

W4-Groups: Modeling the Who, What, When and Where of Group Behavior via Mobility Sensing

AKANKSHA ATREY, University of Massachusetts Amherst, USA

CAMELLIA ZAKARIA, University of Toronto, Canada

RAJESH BALAN, Singapore Management University, Singapore

PRASHANT SHENOY, University of Massachusetts Amherst, USA

Human social interactions occur in group settings of varying sizes and locations, depending on the type of social activity. The ability to distinguish group formations based on their purposes transforms how group detection mechanisms function. Not only should such tools support the effective detection of serendipitous encounters, but they can derive categories of relation types among users. Determining *who* is involved, *what* activity is performed, and *when* and *where* the activity occurs are critical to understanding group processes in greater depth, including supporting goal-oriented applications (e.g., performance, productivity, and mental health) that require sensing social factors. In this work, we propose *W4-Groups* that captures the functional perspective of variability and repeatability when automatically constructing short-term and long-term groups via multiple data sources (e.g., WiFi and location check-in data). We design and implement *W4-Groups* to detect and extract all four group features *who-what-when-where* from the user's daily mobility patterns. We empirically evaluate the framework using two real-world WiFi datasets and a location check-in dataset, yielding an average of 92% overall accuracy, 96% precision, and 94% recall. Further, we supplement two case studies to demonstrate the application of *W4-Groups* for next-group activity prediction and analyzing changes in group behavior at a longitudinal scale, exemplifying short-term and long-term occurrences.

Additional Key Words and Phrases: Group modeling, social interactions, user mobility, next activity prediction

ACM Reference Format:

Akanksha Atrey, Camellia Zakaria, Rajesh Balan, and Prashant Shenoy. 2024. W4-Groups: Modeling the Who, What, When and Where of Group Behavior via Mobility Sensing. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 150 (April 2024), 29 pages. <https://doi.org/10.1145/3637427>

1 INTRODUCTION

Humans are social by nature, and many of our everyday routines involve social interactions with others. For example, activities such as having meals, watching movies, and attending meetings and classes are commonly conducted in groups; with family, friends, and colleagues. Many of these daily routines that involve such group interaction also recur periodically, often involving the same group, yielding four key group interaction characteristics: with *who* is the activity performed, *what* activity is performed, *when* and *where* the interaction occurred. Accounting for these group dynamics – *who-what-when-where* (W4) – for a user reveals many vital aspects of their behavior, including productivity [62], performance [63], and mental and social health [3, 75].

Authors' addresses: Akanksha Atrey, aatrey@cs.umass.edu, University of Massachusetts Amherst, USA; Camellia Zakaria, camellia.zakaria@utoronto.ca, University of Toronto, Canada; Rajesh Balan, rajesh@smu.edu.sg, Singapore Management University, Singapore; Prashant Shenoy, shenoy@cs.umass.edu, University of Massachusetts Amherst, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2024/4-ART150 \$15.00

<https://doi.org/10.1145/3637427>

The rise of the Internet, smartphones, and inexpensive sensing technologies have facilitated the existence of digital nomads, referring to individuals who can work from anywhere [29]. The defining characteristics of “endless mobility” and “location-independent lives” of digital nomads [35] have spurred strong interest among researchers to employ various digital means to understand an individual’s group behavior, which can occur in many community-based spaces. For example, social media data has been subjected to social network analyses to produce social graphs of a user, reflecting their social circle and users with whom physical interactions are likely to occur. Similarly, smartphone sensors such as Bluetooth for proximity sensing and GPS for location sensing have been used to approximate face-to-face interaction among groups [5, 36]. While much work focuses on modeling, detection, and understanding group behaviors, these efforts have focused on some but not all of the *W4* characteristics in tandem. Prior work assumed that groups are known a priori and have focused on what activities are performed, generally, among students on campus [63] or leverage on gathering data from serendipitous group encounters [10]. We argue that a generalized group detection framework must accomplish two essential criteria: (1) distinguish short-term and long-term group formations based on their underlying purposes, and (2) support multiple data types to accommodate diverse applications that rely on group sensing mechanisms.

In this work, we model group interactions of a user by emphasizing all four characteristics, *who-what-when-where*, of the group dynamics to obtain a holistic view of the user’s social interaction. Our approach, grounded by the theoretical understanding of groups in *small group research* [44], utilizes mobile trajectories of users as the basis for inferring *W4* group behavior. *Who*, in this case, is a proxy of the device modality producing these mobility trajectories unless the device owner gives explicit permission to be de-anonymized and detected. These mobile trajectories consist of a sequence of locations that users visit and transition from one location to the next. By intersecting the mobile trajectories of users, we can infer common locations *where* they are present at the same time or common paths *when* they move together. These intersections yield insights into when, where, and who and can be augmented with *what* activity is being performed. A key challenge in designing such a technique is to avoid false positives from chance colocations. For example, strangers visiting a coffee shop where a user is present may be falsely identified as a group. The number of such false positives can be significant in public spaces, requiring modeling techniques to distinguish them from true interactions in short-term or long-term groups. The generalization capabilities of *W4-Groups* in capturing *W4* characteristics using social science theory across sensing modalities distinguishes it from prior work.

In modeling, implementing, and evaluating *W4-Groups*, we make the following contributions:

- (1) We present *W4-Groups*, a group modeling framework, driven by social science perspective to derive four key characteristics distinguishing short-term and long-term group formations based on their underlying purposes. Via any mobility sensing means, the system captures group interactions beyond mere colocation by recursively aggregating users into larger groups based on their mobility rhythms.
- (2) We evaluate and validate our approach using multiple mobility datasets and show that *W4-Groups* yields high performance and generalizability in inferring individual and group behaviors even using noisy, coarse-grained WiFi data and incomplete social network check-in data. Through a user study, our method achieves up to 93.49% accuracy and outperforms various baselines and state-of-the-art techniques. Overall, on three separate datasets, *W4-Groups* yields an average of 92% overall accuracy, 96% precision, and 94% recall.
- (3) We show the utility of *W4-Groups* in supporting a next-activity prediction application for short-term group interaction and demonstrate the value of *W4-Groups* in analyzing longitudinal behavioral change as another important application use.

The rest of the paper is divided as follows. In Section 2, we motivate the need and design rationale for *W4-Groups*. Section 3 discusses the system architecture and modeling approach. We describe *W4-Groups*'s prototype implementation in Section 4. Sections 5 and 6 are system evaluation and case studies demonstrating the applicability of *W4-Groups*. We discuss the privacy and ethical implications of our work affecting the CSCW community in Section 7. Section 8 lists the literature review of group modeling technologies, before concluding the paper in Section 9.

2 MOTIVATING W4-GROUPS

This section motivates our approach to automatically model groups based on the theoretical perspectives in small group research. We describe how our work extends to provide an understanding of the ontological relationships between users and their detected group.

2.1 Supporting Group Modeling Applications

There has been increasing research in the CSCW space where group modeling techniques are applied to gather behavioral data streams of individuals and their groups on topics related to health, engagement, and performance. Alder et al.'s study on how passive personal sensing technologies could help resident physicians with their burnout symptoms revealed the potential for tracking a resident's work deviations from their cohort's norm [1]. Separately, Zakaria et al. reported a significant predictor of stress is the deviations between the student and their group's norm, distinguishing casual groups from workgroups [75] and their social identification to respective groups [76]. In other applications, Fosh et al. demonstrated how groups often reconfigured themselves during a museum visit, balancing their engagement with exhibits with the group's social dynamic [20]. Beyond CSCW scholarships are works that uncover group membership in a crowd or public environment to strategize business opportunities [12, 20, 56], facilitate surveillance of work dynamics [17], and pandemic-preparedness [77].

These cases exemplify the need to detect formed groups across a wide range of scenarios and sensing modalities, similarly motivating the works by de Freitas et al. to propose a group context framework that allows devices to form groups autonomously and share context [10, 11]. However, this framework offers better support for applications intended to increase the opportunities for serendipitous group formation.

2.2 Defining a Group

We argue for group modeling techniques to account for critical parameters that reflect real-world formations – not only effective in modeling serendipitous groups but deriving categories of relation types among users – driven by social science perspectives.

By definition, a 'Group' is two or more individuals connected by social relationships. No one description can capture the nuances of what constitutes a group as the relationship can be on a spectrum of properties from primitive forms, such as sharing a family name to more complex factors, such as the same psychological significance or the skills needed to achieve an agreed-upon goal [19]. Unlike prior work that focuses on opportunistically creating group membership by automatically searching for similar contexts [10], our work begins by modeling groups grounded on the theoretical understanding of small groups. We adopt the definition of group as per McGrath et al., where "a group is an aggregation of two or more people who are to some degree in dynamic interrelation with one another" [44].

Here, we define interrelation as the involvement of an individual's routine behaviors and the person around them, and the context of the environment in promoting and terminating these routines. In truth, an individual maintains regularity in daily activities throughout the week, either alone, with one or more persons [45]. The more one **repeats** the activity, the higher the likelihood

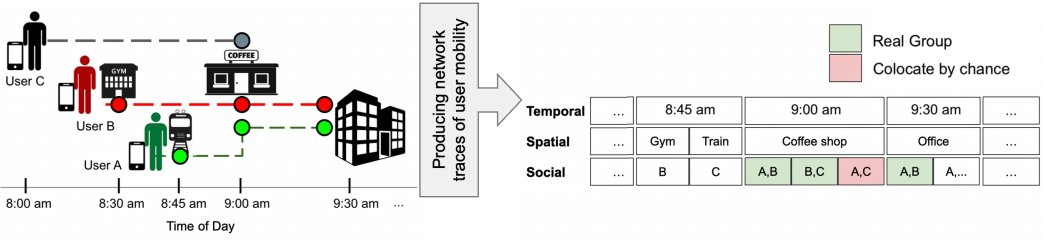


Fig. 1. From left to right, network trace of user mobility can produce three key attributes *who-what-when-where* in the group, reflecting the changes in user location throughout the day. We can infer the pairwise interaction of users collocating by processing these traces collectively at each place over time. Multiple collocations of the same users increase their certainty of existing as a group rather than collocating by chance.

they are routine activities. The higher the **variability** in these activities, the higher the likelihood they are non-routine activities. We pay attention to daily routine activities that only occur with one or more persons for our group detection mechanism. In terms of size, a group can be in the range of two (e.g., small collectives) to the thousands (e.g., crowd) [58]. Small collective groups typically range from two to seven members [28]. Arguably, crowds constitute sets of interlocked small collective groups, albeit gravitating to the smallest size of two [25]. We focus on four characteristics:

- *When* an interaction occurs, defined by **temporal** features such as duration and, accordingly, its recurrence.
- *Where* an interaction occurs, defined by **spatial** features such as the location and, accordingly, the characteristics of a place.
- *Who* is involved in the interaction, defined by **social** features the users themselves represent, the connections they make with one another, and the frequency of these connections.
- *What* type of interaction will likely occur, defined by the activity involved, combining and assessing the three attributes above.

To formally capture these characteristics, we define a group interaction as:

DEFINITION 1. (Group) A group session g is a co-occurrence of multiple users (≥ 2) over a spatiotemporal scale for the same interaction type and defined as

$$g = \{\mathcal{U}, e, d, \mathcal{L}, a\} \quad (1)$$

where \mathcal{U} represents the set of users in the group, e is entry time of all the group members, d is departure time of all the members, $\mathcal{L} = \bigcup_{i \in \mathcal{U}} l_i$ represents all the unique locations \mathcal{U} visited, and $a = \text{mode}_{i \in \mathcal{U}}(a_i)$ represents the most common activity being conducted.

