Received 4 May 2023, accepted 17 June 2023, date of publication 21 June 2023, date of current version 28 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3288283

RESEARCH ARTICLE

Inferring Trips and Origin-Destination Flows From Wi-Fi Probe Data: A Case Study of Campus Wi-Fi Network

THANISORN JUNDEE[1], SANTI PHITHAKKITNUKOON[1,2,3], AND CARLO RATTI[4]

[1]City Context Laboratory, Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand
[2]Excellence Center in Infrastructure Technology and Transportation Engineering (ExCITE), Chiang Mai University, Chiang Mai 50200, Thailand
[3]Chiang Mai Research Center for Carbon Capture and Storage (Chiang Mai CCS), Chiang Mai University, Chiang Mai 50200, Thailand
[4]SENSEable City Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Corresponding author: Santi Phithakkitnukoon (santi@eng.cmu.ac.th)

This work was supported in part by Chiang Mai University.

ABSTRACT This work introduces an alternative solution to costly conventional approaches for large-scale travel behavior data collection by utilizing an opportunistic sensing data source i.e., Wi-Fi probe data. Through our case study of Chiang Mai University campus as a city, we developed a framework for inferring and visualizing Wi-Fi data-based travel behavior by demonstrating how a Wi-Fi probe data can be analyzed to infer trips and origin-destination flows. Specifically, our contributions include algorithms developed for inferring spatial presence, residence, stay, trip, and trip distribution among places in the campus, as well as campus inflow and outflow. Moreover, to handle the Wi-Fi access point data for the analysis, and visualize the inferred trips and flows, an online visual analytics tool called Wi-Flow is developed as part of this work. Our framework differs from the other studies with our residence and trip detection algorithms that produce the result at the individual level as opposed to the overall network. The experimental results are intuitive and insightful, providing useful information for area management. Our research underscores the significance of utilizing Wi-Fi probe data in mobility modeling. Additionally, it introduces an opportunistic sensing approach for estimating mobility flows, which not only contributes to our understanding of transportation dynamics but also holds significance in comprehending the implications for carbon capture efforts.

INDEX TERMS Human mobility, origin-destination flow, travel behavior, trip inference, urban informatics, visual analytics tool, Wi-Fi probe data.

I. INTRODUCTION

Human mobility has become a focus of urban planning as urbanization accelerates. Movement of human beings in space and time reflects the spatial-temporal characteristics of how urban areas are utilized, which inherently describes travel behavior and demand of the region. Information about travel behavior is thus important for transport travel demand modeling. Traditionally, household survey, roadside interview, and traffic count have been used to collect travel behavior data in a region or traffic analysis zone (TAZ). Travel demand is then quantified based on these travel surveys. Typically, each subject recruited for a household survey is asked to complete a diary of activities and travel on a given day or week. The cost of a complete household survey ranges between $150 and $300 per household, depending upon administration and technology used. Due to its highly laborious and costly effort, such household travel surveys are conducted once every 10 years in most countries. The surveyed data becomes as an input for transport demand models that is typically used to predict future travel behavior and demand, which have played an essential role in in managing and planning for urban transportation systems [1], [2], [3].

Unfortunately, the data collected from a roadside interview and a traffic count only provides a snapshot of travel demand, while a household survey is costly and time-consuming as

The associate editor coordinating the review of this manuscript and approving it for publication was Yanli Xu.

well as erroneous due inaccurate response from the survey participants that relies on recall of their past journey details. A recent alternative approach to travel data collection is to make use of information and communication technologies (ICTs) such as ubiquitous technology like global positioning system (GPS) sensors, which have been used in travel surveys [4]. However, due to the privacy concerns and regulations for revealing such locational information to third parties can pose user privacy violation issues as stipulated by GDPR [5] and HIPAA [6], collecting GPS-based travel behavior data at large scale thus becomes difficult and challenging. Recent attempts in GPS-based travel behavior data collection have been limited to specific type of tracked individuals, such as university students [7] and customers of a particular service provider where the data was obtained in exchange of some incentives [8]. Privacy concerns largely prevent this type of detailed mobility data to be available and utilized extensively.

Opportunistic sensing data has become a valuable alternative source of travel behavior data. As opposed to the active sensing paradigm where data is purposely or explicitly collected for a very specific purpose, such as a patient data is collected for a clinical diagnosis, an opportunistic sensing data is the data that is exploited beyond its primary purpose of collection. An example of an opportunistic sensing data is call detail record (CDR) data, i.e., communication logs of individual mobile device, that is collected for billing purposes by a service provider i.e., telecom operator, but it can also be analyzed to study people's social behavior [9]. Locational information of mobile users is collected whenever they connect to the cellular network for a service, such as making or receiving a call, surfing the Internet, or using SMS – i.e., the connected cell tower location, rearrested to the mobile user is recorded. Being a massive dataset in both cross-sectional and longitudinal perspectives allows the CDR data to draw lots of attention from the travel behavior research community. It has been used in human mobility studies such as trip end detection [10], [11], trip distribution modeling [12], [13], transport mode inference [14], [15], and route choice assignment [16]. The main advantage that opportunistic sensing data has over its counterpart is its relatively low cost, which is nearly no cost because it's already been collected for its main purpose. However, CDR's availability is rather rare. Most CDR datasets are only available for those with a non-disclosure agreement (NDA). Publicly available datasets such as Nokia Mobile Data Challenge dataset [17] and Orange Telecom Data for Development Challenge (D4D) [18], [19] are becoming dated, while privacy regulations are preventing a new dataset from being available.

Here we explore another alternative opportunistic sensing data source that is more accessible than the CDR, while being able to provide such detailed locational information. Wi-Fi probe data is a collection of user connectivity logs of different access points (AP) across the network. A mobile device is by default configured to steadily scan and connect to an available Wi-Fi AP. As such, a series of connected AP locations and corresponding timestamps of the Wi-Fi users produce locational traces of individuals, which can then be used for mobility analysis both at individual and aggregate levels, reflecting travel behavior in different aspects.

An early study by Sevsuk et al. [20] analyzed a Wi-Fi user logging data from 3,000 APs in buildings across MIT campus to describe occupancy and movement patterns of the users in order to understand space usage and users of the space. Understanding space usage can help with space planning and designation, such as computational frameworks called Eigenplaces [21] and Xplaces [22] that applied eigendecomposition and clustering algorithms on campus Wi-Fi network data to understand how space is utilized and how it can be segmented accordingly. With using a university's campus environment as a testbed, Wi-Fi locational data across two campuses from 550 APs over six months has been used to highlight relationship between space environment and movement of the users across different buildings and campuses [23]. Since Wi-Fi network is mostly provided for indoor usage, it thus becomes a good source for extracting indoor mobility information. So, Trivedi et al. [24] used Wi-Fi logs of a university's campus network of 2,500 users over two months to construct a multi-modal embedding transformer to predict indoor mobility with analyzed mobility features such as building type, user type (e.g., student, academic, staff). On the other hand, the outdoor mobility has also been explored with the Wi-Fi data. As human mobility is highly predictable as we often repeat our travel patterns by revisiting the same set of places over and over again [25], Wi-Fi data has also been used to discover such significant places in typical daily routines as well as usual locations of 191 mobile users during 18 months in Lausanne, Switzerland [26]. Wi-Fi probe data has been used to estimate population in a city by different categories e.g., workers, residents, and visitors, from 53 APs in lower Manhattan in New York City [27]. The same New York City dataset has also been explored to analyze human mobility [28] by using hourly aggregate population count and applying a spatial network analysis with edge frequency and direction representing mobility flows, which are then mapped onto a road network by assigning APs to nearest road segment.

This study is an attempt to recognize the potential usability of Wi-Fi probe data in travel behavior research by demonstrating how it can be analyzed to infer trips and flows in a case study of campus network, realizing the vision of smart city whose area uses different types of electronic data collection sensors to supply information which is utilized in managing assets and resources efficiently – i.e., transportation system in our case. Mobility estimation has a direct impact on carbon capture efforts. By accurately estimating mobility patterns, such as the movement of vehicles and people, we can better understand the sources and levels of carbon emissions. This information is crucial for implementing effective carbon capture strategies and targeting areas with high emissions.

