

DrinkSense: Characterizing Youth Drinking Behavior using Smartphones

Darshan Santani^{*†}, Trinh-Minh-Tri Do^{*}, Florian Labhart[‡], Sara Landolt[§],
Emmanuel Kuntsche[‡], Daniel Gatica-Perez^{*†}

^{*}Idiap Research Institute, [†]Ecole Polytechnique Federale de Lausanne (EPFL)

[‡]Addiction Switzerland, Research Institute, [§]University of Zurich

Abstract—Alcohol consumption is the number one risk factor for morbidity and mortality among young people. In late adolescence and early adulthood, excessive drinking and intoxication are more common than in any other life period, increasing the risk of adverse physical and psychological health consequences. In this paper, we examine the feasibility of using smartphone sensor data and machine learning to automatically characterize and classify drinking behavior of young adults in an urban, ecologically valid nightlife setting. Our work has two contributions. First, we use previously unexplored data from a large-scale mobile crowdsensing study involving 241 young participants in two urban areas in a European country, which includes phone data (location, accelerometer, Wifi, Bluetooth, battery, screen, and app usage) along with self-reported, fine-grain data on individual alcoholic drinks consumed on Friday and Saturday nights over a three-month period. Second, we build a machine learning methodology to infer whether an individual consumed alcohol on a given weekend night, based on her/his smartphone data contributed between 8PM and 4AM. We found that accelerometer data is the most informative single cue, and that a combination of features results in an overall accuracy of 76.6%.

Index Terms—Mobile Crowdsensing, Ubiquitous Computing, Youth, Nightlife, Alcohol, Prediction, Smartphone Data

1 INTRODUCTION

ALCOHOL consumption is the number one risk factor for morbidity and mortality among young people in many countries in the developed world. Heavy drinking and related incidents in public on weekend nights are a concern for city councils, policy makers, and a nuisance for the general public [1], [2]. In late adolescence and early adulthood, excessive drinking and intoxication is more common than in any other life period [3]. In addition, such behavior carries a significant risk of adverse psychological, social, and physical health consequences, including academic failure, unplanned pregnancy, sexually transmitted diseases, suicide attempts, violence, and accidents [3].

Advances in both alcohol epidemiology and human geography have recently provided initial evidence on young people’s going out and drinking behaviors in urban contexts. It is known that young adults consume alcohol both indoors (home, bars, or night-clubs) and in outdoor public places such as parks or openly on the streets [4], [5]. However, detailed knowledge on youth alcohol consumption habits, such as the types and frequency of alcohol they consume and the specific places where they drink, as well as contextual factors driving their consumption is still limited. To examine drinking habits, most studies in alcohol research have relied on self-reported retrospective assessments, which involves participants recalling their consumption in the past. This setting has three limitations. First, self-reported data on past alcohol consumption is prone to recall biases [6]. Studies have found that people forget to report up to half of their actual consumption [7]. Second, the environment in which drinking happens often differs from the one under which the self-assessment of drinking happens. Third, the collected data is limited by study design, focusing on single aspects of this complex phenomena.

There is a recent push to study alcohol consumption patterns

using mobile phones [8] to increase research validity and reliability and widen the understanding of contextual factors. In particular, smartphones, which can seamlessly collect a wide range of sensor data (location, motion, proximity), event logs (including all phone applications), and media (photos, audio, video) in everyday settings [9], offer a promising alternative to improve the ways in which drinking-related phenomena can be captured and studied. This is further complemented by the increasing use of wearable devices, such as body-worn devices, as means to monitor physical activity and promote self-reflection based on behavioral data [10].

