covariance matrix, respectively. Both NMF and PCA decompose a matrix as the product of two matrices, but PCA works on the non-weighted sum of the squares of the residuals, causing high values to dominate the analysis (and low values getting ignored). In contrast, NMF stresses on each data sample's information, "by weighting the residual squares with the reciprocals of the squares of the standard deviations of the data values" [14].

NMF decomposes a $(n \times m)$ matrix V into a product of two other non-negative matrices: a feature matrix W of dimensions $(n \times k)$ and a coefficient matrix H of dimensions $(k \times m)$ [24], where k is the number of base elements to be found:

$$V \approx WH \tag{1}$$

If V is organised to have each column be a feature vector of length m, then the reconstruction process can be thought of as each feature vector being reconstructed by a linear combination of k feature vectors in W, each weighted by k coefficients in H, as illustrated in Fig. 5.

The values of the matrices W and H are found by minimizing the Frobenius norm of the difference between V and WH, i.e.:

$$0.5 \times ||V - WH||_F^2 \tag{2}$$

where

$$||X||_F = (\sum_{ij} |X_{ij}^2|)^{-1/2} \tag{3}$$

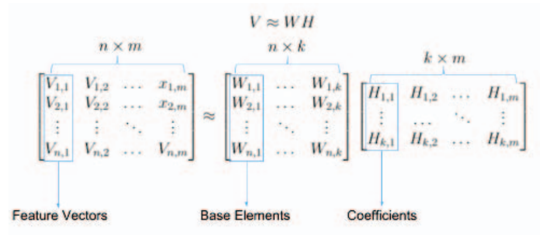

Fig. 5. The reconstruction process of NMF

a) Deriving Temporal Lifestyles using NMF: To find the temporal lifestyle patterns in the data, the lifestyle of an individual user is described using a M dimensional activity vector, a. To extract daily patterns, M is set to 24 (mapped to 24 hours in the day), with the values of a being the check-ins that the individual user performed per hour.

To find patterns across all users, the activity matrix for each individual user is concatenated to create a $M \times u$ matrix, where u is the number of users we have data for. This activity matrix is transposed and then decomposed using NMF into 2 matrices, with the derived lifestyle matrix of dimension $k \times M$, where k is the number of latent lifestyles and is empirically adjusted to obtain interpretable results. The NMF algorithm is implemented using the Python sklearn NMF decomposition library[3], on a laptop with an AMD Ryzen5 5600X CPU and 16GB RAM, with 11% CPU and 205MB RAM usage during runtime. NMF decomposition for temporal lifestyle patterns had an execution time of 26.7 seconds, with the spatial lifestyle patterns executing in 17.9 seconds.

b) Deriving Spatial Lifestyles using NMF: In addition to uncovering temporal lifestyles related to activity levels during different times of the day, finding users' preferences for frequenting specific types of POIs (venues) is another important indicator of their lifestyles. For uncovering the spatial lifestyle patterns, the venue category of each user check-in is used. The spatial lifestyle is formulated as a combination of several categories of venues, with each combination and its corresponding weights for each category, being specific to that lifestyle.

The categories for the venues that Foursquare uses are organized in a hierarchy. Altogether, our dataset has 923 different categories organized in a tree structure. To narrow this down, venues for each check-in in the dataset were designated by their corresponding top-level venue category. At this top level, there were 10 categories. Hence, when looking for latent spatial lifestyles, instead of the temporal activity matrix as before, we decompose a spatial activity matrix A (representing all users in the dataset) into a lifestyle matrix of dimension $k \times M$, where k is the number of lifestyles and M is the number of categories of venues, and a $u \times k$ coefficient matrix that provides information about individuals' preferences for these lifestyles. The preference for a particular lifestyle is defined as the average of the coefficients of users in the group

[3]https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF

for the lifestyle. The coefficient matrix is also split into smaller matrices, each of which records the lifestyles for different genders. Taking into consideration the number of top-level categories (of each visited venue), M was set to 10, with the values being the number of check-ins performed for a given venue category. The value of k was empirically determined to find a good balance between granular and interpretable results.