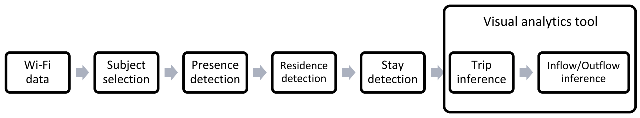


FIGURE 1. **System overview of trip/flow inference and visualization.**

Our contributions include new algorithms for inferring individual trips and Origin-Destination (O-D) flows based on Wi-Fi probe data, as well as a web-based visual analytics tool for handling, visualizing, and analyzing Wi-Fi probe data. Our work differs from the previous studies with our proposed method for inferring outdoor mobility in the notion of O-D flows (i.e., the amount of trips made from an origin to a destination) based on our newly developed residence detection, and our proposed trip detection method that is at the individual level as opposed to overall flows in a spatial network. Moreover, we propose a new and different approach to road network-based trajectory detection, which provides another feasible alternative.

Note that the O-D flow is a term used in this study to reflect on the quantitative aspect of movement flowing between areas (origin-destination pairs) as described by an O-D matrix that is used to assess the demand for transportation i.e., the higher number of trips in the O-D matrix cell the more this route is in demand.

II. METHODOLOGY

This section describes the data used in this study and our methodology for selecting subjects for the analysis, detecting presence, residence, and stay location, as well as inferring inflow/outflow trips, and the development of our visual analytics tool. System overview of our methodology is depicted in Fig. 1. Subject selection is done to ensure granular mobility details. Presence detection process is developed to identify the subject's appearance as a series of visited places. Residence detection is for determining the main origin of each device i.e., campus dorm. Stay detection is to identify locations at which the subject spent a substantial amount of time i.e., making a stop for some activity such as reading at a library, eating at a canteen, and so on. Once stay locations are determined, a trip can be inferred as going from one stay location (i.e., origin) to another (i.e., destination), between which there may be intermediate locations or waypoints. Consequently, inflow and outflow can then be inferred based on the origins and destinations of aggregate trips. To visualize these trips and inflows/outflows, a visual analytics tool is design and developed.

A. WI-FI PROBE DATA

In this study, we used Wi-Fi connectivity data collected from a Wi-Fi network of Chiang Mai University (CMU)'s main campus. There were 3,116 access points (APs) in total, serving an area of 2.93 km^2 that houses the university's administrative center, 17 faculties, graduate school, dormitories, campus resource facilities and services, as well as

FIGURE 2. **Locations of 5,872 Wi-Fi access points across the CMU's main campus (enclosed in blue lines), where color differentiates groups of APs based on their classified places such as building, office, parking lot, and footpath. Main gates located in the north, east, and south sides of the campus are labelled with markers.**

sports facilities. The data was collected at 5-minute sampling rate from each access point (AP) across the network. Each record includes the AP's MAC address, AP's geo-coordinates (latitude, longitude), AP's classified place ID, connected device ID (hashed MAC address), received signal strength indicator (RSSI) between the connected device and the AP, and timestamp. APs were grouped based on their locations and classified into different place IDs, such as building, sports ground, parking lot, and footpath. A (student, staff, or guest) CMU account is only required for each device to begin its connection with the Wi-Fi network for the first time. Then, the device can auto-join the network. Since the network uses a mesh topology, so each device continues to stay connected to the network through APs with a strongest signal as the user's roaming location to location within the campus with available 2.4-GHz and 5.0-GHz wireless bands. AP locations within the CMU's main campus (enclosed in blue lines) are shown in Fig. 2, where color is used to differentiate groups of APs based on their classified places. The APs located inside buildings were strategically placed by the campus network engineers for best service coverage.

Wi-Fi connectivity logs were collected for 26 days for this study, from 9th January – 3rd February 2020, which constitutes ∼40 GB of data including 133,754,260 records from

3,116 unique APs classified into 74 places that altogether served 291,124 different devices. Note that each user account was allowed to register up to five devices.

B. SUBJECT SELECTION

Our interest was to extract trips made by users within the network from Wi-Fi logs. We also would like to select users who lived in the campus dorms to assemble detailed trajectories. So, we selected device IDs that satisfy the following criteria (during the study period).

1. Device connected to the network from more than one place ID.
2. Device made an overnight connection from a campus dorm (10 p.m. – 6 a.m. next morning). Additionally, the device must establish a connection at least once every hour during this period.
3. Criterion #2 must be satisfied for all weekdays.

Note that the period from 10 p.m. – 6 a.m. is a night-time curfew during which no entering or exiting dorm is allowed. While not being used, a device switches to 'sleep' mode in which its connection to the Wi-Fi network becomes less frequent than when it's in use. Observing a connection once an hour is to ensure the device's presence in the dorm – hence, carried by a user who resides in the dorm. After applying these criteria, we retained 4,803 device IDs for our analysis.

C. PRESENCE DETECTION

The next process was to determine the presence of each subject as a series of places visited. To do so, consecutive Wi-Fi connections of each subject were grouped together based on the place IDs and a duration threshold (ε). Each presence includes a starting timestamp, presence duration, and place ID. Let X_j denote a set of all Wi-Fi connection logs of subject j, such that $X_j = \{x_j(1), x_j(2), x_j(3), \ldots x_j(n_j)\}$, where n_j is the total number of connections and $x_j(i)$ is the i^{th} log defined as a set $x_j(i) = \{ap_j^i, place_j^i, t_j^i\}$ containing information about its connected AP's ID (ap_j^i), place ID ($place_j^i$), and connection timestamp (t_j^i). So, the subject j's presence can be defined as $s_j(i) = \left\{\tau_j^i, d_j^i, place_j^i\right\}$, where τ_j^i is the starting time at the visited place $place_j^i$ and d_j^i is the visit duration. Methodologically, the presence detection is described by the Algorithm 1.

The output is a series of visited places by the subject j ($\boldsymbol{S_j}$) or the *presence* of the subject j, where each $s_j(i)$ includes the information on starting timestamp, presence duration, and place ID. The threshold ε was relaxed to one hour during the night-time curfew as the device may switch into sleep mode. This threshold value may vary from case to case, depending upon the network and device configurations. An optimal threshold value is worth a future exploration.

D. RESIDENCE DETECTION

The place of residence is the main origin of each subject's commuting trip, which collectively constitutes the most common trips in the road network in general [29]. In addition, most trips are circular – i.e., a journey that begins and ends at the same location, forming a circular route. This type of trip can involve visiting multiple destinations along the way and returning to the starting point without retracing the exact same path. In our case, students who reside in a dorm travel to places and eventually return to their dorm – i.e., the main origin. So, it's important to be able to identify the main origin for each subject. This main origin will then serve an origin for destinations defined by other significant place IDs such as faculties, offices, and other dorms, which collectively forms an O-D matrix describing the amount of trips flowing across the campus i.e., O-D flow. As we had gathered subjects who were campus dorm residents, so in this step we further determined which dorm each subject was likely to live in. We utilized the detected 'presence' of the subjects from the Algorithm 1, especially during the night-time curfew i.e., 10 p.m. – 6 a.m. The residence is determined as the place visited and spent most time in during the curfew period. Let $R_j = \{r_j(1), r_j(2), \ldots, r_j\left(m_j^s\right)\}$ denote a set of time spent by the subject j in residential places. There were 15 places that were classified as residential in our case, including five male and 10 female dorms. Residence detection is described methodologically in Algorithm 2.

Algorithm 1 Presence Detection

Input: a set of Wi-Fi connection logs of subject j (X_j)
Output: a set of sequential places that subject j has visited (S_j)

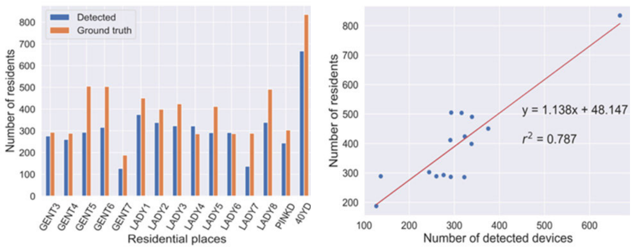
$c = 0$
for $i = 1$ **to** $n_j - 1$ **do**
  if $t_j^{i+1} - t_j^i > \varepsilon$ **or** $place_j^{i+1} \neq place_j^i$
    $d_j^i = t_j^i - t_j^c$
    $s_j(i) = \left\{t_j^c, d_j^i, place_j^i\right\}$
    $c = i + 1$
  end if
end for
$\boldsymbol{S_j} = \left\{\boldsymbol{s_j(1), s_j(2), \ldots, s_j(n_j^s)}\right\}$
return S_j

Algorithm 2 Residence Detection

Input: a set of residential places that subject j has visited $\left(S_j^{res}\right)$
Output: residence of subject j (Res_j)

for $i = 1$ **to** m_j^s **do**
  $r_j(i) = \sum_k (d_j^k | place_j^k \in \text{residential place}\, i)$
end for
$R_j = \left\{r_j(1), r_j(2), \ldots, r_j\left(m_j^s\right)\right\}$
$Res_j = \mathbf{arg\,max}_i \left(r_j(\mathrm{i})\right)?$
return $\boldsymbol{Res_j}$

Dorm residents were detected using the Algorithm 2. The result is depicted in Fig. 3, where Fig. 3(a) shows a bar chart comparing the number of detected residents in



(a) (b)

FIGURE 3. Residence detection result validation: (a) number of detected residents compared against the actual number of dorm residents, (b) correlation between the number of detected and actual numbers of dorm residents.