In this paper, we examine the feasibility of using smartphone sensor data to automatically classify drinking behavior of young adults in an urban, real nightlife setting. The data was collected as part of a large-scale mobile crowdsensing study, called Youth@Night, to capture and examine the nightlife patterns of youth populations in Switzerland. Our goal is to study the discriminative power of smartphone-derived cues related to nightlife physical motion, location, connectivity, and application usage, to automatically infer whether a person has had any alcoholic drink that night. We focus our study on weekend nights (Friday and Saturday) as it has been shown that social drinking principally occurs on weekend nights, when most people do not have any work obligations or study responsibilities the next day [7]. We conducted our study by analyzing data collected from over 240 young people in two major cities in Switzerland, as a case study of a developed western country. Our research questions are the following:

RQ1: Is it possible to infer whether a person has consumed alcohol from smartphone sensor and log data in an uncontrolled, real-life setting?

RQ2: If so, which features are more predictive of alcohol consumption?

The contributions of the paper are the following:

- 1) Using a large-scale mobile crowdsensing study, we collected self-reported data on alcoholic drinks (type, size and alcohol volume) along with information gathered by phone sensors and application logs to examine the relationship between alcohol consumption and associated contextual features (motion, location, connectivity, and application usage). The study was conducted with a total of 241 participants aged 16–25 years old in two Swiss cities (Zurich and Lausanne), where each participant contributed data through over multiple weekend nights, resulting in a total of 2,934 user nights. We believe that the developed methodology provides an attractive alternative opportunity to monitor in-situ alcohol intake for experts specializing in alcohol research.
- 2) As a second contribution, we examined the relationship between alcohol consumption and various features extracted using smartphone sensor data using a penalized logistic regression model. Further, we build a machine learning methodology to infer the binary state of alcohol consumption for single user nights (i.e., whether a participant consumed any alcohol in a given night), and determine what are the most informative phone-derived cues. We found that accelerometer data is the most informative single cue, and that a combination of features results in an accuracy of 76.6%. To contextualize these findings, we explored the role of two potential confounding variables – going out (home vs. away) and gender – on the reported alcohol consumption and found that these two variables alone cannot explain the more nuanced notions of drinking behavior observed in our data. This opens up a number of opportunities for alcohol researchers to pose new research hypotheses.

The paper is organized as follows. We begin with a review of related work (Section 2). Next, we outline the design of the mobile application and the field study as part of the data collection framework (Section 3). All the collected data and subsequent filtering is reported in Section 4. We then describe the data pre-processing pipeline and feature extraction from the collected data (Section 5). After the feature extraction phase, we performed both regression analysis (Section 6) and automatic inference of alcohol consumption (Section 7). Section 8 contextualizes the regression and inference results and examines the role of going out behavior and gender on the reported alcohol consumption. Finally, we discuss and conclude with a summary of our findings and possible research directions for future work in Section 9. In the rest of the paper, “weekend nights” refers to Friday and Saturday nights between 8PM until 4AM.

2 RELATED WORK

Given the multifaceted nature of our research questions, we review the related work along the following two domains: alcohol epidemiology and ubiquitous computing.

2.1 Alcohol Epidemiology

Alcohol use is commonly considered a social activity, especially among adolescents and young adults, where peers are seen as the most consistent and strongest factor in the initiation and maintenance of alcohol use in adolescents and young adults [11]. Recent in situ studies suggest that young adults tend to drink more alcohol when in company than alone [12], that the number of persons present tended to be higher in heavier drinking nights [13] and that the higher the number of drinks consumed at a given time during the course of the evening [14]. Other factors such as

gender, composition of the group, activities, or drinking norms are also known to influence alcohol use among groups of peers [15], [16]. To our knowledge, no previous study has investigated the link between engaging in alcohol use and an automatic measure of social context.

Only a few studies have investigated the link between alcohol use and the number of drinking locations visited. For example, in a study of persons arrested for driving while intoxicated, Wiczorek et al. found that multi-location drinkers had higher blood alcohol levels at arrest than those who drank at a single location [17]. More recently, Dietze and colleagues reported that about two third of young adults visited at least 2 different locations, while 39% drank in only one location on their last big night out [18]. However, since these studies endorsed a public health perspective, they focused on the potential harms related to the quantity of alcohol consumed through self-reported data. To the best of our knowledge, ours is a first study which examines the link between automatically measured mobility and drinking behavior.