B. Third-Order Tensor Decomposition

Although NMF is a proven, effective method to extract patterns from data, it is limited by its inability to find patterns across multiple dimensions. To find patterns in multiple dimensions, we aggregate data related to the check-ins, hour of day and venue category of check-in into a third order tensor. Tensors are a way to structure data across multiple dimensions and can be thought of as a multi-dimensional array. Tensors are a more general form of vectors and matrices, with vectors being a tensor of order 1 and matrices of order 2.

In the same way that vectors of a user's activity can be combined to create a matrix representing the activity of multiple users, multiple matrices that represent the activity of a user across two dimensions can be combined to create a tensor of the activity of multiple users. In a similar fashion to how NMF can be used on matrices to find latent structures within the data, tensor decomposition can be used to find patterns in tensors. One such method of decomposition is CANDECOMP/PARFAC (CP) [16], [25] decomposition. This technique decomposes a tensor into a sum of component tensors, these component tensors are rank-one tensor.

A tensor $T \in \mathbb{R}^{I \times J \times K}$ can be decomposed into three matrices $A \in \mathbb{R}^{I \times R}$, $B \in \mathbb{R}^{J \times R}$, and $C \in \mathbb{R}^{K \times R}$, where R is the number of component tensors. The component tensors can be obtained by taking the outer product of the columns in A, B, and C. These matrices approximate T in the form:

$$T_{(1)} \approx A(C \odot B)^T \tag{4}$$

where $\odot$ is the Khatri-Rao product and $T_{(1)}$ is the mode-1 matricization of T. The matricization of a tensor is simply the unfolding of a tensor into matrix form. The CP decomposition is formulated by minimizing the approximation of the tensor T, formally:

$$\underset{A,C,B}{min} \parallel T_{(1)} - A(C \odot B)^T \parallel_F \tag{5}$$

The alternating least-square (ALS) algorithm can be used to solve the optimization. ALS works by fixing two of A, B, and C whilst solving for the other; this process is done for each of A, B, and C repeatedly until a convergence criterion is fulfilled. Using this method, the number of rank-1 tensors to decompose T into R, must be provided.

This tensor decomposition technique was used to find spatio-temporal lifestyle patterns of individuals from Foursquare check-in data. The ability of third order tensor decomposition to decompose tensors of a larger order makes it an effective technique to find the composite spatio-temporal patterns that extend over multiple dimensions. The first stage of this process requires reorganizing the dataset into a third order tensor, this tensor is of size $u \times M \times c$ where u is the number of users in the dataset, M the number of hours in a day, and c the number of venue category types. This tensor represents each user's activities through a 24 × 10 matrix (again, considering the top-level category of each check-in's corresponding venue), where an entry at row i and column j represents the number of check-ins made at a venue of type j at hour i. These matrices can be combined to build up the complete third order tensor.

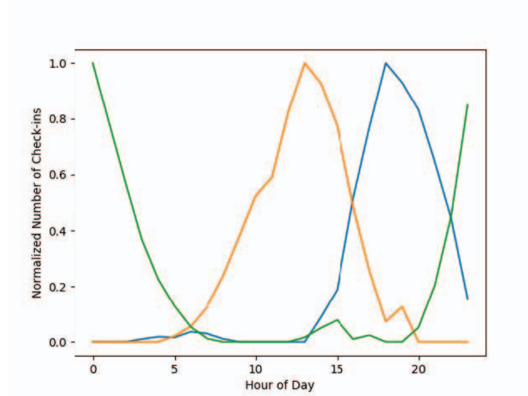

Fig. 6. Temporal lifestyles discovered from Foursquare check-in data

CP decomposition is performed using a Python 'tensor factorization' library [26], which implements a solution using an alternating least squares method as outlined in [16]. The runtime for the CP decomposition was 1.2 seconds. The decomposition decomposes the tensor into R rank-1 tensors, where R is the number of lifestyles. A range of values was used to find multiple lifestyle patterns.