(a) (b)

FIGURE 4. Examples of inferred trips: (a) a few examples of complete and incomplete trips represented with a solid and dash lines, respectively with markers indicating an origin (O), destination (D), and waypoint (w) of each trip; (b) all trips generated by subjects who are residents of a female dorm.

Algorithm 3 Stay Detection

Input: S_j
Output: a set of subject j's stays ($Stay_j$)

$\boldsymbol{c = 0}$
for $i = 1$ **to** n_j^s **do**
  if $d_j^i \geq \varepsilon$ **do**
    $\boldsymbol{\sigma\,(c) = s_j\,(i)}$
    $\boldsymbol{c = c + 1}$
  end if
end for
$\boldsymbol{Stay_j = \{\sigma\,(1)\,, \sigma\,(2)\,, \ldots\}}$

each dorm against the actual number of residents reported by the University, while Fig. 3(b) shows a linear correlation between the two. The coefficient of determination or R^2-value is 0.787 and R-value is 0.887, which indicates a high correlation and hence, our residence detection is reasonable.

E. STAY DETECTION

Since a trip is a journey made from one location to another, thus it necessitates identification of potential locations for the trip's origin and destination, which are likely the place where the traveler spends a considerable amount of time in or stays while engaging in some activity. In this step, stay locations of each subject are determined from which classification of individual trip origins and destinations will be performed in the next step (Section II-F).

By adopting our previous approach described in [30], the place in which the subject's presence lasts at least ε minutes is considered as a location that the subject makes a 'stay', otherwise the place is considered as a 'waypoint'. In addition to the origin and destination, a trip is also composed of intermediate points or waypoints that provides a detail about trajectory of the trip. So, a trip is a series of waypoints that starts with an origin followed by intermediate points and ended with a destination. Algorithm 3 describes our stay detection methodologically.

F. TRIP INFERENCE

Each trip starts with an origin and ends with a destination, while intermediate places collectively add a trajectory detail into the trip as waypoints. So, an individual trip is a sequence of visited places, i.e., subject's presence, each classified into origin, destination, or waypoint. A waypoint is defined as a place in which the subject spends considerably a short amount of time, i.e., presence duration is less than ε, as opposed to the stay location (described in the Algorithm 3). The stay location, on the other hand, can be identified as an origin or destination or both. A set of each subject's stays obtained from the Algorithm 3 is used to guide our trip reconstruction.

The basic idea is to reconstruct a trip based on continuity of the subject's presence (ε is used as a threshold). A trip can be classified as complete or incomplete. A complete trip is a trip that can be reconstructed with all three components i.e., an origin, waypoint(s), and a destination, or at least an origin and destination, while an incomplete trip is a trip without either its origin or destination or both. Trip inference method is described in the Algorithm 4. In terms of the notations, any trip c of subject j ($trip_j\,(c)$) is defined as a set that can contain an origin ($orig_j^c$), destination ($dest_j^c$), and a set of waypoints (w_j^c), i.e., $trip_j\,(c) = \{orig_j^c, w_j^c, dest_j^c\}$, where set w_j^c may contain intermediate waypoints of the trip, i.e., $w_j^c = \{w_j^c\,(1)\,, w_j^c\,(2)\,, \ldots, w_j^c\left(n_j^c\right)\}$.

As a result of the Algorithm 4, a total of 440,997 trips were inferred, including 357,287 complete trips and 83,710 incomplete trips. Examples of inferred trips are shown in Fig. 4, where Fig. 4(a) shows a couple of examples of complete and incomplete trips, and Fig. 4(b) shows all trips generated by residents of a female dorm. Building footprint of the female dorm is highlighted in blue, based on which its centroid is used as a representative geo-coordinate.

In Fig. 4(a), trip #1 is a complete east-to-west trip with three waypoints, while trip #2 is an incomplete trip (without detected destination) and its origin shares a common location with the trip #1's destination. Trip #3 is an example of an incomplete trip without its origin detected, while trip #4 is a complete trip traveling from a male dorm situated near

Algorithm 4 Trip Inference

Input: S_j and $Stay_j$
Output: a set of subject j's trips ($Trip_j$)

$orig, dest = 0, c = 1, w = \emptyset$
for $i = 1$ **to** $|S_j| - 1$ **do**
  if $t_i(i+1) - [t_j(i) + d_j(i)] < \varepsilon$ **do**
    if $s_j(i) \in Stay_j$ **do**
      if $orig \neq 0$ **or** $w \neq \emptyset$ **do**
        $dest = s_j(i)$
        $trip_j(c) = \{orig, w, dest\}$
        $dest = 0, w = \emptyset$, c= $c + 1$, $orig = s_j(i)$
      else
        $orig = s_j(i)$
      endif
    else
      $w.append(s_j(i))$
    endif
  else
    if $orig \neq 0$ **or** $w \neq \emptyset$ **or** $dest \neq 0$ **do**
      if $s_j(i) \in Stay_j$ **do**
        if $orig \neq 0$ **or** $w \neq \emptyset$ **do**
          $dest = s_j(i)$
        else
          $orig = s_j(i)$
        endif
      else
        $w.append(s_j(i))$
      endif
      $trip_j(c) = \{orig, w, dest\}$
      $c = c + 1$
    endif
    $orig, dest = 0, w = \emptyset$
  endif
end for
$Trip_j = \{trip_j(1), trip_j(2), \ldots\}$

the center of the campus to a bank in the north direction. In Fig. 4(b), most trips are intuitively generated from the dorm as the primary origin. A relatively large number of trips are made to nearby places in the west side of the dorm, which are the areas where other female dorms are situated. These presumably are friend visit trips or going for food in other female dorms' canteens. At CMU, student dorms are restricted to gender-exclusive housing policy, including both stay and guest visit. Although students are allowed to eat at any dorm's canteen, female students have a tendency to eat at their female dorms unlike the male students. There are also trips to different faculties, which are presumably going to lectures.

G. INFLOW AND OUTFLOW INFERENCE

A set of inferred trips from the Algorithm 4 allows us to further examine trip distribution – i.e., the number of trips

Algorithm 5 Inflow and Outflow Inference

Input: a set of subject j's incomplete trips ($Trip_j^{inc}$), and G
Output: a set subject j's inflow and outflow trips ($inflow_j$, $outflow_j$)

$inflow_j, outflow_j = \emptyset$
for $i = 1$ **to** $\left|Trip_j^{inc}\right|$ **do**
  if $trip_j^{inc}(i) : dest_j^i$ **exist and** $w_j^c(1) \in G$ **do**
    $inflow_j.append(trip_j^{inc}(i))$
  else if $trip_j^{inc}(i) : orig_j^i$ **exist and** $w_j^c\left(n_j^c\right) \in G$ **do**
    $outflow_j.append(trip_j^{inc})$
  end if
end for
return $inflow_j, outflow_j$

that occur between each origin and each destination zone (or place). The complete trips can be used to explore trip distribution between places on campus, while the incomplete trips can be further processed and utilized to discern the incoming and outgoing movement of the whole campus's population.

Trips that are originated from outside of the campus are considered incoming trips. So, to infer about campus inflow, incomplete trips are taken into consideration, especially those with no origins. A trip is inferred as an incoming trip if it is an incomplete trip that has no origin and its first waypoint is one of the places nearby the campus gates. Likewise, a trip can be inferred as an outgoing trip if it is an incomplete trip that has no destination and its last waypoint is one of the places nearby the gates. Algorithm 5 describes our method for inferring the campus inflow and outflow, where G denotes a set of places that are in vicinity of the campus gates. These places that are considered near the gates and used for inferring the campus inflow and outflow are shown in Fig. 5. There is a total of 238 APs that belong to 13 places and outdoor facilities are used as the gatekeepers.