Self-reported data is potentially prone to recall biases. In the alcohol research literature, it has been shown that people forget to report about half of their actual consumption. In Switzerland, each person drinks about 3.4 liters of pure alcohol annually as per the survey data, while the alcohol sales data indicate an average of 6.8 liters of alcohol consumption per person annually [7]. In [7], the authors reported statistically significant differences between self-reported retrospective alcohol consumption and mobile phone based self-reports among young people. As a result, the feasibility to automatically characterize drinking behavior using smartphone data presents unique opportunities to study the urban nightlife of young adults.

2.2 Mobile Phones and Ubiquitous Healthcare

In ubiquitous computing, few groups worldwide have collected mobile phone sensor data that is at the same time rich, longitudinal, and covers a large population. One of the earliest work was done as part of the MIT’s Reality Mining initiative [19]. Nokia Mobile Data Challenge showed the feasibility of collecting continuous smartphone data from close to 200 users over the period of a year [9].

Research has also started to examine the role of smartphones and wearable devices to monitor health-related variables. In these domains, the recognition of stress levels has been an area of active research. In *StressSense*, the authors identified stress rate from human voice recorded using smartphones with an accuracy of 76% and 81% resp. in outdoor and indoor environments [20]. Another stress related study was conducted using five days of data from 18 participants in [10]. Using features derived from a variety of in-built smartphone sensors including accelerometer, mobile phone usage (call, SMS, location and screen events), their method achieved an accuracy of 75% in classifying the stress levels of participants (under stress or not). Recently, the authors have developed a stress model (called *cStress*), to establish a gold standard for continuous stress assessment in the mobile environment [21].

Another application which has received attention in ubiquitous healthcare research is the detection of smoking gestures [2], [22], [23]. In [22], the authors conducted a study with six participants, all wearing activity monitor devices in their wrists, to automatically detect puffing and smoking behavior. The authors developed a detection model incorporating temporal and high-level smoking

topographic features. In a similar work, Parate et al. proposed a machine learning pipeline to capture changes in arm orientations to detect smoking gestures in real-time [23].

In the field of alcohol consumption detection, there have been relatively few works. In [24], the authors proposed a phone-based system to detect the gait anomalies of walking under the influence of alcohol. Based on accelerometer data from three participants, the authors extracted gait features to differentiate intoxicated walking patterns from regular patterns. In the domain of drunk driving detection, a study on early detection of drunk driving using a smartphone placed inside the vehicle was described in [25]. Using accelerometer readings, the authors developed a mobile application to detect whether the driving patterns match with the cues for drunk driving gathered from real driving tests (such as lateral acceleration, lane positioning, speed control, etc.). To support people with alcohol addiction, Wang et al. proposed SoberDiary, a phone-based support system to help users monitor and manage their alcoholic intake and remain sober in their daily lives [26]. The study was conducted on 11 clinical patients over 4 weeks, using a Bluetooth-enabled breathalyzer which was paired with their smartphones. The authors found SoberDiary system helped patients reduce their alcohol consumption.

In [27], we present a first analysis of the Youth@Night data, focusing on in-situ and self-reported survey data to examine places, social context and nightlife activities of young people. In contrast, in this work we present a computational analysis of previously unexplored mobile sensor and logs data to infer alcohol consumption in an urban, ecologically valid nightlife setting, which to the best of our knowledge is a novel idea. Note that the sensor and logs data was not used in [27].

3 DATA COLLECTION FRAMEWORK

In this section, we describe our data collection framework and the various types of data collected. We conducted a large-scale field study to collect a large amount of real-life behavioral and mobile sensor data originally described in [27]. Our data collection methodology was approved by the ethical review board of Vaud and Zurich cantons, respectively for the cities of Lausanne and Zurich in Switzerland.

3.1 Mobile Application Design

As described in [27], to collect data for the Youth@Night study, we developed two Android-based smartphone applications: a survey logger and a sensor data logger application. The survey logger was designed to allow participants to respond to various in-situ surveys including reporting their drink consumption, while the sensor logger was designed to collect sensor and log data in a non-intrusive and privacy-preserving manner. Below, we describe these two applications.