2.3 Design Rationale

With the prerequisite of a sensing modality, in this case, we assume the user's smartphone with location-based information (i.e., Bluetooth, GPS, WiFi), we can observe the changes in location as they move from one place to the next. Figure 1 describes how mobility data of a typical nomadic user can be interpreted at every interval through 24 hours to infer pairwise group interaction with another user.

As users move or remain stationary in a particular place, their device simply logs the present location. To that degree, the device stands as a proxy of the user. The trajectories of multiple users put together allow us to identify intersecting movements, thereby constructing the following three key attributes; *spatial*, *temporal*, and *social*. The short-lived synchrony between Users A and C and Users B and C at the coffee shop can be interpreted as chance collocation compared to Users A and B, as

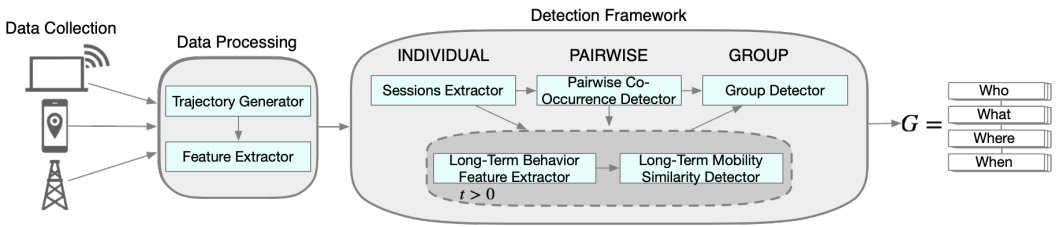


Fig. 2. *W4-Groups* overview consisting of two modules: data processing and group detection.

they would later transition together to an office location and be interpreted as a real group. However, over different time scales of days, weeks, and months, the interpretation of these interactions can change for each pair with repeatable occurrences and their variances over time. The above example yields several critical insights to consider in designing a group detection system. Particularly, an inferred pairwise interaction is more likely to be a real group when their colocations persist across diverse locations (*spatial similarity*), time periods (*temporal similarity*), and involve individuals from the same social circles (*social similarity*). Thus, by capturing the functional perspective of *variability* and *repeatability* when automatically constructing a group, these parameters can help provide a distinctive view of mobility-defined groups and group processes.

3 W4-GROUPS SYSTEM DESIGN

This section provides an overview of the system's architecture, followed by a detailed description of the design of each component.

3.1 System Overview

The architectural overview of our system is depicted in Figure 2. In the design of *W4-Groups*, we overcome several non-trivial challenges associated with mobility traces.

- (1) *Incomplete and Noisy Data*: Depending on its source, mobility data can range in accuracy and completeness. For example, WiFi traces of connected devices at different access points (AP) are coarse-grained [54]. However, Received Signal Strength Indicator (RSSI) values can detect more precise information [59]. Location-based social networks can be used to gather accurate check-in data but often lack long-term temporal completeness. Thus, our system must consider flexibility in working with varying data granularity.
- (2) *Accurate Detection of Real Groups*: Building on the aforementioned, the type of mobility data can potentially restrict the system's ability to determine the user behind these trajectories. This limitation will result in our system grouping users that may colocate by chance and not for an actual purpose. Our approach must filter out these unintentional groups to accurately detect real groups.
- (3) *Computational Complexity*: Since our system looks at every pairwise interaction, determining if a pair belongs to a larger group will increase computational complexity (e.g., $O(n^2)$ in the simplest case). However, this does not scale for large systems with many users. Our approach reduces the complexity through careful data selection.

The understanding of conceptualizing groups and the four key attributes, described in Section 2, introduced two important factors: *repeatability* and *variability*. Prior works in group detection systems have considered repeated colocations in the time scale of minutes and hours as a measure of real groups [54]. Our system extends this approach to consider long-term dependencies.

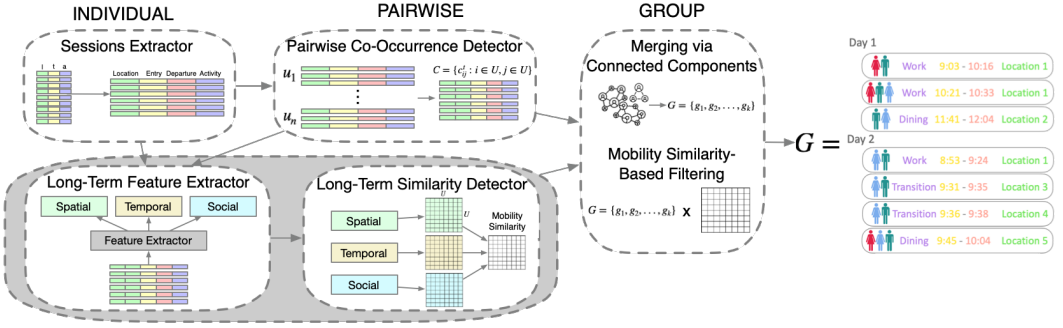


Fig. 3. The inner workings of *W4-Groups*'s modeling approach to produce *who-what-when-where* as its output.

3.2 Data Processing Module

Let U be a set of users. We represent user i 's mobility trajectories as

$$M_i = \{\{l_1, t_1\}, \{l_2, t_2\}, \dots, \{l_k, t_k\}\} \quad (2)$$

where l represents a location (e.g., AP) and t represents a timestamp. These mobility trajectories can be used to extract contextual features such as location type (b) and activity (a). Location types, b_k , are extracted manually using the buildings' primary purpose. For instance, we can map all campus buildings in WiFi mobility traces into the following categories: administration, dining, health, labs, landmark, library, parking, police, recreation, residential, school, student organizations, and other. This information can be easily extracted from public sites of colleges, workplaces, and/or cities.

We define activity type as the primary activity performed by the user. For each log event in M_i , an activity is extracted as a function of time (t_k) and location (l_k). We extract six activity types: dining, gym, home, work, transition, and others. *Dining*, *gym*, and *home* activities are extracted using location being a dining location, recreation center, and housing, respectively. *Work* activities are detected using extended periods spent in a particular office building between 8:00 am and 6:00 pm. For a campus setting, we identify facilities that fall into the library, administration, labs, and school type as office buildings. Lastly, *transitions* are detected if the user spends less than ten consecutive minutes at a given location. If the activities do not fall in the remaining five, they are labeled as *other*. This breakdown allows practitioners to associate individual trajectories with tasks and activities and becomes more relevant when dealing with noisy location data such as WiFi signals. Note that the definition of activity is kept generic to protect the user's privacy and can be easily expanded on if explicit permission is received from the user.

3.3 Group Detection Module

As described in Section 2.3, modeling groups to consider *who-what-when-where* is primarily based on three primary metrics: *temporal*, *spatial*, and *social similarities*. We now present the modeling approach used by *W4-Groups*'s group detection module, depicted in Figure 3.

3.3.1 Modeling Sessions. The first component, sessions extractor, processes features of time and location to output sessions of periods in a vicinity. It aims to minimize errors produced due to poor-quality data by dividing a user's trajectories into periodic sessions. We adopt a strategy proposed by Trivedi et al. to process WiFi-based mobility trajectories, M , by considering all consecutive locations on a single floor of a building (i.e., indoor localization) or all consecutive building locations (i.e., outdoor localization). Note that the degree of information (i.e., room, floor, and building granularity)

depends on the localization capability. In doing so, this sub-component outputs **sessions**, which represent extended periods spent in a particular vicinity [64].

DEFINITION 2. (Session) Given M_i , we define each session for user i at time instance t as $s_i^t = \{l, e, d, a\}$ where l represents location, e represents entry time, d represents the departure time and a represents activity type for user i .

A key factor in determining periodic sessions requires temporal information to estimate how long a user spent in a particular location. For incomplete datasets such as check-in data, such information is not available. Thus, when continuous temporal information is unavailable and location information is sparse, each location can be considered a single session. The temporal information can be divided into periods (i.e., 10-minute windows) or time estimates using external means such as location type.

Note that the features in each session are represented as a set since each session can consist of more than one location (i.e., a user who fluctuates between two nearby locations will be classified into a single session). We aggregate these sessions to form a session trajectory over time for user i

$$S_i = \{s_i^1, s_i^2, \dots, s_i^k\} \quad (3)$$

3.3.2 Modeling Pairwise Co-Occurrence. With each user's session computed, the pairwise co-occurrence detector seeks to discover intersecting trajectories between two users. We define a **co-occurrence event** m_{ij} between users i and j .

DEFINITION 3. (Co-occurrence) Given sessions s_i^t and s_j^t of users i and j respectively, if the session locations and times overlap, $m_{ij}^t = 1$ else $m_{ij}^t = 0$.

$$m_{ij}^t = \begin{cases} 1, & \text{if } |s_i^t[l] \cap s_j^t[l]| > 0 \text{ and } s_i^t[e] \leq s_j^t[d] \text{ and } s_j^t[e] \leq s_i^t[d] \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Based on the positive co-occurrence events, we can extract all **pairwise co-occurrences** $C = \{c_{ij}^t : i \in U, j \in U\}$. Let c_{ij}^t represent a co-occurrence between user i and j at time instance t .

$$c_{ij}^t = \{e, d, l_i, l_j, a\} \quad (5)$$

As mentioned earlier, this step presents scalability challenges when dealing with a large number of users. Specifically, co-occurrence detection exhibits a $O(n^2)$ complexity where each user's mobility trajectory is compared with every other user. To support a wide range of users, we develop an alternate method to reduce the complexity by carefully selecting pairs for comparison and terminating early when the overlap between pairs is found. Appendix A.1 provides a detailed description of this algorithm.

Note that this component is sufficient in detecting accurate pairwise co-occurrences when dealing with fine-grained location trajectories such as continuous GPS coordinates. However, coarse-grained and incomplete location trajectories will lead to the detection of more false positive pairwise groups. Thus, we employ long-term mobility behavior, described in Sections 3.3.3 and 3.3.4, to filter out such false positives and understand groups beyond mere co-occurrence.

3.3.3 Modeling Long-Term Mobility Behavior and Feature Extraction. Prior works identify longitudinal similarity between two mobility trajectories using distance measures (e.g., Euclidean distance) [56] or time-series correlations [54]. However, these techniques can be inefficient when processing coarse-grained mobility measures like AP connectivity. Moreover, such methods do not allow the inclusion of additional contextual features that are indicative of patterns or routines. Lastly, most existing approaches do not account for the dynamic nature of mobility trends; human interaction

Table 1. Behavioral factors extracted at spatial, temporal, and social scales represent repeatability (Rep) and variability (Var) concepts. Each factor is extracted at the user level and pairwise level.

Feature	Scale	Description	Concept		Level	
			Rep	Var	User	Pair
f^1	Spatial	Number of unique locations visited	✗	✓	✓	✓
f^2	Spatial	Average number of location visit counts	✓	✗	✗	✓
f^3	Temporal	Total time spent in minutes	✓	✗	✓	✓
f^4	Temporal	Number of unique minutes spent	✗	✓	✓	✓
f^5	Social	Number of unique users met	✗	✓	✓	✓
f^6	Social	Number of interactions across all users	✓	✗	✓	✓

and group behavior change and evolve over time. We argue that this is a key factor in identifying generalizable group behavior.

In this work, we define the long-term mobility behavior of individual users and pairs of users in terms of their *spatial*, *temporal*, and *social* attributes, as reasoned in Section 2.2. This component takes as input user sessions from the sessions extractor component (see Section 3.3.1) and the pairwise co-occurrence matrix from pairwise co-occurrence detector (see Section 3.3.2). The output is a set of behavioral factors. All factors are calculated for the period of the long-term data at a user level and pairwise level. We examine different time periods in Section 5.3.