H. VISUAL ANALYTICS TOOL

To carry out this study, there was a need to visualize and manage our geospatial data, which included AP geolocations, trips, and flows. Conventional GIS tools, such as ArcGIS[1] (closed source) and QGIS[2] (open source), are capable of visualizing and managing our data, however sharing results across platforms and online capabilities are troublesome. An online tool like Kepler.gl[3] is a better option for our study with its open-source and online solutions, but data labeling and grouping capabilities are lacking. So, we developed our own web-based tool that can visualize and manage geospatial data in an interactive way, called *Wi-Flow*.

[1] https://www.arcgis.com
[2] https://qgis.org
[3] https://kepler.gl

(a) (b) (c) (d)

FIGURE 5. Places (highlighted) that are used as the gatekeeps include: (a) three places and outdoor APs near the North gate composed of 10 APs, (b) three places near the East gate consisting of 27 APs, (c) three places near the road and South gate #1 consisting of 150 APs, and (d) four places near the South gate #2 consisting of 51 APs.

Wi-Flow was developed with React.js,[4] which is an open-source JavaScript framework, widely used for building interactive user interfaces and web application. While React.js was used for web development, we used Deck.gl[5] to handle 2D and 3D data visualization as it is a widely used WebGL[6]-powered library for visualizing large spatial datasets on the fly and with minimal complexity. Deck.gl utilizes WebGL library that provides access to the GPU on the user's computer asynchronously, so it is able to handle millions of data points fast. As such, Deck.gl is very suitable for rendering a large number of markers, such as AP locations, on a map. In addition, Deck.gl provides a set of view layers such as scatter plot, heatmap, and 3D renderer that serve our purpose. Furthermore, Deck.gl works perfectly with Nebula.gl[7] that allows us to use Selection Layer for the data point selection with mouse for our AP labeling – i.e., selecting multiple APs on the map by dragging the mouse. This capability allowed us to label and group all APs into places (as described in Section II-A).

A screenshot of our developed tool, Wi-Flow is shown in Fig. 6, where Mapbox Dark[8] is used as the background

[4]https://reactjs.org
[5]https://deck.gl
[6]https://get.webgl.org
[7]https://nebula.gl
[8]https://www.mapbox.com/maps/dark

FIGURE 6. Snapshot of our developed visual analytics tool namely, Wi-Flow while the user is in the data labelling mode.

map. The user can interact with the tool with mouse to select different options. In the upper right area of the tool, the top bar (#1) is the area where the user can choose between labeling or O-D modes. In the labelling mode as shown in Fig. 6, the second lower bar pointed by #2 marker is where the user can choose to import an AP geolocation data (can be either CSV or JSON formats), export a resulting data in CSV format, export a resulting data in JSON format, or select data points for labelling task, from left to right icons, respectively. In #3 bar area there is an icon that allows the user to create a new group (e.g., place ID) for selected data points e.g., grouping new APs. Area pointed by #4 marker is the list of all place IDs from which the user can select to assign, edit, or delete AP labels by their place IDs. Marker #5 points at the main graphics display.

In the O-D mode (when the OD icon is clicked on), the user can choose to display the individual trips or aggregate O-D flows from any selected place of origin (i.e., dorms). In the O-D mode, the user can choose the origin by its place ID, subject by its device ID, and start/end times from which the tool displays the result with 3D graphics as well as statistical information regarding top destinations, as shown in Fig. 7.

When the user selects a particular subject by its device ID, the tool then allows the user to choose to view the selected subject's trips in the forms of most likely routes used, O-D flows, waypoints, and/or heatmap by clicking on Routes, O-D, Waypoints, and Heatmap options, respectively as shown in Fig. 8. The most likely route used was estimated using the Google Directions API[9] based on the origin, destination, and waypoint geolocations (generated by the Algorithm 4). For demonstration, a video clip showing how Wi-Flow tool works is available at https://youtu.be/K-k41aTmOPA. The actual Wi-Flow application is available at https://wiflow.citycontext.info. Note that there might be some slight differences in the user interface and/or results from what's shown in this paper as the tool may have progressed further in its development, such as extending its coverage beyond the main campus (there are three

[9]https://developers.google.com/maps/documentation/directions/overview

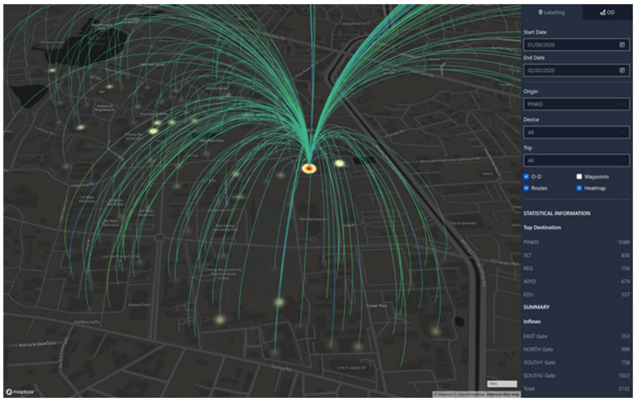
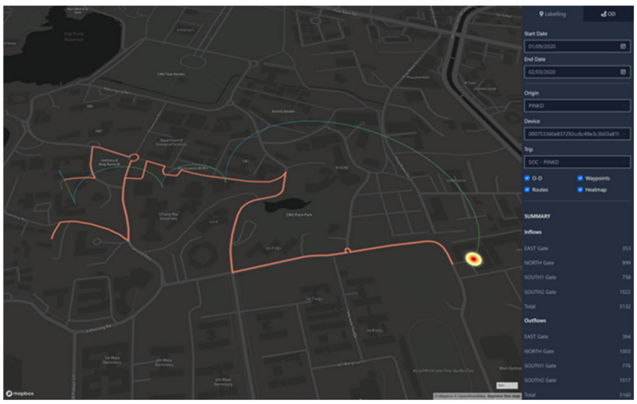


FIGURE 7. **Snapshot Wi-Flow while the user is using the O-D mode with O-D flow option. This example shows outflow from a female dorm (place ID: PINKD).**



(a)

(b)

FIGURE 8. **Snapshots of the O-D mode displaying both routes and O-D flows: (a) an individual trip and (b) all trips made by a selected subject.**

campuses: main, medical science, and agro-industry campuses). For demonstration purposes, a processed result has been loaded into the tool for the user to view the result of the CMU Wi-Fi network. A sample file containing locations of the CMU Wi-Fi APs is available at https://wiflow.citycontext.info/sample if one wants to try its labelling mode.

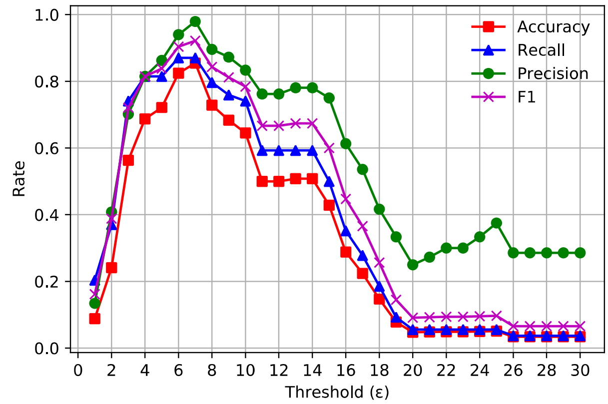


FIGURE 9. **Experimental result of varying the threshold value (ε) from which $\varepsilon = 7$ yields the highest result in terms of accuracy, recall, precision, and F1 value.**

III. RESULTS

To evaluate our developed trip inference, a set of ground truth data was collected using our own developed mobile app that can collect GPS trajectory of the user at a sampling rate of one second. The user identified each trip through the app i.e., the origin and destination were marked. This origin and destination information was later labelled with the place ID for our evaluation. A total of 54 labelled trips were collected as a ground truth, which includes trips that were different in average speed, travel distance, and transport mode. Four transport modes were considered including walking, cycling, motorcycling, and driving.