3.1.1 Drink Logger

As part of the survey logger application, the drink logger module was designed to allow participants to report their drink consumption in real-time on weekend nights (Friday and Saturday) from 8PM to 4AM (the next morning.) The participants were asked to document their drinks (alcoholic and non-alcoholic), as they consume them during the night. In addition to in-situ reporting, the participants also had the opportunity to report forgotten drinks retrospectively any time during the night as part of the application design. Figure 1 shows the primary user interface of the drink

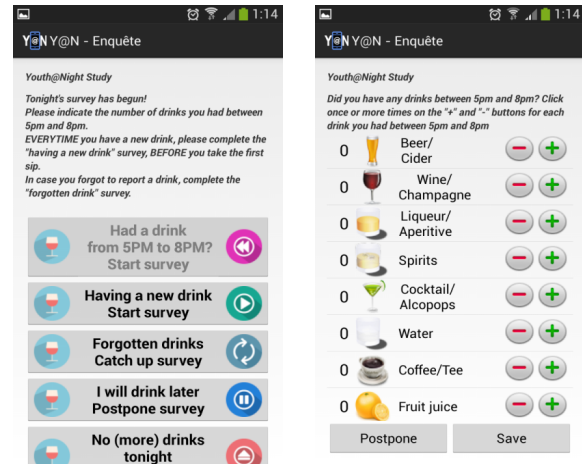


Fig. 1: Screenshots of the drink logger mobile application

logger application. The application was designed to facilitate users to report their alcohol consumption in a survey-based format. We report the list of survey questions in Section 3.3. Since the focus of the paper is to characterize drinking behavior, we are only describing the drink logger module. We refer the readers to [27] for a detailed description of the different survey modules of the survey logger application.

3.1.2 Sensor Logger

We developed a second mobile application to collect different types of mobile sensors and log data. It was designed to run as a background process without any user interaction. As with the survey logger application, all data was recorded only during the weekend nights from 8PM until 4AM in a non-intrusive manner. By design, it did not appear in the list of running applications on the user's device. So, if users wanted to close the service they had to manually terminate the application. We describe the type of data collected using the sensor logger in Section 3.3.

3.2 Field Study

For our field study, we recruited 241 participants across two major cities known for their nightlife activities in Switzerland, namely Zurich (German-speaking region) and Lausanne (French-speaking region). All participants are young, aged between 16 and 25 years (Note that the minimum age to purchase alcohol in Switzerland is 16 years). Before the study began, participants were asked to install the two applications on their own smartphones and contribute data for 10 weekend nights. During a weekend night, participants were asked to document every drink (alcoholic and non-alcoholic) using the drink logger application; at the same time, we collected sensor data and app logs using the sensor logger. All participants were informed of the different data types collected for the study, as well as on all other aspects of the data collection.

3.3 Data Types

In this section, we describe the data types collected in our study.

3.3.1 Drink Survey Data

This data was collected using the drink logger module of the survey logger application. This survey was designed to let users report their alcoholic and non-alcoholic beverages. Users were

asked to answer this survey any time during a weekend night, with an automatic reminder every hour prompting users if they wanted to report a new drink. While answering the survey, users were asked to describe the attributes of their current drink, that include the type of drink (beer/cider, wine/champagne, whisky, tea/coffee, fruit juice, water, etc.), the size of the drink (small, medium and big), and the alcohol quantity (light, medium, strong, and alcohol-free). The list of the drink attributes were selected based on existing literature on alcohol consumption [28].

It is common for users to forget reporting every consumed drink. As a result, we designed an additional survey for users to report any forgotten drinks, which otherwise should have been reported using the drink survey. As with the drink survey, this survey could be answered any time during the weekend night from 8PM until 4AM, with an automatic reminder every hour prompting users to add their forgotten drink. In this survey, users were asked to indicate only the number of forgotten drinks per drink type. As a result in this survey, we have no information of the drink attributes (other than the drink type), as in the drink survey.