The computation for each group attribute (*spatial*, *temporal*, and *social*) is based on repeatability and variability over the day, as per Table 1. For *spatial*, *num_unique_locations* is extracted to represent the variability in the locations visited, and *avg_location_counts* is extracted to represent the average repeatability in the number of visits to each location. For *temporal*, we extract *num_unique_minutes* to represent the variability in the times of the day an individual or pair of users is seen together. We extract *total_time_spent* to represent repeatability in time spent (individual-level) and together (pairwise-level). For *social*, *num_unique_users* and *num_interactions* are extracted to represent variability in the number of unique users met and repeatability in the number of interactions across users at the social scale.

3.3.4 Long-Term Mobility Similarity Detection. In this component, we develop similarity metrics using the features computed in Section 3.3.3 and the Jaccard similarity measure. Jaccard similarity is a statistic used to understand similarity and diversity among sets as follows.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

We employ Jaccard similarity as a measure to detect homogeneity among user sessions S at spatial, temporal, and social scales using the features extracted in Section 3.3.3. Each feature in Table 1 represents a count of some characteristic in S . For spatial features, this characteristic is location l . For temporal features, this characteristic is entry and departure times e and d . For social features, this characteristic is the relationship with other users in pairwise co-occurrence sessions C .

Let S_i^k and S_j^k represent the sets of k^{th} feature of the mobility trajectories for i^{th} and j^{th} users respectively. The intersection between sessions can be determined using the pairwise-level features such that $|S_i^k \cap S_j^k| = f_{ij}^k$. The union between sets can be calculated using simple set operations between individual and pairwise-level features such that $|S_i^k \cup S_j^k| = f_i^k + f_j^k - f_{ij}^k$ where $f_i^k = |S_i^k|$ and $f_j^k = |S_j^k|$. Thus we define the similarity between users i and j using behavioral features as:

$$J(S_i^k, S_j^k) = \frac{f_{ij}^k}{f_i^k + f_j^k - f_{ij}^k} \quad (6)$$

Accordingly, the definitions of spatial, temporal and social similarities are:

DEFINITION 4. (Spatial Similarity) Let spatial similarity $\Phi_{ij}^{spatial}$ between users i and j represent the homogeneity in spatial features as represented by the number of unique locations visited together versus alone weighted by the frequency of visits.

$$\Phi_{ij}^{spatial} = J(S_i^1, S_j^1) f_{ij}^2 \quad (7)$$

DEFINITION 5. (Temporal Similarity) Let temporal similarity $\Phi_{ij}^{temporal}$ between users i and j represent the homogeneity in temporal factors as represented by the total time and number of unique time instances spent together versus alone.

$$\Phi_{ij}^{temporal} = \frac{J(S_i^3, S_j^3) + J(S_i^4, S_j^4)}{2} \quad (8)$$

DEFINITION 6. (Social Similarity) Let social similarity Φ_{ij}^{social} between users i and j represent the homogeneity in social factors as represented by the number of unique common users versus total unique users and number of interactions together versus alone.

$$\Phi_{ij}^{social} = \frac{J(C_i^5, C_j^5) + J(C_i^6, C_j^6)}{2} \quad (9)$$

where C_i, C_j and C_{ij} represent the pairwise co-occurrences including user i, j and ij together respectively.

These metrics capture long-term similarities in *spatial, temporal, and social* scales, collectively forming a robust measure of long-term mobility behavior as follows.

DEFINITION 7. (Mobility Similarity) Let mobility similarity $\Phi_{ij}^{mobility}$ between users i and j represent the aggregation of spatial, temporal, and social similarities:

$$\Phi_{ij}^{mobility} = \alpha \Phi_{ij}^{spatial} + \beta \Phi_{ij}^{temporal} + \gamma \Phi_{ij}^{social} \quad (10)$$

where α, β and γ represent the weight of the spatial similarity, temporal similarity and social similarity respectively such that $\alpha + \beta + \gamma = 1$.

This definition allows us to weigh each similarity scale based on the input type. For instance, the temporal similarity would be weighted lower in incomplete datasets due to the lack of continuous spatial input. The ability to set α, β , and γ makes *W4-Groups* a generalizable group detection system.

3.3.5 Detecting W4-Groups. The final component recursively extracts groups using the pairwise co-occurrences from Section 3.3.2 and long-term mobility similarities from Section 3.3.4. First, the dynamic group detector clusters users connected via pairwise sessions and sharing the same location. That is, if two sets of users (u_i, u_j) and (u_i, u_k) are found at the same location with overlapping sessions, connected components will form the group (u_i, u_j, u_k) . To avoid losing information in pairs during recursive merging, all pairs are also included as groups. This outputs all **groups** $G = \{g_1, g_2, \dots, g_k\}$. However, whether groups detected in this step are true groups, or false positive detection is unclear. To filter out such false positives, we identify the group mobility similarity.

DEFINITION 8. (Group Mobility Similarity) Let group mobility similarity $\Phi_{g_k}^{group}$ for all users in group g_k represent the aggregation of $\Phi_{ij}^{mobility}$ of all subset of pairs in $g_k[\mathcal{U}]$:

$$\Phi_{g_k}^{group} = \text{median}(\Phi_{ij}^{mobility} \forall \{i \in g_k[\mathcal{U}], j \in g_k[\mathcal{U}]\}) \quad (11)$$

Filtering is defined as a function of $\Phi_{g_k}^{\text{group}}$ and group session duration. Particularly, we accept higher duration group sessions (> 60 minutes), reject low similarity group sessions, and set a higher similarity confidence for smaller duration (< 15 minutes) for acceptance. This is summarized as:

$$m_{g_k} = \begin{cases} 1, & \text{if } g_k[d] - g_k[e] > 60 \\ 0, & \text{if } \Phi_{g_k}^{\text{group}} < \phi_l \\ 1, & \text{if } g_k[d] - g_k[e] > 15 \text{ or } \Phi_{g_k}^{\text{group}} > \phi_u \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where ϕ_u and ϕ_l represent upper and lower bound. The output of *W4-Groups* is all positive group sessions $m_{g_k} = 1$. Although this filtering step requires manual setting of upper and lower similarity bounds, these settings allow for an adaptive system based on different input types.

4 PROTOTYPE IMPLEMENTATION

We have implemented a complete prototype of *W4-Groups* in Python. Our prototype currently runs on a Dell server with a 2.40GHz Intel Xeon E5-2680 processor and 128GB RAM running Ubuntu 20.04 and can run on a cluster of servers to scale to a large number of users.

As explained earlier, our approach works with any sensing modality that produces a location and mobility trace of a user. Our current prototype supports two data types: WiFi mobility traces and location-based social network (LBSN) check-in data. A detailed description of the data formats and associated pre-processing steps are provided in Appendix B.1.

The subcomponents implementation details of *W4-Groups* are as follows.

4.1 Sessions Extractor

The session extraction module computes location-based sessions of each user in two different ways, depending on the dataset. For *temporally continuous datasets* such as WiFi mobility traces where indoor localization is available, we define sessions at floor-level granularity. That is, all consecutive AP associations in the same building and on the same floor are interpolated into the same session. In contrast, for *temporally non-continuous datasets*, such as LBSN check-in mobility traces where localization is at a geo-location level, the module converts each event in trajectory M into 10-minute periods. This discretization can be altered based on the type of LBSN.

4.2 Pairwise Co-Occurrence Detector

In computing trajectory intersections between pairs and their corresponding sessions, the module implements Algorithm 1 to extract pairwise co-occurrences. In speeding up this computation process, the module uses the Pandas library, allowing the utilization of in-built optimized functions such as `pd.groupby()` and `pd.apply()`. Additionally, since users can have more than one device, a common theme with WiFi-based mobility traces, we add a post-processing step that merges overlapping times across devices for pairs.

4.3 Long-term Mobility Similarity Detector

The following two modules extract long-term behavioral features and determine long-term mobility similarity for each pair of users, respectively. We only gather this long-term information for pairs found in the second component to avoid unnecessary computations.

Our module uses the mobility similarity metric's α , β , and γ values in the following ratios:

- (1) **Equal dependencies:** 0.33-0.33-0.33
- (2) **Dependency on spatial and temporal:** 0.5-0.5-0.0
- (3) **Dependency on spatial and social:** 0.5-0.0-0.5

Table 2. Summary of three datasets used to evaluate *W4-Groups*.

	<i>Campus #1</i>	<i>Campus #2</i>	<i>Gowalla</i>
Data Usage	System evaluation, Sec. 5.3, 5.4, 5.5 Case studies, Sec. 6	System generalization, Sec. 5.4	
Period/ Duration	System eval: Feb-Mar 2022, 3 weeks Case Studies: Aug 2019 - May 2022	Aug-Dec 2019, 1 semester	Jan 2009-Aug 2011, 32 months
Location	University of Massachusetts Amherst	Singapore Management University	Worldwide (anonymized)
No. of Pax.	72 (system eval.), 200 (case study)	57	1000
User Details	<i>Demographics:</i> 8 M 4 F (deanon. subset) <i>Grade:</i> Graduate (n=11), Staff (n=1) <i>Routine:</i> Non-targeted (daily work) <i>Group:</i> Self-determined teams of 2-4 <i>Self reports:</i> groupwork schedule as part of graded material, team identification. <i>Device MAC Addr:</i> Smartphones, laptops <i>Approval:</i> Voluntary, IRB-approved <i>Compensation:</i> None	<i>Demographics:</i> 28 M 29 F <i>Grade:</i> Sophomore (n=57) <i>Routine:</i> Targeted (computing course) <i>Group:</i> Pre-assigned teams <i>Self reports:</i> course project schedule as graded material, team identification assessment <i>Device MAC Addr:</i> Smartphones <i>Approval:</i> Voluntary, IRB-approved <i>Compensation:</i> US\$30	N.A.
Ground Truth	Individual and group locations (datetime,member,activity,relation type)		Social network friendship data
Dataset	.csv files of WiFi traces		.csv files of check-ins

(4) Dependency on temporal and social: 0.0-0.5-0.5

By default, the α , β , and γ values are set to be 0.33, 0.33, and 0.33, respectively. We also compare the length of the long-term data, which by default is set to the full three weeks. Note, this duration is based on our user study data, explained in the following Section 5.1.

4.4 Dynamic Group Detector

The final module is responsible for extracting *W4* groups by recursively merging pair-wise groups and filtering them by their group similarity. In implementing recursive merging, the module expands each session by its location and then groups each group session by session time and location. We employ the NetworkX library for each sub-group found to create a graph and gather connected components. We then work backward to combine each expanded session into groups based on the connected components found. The lower and upper bound similarity values are set to $\phi_l = 0.05$ (75th percentile) and $\phi_u = 0.1$ (95th percentile) respectively in Equation 12, unless otherwise stated.

5 EMPIRICAL EVALUATION

In this section, we evaluate the *W4-Groups* prototype. We describe the evaluation setup and present performance, latency, robustness, and generalizability results.

5.1 Study Procedure

We employ the following three datasets, with primary analyses using *Campus #1* dataset. Key details of these datasets are summarized in Table 2.

We conducted an IRB-approved user study among 72 users, collecting WiFi network traces as mobility data between February and March 2022 from the University of Massachusetts Amherst, labeled as *Campus #1*¹. A subset (n=12) of students participated in groups, self-determining their

¹All WiFi data collection was based on a data usage agreement that includes additional safeguards for anonymization and terms of use. Specifically, device network MAC addresses and usernames used to authenticate the devices to the WiFi network via RADIUS authentication were hashed using strong cryptographic hash functions, anonymizing their identities.

teams between 2 to 4 members. They self-reported daily activities of work and social routines, including when they got out of bed, physically met with their group member/s for that day, start and end time on campus, and meal times. For all the times the group met, participants logged the extent to which each group member was engaged in the activity, from feeling alone to very stimulated. We collected other information, including classes the participant took, their primary work location and device MAC addresses to collect WiFi data.