V. Results

A. Temporal and Spatial Lifestyle Aspects

We first present the temporal and spatial lifestyles derived by applying NMF to the Foursquare data. Fig. 6 shows the three different temporal lifestyles obtained from the data, with the 24-hour scale on the x-axis and the normalized number of check-ins on the y-axis. The three lifestyles show peaks in activity during different times in the day. These relate to people who are most active early in the day, in the afternoon, or in the evening. The nature of NMF means that an additive combination of these lifestyles can be used to represent the lifestyle of an individual. The findings corroborate those from literature [8] and the theory on circadian rhythms [15], which assigns individuals to be a weighted combination of three temporal behaviors: 'early bird' (morning types, who get up early in the day and go to bed early in the evening), 'nightowl' (evening types, who stay up until late in the night), and 'intermediate' (individuals with work/rest patterns in-between the early birds and the night owls).

Table I shows the normalized average weight of each lifestyle. It shows that the afternoon lifestyle is the predom-

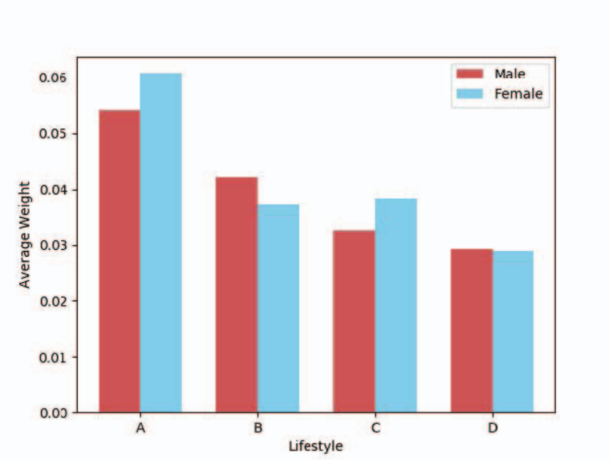

Fig. 7. Comparison of average weight of lifestyles for males and females

inant lifestyle followed by the evening and early morning lifestyles.

TABLE I
SCALED AVERAGE OF THE WEIGHT OF EACH TEMPORAL LIFESTYLE

Lifestyle	Normalized Average
Early	0.27
Afternoon	0.43
Night	0.30

The spatial lifestyle patterns obtained using NMF, by setting $k = 4$, are displayed in Table II; this is displayed by listing the top spatial (i.e. venue) categories for each lifestyle pattern along with the normalised weight of that category. It can be seen from Table II that lifestyle A consists of four spatial categories, but lifestyles B-D are composed of only three prominent venue categories each. These discovered lifestyles show discernible patterns, for example lifestyle C has the highest weighted categories being 'Travel' and 'Professional', indicating a commuter lifestyle whilst pattern A, B, and D show outgoing and consumer lifestyles, with lifestyle B showing more of an active lifestyle whilst lifestyle A and D show more of a partying lifestyle. Interestingly, across all lifestyles, the 'Professional' category is one of the top categories, however with a low weighting, this would make sense as most people work, however their lifestyle isn't defined by this, but rather what they do in their spare time.

Fig. 7 shows the difference in weights for each lifestyle for the different genders. The data shows that there are slight differences in lifestyles between the genders. The graph shows that females seem to have a slight preference for lifestyle A, which is weighted highly towards 'Food' and 'Nightlife' venues whilst males seem to have a preference for lifestyle B which is weighted higher with 'Outdoor and Recreation' locations. It also seems to show that females travel more often, as seen by their slightly higher average in lifestyle C.

B. *Composite Lifestyle Aspects*

Composite patterns, which reflect a combination of individuals' daily and spatial habits, have been derived using the tensor decomposition method. Changing the number of patterns, R, to find patterns, resulted in some distinct lifestyle patterns to emerge. The time of day component of each pattern is shown in Fig. 8, with the corresponding top weighted component spatial locations (venue categories) shown in Fig. 9.