3.3.2 Sensor Data

In addition to collecting the participants’ self-reported alcohol consumption, we also collected data from various mobile sensors and application logs using the sensor logger application (Section 3.1). The data types included accelerometer, application, battery, bluetooth, location, screen and wifi logs. Table 1 lists all data types, along with their respective sampling rate. Sampling rate for various sensors was optimized using experimentation to optimize battery life, following prior work [9]. We conducted internal tests to select sampling rates for different sensors which would not drain the battery more than 50% during eight hours of data collection (8PM-4AM), assuming that the battery was fully charged before the start of a night. Participants were sent reminders during weekends to charge their phones six hours before the night began. For user safety and convenience, the sensor sampling was terminated completely when the battery level fell below 20%. All the experiments done during the app design phase were conducted using Samsung ACE 3 and Samsung Galaxy S3 phones running Android version 4.3. Note that due to the diversity of phone vendors and mobile versions (eight mobile vendors and 51 model versions in our study), it is difficult to optimize the battery profiles for all possible devices during an “in-the-wild” crowdsensing study [29], [30].

In addition to collecting the survey and sensor data, we interviewed 40 participants about their nightlife experiences, including their experiences with the data collection by the mobile application, as reported in [27].

4 DATASET

In this section, we describe the data collected as part of the “in-the-wild” field study.

4.1 Collected Dataset

We ran our field study from September to December 2014. We collected a total of 2,546 *drink surveys* and 942 *forgotten drink surveys*, while at the same time we gathered over 8 million sensor data records from 241 participants from the two Swiss cities. For the rest of the paper, we conduct our analysis on a *user-night*. A user night indicates a night-out per user. If the field study run for m nights, where for each night we have n participating users,

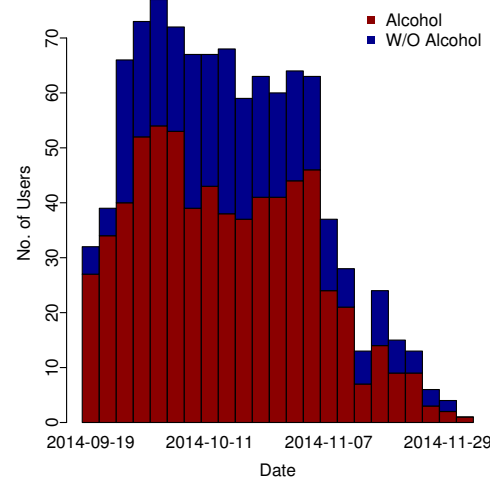


Fig. 2: Distribution of participants contributing data to the field study. Numbers in red indicate users who reported consuming alcohol on a specific night, while in blue we show the participants who had reported no consumption of alcohol throughout the night.

the total number of user-nights will be mn . In our analysis, we collected data from 2,934 user-nights from the 241 participants. On average, a participant contributed data for 12 nights, with more than half contributing 14 nights or more. The maximum number of nights contributed by a single user is 30.

The data collected provides numerous opportunities to explore different angles of drinking behavior. In this paper, we used a subset of the data to analyse and infer drinking behavior. We plan to explore other angles of drinking activity as part of future work.

4.2 Data Filtering

In our data, we observe that many user nights miss one or more sensor data type. This is inevitable given the real-time and in-situ nature of our study; similar trends have been reported in previous large-scale mobile data campaigns [31]. In [31], using data collected over five months in a real world setting, the authors reported that the availability of GPS data ranges on average between 5–30% of the time for a mobile device. Table 1 lists the number of user nights with missing records for all sensor data types. Missing sensor data can be attributed to a variety of reasons. First, in our study, we have encountered a significant diversity of mobile devices – eight mobile manufactures (including, Samsung, HTC, Sony, LG, Huawei, etc.) and 51 different model versions. Different phone manufactures and model versions behave differently even when they are running the same version of Android OS. As a result it is non-trivial to control for phone sensor biases including sampling rate heterogeneity and instability [32]. Second, the Youth@Night participants were instructed to use their phones in a normal way following their usual practices. So, it was common for participants to run out of battery, switch off their smartphones, terminating the sensor logger altogether, selectively disabling GPS or wireless interfaces (WiFi and Bluetooth) due to battery constraints, forgetting their phones at home or devices proactively going to power saving mode, etc. Furthermore, for each weekend night, the sensor data collection began at 8PM (Section 3.1), so it is likely that the participants were running out of battery more often. Third, for participants safety and convenience, the sensor sampling was terminated completely when the battery level fell below 20% (Section 3.3.2).