The second dataset, *Campus #2*, also IRB-approved, is acquired from Singapore Management University, similarly consisting of WiFi network traces as mobility data. Compared to *Campus #1*, student recruitment was targeted at a computing course cohort, where high groupwork duties were expected as part of the course requirement. The course practices included students pre-assigned to teams of 5 based on their performance in a prerequisite course. While group participation amongst team members was encouraged, opting in/out of the study did not impact class grading and functions. Participants were instructed to have their mobile devices connected to the campus WiFi throughout the semester. They reported their team identification [52] with other members on a biweekly basis and supplied their group project schedule, which recorded the time, location, and members present to complete different group tasks. Note participants anonymized the group project schedule, which contained the names of non-consented students.

Third and finally, we employed a publicly available dataset collected by Yang et al. from *Gowalla* social networking service of check-in information [72]. It was reported to have 600,000 users in November 2010 and was acquired by Facebook (now Meta) shortly after. This dataset comprises approximately 36 million check-ins made by 300,000 users over 2 million locations. We extract the top 1000 users with the most check-ins for experiments. Ground truth group structure is measured via the users' social network friendship network.

With WiFi data reflecting users' mobility behavior in datasets *Campus #1* and *Campus #2*, the self-reports allow us to verify collocating users' actual relationships and their intentions to be at an area with their group. Employing *Campus #2* and *Gowalla* allow us to evaluate model generalizability on datasets collected via varying modalities.

5.2 Evaluation Baseline and Metric

Our evaluation compares *W4-Groups* with eight other methods described below. Several of these methods are chosen as simplified versions of *W4-Groups* where only a component of *W4-Groups* is used standalone for detection. Two of these methods are explicitly selected from prior work that pursued similar goals to our effort. Specifically, GruMon demonstrated high performance on detecting the *who-when-where* groups in dense and complex urban spaces such as malls and airports [54]. While GruMon enables location data with additional sensors to enhance group detection, it is one of the few systems that has shown state-of-the-art performance using solely WiFi data (as in our work). We employ GruMon only on raw AP connection events, which are inherently noisy at hour and day-level time scales. Second, we employ PGT to extract the strength of relationships between pairs by weighting co-location frequency with personal, global, and temporal factors [67]. While PGT is not traditionally used for group detection, we aim to evaluate how PGT performs in replacement of our long-term similarity detector. Note that these comparisons are drawn based on a re-implementation of both systems.

- (1) *Trajectory Co-Occurrence*: This method includes identifying spatio-temporal co-occurrence using raw trajectory data as represented in Equation 2.
- (2) *Session Co-Occurrence*: This method employs only spatio-temporal co-occurrences over sessions. This is equivalent of not employing long-term behavioral data in group detection.

Table 3. Performance of the proposed method against baselines and state-of-the-art methods.

Method	Acc.	Prec.	Rec.
Trajectory Co-Occurrence	38.90	95.46	39.63
Session Co-Occurrence	61.69	61.69	100
GruMon* @ Hour-Level	33.33	96.15	32.05
GruMon* @ Day-Level	72.22	85.71	80
Session Co-Occurrence with PGT*	45.18	78.21	26.99
Session Co-Occurrence with Spatial Similarity	82.38	99.15	72.05
Session Co-Occurrence with Temporal Similarity	83.14	78.54	100
Session Co-Occurrence with Social Similarity	75.48	71.75	99.38
Session Co-Occurrence with Mobility Similarity (<i>W4-Groups</i>)	93.49	97.37	91.93

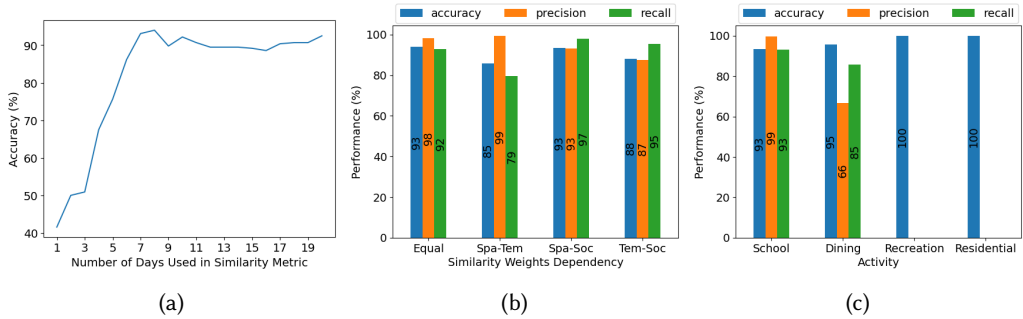


Fig. 4. Results of analyzing system parameters and performance tradeoffs: (a) impact of the length of long-term data; (b) impact of the similarity metric weights where 'spa' refers to spatial, 'tem' refers to temporal, and 'soc' refers to social; and (c) performance across group activity types.

- (3) *GruMon* @ Hour-Level*: This method employs an implementation of GruMon, a group monitoring system, at hour-level time windows [54].
- (4) *GruMon* @ Day-Level*: This method employs an implementation of GruMon, a group monitoring system, at day-level time windows [54].
- (5) *Session Co-Occurrence with PGT**: This method employs session co-occurrences with an implementation of PGT, a method that measures the strength of relationship between pairs.
- (6) *Session Co-Occurrence with Spatial Similarity*: This method employs session co-occurrences with only long-term spatial similarity. This is equivalent of setting $\alpha = 1$, $\beta = 0$ and $\gamma = 0$.
- (7) *Session Co-Occurrence with Temporal Similarity*: This method employs session co-occurrences with only long-term temporal similarity. This is equivalent of setting $\alpha = 0$, $\beta = 1$ and $\gamma = 0$.
- (8) *Session Co-Occurrence with Social Similarity*: This method employs session co-occurrences with only long-term social similarity. This is equivalent of setting $\alpha = 0$, $\beta = 0$ and $\gamma = 1$.

We use accuracy, precision, and recall to measure the effectiveness of our system with the goal that all these metrics should be balanced in a robust system.

5.3 Performance Results

While *W4-Groups* can be run proactively in a real-time manner as more data arrives, for evaluation purposes, we pass the entire dataset as described in Section 5.1 to *W4-Groups*. Our first experiment compares *W4-Groups* performance with the baseline methods described above. As tabulated in Table 3, *W4-Groups* significantly outperforms all the other methods. Specifically, by including long-term behavior, *W4-Groups* achieved $\approx 21\%$ and $\approx 48\%$ improvements in overall accuracy compared to

Grumon and PGT, respectively. While precision and recall are higher for other baseline methods, *W4-Groups* yields the most optimal performance on all three metrics. For instance, although session co-occurrence with spatial similarity has 99.15% precision, the recall is only 72.05%, implying only a subset of groups is detected. Similarly, although session co-occurrence with social similarity identifies all correct groups by 99.38%, it includes many false positives as represented by a 71.75% precision. Additionally, we note that despite Grumon's high performance at the day level, the output is a coarse cluster of groups instead of fine-grained session-level groups as in *W4-Groups*.

5.3.1 Tradeoff between length of long-term data and performance. Finding the right balance in detection is not straightforward due to the differing nature of groups. While long-term representations are captured through the notion of repeatable colocations and their variances over time, these factors broadly vary across users and groups – particularly causing errors among idiosyncratic groups when we set general thresholds.

As illustrated in Figure 4a, computing similarity metric based on one day's historical data merely achieves 40% accuracy across the full dataset. However, it will correctly identify real groups with 90% accuracy over a week-long history. Performance in detecting real groups tends to stabilize over time, suggesting regularity of users colocating with the same people over the observed duration. This finding implies that while group detection mechanisms can occur in real-time, detecting real groups using coarse-grained data is less likely to be accurate in a short duration; this is where long-term behavioral dependencies prove necessary to proactively update previous results.

5.3.2 Tradeoff between similarity metric weights and performance. Figure 4b charts system's performance across different mobility similarity ratios as described in Section 4.3. Our results yield $\approx 94\%$ accuracy with equal weights on *spatial*, *temporal*, and *social* attributes to detect groups.

Our observation found most of our student groups colocating in study-related areas and dining halls during these three weeks. However, the frequency of their colocating in these areas also depends on their social relations with their pairs. It should be noted that utilizing *equal dependencies* as weights is optimal for our student population on campus and may not be the same in other settings. In an environment such as a shopping mall, we expect a high distribution of locations grouped by distinct types (e.g., dining, entertainment, groceries, medical, and shopping). In such cases, the weightage on *spatial* similarity may be set more than other factors.

5.3.3 Performance across activity types. As explained above, during the three weeks of study, the activities of our student groups were highly concentrated in study-related areas and dining halls. For example, the four student groups who went for lunch together were detected by *W4-Groups* with 85.71% recall. The computation of sessions, as defined in Section 4.1, is based on consecutive AP associations on the same building floor. Accordingly, what type of activity is being performed is determined by *W4-Groups* at floor-level granularity.

While *W4-Groups* operates to detect groups in residential and recreational locations, as shown in Figure 4c, our results found no positive detection of group activities occurring within recreational and residential typed locations; thus, yielding 0% on precision and recall. It should be noted that the continuation of indoor masking and social distancing remained as requirements during the study. None of our student groups reported going to the recreational sites.

5.3.4 Long-Term Behavior Robustness Across User Types. Building on our findings for using long-term data, Figure 5a plots the probability distribution of student participants for different interaction levels. Among those working in the same lab, there is a higher probability that members in this group take on a mobility similarity value of 0.15 and a lower probability that their mobility similarity value is 0.05. Inversely, two users located on the same floor but do not share the same

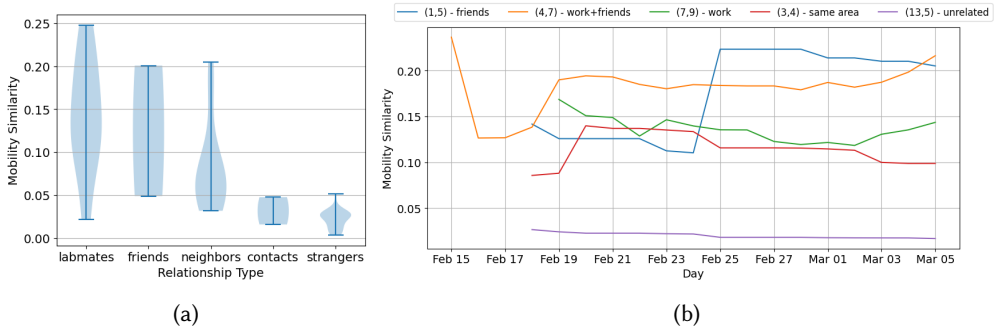


Fig. 5. Results of analyzing the robustness of long-term behavior across user types: (a) distribution of mobility similarity across pair types; and (b) change in mobility similarity over time for a subset of pairs.

workspace (e.g., work neighbors) are more likely to have a lower mobility similarity. Through this measure, *W4-Groups* can filter infrequent contacts and strangers who colocate by chance.

The mobility similarity metric among users who are friends can range from 0.05 to 0.20, likely to be so since there tend to be overlaps in individual perceptions of relationships among members. For example, we learned that five students regarded the interaction with their pair as “friends”, while three others considered themselves as sharing working and social relationships. Users with overlapping relationships are more likely to share higher similarity metrics as they colocate in the same area for work. We illustrate this example through Figure 5b, which plots the changes in the mobility similarity metric of these user groups over time. For groups sharing working and social relationships (i.e., work + friends), their mobility similarity quickly stabilizes near 0.20 over two days. Student groups who regarded themselves as friends met less frequently in work-related areas but more often in dining places. Their mobility similarity, too, stabilizes around 0.20, but it takes approximately a week before *W4-Groups* makes this determination.