Lifestyle A is mainly weighted with check-ins at 'Food' venues and 'Shops and Services', as shown in Fig. 9. The daily pattern of this lifestyle (Fig. 8, left-most graph), shows that people partake in these activities later in the day rather than in the morning, with the peak being at 6 pm. A sharp increase can also be seen at around 1 pm, which corresponds to lunch time for many people.

Lifestyle B is weighted very highly in the 'Transport and Travel' location type and the middle graph of Fig. 8 shows the daily pattern for the lifestyle. An extremely high peak in the morning is most likely an indicator of the morning commute, whilst a smaller peak at 1 pm most likely indicates lunchtime travels. This pattern also shows relatively high activity levels during the early evening, between 5-8pm. Here, in contrast to the sharp peak in the morning, the higher activity is more spread out, suggesting that although individuals may travel to work at a similar time, the time for travelling back is a lot more varied.

Lifestyle C is also heavily weighted in 'Travel and Transport', however it also has a significant weighting on 'Outdoor and Recreation' check-ins. As the right-most graph of Fig. 8 shows, the daily pattern for this lifestyle has a peak in the morning and one in the afternoon, with very low weightings at all other times. This would suggest that for commuting, along with recreation, the timings of individuals of this lifestyle are very consistent and set.

VI. Conclusions and Future Work

In this paper, we extract latent lifestyle behaviors of urban residents using Foursquare LBSN data. To find lifestyle patterns across a single dimension, NMF was used. The dimensions explored using this technique were the temporal and spatial dimensions. The temporal dimension analysis gave 3 patterns that showed variances in diurnal activity. The 3 patterns that were found all displayed a peak in activity at a different time of the day, one in the early morning, at midday, one in the early evening and one later in the evening. These four peaks show that individuals may be more active at different times during the day. This agrees with research into CT which indicates that there are 3 circadian categories. From the three patterns, the midday peak has the largest width suggesting people of this type may be the most active, however this may also be because this activity pattern is aligned with the opening times of most business establishments.

The same NMF technique was used to find patterns in the spatial dimension. This analysis resulted in 4 distinct patterns that can be understandable as different lifestyles. One interesting fact garnered from looking at the obtained patterns

TABLE II
SPATIAL LIFESTYLES DISCOVERED FROM FOURSQUARE CHECK-IN DATA

Lifestyle	Category 1	Category 2	Category 3	Category 4
A	Food(0.65)	Nightlife Spot (0.15)	Shops and Services (0.13)	Professional (0.03)
B	Outdoors and Recreation (0.61)	Shops and Services (0.21)	Professional (0.1)	
C	Travel and Transport (0.8)	Professional (0.1)	Shops and Services (0.05)	
D	Arts and Entertainment (0.67)	Nightlife Spot (0.18)	Professional (0.14)	