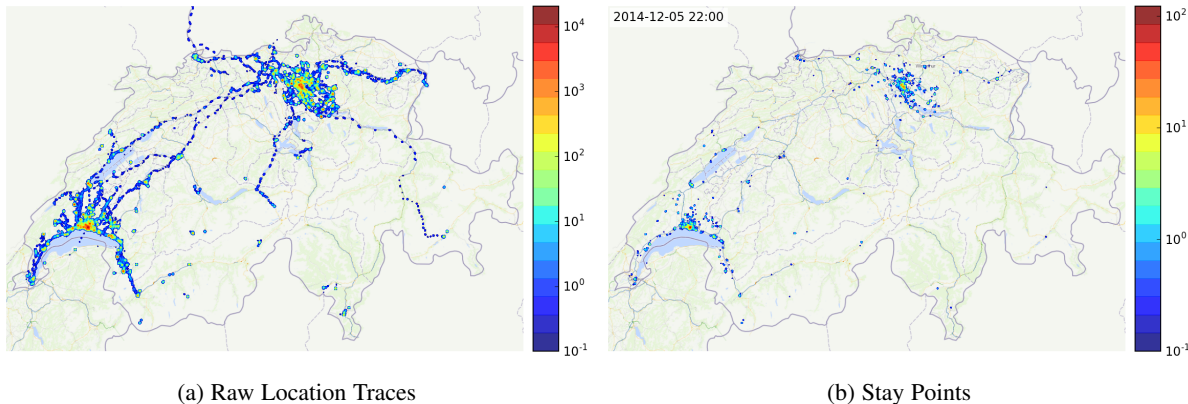


Fig. 3: Heatmap showing spatial distribution of raw location traces (left) and stay-points (right) of participants. Zurich (north-east of the country) and Lausanne (south-west) are clearly represented in the map. Many other places in the country were also visited. Red color indicates high density of stay-points, while blue represents low density. For scale purposes, note that the driving distance between Lausanne (southwest red spot of the map) and Zurich (center north red spot of the map) is about 220 km (according to Google Maps). Best viewed online in color and high-resolution.

Sensor Data Group	Description	Sampling Rate	% User Nights with Available Records in Raw Data	% User Nights with Available Records in Filtered Data
Accelerometer	Set of acceleration values along three axes	10 seconds continuously every minute at 50Hz	76.18%	99.51%
Application	List of all background applications	Once per minute	38.24%	99.31%
Battery	Battery charging status	Whenever a change in battery status is detected	76.24%	100%
Bluetooth	List of bluetooth devices in range	Once every 5 minutes	75.66%	98.52%
Location	Set of location estimates using GSM or GPS	1 minute continuously every 2 minutes	55.69%	100%
Screen	Status of the phone screen (on/off)	Whenever the screen state is changed	69.33%	92.78%
WiFi	List of all visible WiFi hotspots	Once every 5 minutes	65.13%	93.67%

TABLE 1: List of all types of sensor and log data collected during the study (in alphabetical order). The last two rows indicate the percentage of user-nights with available data in the raw and filtered data respectively.

Due to inherent sparsity, we filtered the data to select only those user-nights for which sufficient sensor data is available. Consequently, for our analysis and inference framework (Sections 5, 6 and 7), we select only those user-nights which satisfy the following three criteria: a) the participant had responded to either the *drink* or the *forgotten drink* survey at least once during a given night; b) the participant had at least one data sample for any sensor data type during the night; and c) the participant had at least one stay-point based on location sensor logs. (A stay-point is a stable location defined in detail in Section 5.5.)