Our first experiment was to vary the duration threshold (ε) to obtain a value that yields the best performance rates in terms of the accuracy, recall, precision, and F1 value, based on the number of exact matches of the actual and inferred trips (i.e., origin and destination). The result is shown in Fig. 9, where the threshold varies from 1 – 30 minutes, from which the best performance (accuracy = 0.855, recall = 0.870, precision = 0.979, and F1 = 0.922) was found at $\varepsilon = 7$ minutes. So, this ε was used for the rest of our experiment as well as the implementation of our online tool, Wi-Flow.

To further investigate on what may affect the performance of our approach, we examined the performance from the point of view of trip speed, distance traveled, and transport mode. When trips are separated by their average speed, it can be observed that the performance increases with the speed, as shown in Table 1. Intuitively, this may be due to the fact that some places (i.e., faculty, facility, etc.) cover a large area which consequently requires an extended period of time to travel through – and in some cases, it surpasses the threshold (ε) and so causes an incorrect inference of the stay location.

When the performance is considered from the perspective of trip distance, Table 2 shows that the performance values increase with the distance. Trips that were at least 1 km can be detected perfectly. This may be due to the overlapping of nearby places' Wi-Fi coverage areas, as well as the trip speed that is likely to be associated with the distance

TABLE 1. Experimental result based on average trip speed.

Average trip speed (km/hour)	No. of trips	Performance index			
		Accuracy	Recall	Precision	F1
[0, 8)	23	0.680	0.895	0.739	0.810
[8, 16)	11	0.909	1.000	0.909	0.952
[16, 24)	11	0.909	1.000	0.909	0.952
[24, 32)	4	1.000	1.000	1.000	1.000
[32, 40)	5	1.000	1.000	1.000	1.000

TABLE 2. Experimental result based on trip distance.

Trip distance (m)	No. of trips	Performance index			
		Accuracy	Recall	Precision	F1
[0, 500)	14	0.600	0.900	0.643	0.750
[500, 1,000)	17	0.778	0.933	0.824	0.875
[1,000, 1,500)	14	1.000	1.000	1.000	1.000
[1,500, 2,000)	5	1.000	1.000	1.000	1.000
[2,000, 2,500)	4	1.000	1.000	1.000	1.000

TABLE 3. Experimental result based on transport mode.

Transport mode	No. of trips	Performance index			
		Accuracy	Recall	Precision	F1
Walking	17	0.611	0.917	0.647	0.759
Cycling	11	0.833	0.909	0.909	0.909
Motorcycling	13	0.923	1.000	0.923	0.960
Driving	13	1.000	1.000	1.000	1.000

traveled – i.e., a longer trip tends to trigger a higher travel speed [31].

The most common modes of transportation in the CMU campus are walking, cycling, motorcycling, and driving. When trips are separated into these common modes, the performance values gradually increase from walking to cycling to motorcycling and eventually driving. This is in line with the result observed in Table 1, as the trip speed is highly determined by the transport mode used.

These results suggest that generally our trip inference method performs reasonably well with an accuracy rate of 0.855 overall. It works particularly well with motorized modes of transportation such as motorcycling and driving, that can travel at least on average speed of 24 km/hour, which is still under the campus speed limit of 40 km/hour.

So, we implemented our framework on the entire Wi-Fi probe data, which then allows us to observe interesting and useful insights concerning trips made on campus. As shown in Fig. 10(a), when considering the number of trips generated in each day of the week, the result shows that the lowest number of trips per weekday on average occurs on Wednesday (15,050.33 trips/day), which is the day that holds 180-minute lectures that are less offered than 60 or 90-minute lectures on other weekdays, i.e., the 2-day pattern classes (Monday-and-Friday (M-F) and Tuesday-and-Thursday (T-H) classes). Tuesday and Thursday have the highest average numbers of generated trips 20,480 trips/day and 21,244 trips/day, respectively), which is likely due to T-H classes. As reflected by less

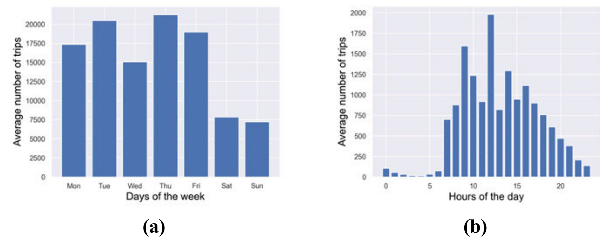


FIGURE 10. Average number of generated trips: (a) the amount of trips observed per day in each day of the week, (b) the amount of trips observed per hour in each hour of the day.

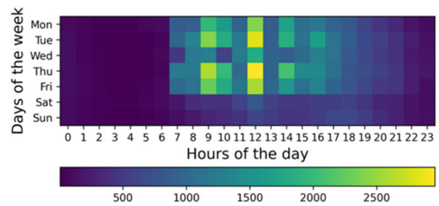


FIGURE 11. Average number of trips observed in each day of the week and in each hour of the day.

crowded trips on Mondays and Fridays (17,346.25 trips/day and 18,961.50 trips/day, respectively), the M-F classes were seemingly less enrolled by the students. Intuitively, the weekends have much lower number of trips than the weekdays as most classes are offered on weekdays. This also tells us that most students tend to stay at their dorms and so don't travel much on weekends or when there isn't a class.

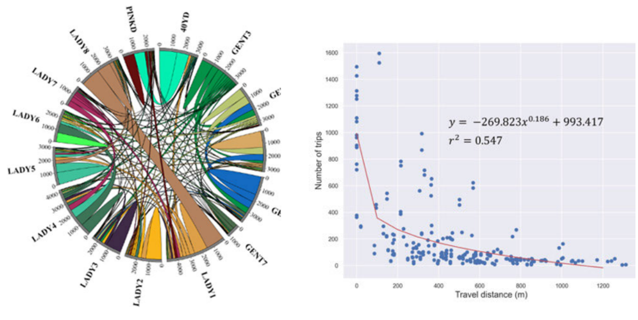
Hourly, trips are observed mostly around noon (on average 1,980.54 trips), which is a lunch time. Second most traffic hour is 9 a.m. period (1,595.42 trips), during which many students are presumably rushing to classes. There were expectedly very few trips made during the curfew period.

Overall, as summarized in Fig. 11, there is a heavy traffic around 9 a.m. hour and noon on Tuesday and Thursday most likely due to the T-H classes. Likewise, the M-F classes seem to also create a similar traffic pattern for Monday and Friday. Expectedly, a much low traffic is on the weekends than the weekdays.

A. TRIP DISTRIBUTION (INTERNAL FLOWS)

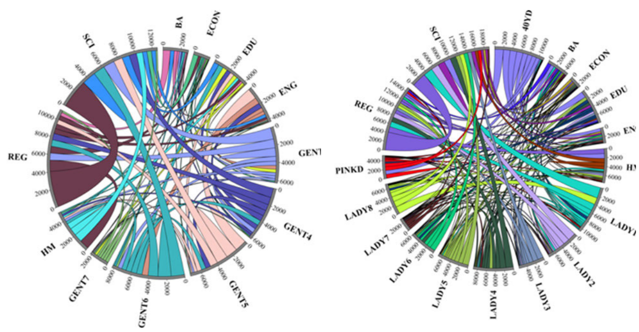
Trip distribution is the second step in the four-step transport planning process [32], which matches the origins with destinations and subsequently models the number of trips that occur between each origin zone and each destination zone (or TAZ). A trip distribution model is then used to describes how trips are distributed across TAZs and predict the spatial pattern of trips or flows between them.

When dorms are considered as TAZs, Fig. 12(a) shows a chord diagram illustrating how trips are distributed across 15 dorms (five male and 10 female dorms). Place IDs for female dorms are LADY1 – LADY8, PINKD, and 40YD, while five male dorms are GENT3 – GENT7. The result shows that there is a high number of cycle trips i.e., trips



(a) (b)

FIGURE 12. Resulting trip distribution among 15 dorms: (a) chord chart showing how trips are distributed among dorms, (b) statistical relationship between travel distance and number of trips.