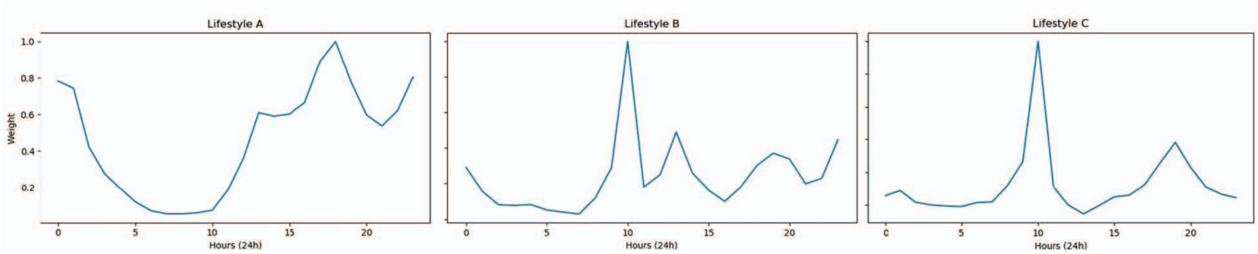

Fig. 8. Spatio-temporal lifestyle trends along 24 hours of the day

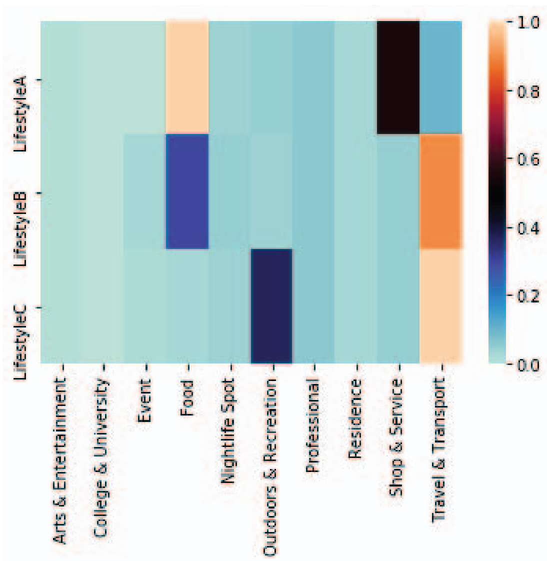

Fig. 9. Spatio-temporal lifestyle trend by component venue visit weights

is that the 'Professional' category is seen in all the lifestyle patterns, showing that it is an important and key part of most people's lifestyle, but not the most important or defining factor. The range of categories that show significant weightings in each of the four lifestyles discovered shows that individuals lifestyles aren't one dimensional and show a wide amount of variety, being composed of multiple aspects.

From the tensor analysis uncovering joint spatio-temporal patterns in the data, it is evident that lifestyle patterns may be seen as a combination of individuals' daily (delineated by hour of the day) and spatial habits. Moreover, we are able to provide a precise description of activity level along time of day for each lifestyle. It's natural to see one's behaviors as a combination of several lifestyles. For example, people may exhibit one lifestyle during the weekdays and another during the weekends.

The next phase of this research will look at analysing the density of different venue types in different areas of the city and how this maps to the different lifestyles. For instance, how the lifestyles relate to different 'neighborhoods' of the city, showcasing the interaction of citizens with the built environment. We also plan to extend the analysis with multiple social network data sources such as the London Santander bike trips data, in order to gain a broader and comprehensive understanding of human activities using large-scale CPSS data.

REFERENCES

[1] K. L. S. Machado, A. Boukerche, E. C. Cerqueira, and A. Loureiro, "A data-centric approach for social and spatiotemporal sensing in smart cities," *IEEE Internet Computing*, vol. 23, no. 1, pp. 9–18, jan 2019.

[2] J. Zeng, L. T. Yang, H. Ning, and J. Ma, "A systematic methodology for augmenting quality of experience in smart space design," *IEEE Wireless Communications*, vol. 22, no. 4, pp. 81–87, 2015.

[3] S. De, Y. Zhou, I. Larizgoitia Abad, and K. Moessner, "Cyber–physical–social frameworks for urban big data systems: A survey," *Applied Sciences*, vol. 7, no. 10, 2017.

[4] Y. Zhang, B. Li, and J. Hong, "Using online geotagged and crowd-sourced data to understand human offline behavior in the city," *ACM Transactions on Intelligent Systems and Technology*, vol. 9, no. 3, pp. 1–24, may 2018.

[5] S. K. Rumi, K. Deng, and F. D. Salim, "Crime event prediction with dynamic features," *EPJ Data Science*, vol. 7, no. 1, oct 2018.

[6] S. De, W. Wang, Y. Zhou, C. Perera, K. Moessner, and M. N. Alraja, "Analysing environmental impact of large-scale events in public spaces with cross-domain multimodal data fusion," *Computing*, vol. 103, no. 9, pp. 1959–1981, 2021.

[7] E. Celikten, G. L. Falher, and M. Mathioudakis, "Modeling urban behavior by mining geotagged social data," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 220–233, jun 2017.

[8] T. Hu, E. Bigelow, J. Luo, and H. Kautz, "Tales of two cities: Using social media to understand idiosyncratic lifestyles in distinctive metropolitan areas," *IEEE Transactions on Big Data*, vol. 3, no. 1, pp. 55–66, mar 2017.