As a result of filtering, we are left with 1,011 user-nights from 160 users consisting of 12 Fridays and 11 Saturdays nights between September and December 2014. In Figure 2, we show the distribution of participants for each night, in addition to showing the overall participation rate. We observed 12 weekend nights where we had more than 50 participants contributing data with drinking events to our study. Figure 3 shows the spatial movement of the participants based on their accumulated stay-points and individual location traces, plotted at the country level. The study participants restricted their movements mostly to the two studied cities – Lausanne and Zurich; but sometimes they also spent their nights in neighbouring cities as a reflection of real-life mobility. Even after filtering, the data contains user-nights with missing data for some of the sensor data types. In Table 1, we show the percentage of user-nights with available data in the raw and filtered data. For instance, 64 user-nights have missing WiFi sensor

logs in the filtered dataset. It points towards the challenges of in-situ studies. For the subsequent analysis (pre-processing, feature extraction, regression analysis and classification tasks), we use the filtered 1,011 user-nights dataset, unless otherwise stated.

5 FEATURE EXTRACTION AND ANALYSIS

In this section, we describe the preprocessing of data and the procedure for extraction of features. As mentioned in the previous section, our unit of analysis is a *user-night*. Consequently, we aggregated all the data informing different aspects of phone usage per user on a nightly level. Table 3 lists all the different features extracted. Features were selected based on prior work that could potentially characterize levels of drinking activity [10], [31], [33], [34], [35], [36]. Note that, the feature extraction process is driven with the objective to test the feasibility of inferring a binary state of alcohol consumption using features extracted from raw phone sensor data, without inferring first any mid-level representation from phone sensors. Below, we describe each data type and list all the extracted features.

5.1 Accelerometer Logs

Using the *sensor logger*, we obtained accelerometer data every minute for 10 seconds, with a sampling rate of 50Hz resulting in a total of 500 readings per minute. Each individual reading contained the raw acceleration values along the three axes (x_i, y_i, z_i).

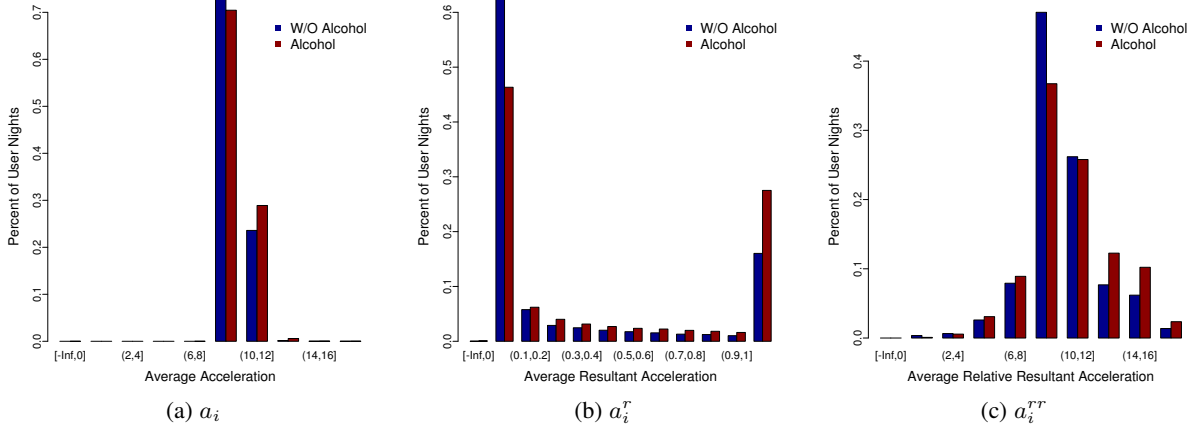


Fig. 4: Plots showing the histograms for a) Acceleration, b) Resultant Acceleration, and c) Relative Resultant Acceleration for both night types.