(a) (b)

FIGURE 13. Chord charts showing how trips are distributed between campus facilities each gender dorm type: (a) male and (b) female dorms.

begin and end at the same TAZ. The number of trips seems to vary with distance between dorms. More trips are distributed to dorms nearby than further away. As shown in Fig. 12(b), the number of trips somewhat has a reverse proportional relationship with the travel distance that can be described by a negative exponential function, $y = -269.823x^{0.186} + 993.417$ with $r^2 = 0.547$, which is in line with previous studies in transportation research [33], [34]. The r^2 value isn't as high as one may expect (though its correlation coefficient value, $r = 0.740$ can already be considered high [35]). This is due to the fact that there are also other influential factors for how trips are generated and distributed, nonetheless the travel distance that often requires a varying level of effort (mostly time spent) as its cost remains one of the top factors.

Commuting is a common and recuring trip made between one's residence and workplace, which constitutes the majority of trips made by individuals and hence is one of the major causes of traffic congestion [29]. As our subjects are most likely students who live in the campus dorms, which can be considered as place of residence and subsequently faculties where they study can be reasonably considered as workplaces – as campus deemed as city.

Flow of trips between male dorms and different faculties across the campus is shown in a chord chart in Fig. 13(a). There are a lot of trips made to and from male dorms and SCI,

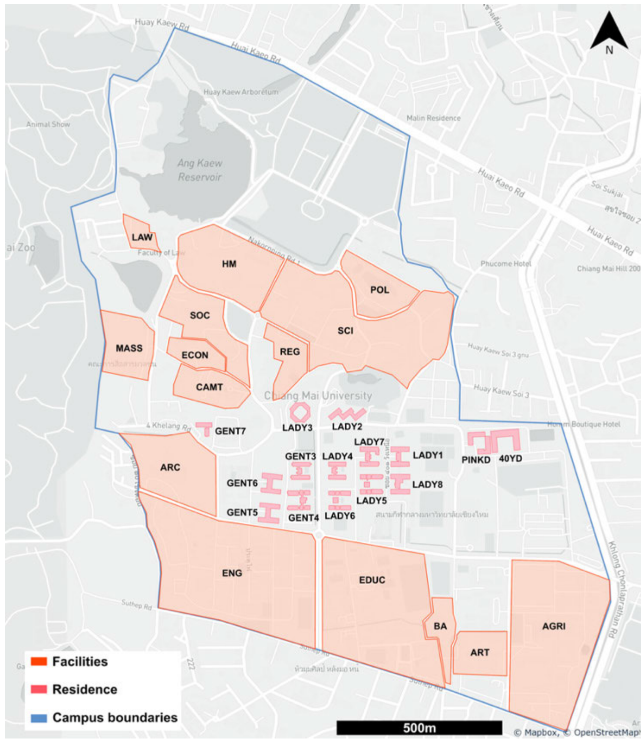
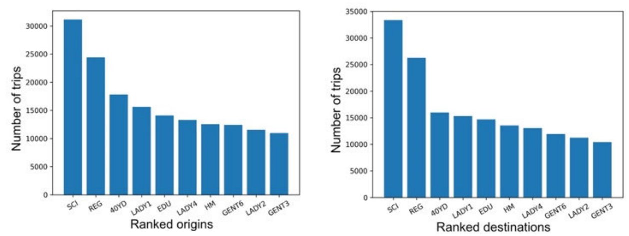


FIGURE 14. Campus map with labelled places.

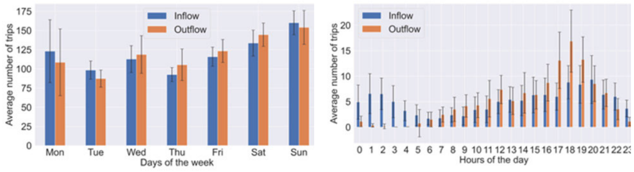


(a) (b)

FIGURE 15. Top 10 (a) origins and (b) destinations.

which is the Faculty of Science, which offer most mandatory introductory courses for all majors. Huge incoming and outgoing flows are also observed for REG, which is the place that hosts the registration office as well as two 3-story buildings that are used for multidisciplinary course lecturing. On the other hand, Fig. 13(b) shows how female trips are distributed. Similar observation with the male trips can be made here where large flows are observed for both SCI and REG. The main difference is the female trip volume is less than male one to ENG, which the Faculty of Engineering where there are many more male than female students. As a spatial reference, Fig. 14 illustrates the campus map with labelled place IDs.

The top 10 places that generate most trips are SCI, REG, 40YD, LADY1, EDU, LADY4, HM, GENT6, LADY2, and GENT3, respectively. Note that EDU is the Faculty of Education, HM is the Faculty of Humanity. On the other hand, the top 10 places that attract most trips are SCI, REG, 40YD, LADY1, EDU, HM, LADY4, GENT6, LADY2, and GENT3, respectively. The number of trips generated and



(a) (b)

FIGURE 16. Residence detection result validation: (a) number of detected residents compared against the actual number of dorm residents, (b) correlation between the number of detected and actual numbers of dorm residents.

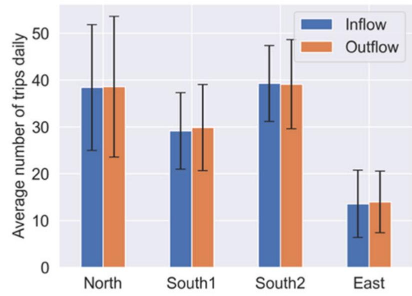


FIGURE 17. Average number of trips through campus gates.

attracted by these top places is shown in Fig. 15. Largest flows are centered around SCI and REG in both directions. Both lists contain the same places within slightly different orders, which intuitively suggest that places that generate lots of trips also attract a comparably large volume of trips as well.

B. CAMPUS INFLOW AND OUTFLOW (EXTERNAL FLOWS)

Managing traffic in and out of the campus or a city area is important for transport planning, which generally requires traffic count information that is costly and laborious. With our ubiquitous sensing approach using Wi-Fi probe data, campus inflow and outflow can be reasonably estimated (with the Algorithm 5). The amount of campus inflow and outflow in terms of the average number of trips in each day of the week is shown in Fig. 16(a), along with a corresponding standard deviation bar. In general, the levels of inflow and outflow are relatively comparable in each day – and hence, intuitively exhibiting an expected balance of flows in both directions. Larger flows are observed on weekends than weekdays, which are likely due to the regular weekday-only-class schedules of most students who then travel in and out the campus more often on the weekends. The M-F, T-H, and W class patterns also play a role in the inflow-outflow travel behavior as the matching patterns also are observed in the result. When flows are inspected hourly, as shown Fig. 16(b), the inflow level is higher than the outflow during the nighttime (late night to early morning periods i.e., 10 p.m. – 5 a.m.), while the opposite trend is observed in the rest of the hours of the day i.e., 6 a.m. – 9 p.m. There is a gradual change from one trend to the other. Outflow has the highest deficit at 6 p.m. period, while the inflow has its highest deficit observed during 2 a.m. period. The results make sense as students may go out of the campus for a dinner in the evening, while late night inflow could be trips made by male students whose dorms have no curfews.

Regarding the flows coming in and out through each gate, Fig. 17 shows the average number of trips daily through each of the four gates. South gate #2 has the highest flow level closely followed by the North gate, then South gate #1, while the East gate has the lowest flow level. There are several street food vendors, restaurants, and shops near the North and South gates located just outside the campus, so these shops are much more accessible to students by using the North and South gates than the East gate, which is mainly used for getting on a highway or traveling to the city center. This result seemingly suggests that most inflow and outflow are presumably associated with food and shopping. We'd like to note that there are also other gates, but smaller than these four main gates and with some restrictions such as opening hours and allowed vehicle types. Inflow and outflow detected by our algorithms are by no means all the trips coming in and out from the campus. Speed of the vehicle can certainly play a big role in the trip detection at the gates. Device connectability is also among other factors for a gate-passing trip to be detected. Nonetheless, the trends observed from the detected inflow and outflow is intuitive and useful for the management's point of view.

IV. CONCLUSION

Conventional approaches to travel behavior data collection such as household survey, traffic count, and roadside interview are costly, laborious, and time-consuming. With today's ubiquitous technology, the opportunistic sensing approach has emerged as an exciting and promising alternative. A ubiquitous computing environment that aims to seamlessly connect people and things anywhere such as Wi-Fi network allows us to capitalize on the opportunistic sensing data for gaining insights into spatial distribution of human mobility. In this work, we developed a framework for inferring and visualizing Wi-Fi data-based travel behavior by demonstrating how a Wi-Fi probe data can be analyzed to infer trips and O-D flows in a case study of campus network. Particularly, we proposed algorithms for inferring spatial presence, residence, stay, trip, and trip distribution among places in the campus, as well as campus inflow and outflow. In addition, to manage the Wi-Fi access point data for the analysis, and visualize the inferred trips and O-D flows, an online visual analytics tool called Wi-Flow was developed with React.js for building our web application with an interactive user interface, along with Deck.gl for rendering a large number of spatial data points (i.e., APs) on a map, as well as providing other geospatial data visualization functionalities such as scatter plot, heatmap, and 3D renderer.