[9] K. S. Jankowski, "Morningness/eveningness and satisfaction with life in a polish sample." *Chronobiology International*, vol. 29, pp. 780–5, 2012.

[10] N. Wijerathne, S. K. Viswanath, M. S. Hasala, V. Beltran, C. Yuen, and H. B. Lim, "Towards comfortable cycling: A practical approach to monitor the conditions in cycling paths," in *Proc. IEEE 4th World Internet Things (WF-IoT)*, 2018, p. 778–783.

[11] S. K. Viswanath, C. Yuen, X. Ku, and X. Liu, "Smart tourist - passive mobility tracking through mobile application," in *Internet of Things. IoT Infrastructures, IoT360 2014. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, R. Giaffreda, D. Cagáňová, Y. Li, R. Riggio, and A. Voisard, Eds. Springer, 2015, pp. 183–191.

[12] S. H. Marakkalage, S. Sarica, B. P. L. Lau, S. K. Viswanath, T. Balasubramaniam, C. Yuen, B. Yuen, J. Luo, and R. Nayak, "Understanding the lifestyle of older population: Mobile crowdsensing approach," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 1, pp. 82–95, feb 2019.

[13] S. Jiang, J. Ferreira, and M. C. Gonzalez, "Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 208–219, jun 2017.

[14] S. De, U. Jassat, W. Wang, C. Perera, and K. Moessner, "Inferring latent patterns in air quality from urban big data," *IEEE Internet of Things Magazine*, vol. 4, no. 1, pp. 20–27, mar 2021.

[15] C. S. Smith, C. Reilly, and K. Midkiff, "Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness." *The Journal of applied psychology*, vol. 74, pp. 728–38, 1989.

[16] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[17] X. Wang, L. T. Yang, J. Feng, X. Chen, and M. J. Deen, "A tensor-based big service framework for enhanced living environments," *IEEE Cloud Computing*, vol. 3, no. 6, pp. 36–43, 2016.

[18] X. Wang, L. T. Yang, X. Chen, M. J. Deen, and J. Jin, "Improved multi-order distributed HOSVD with its incremental computing for smart city services," *IEEE Transactions on Sustainable Computing*, vol. 6, no. 3, pp. 456–468, 2021.

[19] J. Cranshaw, R. Schwartz, J. Hong, and N. Sadeh, "The livehoods project: Utilizing social media to understand the dynamics of a city," in *Proc. International AAAI Conference on Web and Social Media*, vol. 6, no. 1, 2012, pp. 58–65.

[20] A. X. Zhang, A. Noulas, S. Scellato, and C. Mascolo, "Hoodsquare: Modeling and recommending neighborhoods in location-based social networks," in *Proc. IEEE Int. Conf. Social Comput.*, Washington D.C., USA, Sep. 2013.

[21] T. H. Silva, P. O. S. V. de Melo, J. M. Almeida, J. Salles, and A. A. F. Loureiro, "Revealing the city that we cannot see," *ACM Transactions on Internet Technology*, vol. 14, no. 4, pp. 1–23, 2014.

[22] D. Quercia and D. Saez, "Mining urban deprivation from Foursquare: Implicit crowdsourcing of city land use," *IEEE Pervasive Computing*, vol. 13, no. 2, pp. 30–36, apr 2014.

[23] A. Komninos, V. Stefanis, A. Plessas, and J. Besharat, "Capturing urban dynamics with scarce check-in data," *IEEE Pervasive Computing*, vol. 12, no. 4, pp. 20–28, oct 2013.

[24] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proceedings of the 13th International Conference on Neural Information Processing Systems*, ser. NIPS'00, 2000, p. 535–541.

[25] S. Rabanser, O. Shchur, and S. Günnemann, "Introduction to tensor decompositions and their applications in machine learning," arXiv:1711.10781, 2017.

[26] M. Nickel, "Python library for multilinear algebra and tensor factorizations:mnick/scikit-tensor," 2019.