Key Takeaway: The experimental results yield compelling evidence of *W4-Groups*’s achieving up to 93.49% accuracy, even in densely populated areas like dining halls. Including long-term behavioral dependencies contributes to the stability of group detection over time. Furthermore, the analysis reveals the system’s ability to accurately differentiate between group relationship types across time. The modular nature of *W4-Groups* enables the customization of similarity metrics to align with specific group settings (e.g., school), further enhancing its flexibility and adaptability.

5.4 Model Generalization

Our next experiment looks to investigate the generalization of our approach by employing the *Campus #2* and *Gowalla* datasets. Since *Campus #2* dataset was gathered as part of a user study that heavily emphasized the analysis of students’ interaction during group meetings, we assign less weight to *social* similarity when computing mobility similarity. Particularly, we set α, β, γ values to be 0.45, 0.45, 0.1, respectively. We set the lower bound similarity to 0.24 (75th percentile) and upper bound similarity to 0.47 (95th percentile) in Equation 12. For *Gowalla* check-in dataset, we discretize each check-in period into 10 minutes sessions (see Section 4.1), which inadvertently limits our system’s capability to extract temporal information such as the time spent in a particular location. As such, our parameters could only be set for *spatial* and *social* dependencies. We considered tweaking *spatial* similarity with more weight due to its large distribution of locations. Particularly, we set α, β, γ values to be 0.8, 0, 0.2, respectively.

Table 4. Generalized performance across datasets.

Dataset	Acc.	Prec.	Rec.
<i>Campus #1</i> : WiFi	93.49	97.37	91.93
<i>Campus #2</i> : WiFi	93.04	100	93.04
<i>Gowalla</i> Check-ins	92.40	92.93	99.30

Table 5. Latency of *W4-Groups*'s sub-module by runtime (in seconds) and compute cycles for *Campus #1*.

Module	Runtime	Cycles
Session Extractor	517.16	1.240 e+12
Pairwise Co-occurrence Extractor	17.03	4.417 e+10
Long-Term Behavior Feature Extraction	141.71	3.397 e+11
Long-term Mobility Similarity Detector	0.15	3.510 e+8
Dynamic Group Detector	186.42	4.725 e+11

Table 4 summarizes the performance for detecting groups using *W4-Groups*. Our experiments yield equivalently high performance of 93.04% accuracy, 100% precision, and 93.04% recall for *Campus #2* dataset, demonstrating that WiFi-based mobility traces, regardless of their setting, can be used to determine groups effectively. Separately, evaluation on the *Gowalla* check-in dataset achieves 92.40% accuracy, 92.93% precision, and 99.30% recall. Despite the incomplete check-in dataset lacking continuous spatial information, the proposed system can detect friend groups accurately using their spatial and social similarities only.

Key Takeaway: *W4-Groups*'s ability to emphasize *spatial*, *temporal*, and *social* similarities according to the nature of interactions contributes to its enhanced generalizability. The results provide quantitative evidence supporting the accurate group detection capabilities of *W4-Groups* across various data granularities.

5.5 System Overhead

Table 5 demonstrates the run times of each module in seconds and CPU cycles, respectively, for 72 users over three weeks. The runtime for *W4-Groups* to detect all groups totals up to 862 seconds, assuming each module is executed sequentially.

By implementing Algorithm 1, the pairwise co-occurrence extractor reduces computation time to 17 seconds. As noted, the system overhead is primarily attributed to modules 1, 3, and 5. However, *W4-Groups* parallelizes its operations for modules 1, 2, and 5 in the actual deployment. All three modules operate at a day-wise level, allowing consequent days to be processed through the system concurrently. Modules 3 and 4 require cumulative processing since they work at a long-term level.

Key Takeaway: *W4-Groups* achieves a total runtime of approximately 14 minutes when processing data from 72 users over three weeks. Due to its parallelizability, modules within *W4-Groups* can be processed concurrently, further reducing the overall runtime.

6 CASE STUDIES

In this section, we present two case studies to demonstrate the potential application of *W4-Groups* detection framework. From *Campus #1*, we meticulously picked 200 anonymous students who frequented a freshmen-designated dormitory. Note the utility of an anonymized dataset meets our computing agreement to only report results at the aggregated level.

The case studies are motivated by a common theme of analyzing the influence of short-term and long-term groups on individual user behavior. We first model the effect of short-term groups on the mobility of individual users, naturally extending from the *W4-Groups* modeling approach. We then examine the potential of *W4-Groups* in enhancing individual health monitoring systems by analyzing longitudinal changes in group behavior.

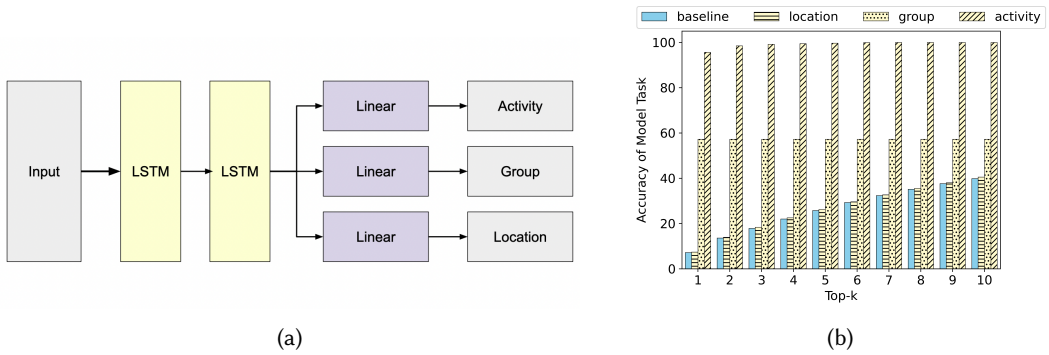


Fig. 6. (a) LSTM architecture for multi-task training; and (b) performance of each task in the model (in yellow) as compared with a traditional next location prediction model (in blue).

6.1 Next Activity Prediction

Next location prediction has been fundamental in domains ranging from recommendation systems to influencing urban designs. These models use historical spatiotemporal contexts to predict the following location a user will visit. Such predictions can be used in a wide variety of manners. For instance, a mapping service may predict the travel time to the user’s next predicted location.

Prior works have employed Markov and hidden Markov models [21, 43], sequential neural networks [18, 34, 61, 79] and more recently, transformer-based models [65] to build human mobility models. However, none of these works have explored predicting beyond the ‘where’ and ‘when.’ One such work by Zhang et al. proposed using hidden Markov models to simultaneously learn mobility models for large group clusters (i.e., who-where-when) [78]. Conversely, Liao et al. proposed using context-aware recurrent neural networks to predict activity and location for individual users (i.e., what-where-when) [38].

This case study examines the influence of groups on individual users in the context of *who-what-when-where*. Each individual establishes a daily routine consisting of various activities, which can be performed independently or in the company of one or more individuals [45]. We exploit this aspect of human nature to predict the *who-what-when-where* of a nomadic user at a fine-grained level using their coarse-grained location trajectories. We argue that predicting *when* and *where* a user will visit along with *with-who* and *what* activity they will perform is more valuable from urban, campus, and office design perspectives.

6.1.1 Model Training. We begin by extracting the mobility trajectories of 100 users to train the next activity prediction model. The output of *W4-Groups* is used to annotate individual session trajectories. That is, we augment the session trajectories of each user i to include their group and activity information during time instance t as follows: $s_i^t = \{l, e, d, a, g\}$.

We use this data as input to a long-short term memory (LSTM) model [26]. LSTMs have shown exemplary performance in predicting human mobility due to their ability to learn and remember long-term dependencies in sequences [18, 34, 64, 79].² Particularly, we train a two-layer multi-task LSTM model M where input is the previous five session sequences. Output is the next location,

²While other state-of-the-art models may be employed, such as in Trivedi et al. [65], LSTMs align with the prevailing approach in the field, allowing for better comparison and benchmarking. It is important to note that our primary goal is to showcase the efficacy of *W4-Groups* for predicting the *who-what-when-where* of a nomadic user, and the specific model choice does not affect this objective.

next group, and next activity: $M : s_i^{t-5}, s_i^{t-4}, \dots, s_i^{t-1} \rightarrow l_i^t, g_i^t, a_i^t$ ³. The model architecture is depicted in Figure 6a. The distinct linear layers allow for simultaneous training of the three learning tasks by fine-tuning the representations learned by the LSTM layers. For each task, we compute its unique cross-entropy loss and aggregate it for learning. The model is implemented using PyTorch and is trained using a learning rate of $1e-4$ with a weight decay of $1e-6$ and a hidden layer size of 128. We use batches of size 128 with a dropout rate of 0.2 between the LSTM layers and train for 15 epochs.

6.1.2 Model Performance. Figure 6b presents the results of next activity prediction at building-level localization. We compare the model's performance with a traditional next-location prediction model where the output modality is only 'where.' The models are evaluated using the top-k accuracy metric, which counts the instances where the correct label is a subset of the model's top-k most probable classes. There is no notion of top-k for group prediction since it is a multi-class problem.

Location prediction in both models has similar performance; for $k = 10$, the models can accurately predict the next location with approximately 40% accuracy. However, with the multi-output model, we can also identify the next group the user will be with 58% accuracy and the next activity the user will perform with 99% accuracy. Further optimization via advanced LSTM cells that contain task-specific parameters for task-specific learning can improve the model's performance [40].

Key Takeaway: The results demonstrate the applicability of *W4-Groups* in predicting *who-what-where* for the next timestep (*when*). Employing *W4-Groups* to augment traditional next location prediction models with group-level information offers unique capabilities that can enhance our understanding of the underlying reasons behind location visits. Since groups are integral to daily routines, this capability holds significant implications for future research and advancements in various fields, including location-aware mobile services and recommendation systems.

6.2 Longitudinal Change in Behavior

The subject of health and well-being is a multifaceted outcome resulting from the interplay of individual, social, and environmental factors. Much work in ubiquitous computing has focused on developing health risk management systems to help the population cope with behavioral, mental, and social health-related issues. These systems enable accurate monitoring of psychological measures and estimation of mental states via various modalities, such as smartphone sensing [6, 53, 68] and passive WiFi sensing [42, 69, 75]. However, the emphasis on sensing individual factors overlooks social interaction that significantly influences one's health. Jetten et al. present compelling evidence emphasizing the importance of social groups in preserving health and well-being, yet they assert that group identities are often disregarded in practical applications [30]. We believe *W4-Groups* is highly applicable to the extension of such systems, as it has the potential to provide accurate measures related to group behaviors and their social routines.

One aspect of supporting health monitoring systems is the ability to understand how user behavior changes over time. This case study demonstrates how *W4-Groups* supports this goal by effectively monitoring longitudinal changes in group behavior. By integrating group information, our study highlights the potential of enhancing health monitoring systems in capturing and analyzing shifts in individual and collective behavior patterns.

The WiFi-derived mobile trajectories were passed through the *W4-Groups* pipeline. For pairwise groups, we aggregate their behaviors at per week granularity. We report the total time spent per week in groups across the semester. Since group sessions depend on WiFi connectivity, these sessions may not precisely represent users spending long hours in one place but relatively shorter sessions of users in the vicinity.

³The choice of architecture is selected based on existing state-of-the-art models for next location prediction.