The main difference between our work and other previous studies is our proposed methodology for inferring outdoor mobility in the notion of O-D flows, which is based on our residence detection and trip detection algorithms that produce the result at the individual level as opposed to the overall flow in a spatial network as previously described in the literature. The residence detection result was validated against the actual dumber of dorm residents with a relatively high correlation. Based on a ground truth, our framework performs generally well with a reasonably high accuracy. Particularly, it performs well for trips with an average trip speed of at least 24 km/hour. Based on our framework, the amount of trips detected in different days of the week and hours of the day show the intuitive trends that are in line with the campus teaching schedules. Our internal trip distribution result is in line with the transport modeling principle, where the relationship between the travel distance and the number of trips is described well by a negative exponential function. The internal trip distribution statistically informs us of places with large and small flow magnitudes, so that a better area utilization can be considered. Lastly, the results of the campus inflow and outflow uncover insightful trends temporally on how trips are made in and out of the campus, as well as spatially on how each campus gate is used, which could potentially lead to a more in-depth analysis of trip purpose and activity-based mobility.

There are, nonetheless, some limitations of our work that can be further invested in the future, such as issues that could undesirably affect our trip detection like network connection and device connectability that may cause disconnections and hence information loss in the Wi-Fi logs, and issues with users carrying multiple devices while traveling that may cause over detected trips.

Network disconnection can unpleasantly cause some errors in trip inference in two main scenarios; (1) disconnection occurs at stay location and (2) disconnection occurs at waypoint location(s).

In the Scenario 1, there are two situations – firstly when the disconnection time $< \varepsilon$, and secondly when the disconnection time $\geq \varepsilon$. First situation doesn't affect the trip inference. However, the second situation may affect the stay duration and consequently may have an impact on the trip inference in two possible cases.

Firstly, Case 1.1 is when there is a partial loss of stay presence. This won't affect the trip inference if the remaining stay presence lasts at least ε. Otherwise, this will cause the stay to become misinterpreted waypoint(s). One of the possible approaches to mitigate these incorrect waypoints is to make use of the person's past trips to seek for repeated patterns and consider its likelihood that the person repeats the same trip, such as eating at the canteen during the lunch time or having a class in the lecture hall, for instance. Yet, this issue is interesting and challenging, and hence worth a future investigation on systematic approaches to mitigate or even avoid this kind of erroneous waypoints.

Secondly, Case 1.2 is when there is an entire loss of stay presence. This case will cause an incomplete trip (as described in Section II-F). Depending upon the disconnection time, if the adjacent (nearest) waypoint is still present, then the stay location may be estimated from the trip's direction i.e., heading toward a particular place (considered a destination) or leaving from a particular place (considered an origin). However, if the disconnection time is extensive, then the recovery becomes much more challenging and hence it's worth a future investigation. One of the possible approaches could be using the individual historical trip patterns and its likelihood of repeating the same journey. Nevertheless, a related issue is to first recognize that there's actually a loss of entire stay presence and its adjacent waypoints.

In the Scenario 2, there are two situations to be considered. First one is when the disconnection time $< \varepsilon$, which will not affect the trip inference. In the second situation, however, when the disconnection time $\geq \varepsilon$, this will undesirably affect the trip inference in two cases.

The first case, Case 2.1 is when there is a loss of some waypoints of the trip. There will be a need to reconstruct the trip. So, it'd be an interesting future investigation into how to reconstruct the lost waypoints and recomplete the trip. One of the plausible approaches is to interpolate the trip's waypoints by implementing a trajectory data interpolation method such as Spline, Neural Neighbor, and Local regression (LOESS). This could be complemented by the road network information (e.g., road direction, road type, etc.) as well as taking consideration of other trips' waypoints.

The second case, Case 2.2 is when there is a loss of all waypoints of the trip. In this case, there will be a need for constructing the whole trip's waypoints, which is also worth exploring as a future study. One of the possible solutions could be taking probabilistic approach by utilizing other trips' waypoints along with the road network's spatial data that could provide a good guideline.

In a real-world city environment where there are challenging areas such as tunnels, hills, and regions with precarious connections, these connectivity issues could arise. On the bright side, these issues open up unique opportunities for research and intriguing directions for future development of related topics around trajectory data mining.

Furthermore, collecting a large-scale group truth data is one of the main challenges for the opportunistic sensing approach particularly for a large-scale travel behavior data collection. It is still an open question of how to sufficiently validate results obtained from such data. Recent studies have employed intermediate result validations, such as comparing an intermediate result against publicly available data like census data (as we did in Section II-D), matching the result with the known principles and theories (as we did in Section III-A), and using a small-scale ground truth (as we did in Section III). We hope that these limitations motivate future studies and development of better solutions.

This study introduces an alternative approach to travel behavior data collection in a more convenient, cost efficient, and sustainable way than the conventional approaches. As a campus is like a miniature city (with high-resolution spatial information) that is self-contained community with a variety of buildings, facilities, and amenities that serve the needs of the people who live, work, and study there, so this work is an attempt to realizing the vision of smart city where different types of electronic data collection sensors are employed to supply information which is used to manage assets and resources efficiently – i.e., transportation system in our case.

Accurate mobility estimation also allows us to identify traffic congestion hotspots and areas with high transportation-related carbon emissions. This knowledge enables us to prioritize the deployment of carbon capture technologies in these areas, effectively capturing and mitigating emissions at the source. Furthermore, mobility estimation helps in evaluating the effectiveness of carbon capture initiatives. By monitoring changes in mobility patterns over time, we can assess the impact of various interventions, such as the introduction of electric vehicles or improvements in public transportation infrastructure. This data allows us to optimize carbon capture strategies and make informed decisions to reduce overall carbon emissions.

In the future, such Wi-Fi probe data might be able help us understand semantics attached to trips such as what people are doing through their Wi-Fi activity. Other opportunistic sensing data sources could also be exploited in the future, such as GPS traces from a running app like Strava or Nike Run, which are also increasing available. We hope that this work helps pave the way for cities as well as miniature cities like university campuses to better understand people's behavior and life, especially after the COVID that has changed the way we live and work.

ACKNOWLEDGMENT

The authors would like to thank the Information Technology Service Center (ITSC), Chiang Mai University, for providing the Wi-Fi probe data for this research.

REFERENCES

[1] M. L. Manheim, *Fundamentals of Transportation Systems Analysis*. Cambridge, MA, USA: MIT Press, 1979.

[2] M. Ben-Akiva and S. Lerman, *Discrete Choice Analysis*. Boston, MA, USA: MIT Press, 1985.

[3] J. Ortúzar and L. G. Willumsen, *Modelling Transport*, 4th ed. Chichester, U.K.: Wiley, 2011.

[4] L. Shen and P. R. Stopher, "Review of GPS travel survey and GPS data-processing methods," *Transp. Rev.*, vol. 34, no. 3, pp. 316–334, 2014, doi: 10.1080/01441647.2014.903530.

[5] GDPR.EU. *Complete Guide to GDPR Compliance*. Accessed: Mar. 15, 2022. [Online]. Available: https://gdpr.eu

[6] U.S. Department of Health & Human Services. *Summary of the HIPAA Privacy Rule*. [Online]. Available: Accessed: Mar. 15, 2022. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

[7] A. Cuttone, S. Lehmann, and M. C. González, "Understanding predictability and exploration in human mobility," *EPJ Data Sci.*, vol. 7, no. 1, pp. 1–7, Dec. 2018, doi: 10.1140/epjds/s13688-017-0129-1.

[8] S. Phithakkitnukoon, T. Horanont, A. Witayangkurn, R. Siri, Y. Sekimoto, and R. Shibasaki, "Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan," *Pervasive Mobile Comput.*, vol. 18, pp. 18–39, Apr. 2015, doi: 10.1016/j.pmcj.2014.07.003.