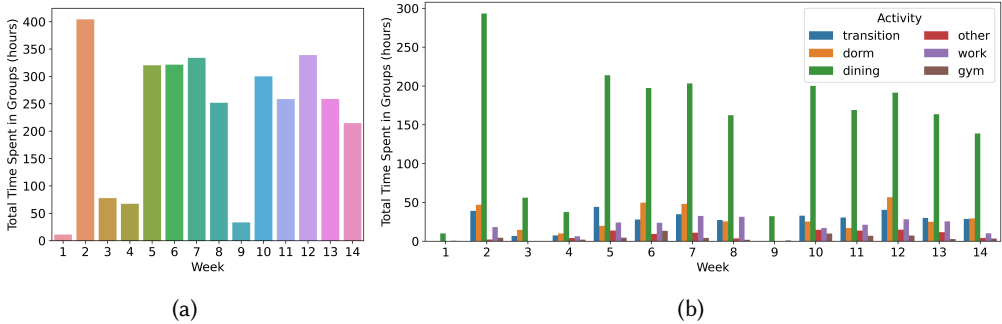


Fig. 7. Group behavior over a semester: (a) total time spent weekly in groups; and (b) the respective activities.

6.2.1 Time Spent and Activity Types in Groups. The statistics in Figure 7a show the variation of group sessions over 14 weeks, beginning one week before classes commenced and ending shortly before the final exams. We observe a sharp increase in group behavior corresponding to the first week of the semester (i.e., week 2). We expect this behavior to naturally reduce as students settle into their routines the following week but fluctuate as they move along with course requirements and midterms (i.e., weeks 6-8, weeks 8-11). In week 9, the significant reduction in group behavior is likely a result of Spring break, and this behavior fluctuates once again upon resuming the second half of the semester.

Figure 7b demonstrates the group activities that are most common on campus, along with how they change throughout the semester. From the plot, we can infer that students mostly gather over dining activities, consistent throughout the semester. Interestingly, we note that groups in dorms and work environments increase as we head toward midterm weeks 8 and 11.

Figure 8a reveals the pattern of group behavior at an aggregated level on an hourly scale across the semester. The cumulative time for group sessions is significantly high throughout the morning, between 00:00 am to 5:00 am, likely because students are in their dorms and their devices are connected to the same WiFi AP.⁴ It should be emphasized that the variance for these group sessions is also high, suggesting many combinations of groups with varying levels of continuous WiFi connectivity while the device is not in use. Duration for these group sessions drops tremendously during the day as they leave their dormitory to fulfill their academic routines. Note that this drop in duration is also an artifact of the nature of passive sensing, in this case, WiFi, which often contains shorter continuous sessions after denoising and processing. *W4-Groups* can still successfully detect groups despite such noisy input.

6.2.2 Activity Comparison Pre-Post COVID-19. One of the earliest significant steps to curb the virus spread during the COVID-19 outbreak was for educational institutions to fully operate their academic schedules virtually, changing students' daily routines around campus. Over time, campuses took careful steps in reopening and implementing social restrictions at different phases. In this case, our analysis is to analyze the difference in group behaviors before and after COVID-19.

Figure 8b compares two freshmen batches in 2019 and 2022. The violin chart illustrates the time spent weekly doing various group activities and its probability distribution. We observe a general trend of students spending less time in groups post-COVID. The time spent in groups pre-COVID has higher variability in dining and work activities. There is a higher distribution of time spent

⁴Groups that repeatedly spend extended periods together in common locations such as dormitories or offices are not considered false positives as these groups are present for a common activity such as sleeping or working.

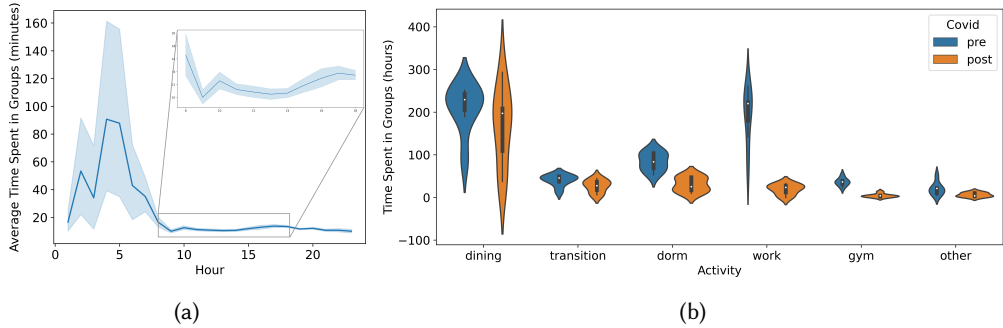


Fig. 8. Group behavior over a semester: (a) average time spent in groups across a day; and (b) pre versus post COVID-19 comparison by activities.

in different activities around the median post-COVID. Lesser time spent in dormitories and work sites implies fewer people live and work on campus, likely due to precautionary reasons.

Key Takeaway: The results demonstrate how *W4-Groups* can be employed to analyze group behavior on a longitudinal scale. While the figures present intuitive information about student life on an aggregated scale, the ability to augment behavioral monitoring systems with information about social interactions can enable more robust psychological measures. This capability holds intriguing implications for behavioral, mental, and social health-related studies.

7 DISCUSSION

Our study's objectives were to develop *W4-Groups*, which quantitatively determines key characteristics of *who-what-when-where* to distinguish group formations based on their purposes existing as short-term and long-term. Here we discuss the privacy and ethical implications of our system.

7.1 Strengthening Privacy Safeguards

W4-Groups's primary strategy to detect dynamic groups is utilizing mobility-typed data of any nature. It is worth noting that modern systems (e.g., WiFi networks, LBSNs) often rely on user authentication, which facilitates accessible data collection. Our user studies employed privacy safeguards of hashing device MAC and usernames, disallowing any reverse-engineering of the anonymization (see Section 5.1, active users explicitly consented to be de-anonymized).

Beyond the scope of this study, autonomous detection systems of digital users, such as ours, can raise privacy concerns. Further aggravating is the tendency for implicit consent via default opt-in, a widely used behavioral economics tool, requiring users to actively opt out should they choose differently [47]. Cumulatively, this often leads to the construction of detailed user profiles for targeted advertising and personalized content, with decisions about its collection and use made by powerful corporations [7]. Thus, we stress the significance of enabling users with greater control over their data for informed consent, aligning with the recommendations by Cheney-Lippold [7]. For example, users can opt into data sharing in particular vicinities, eliminating the inadvertent collection of sensitive data such as visit to a doctor's office, while still enabling group detection in many social settings.⁵ Nouwens et al. propose the usage of more durable design patterns for setting data preferences, effectively empowering users with the option to set

⁵Such practices are automated in many LBSNs where location data is only shared across the network upon the user's explicit permission (e.g., Gowalla check-in dataset used in our evaluation).

controls upon initial authentication. Further abiding by Cheney-Lippold's recommendations on data security [7], we recommend employing cryptographic techniques to standardize anonymized data collection in target environments. As our case studies have demonstrated, *W4-Groups*'s framework depends on the granularity of the data supplied. For example, in scenarios centering on large-scale behavioral monitoring, maintaining anonymity in data is sufficient to preserve user privacy while enabling meaningful aggregated analyses. Deanonimization is a second-step process necessary in applications requiring targeted user tracking. In such cases, while targeted users may be de-anonymized with their permission, the system can provide high-level insights on interactions with anonymized users relevant to the application at hand.

While cryptographic techniques stand to be well-accepted modes of data protection, we also consider potential improvements in modeling techniques that may enable model-dependent privacy protection. Specifically, we note that temporal discontinuity in mobility trajectories will produce shorter sessions in *W4-Groups* by design, which may not align perfectly with true group interactions. Our work continues to employ more sophisticated techniques to address these inaccuracies, such as adapting transformer-based models from natural language processing (NLP) to incorporate surrounding contexts and learn robust mobility representations that extend beyond spatiotemporal data [66]. Utilizing ML methods facilitates on-device and collaborative learning [33], enhancing privacy by granting users greater control over their data and the ability to selectively retract specific portions of it while supporting community-driven group initiatives.

7.2 Application of Ethics

Consider our case study of analyzing longitudinal behavioral change among students on campus in Section 6.2 to separately analyze the behaviors of resident physicians in their workplaces [1] as an example. Such a system would require individual stakeholders to trust the institution enormously. Thus, policy solutions are especially crucial as *W4-Groups* can be used without explicit user consent from its inherent mechanisms. A pressing concern requiring interdisciplinary research attention is how the system can provide actual benefits to target users without violating their privacy and feeling unfairly targeted or discriminated against [4]. Monitoring technologies integrating community-based initiatives with personal benefits must be accompanied by appropriate policy frameworks protecting user rights, especially concerning health and well-being.

With *W4-Groups* accommodating different granularities of user permissions for personal health and productivity monitoring (e.g., monitoring employee burnout symptoms and work performance), it can facilitate usage across diverse ethical policies. For example, a multi-tiered crisis care infrastructure can be enabled among residents, supervisors, and educational mentors to foster communication among stakeholders and provide interventions. From a system's perspective, generalizing key group detection parameters from different mobility data sources ensures fair participation opportunities among users via commodity devices and services. Our evaluation of *W4-Groups*'s robustness against ranging data characteristics in Section 5.4 is guided by real-world inaccuracies in spatiotemporal signals. By providing residents the autonomy to share information types (e.g., location granularity, frequency, unique device identifier), *W4-Groups* readily adapts its grouping operation with a less precise one using anonymization techniques – effectively storing the cohort's spatial, temporal, and social similarities at coarse granularity over time. Behavioral deviations from the cohort's norm (indicating a likelihood of burnout symptoms among unknown users) would reveal where and when these behaviors are observed, approximating what activities are typically being engaged. Without violating individual privacy, aggregated insights ensure consistent deployment of community crisis resources is within reach of potential target users. Aside from direct interventions, enabling a data-sharing use case to encourage training programs better manage residents' well-being [1] should only be at the discretion of the individual, coupled with policies for data accountability.

Table 6. Comparison of prior work with ours. ‘Who’ detects dynamically-changing small groups across time. ‘What’ defines a detected activity. ‘When’ analyzes whether group output was computed in time windows (periods) or at real-time. ‘Where’ provides location information at varying granularity.

Work(s)	Who			What Activity	When		Where		
	Dynamic	Small Group	Long Term		Periods	Real-Time	Coarse	Fine	Incomplete
Wearables [48, 49]	✓	✓	×	✓	×	✓	×	✓	×
RFID Tags [5]	✓	✓	×	×	✓	×	✓	×	×
Accelerometers [22]	✓	✓	×	×	✓	×	✓	×	×
BT Proximity [17, 60, 74]	✓	✓	×	×	×	✓	×	×	✓
BT Proximity + Audio [36, 41]	✓	✓	×	×	×	✓	×	×	✓
WiFi RSSI Signals [27, 55, 56, 59]	×	✓	×	×	✓	×	×	✓	✓
WiFi + Sensors [14, 32]	×	✓	×	×	✓	×	✓	✓	✓
WiFi Network Events [54]	×	✓	×	×	✓	✓	✓	✓	✓
GPS [70, 73]	✓	✓	×	×	×	✓	×	✓	✓
LBSN Groups [16, 51, 67, 72]	×	✓	×	×	✓	×	×	✓	✓
<i>W4-Groups</i> (ours)	✓	✓	✓	✓	✓	✓	✓	✓	✓

In that order, *W4-Groups*’s operationalization complies with the principles of the Belmont Report [46]. A framework capable of distinguishing real-world group formations via (1) common means to promote equal participation (*justice*), (2) adaptation to changes in data granularity from users’ autonomy (*respect for persons*), and (3) maximizing possible benefits (*beneficence*) across stakeholders in communities, can competently support practical and ethical applications of group dynamics.

8 RELATED WORK

The ubiquitous availability of sensing technologies has enabled researchers to comprehensively investigate user mobility and detecting groups at different granularity and modalities. The expansion of the “digital nomad” lifestyle motivated group modeling approaches to primarily seek location-based information [35]. We summarize the key differences of these works in Table 6.

8.1 Location-based social network (LBSN)

LBSN derives group interactions based on user connections on content tied to locations. By gathering location data during user check-ins into the application, prior work has learned the relationships between users where users’ friendship networks serve as the ground truth for evaluation. Researchers have found that spatiotemporal co-occurrence strongly indicates social ties in LBSN datasets [8]. Many efforts went into understanding the relationship between mobility and strength of relationship [9, 15, 16, 37, 51, 67, 72], while Xiao et al. utilized semantic information to estimate the similarity between people without geographical overlap [71]. More recent work introduced commitment and compatibility as properties that any social distance measure should follow to infer social connections [50].

8.2 Indoor Localization

With humans spending most of their time indoors, researchers have increasingly investigated localization orientation by instrumenting IoT technologies within the environment or on-body wearables to derive group proximity. For example, Brown et al. used lightweight RFID tags to detect face-to-face encounters [5]. Researchers also developed sociometric badges to measure interaction, physical proximity between users, and activity [48, 49]. More recently, Gedik and Hung used wearable accelerometers to identify conversational groups [22]. While these methods work well in accurately determining close contacts within an individual’s social circle, a clear drawback is an

insufficiency for scalability since these modalities require the dedicated installation of equipment, which is not feasible on a large scale.

8.3 Smartphone Localization

The increasing reliance on smartphones by everyday users motivated researchers to model groups based on fixed and moving locations from the device or via its network infrastructures, such as Bluetooth, GPS, and WiFi. For example, Feese et al. used a low-power ANT radio in smartphones to measure proximity to other firefighters during an event [17]. Bluetooth (BT) signal strength is another modality for determining intentional and transitional groups [60, 74], including intersecting it with smartphone's microphone signals to detect conversational groups [36, 41]. Through spatiotemporal clustering techniques, prior work successfully used GPS to detect groups of pedestrians moving together [70] and groups of vehicles on the road [73].

Another modality showing promise in group detection is WiFi information, where mobility trajectories of users are estimated from the access point (AP) connection events across time [31] and recursively approximating the distance between two users. Shen et al. proposed techniques using matrix factorization from the users' similarity matrix of raw WiFi logs with Bluetooth RSS data to detect groups [55, 56]. Group-In and SocialProbe, among others, are examples of group modeling using wireless RSSI [23, 24, 27, 57, 59]. However, RSSI-based approaches have been criticized for being unreliable in crowded, urban spaces like malls and college campuses [55]. Since WiFi data is often incomplete and noisy, many similar approaches employ multimodal sensing to augment user mobility data with behavioral semantics [13, 14, 32, 54]. These efforts face similar challenges in scalability as sensor-based group detection methods.

9 CONCLUSION

This work proposed *W4-Groups* as a promising group detection framework that models and detects the *who-what-when-where* characteristics of group dynamics through a theoretical understanding of groups in small group research. *W4-Groups* incorporates the functional perspective of variability and repeatability to construct short-term and long-term groups through a distinctive view of *spatial*, *temporal*, and *social* relationships. These steps collectively allow the technique to more accurately determine users that colocate for a specific purpose and segregate those that colocate by chance. In this regard, our technique handles computational complexity to allow for scalability in detecting larger groups of users. Through two user studies on three datasets, our evaluation demonstrated *W4-Groups* yielding an average of 92% overall accuracy, 96% precision, and 94% recall. Our results proved system robustness to different mobility-type datasets and performance improvements over prior group detection techniques. Finally, we exhibited the practical relevancy of *W4-Groups* in supporting two real-world applications, further establishing the usability of the proposed system in applications requiring intentional group formations.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable comments. This research was supported in part by NSF grants 2211302, 2211888, 2213636, 2105494, 1908536, Army Research Lab contract W911NF-17-2-0196, Adobe, and Ministry of Education, Singapore under its Academic Research Fund Tier 2 (Project ID: T2EP20220-0016). Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

- [1] Daniel A Adler, Emily Tseng, Khatiya C Moon, John Q Young, John M Kane, Emanuel Moss, David C Mohr, and Tanzeem Choudhury. 2022. Burnout and the Quantified Workplace: Tensions around Personal Sensing Interventions for Stress in Resident Physicians. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–48.
- [2] Aruba Networks, Inc. 2013. *ArubaOS 6.3.x Syslog Messages*. https://higherlogicdownload.s3.amazonaws.com/HPE/MigratedAssets/ArubaOS_6.3.x_Syslog.pdf
- [3] Sigal G Barsade. 2002. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly* (2002).
- [4] Ruha Benjamin. 2019. Assessing risk, automating racism. *Science* (2019).
- [5] Chloë Brown, Christos Efstratiou, Ilias Leontiadis, Daniele Quercia, Cecilia Mascolo, James Scott, and Peter Key. 2014. The architecture of innovation: Tracking face-to-face interactions with ubicomp technologies. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*.
- [6] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 1293–1304.
- [7] John Cheney-Lippold. 2017. We are data. In *We Are Data*. New York University Press.
- [8] David J Crandall, Lars Backstrom, Dan Cosley, Siddharth Suri, Daniel Huttenlocher, and Jon Kleinberg. 2010. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences* (2010).
- [9] Justin Cranshaw, Eran Toch, Jason Hong, Aniket Kittur, and Norman Sadeh. 2010. Bridging the gap between physical location and online social networks. In *Proceedings of the ACM International Conference on Ubiquitous Computing*.
- [10] Adrian A de Freitas and Anind K Dey. 2015. The group context framework: An extensible toolkit for opportunistic grouping and collaboration. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1602–1611.
- [11] Adrian A de Freitas and Anind K Dey. 2015. Using multiple contexts to detect and form opportunistic groups. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1612–1621.
- [12] Eyal Dim and Tsvi Kuflik. 2014. Automatic detection of social behavior of museum visitor pairs. *ACM Transactions on Interactive Intelligent Systems (TiiS)* (2014).
- [13] He Du, Zhiwen Yu, Fei Yi, Zhu Wang, Qi Han, and Bin Guo. 2016. Group mobility classification and structure recognition using mobile devices. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE.
- [14] He Du, Zhiwen Yu, Fei Yi, Zhu Wang, Qi Han, and Bin Guo. 2017. Recognition of group mobility level and group structure with mobile devices. *IEEE Transactions on Mobile Computing* (2017).
- [15] Nathan Eagle, Alex Pentland, and David Lazer. 2009. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences* (2009).
- [16] Chao Fan, Yiding Liu, Junming Huang, Zhihai Rong, and Tao Zhou. 2017. Correlation between social proximity and mobility similarity. *Scientific Reports* (2017).
- [17] Sebastian Feese, Bert Arnrich, Gerhard Trfooster, Michael Burtscher, Bertolt Meyer, and Klaus Jonas. 2013. Sensing group proximity dynamics of firefighting teams using smartphones. In *Proceedings of the International Symposium on Wearable Computers*.
- [18] Jie Feng, Can Rong, Funing Sun, Diansheng Guo, and Yong Li. 2020. PMF: A privacy-preserving human mobility prediction framework via federated learning. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2020).
- [19] Donelson R Forsyth. 2018. *Group dynamics*. Cengage Learning.
- [20] Lesley Fosh, Steve Benford, and Boriana Koleva. 2016. Supporting group coherence in a museum visit. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 1–12.
- [21] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. 2012. Next place prediction using mobility Markov chains. In *First Workshop on Measurement, Privacy, and Mobility*.
- [22] Ekin Gedik and Hayley Hung. 2018. Detecting conversing groups using social dynamics from wearable acceleration: Group size awareness. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2018).
- [23] Yu Gu, Lianghu Quan, and Fuji Ren. 2014. Wifi-assisted human activity recognition. In *IEEE Asia Pacific Conference on Wireless and Mobile*. IEEE.
- [24] Yu Gu, Fuji Ren, and Jie Li. 2015. Paws: Passive human activity recognition based on wifi ambient signals. *IEEE Internet of Things Journal* (2015).
- [25] A Paul Hare. 1976. Handbook of small group research. (1976).
- [26] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* (1997).
- [27] Hande Hong, Chengwen Luo, and Mun Choon Chan. 2016. Socialprobe: Understanding social interaction through passive wifi monitoring. In *Proceedings of the 13th international conference on mobile and Ubiquitous systems: Computing, networking and services*. 94–103.

- [28] John James. 1953. The distribution of free-forming small group size. *American Sociological Review* (1953).
- [29] Mohammad Hossein Jarrahi, Gemma Newlands, Brian Butler, Saiph Savage, Christoph Lutz, Michael Dunn, and Steve Sawyer. 2021. Flexible work and personal digital infrastructures. *Commun. ACM* 64, 7 (2021), 72–79.
- [30] Jolanda Jetten, Catherine Haslam, S Alexander Haslam, Genevieve Dingle, and Janelle M Jones. 2014. How groups affect our health and well-being: The path from theory to policy. *Social Issues and Policy Review* (2014).
- [31] Manpreet Kaur, Flora D Salim, Yongli Ren, Jeffrey Chan, Martin Tomko, and Mark Sanderson. 2020. Joint modelling of cyber activities and physical context to improve prediction of visitor behaviors. *ACM Transactions on Sensor Networks (TOSN)* (2020).
- [32] Mikkel Baun Kjærgaard, Martin Wirz, Daniel Roggen, and Gerhard Tröster. 2012. Detecting pedestrian flocks by fusion of multi-modal sensors in mobile phones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*.
- [33] Jakub Konečný, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. 2016. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492* (2016).
- [34] Dejiang Kong and Fei Wu. 2018. HST-LSTM: A Hierarchical Spatial-Temporal Long-Short Term Memory Network for Location Prediction. In *International Joint Conferences on Artificial Intelligence*.
- [35] Ahreum Lee, Austin L Toombs, Ingrid Erickson, David Nemer, Yu-shen Ho, Eunkyung Jo, and Zhuang Guo. 2019. The social infrastructure of co-spaces: Home, work, and sociable places for digital nomads. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–23.
- [36] Youngki Lee, Chulhong Min, Chanyou Hwang, Jaegung Lee, Inseok Hwang, Younghyun Ju, Chungkuk Yoo, Miri Moon, Uichin Lee, and June-hwa Song. 2013. Sociophone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion. In *Proceeding of the International Conference on Mobile Systems, Applications, and Services*.
- [37] Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, and Wei-Ying Ma. 2008. Mining user similarity based on location history. In *Proceedings of the ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*.
- [38] Dongliang Liao, Weiqing Liu, Yuan Zhong, Jing Li, and Guowei Wang. 2018. Predicting Activity and Location with Multi-task Context Aware Recurrent Neural Network.. In *International Joint Conferences on Artificial Intelligence (IJCAI)*.
- [39] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. 2014. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*. 739–748.
- [40] Peng Lu, Ting Bai, and Philippe Langlais. 2019. Sc-lstm: Learning task-specific representations in multi-task learning for sequence labeling. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NACCL)*.
- [41] Chengwen Luo and Mun Choon Chan. 2013. Socialweaver: Collaborative inference of human conversation networks using smartphones. In *Proceedings of the ACM Conference on Embedded Networked Sensor Systems*.
- [42] Priyanka Mary Mammen, Camellia Zakaria, Tergel Molom-Ochir, Ameer Trivedi, Prashant Shenoy, and Rajesh Balan. 2021. WiSleep: Scalable Sleep Monitoring and Analytics Using Passive WiFi Sensing. *arXiv preprint arXiv:2102.03690* (2021).
- [43] Wesley Mathew, Ruben Raposo, and Bruno Martins. 2012. Predicting future locations with hidden Markov models. In *ACM Conference on Ubiquitous Computing*.
- [44] Joseph Edward McGrath. 1984. *Groups: Interaction and performance*. Vol. 14. Prentice-Hall Englewood Cliffs, NJ.
- [45] Timothy K Monk, Joseph F Flaherty, Ellen Frank, Kathleen Hoskinson, and David J Kupfer. 1990. The Social Rhythm Metric: An instrument to quantify the daily rhythms of life. *Journal of Nervous and Mental Disease* (1990).
- [46] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, United States. 1978. *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*. Vol. 2.
- [47] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
- [48] Daniel Olguin Olguin, Peter A Gloor, and Alex Sandy Pentland. 2009. Capturing individual and group behavior with wearable sensors. In *Proceedings of the AAAI Spring Symposium on Human Behavior Modeling*.
- [49] Daniel Olguin Olguin, Benjamin N Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex Pentland. 2008. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* (2008).
- [50] Huy Pham, Ling Hu, and Cyrus Shahabi. 2011. Towards integrating real-world spatiotemporal data with social networks. In *Proceedings of the ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*.
- [51] Huy Pham, Cyrus Shahabi, and Yan Liu. 2013. Ebm: an entropy-based model to infer social strength from spatiotemporal data. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*.