[9] S. Phithakkitnukoon, T. W. Leong, Z. Smoreda, and P. Olivier, "Weather effects on mobile social interactions: A case study of mobile phone users in Lisbon, Portugal," *PLoS ONE*, vol. 7, no. 10, Oct. 2012, Art. no. e45745, doi: 10.1371/journal.pone.0045745.

[10] M. G. Demissie, S. Phithakkitnukoon, T. Sukhvibul, F. Antunes, R. Gomes, and C. Bento, "Inferring passenger travel demand to improve urban mobility in developing countries using cell phone data: A case study of Senegal," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 9, pp. 2466–2478, Sep. 2016, doi: 10.1109/TITS.2016.2521830.

[11] S. Phithakkitnukoon, S. Hankaew, M. G. Demissie, Z. Smoreda, and C. Ratti, "Temporary migration flow inference and analysis from perspective of mobile phone network data," *IEEE Access*, vol. 10, pp. 23248–23258, 2022, doi: 10.1109/ACCESS.2022.3154485.

[12] M. G. Demissie, S. Phithakkitnukoon, and L. Kattan, "Trip distribution modeling using mobile phone data: Emphasis on intra-zonal trips," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 7, pp. 2605–2617, Jul. 2019, doi: 10.1109/TITS.2018.2868468.

[13] S. Hankaew, S. Phithakkitnukoon, M. G. Demissie, L. Kattan, Z. Smoreda, and C. Ratti, "Inferring and modeling migration flows using mobile phone network data," *IEEE Access*, vol. 7, pp. 164746–164758, 2019, doi: 10.1109/ACCESS.2019.2952911.

[14] S. Phithakkitnukoon, T. Sukhvibul, M. Demissie, Z. Smoreda, J. Natwichai, and C. Bento, "Inferring social influence in transport mode choice using mobile phone data," *EPJ Data Sci.*, vol. 6, no. 1, pp. 1–29, Dec. 2017, doi: 10.1140/epjds/s13688-017-0108-6.

[15] H. Huang, Y. Cheng, and R. Weibel, "Transport mode detection based on mobile phone network data: A systematic review," *Transp. Res. C, Emerg. Technol.*, vol. 101, pp. 297–312, Apr. 2019, doi: 10.1016/j.trc.2019.02.008.

[16] P. Sakamanee, S. Phithakkitnukoon, Z. Smoreda, and C. Ratti, "Methods for inferring route choice of commuting trip from mobile phone network data," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 5, p. 306, May 2020, doi: 10.3390/ijgi9050306.

[17] J. K. Laurila, D. Gatica-Perez, I. Aad, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen, "The mobile data challenge: Big data for mobile computing research," in *Proc. Work. Nokia Mobile Data Challenge, Conjunction 10th Int. Conf. Pervasive Comput.*, 2012, pp. 1–8, doi: 10.1016/j.pmcj.2013.07.014.

[18] V. D. Blondel, M. Esch, C. Chan, F. Clerot, P. Deville, E. Huens, F. Morlot, Z. Smoreda, and C. Ziemlicki, "Data for development: The D4D challenge on mobile phone data," 2012, *arXiv:1210.0137*.

[19] Y.-A. de Montjoye, Z. Smoreda, R. Trinquart, C. Ziemlicki, and V. D. Blondel, "D4D-Senegal: The second mobile phone data for development challenge," 2014, *arXiv:1407.4885*.

[20] A. Sevtsuk, S. Huang, F. Calabrese, and C. Ratti, "Mapping the MIT campus in real time using WiFi," in *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*. Hershey, PA, USA: IGI Global, 2008, pp. 326–328.

[21] F. Calabrese, J. Reades, and C. Ratti, "Eigenplaces: Segmenting space through digital signatures," *IEEE Pervasive Comput.*, vol. 9, no. 1, pp. 78–84, Jan./Mar. 2010, doi: 10.1109/MPRV.2009.62.

[22] S. Chaikul, S. Phithakkitnukoon, and C. Ratti, "Xplaces: Segmenting physical space through Wi-Fi traces using eigendecomposition and X-means," *IEEE Access*, vol. 10, pp. 17377–17388, 2022, doi: 10.1109/ACCESS.2022.3150006.

[23] F. Meneses and A. Moreira, "Large scale movement analysis from WiFi based location data," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Nov. 2012, pp. 1–9, doi: 10.1109/IPIN.2012.6418885.

[24] A. Trivedi, K. Silverstein, E. Strubell, P. Shenoy, and M. Iyyer, "WiFiMod: Transformer-based indoor human mobility modeling using passive sensing," in *Proc. 4th ACM SIGCAS Conf. Comput. Sustain. Soc. (COMPASS)*, 2021, pp. 1–18, doi: 10.1145/3460112.3471951.

[25] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, Feb. 2010, doi: 10.1126/science.1177170.

[26] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila, "Towards rich mobile phone datasets: Lausanne data collection campaign," in *Proc. ACM Int. Conf. Pervasive Services (ICPS)*, 2010, pp. 1–7.

[27] C. E. Kontokosta and N. Johnson, "Urban phenology: Toward a real-time census of the city using Wi-Fi data," *Comput., Environ. Urban Syst.*, vol. 64, pp. 144–153, Jul. 2017, doi: 10.1016/j.compenvurbsys.2017.01.011.

[28] M. W. Traunmueller, N. Johnson, A. Malik, and C. E. Kontokosta, "Digital footprints: Using WiFi probe and locational data to analyze human mobility trajectories in cities," *Comput., Environ. Urban Syst.*, vol. 72, pp. 4–12, Nov. 2018, doi: 10.1016/j.compenvurbsys.2018.07.006.

[29] W. Vickrey, "Congestion theory and transport investment," *Amer. Econ. Rev.*, vol. 59, no. 2, pp. 251–260, Feb. 1969, doi: 10.2307/1823678.

[30] T. Horanont and S. Phithakkitnukoon, "Sensing urban density using mobile phone GPS locations: A case study of Odaiba area, Japan," in *Proc. Int. Conf. Nature Comput. Commun.*, in Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2015, pp. 146–155, doi: 10.1007/978-3-319-15392-6_15.
[31] C. D. van Goeverden, "The value of travel speed," *Transp. Res. Interdiscip. Perspect.*, vol. 13, pp. 1–7, Mar. 2022, doi: 10.1016/j.trip.2021.100530.
[32] M. G. Mcnally, "The four step model," in *Handbook of Transport Modelling*, 2007, doi: https://escholarship.org/uc/item/0r75311t
[33] M. Florian and S. Nguyen, "A combined trip distribution modal split and trip assignment model," *Transp. Res.*, vol. 12, no. 4, pp. 241–246, 1978, doi: 10.1016/0041-1647(78)90065-5.
[34] L. Xiao-Qiang, "The proof of gravity model with negative exponential land-mixed entropy and similar to the method of hyman calibration technology," *Open Civil Eng. J.*, vol. 9, no. 1, pp. 1035–1040, 2015, doi: 10.2174/1874149501509011035.
[35] M. M. Mukaka, "Statistics corner: A guide to appropriate use of correlation coefficient in medical research," *Malawi Med. J.*, vol. 24, no. 3, pp. 69–71, 2012.

THANISORN JUNDEE received the B.Eng. degree in information systems and network engineering from Chiang Mai University, Thailand, where he is currently pursuing the degree with the Department of Computer Engineering. His research interests include urban data science and visual analytics.

SANTI PHITHAKKITNUKOON received the B.S. and M.S. degrees in electrical engineering from Southern Methodist University, Dallas, TX, USA, in 2003 and 2005, respectively, and the Ph.D. degree in computer science and engineering from the University of North Texas, Denton, TX, USA. He is currently an Associate Professor with the Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Thailand. Before joining Chiang Mai University, he was a Lecturer in computing with The Open University, U.K., a Research Associate with Newcastle University, U.K., and a Postdoctoral Fellow with the SENSEable City Laboratory, Massachusetts Institute of Technology, USA. His research interest includes the area of urban informatics.

CARLO RATTI received the Ph.D. degree in architecture from the University of Cambridge. He is currently an Architect and an Engineer who practices architecture in Turin and teaches with the Massachusetts Institute of Technology (MIT), where he directs the SENSEable City Laboratory. His research interests include urban design, human–computer interfaces, electronic media, and the design of public spaces. He is a member of the Ordine degli Ingegneri di Torino, the Architects Registration Board, U.K., and the Association des Anciens Élèves de l'École Nationale des Ponts et Chaussées.