- [52] Tom Postmes, S Alexander Haslam, and Lise Jans. 2013. A single-item measure of social identification: Reliability, validity, and utility. *British journal of social psychology* 52, 4 (2013), 597–617.
- [53] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, David C Mohr, et al. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet research* 17, 7 (2015), e4273.
- [54] Rijurekha Sen, Youngki Lee, Kasthuri Jayarajah, Archan Misra, and Rajesh Krishna Balan. 2014. Grumon: Fast and accurate group monitoring for heterogeneous urban spaces. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*.
- [55] Jiaxing Shen, Jiannong Cao, and Xuefeng Liu. 2020. BaG: Behavior-aware group detection in crowded urban spaces using WiFi probes. *IEEE Transactions on Mobile Computing* (2020).
- [56] Jiaxing Shen, Jiannong Cao, Xuefeng Liu, and Shaojie Tang. 2018. SNOW: Detecting shopping groups using WiFi. *IEEE Internet of Things Journal* 5, 5 (2018), 3908–3917.
- [57] Stephan Sigg, Ulf Blanke, and Gerhard Tröster. 2014. The telepathic phone: Frictionless activity recognition from wifi-rssi. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE.
- [58] Georg Simmel. 1902. The number of members as determining the sociological form of the group. I. *American journal of Sociology* 8, 1 (1902), 1–46.
- [59] Gürkan Solmaz, Jonathan Fürst, Samet Aytaç, and Fang-Jing Wu. 2020. Group-in: Group inference from wireless traces of mobile devices. In *19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE.
- [60] Gürkan Solmaz and Fang-Jing Wu. 2017. Together or alone: Detecting group mobility with wireless fingerprints. In *2017 IEEE International Conference on Communications (ICC)*. IEEE.
- [61] Xuan Song, Hiroshi Kanasugi, and Ryosuke Shibasaki. 2016. Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. In *International Joint Conference on Artificial Intelligence*.
- [62] Ralph M. Stogdill. 1972. Group Productivity, Drive, and Cohesiveness. *Organizational Behavior and Human Performance* (1972). [https://doi.org/10.1016/0030-5073\(72\)90035-9](https://doi.org/10.1016/0030-5073(72)90035-9)
- [63] V Das Swain, HyeokHyen Kwon, Bahador Saket, M Bin Morshed, Kathy Tran, Devashru Patel, YEXIN Tian, Joshua Philipose, Yulai Cui, Thomas Plötz, et al. 2020. Leveraging WiFi network logs to infer social interactions: A case study of academic performance and student behavior. *arXiv preprint arXiv:2005.11228* (2020).
- [64] Ameer Trivedi, Jeremy Gummeson, and Prashant Shenoy. 2020. Empirical Characterization of Mobility of Multi-Device Internet Users. *arXiv preprint arXiv:2003.08512* (2020).
- [65] Ameer Trivedi, Kate Silverstein, Emma Strubell, Prashant Shenoy, and Mohit Iyyer. 2021. WiFiMod: Transformer-based Indoor Human Mobility Modeling using Passive Sensing. In *ACM SIGCAS Conference on Computing and Sustainable Societies*. 126–137.
- [66] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* (2017).
- [67] Hongjian Wang, Zhenhui Li, and Wang-Chien Lee. 2014. PGT: Measuring mobility relationship using personal, global and temporal factors. In *IEEE International Conference on Data Mining*. IEEE.
- [68] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 3–14.
- [69] Shweta Ware, Chaoqun Yue, Reynaldo Morillo, Jin Lu, Chao Shang, Jayesh Kamath, Athanasios Bamis, Jinbo Bi, Alexander Russell, and Bing Wang. 2018. Large-scale automatic depression screening using meta-data from wifi infrastructure. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–27.
- [70] Martin Wirz, Pablo Schlapfer, Mikkel Baun Kjaergaard, Daniel Roggen, Sebastian Feese, and Gerhard Troster. 2011. Towards an online detection of pedestrian flocks in urban canyons by smoothed spatio-temporal clustering of GPS trajectories. In *Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks*.
- [71] Xiangye Xiao, Yu Zheng, Qiong Luo, and Xing Xie. 2014. Inferring social ties between users with human location history. *Journal of Ambient Intelligence and Humanized Computing* (2014).
- [72] Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudre-Mauroux. 2019. Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In *The World Wide Web Conference*.
- [73] Yu Yang, Zhihan Fang, Xiaoyang Xie, Fan Zhang, Yunhuai Liu, and Desheng Zhang. 2020. Extending coverage of stationary sensing systems with mobile sensing systems for human mobility modeling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2020).
- [74] Na Yu and Qi Han. 2014. Grace: Recognition of proximity-based intentional groups using collaborative mobile devices. In *IEEE International Conference on Mobile Ad Hoc and Sensor Systems*. IEEE.

- [75] Camellia Zakaria, Rajesh Balan, and Youngki Lee. 2019. StressMon: scalable detection of perceived stress and depression using passive sensing of changes in work Routines and group interactions. *Proceedings of the ACM on Human-Computer Interaction (CSCW)* (2019).
- [76] Camellia Zakaria, Youngki Lee, and Rajesh Balan. 2021. Detection of Social Identification in Workgroups from a Passively-sensed WiFi Infrastructure. *Proceedings of the ACM on Human-Computer Interaction (CSCW)* (2021).
- [77] Camellia Zakaria, Ameer Trivedi, Emmanuel Cecchet, Michael Chee, Prashant Shenoy, and Rajesh Balan. 2022. Analyzing the Impact of COVID-19 Control Policies on Campus Occupancy and Mobility via WiFi Sensing. *ACM Transactions on Spatial Systems and Algorithms* (2022).
- [78] Chao Zhang, Keyang Zhang, Quan Yuan, Luming Zhang, Tim Hanratty, and Jiawei Han. 2016. Gmove: Group-level mobility modeling using geo-tagged social media. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- [79] Pengpeng Zhao, Anjing Luo, Yanchi Liu, Fuzhen Zhuang, Jiajie Xu, Zhixu Li, Victor S Sheng, and Xiaofang Zhou. 2020. Where to go next: A spatio-temporal gated network for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2020).

A MODELING W4-GROUPS

A.1 Fast Pairwise Co-Occurrence Detection

The second module in *W4-Groups* requires the modeling of pairwise co-occurrences across all users. To address scalability concerns, we propose an approach that involves selectively comparing users and terminating the comparison process early when an overlap between pairs is detected, as outlined in Algorithm 1⁶.

We first sort the sessions by time for all users and subsetting sessions by location. These groups are filtered to be greater than size one (i.e., more than one session) and have more than two unique users. This step significantly narrows down the number of pairs that must be iterated. When iterating over groups, we use the sorted nature of the groups to identify when an overlapping session is found by comparing session entry and departure times. That is, for each user in each group, we iterate over sessions until the start time begins after the departure time of another user.

Algorithm 1: Fast Pairwise Co-Occurrence Detection

```

Data:  $S$ 
Result:  $C$ 
 $S \leftarrow \text{sort}(S, \text{by}=e)$ ;
 $S\_groups \leftarrow S.\text{groupby}(l)$ ;
 $S\_groups \leftarrow \text{len}(S\_groups) > 1 \ \& \ S\_groups[u].\text{unique}() > 2$ ;
for  $G$  in  $S\_groups$  do
  for  $i$  in  $G$  do
     $j \leftarrow i + 1$ ;
    found_overlap  $\leftarrow$  False;
    while !found_overlap do
      if  $G_i[e] > G_j[d]$  then
        found_overlap  $\leftarrow$  True;
      else
         $c_{ij}^t \leftarrow \{G_j[e], \min(G_i[d], G_j[d]), G_i[l], G_j[l], G_j[a]\}$ 
         $j \leftarrow j + 1$ 

```

⁶Note that this method is equivalent to brute force comparison in the worst-case scenario where all users interact with each other. However, we work on the assumption that this is unlikely in most public settings.

B PROTOTYPE IMPLEMENTATION

B.1 Mobility Data Types

Our current prototype supports two mobility data formats: WiFi mobility traces and location-based social network (LBSN) check-in data.

B.1.1 WiFi mobility and location data: *W4-Groups* can extract mobility and location traces of each user from logs collected by WiFi access points in the form of syslog files. WiFi syslog data consists of AP connection events across time [2]. As a user's device connects and disconnects to the APs, event logs are generated and aggregated in a centralized syslog server. Such data are routinely logged by commercial and consumer-grade access points. Further, many enterprises collect such information centrally using syslog servers for diagnosis and maintenance. All syslog messages are formatted as:

```
<date> <hh:mm:ss> <controller> <event_ID> <severity>
<AP, MAC and IP addresses> <message text with BSSID and SSID>
```

The most relevant items in the log include <date>, <hh:mm:ss>, event_ID, AP and MAC. <date> <hh:mm:ss> allow us to identify the precise timestamp of the event. The event_ID contains the event type generated. We particularly focus on association, disassociation, reassociation, authentication, de-authentication, and drift events. When a user connects their device to the WiFi network, the device generates an association event. As the user moves with their device to another location, the device gets disassociated and re-associated with the nearest corresponding AP. Additionally, authentication events often contain messages with usernames and user roles (i.e., student or faculty/staff in a campus setting) that allow for the identification of username and device mapping. AP contains a unique identifier involving the building name, floor information, and room number. MAC or Media Access Control addresses contain the unique device ID. Note, the MAC addresses contained in our study dataset are hashed. We describe in Section 5.1, participants consented to being detected through their hashed device MAC addresses.

The association and disassociation of MAC addresses to APs across buildings allow us to build a mobility trace of users. A fundamental assumption is that the users are connected to the same WiFi network. In order to remove inconsistencies and noise in the syslog data, the logs are grouped by MAC addresses, and each event is thoroughly analyzed to generate a sequence of association and disassociation logs. These logs are then interpolated as continuous mobility trajectories where missing entries are labeled with a "UNKN" location type.

For location types extraction, we classify college campus buildings into the following categories: administration, dining, health center, labs, landmark, library, parking, police, recreation, residential, school, student organizations, and others.

B.1.2 LBSN check-in data: *W4-Groups* can also extract location traces of a user using LBSN checkin data, which are a form of social networking data where the user connections and interactions revolve around content tied to locations. For instance, a user can share a message or content tied to a particular location, and other users can access this content when they arrive at the same location. Foursquare, Facebook Places, and Yelp are popular examples of LBSNs. LBSNs are formatted as:

```
<user_ID> <timestamp> <latitude> <longitude> <location_ID>
```

Embedded geo-locations in user check-ins or posts can be converted into mobility trajectories for each user. Since such data consists of location-based events at random intervals from users checking into the application, the data is often incomplete and not temporally continuous. However, *W4-Groups* is designed to support such data formats as well.

For location types extraction, we employ seven categories as in Liu et al. [39]: college campus buildings into the following categories: community, entertainment, food, nightlife, outdoors, shopping, and travel.