Id	Category	Total Apps	Top-3 Apps
C1	Systems/Native	39	Contacts, Accueil TouchWiz, Parametres
C2	Communication	12	Messages, InCallUI, S Voice
C3	Entertainment	8	Youtube, Musique, Shazam
C4	Social	5	Whatsapp, Facebook, Snapchat
C6	Travel	5	Maps, Mobile CFF, SBB Mobile
C5	Other	28	Chrome, Internet, Recherche Google
C7	Below100	1,200	Alben, Applications actives
C8	Y@N	3	Y@N-Enquete, Y@N-Studie, Y@N-Study

TABLE 2: Categorization of mobile applications. Apps in multiple languages appear due to the multi-region nature of the study.

Given the raw readings, we computed basic features including signal magnitude area (SMA_i), raw acceleration (a_i), resultant acceleration (a_i^r), and relative resultant acceleration with respect to the previous reading (a_i^{rr}). Note that feature names are taken from previous literature [33], [37].

We define each scan per minute as a *segment*. Our data contains a total of 391,717 segments. To transform the raw readings per segment, we further computed basic summary statistics (mean, standard deviation, min., max.) over each segment, in addition to calculating the angle between the gravity (g) vector and the vector containing the average acceleration value for each axis ($[\bar{x} \ \bar{y} \ \bar{z}]$). To obtain the aggregated features per user-night, we generated histograms of segment features, as described in Table 3. In total, we generated a total of 79 acceleration features (all variants of five basic features.) Most of the features extracted from accelerometer logs are derived and adapted from prior work [33], [34], [38]. For instance, the SMA feature is designed to distinguish between static and dynamic activities, while the angle provides cues to inform phone’s tilt and orientation. Figure 4 shows the distribution of raw acceleration, resultant acceleration and relative resultant acceleration for both alcoholic and non-alcoholic user-nights.

Accelerometer data has widely been used to recognize physical activities of users ranging from standing, sitting, walking, jogging, and climbing [33], [34], [38]. We are not inferring categories of physical activity, as we did not collect physical activity labels as part of the field study. To complement the existing features, we plan to extract spectral based features (e.g., energy, spectral entropy, etc.) from accelerometer data as part of future work.

5.2 Application Logs

There have been numerous studies which have examined mobile applications logs to understand user context ranging from location, temporal and social context [39], [40], [41]. In this subsection, we

extract features to explore the role of mobile application usage (in terms of spatial and social context) on alcohol consumption. Application usage might be informative of drinking-related activity as it has been shown that differentiated usage of apps occur in real life depending on where and with whom the phone users are.

Through the *sensor logger*, we scanned the list of all running applications every minute (Table 1). For each scan, we obtained a list of up to 50 applications. It is important to note that, as opposed to dedicated apps developed to monitor application usage [41], our *sensor logger* was designed to gather only the list of currently running apps. As a result, our analysis of application logs is limited to understand fine-grained application usage behavior (e.g., time spent on Facebook, number of messages and pictures shared using Whatsapp, etc.). In our dataset, we observed a total of 1,299 unique applications, with Whatsapp being the most frequently used application. Running applications can either be native (i.e., pre-installed by the device manufacturers), default Android apps or user-installed apps. As noted earlier, due to the multi-region nature of our study, the mobile interface was designed in three different languages. Consequently, apps having the same functionality appear in different names due to the translated interface e.g., Telephone and Telefon both refer to the Android application handling the call activity in English and French interface respectively.

Application Categorization: To deal with the diversity of applications, we manually coded the top 100 applications (based on their usage frequency) into eight categories (Table 2) which are: System/Native, Communication, Entertainment, Social, Travel, Other, Below100, and Youth@Night (in short Y@N, our app for this study). Similar app categorization systems were used to group mobile apps in the literature [40], [41]. In Table 2, we show the total number of applications classified in a given category, in addition to showing the Top-3 apps for that category. The Below100

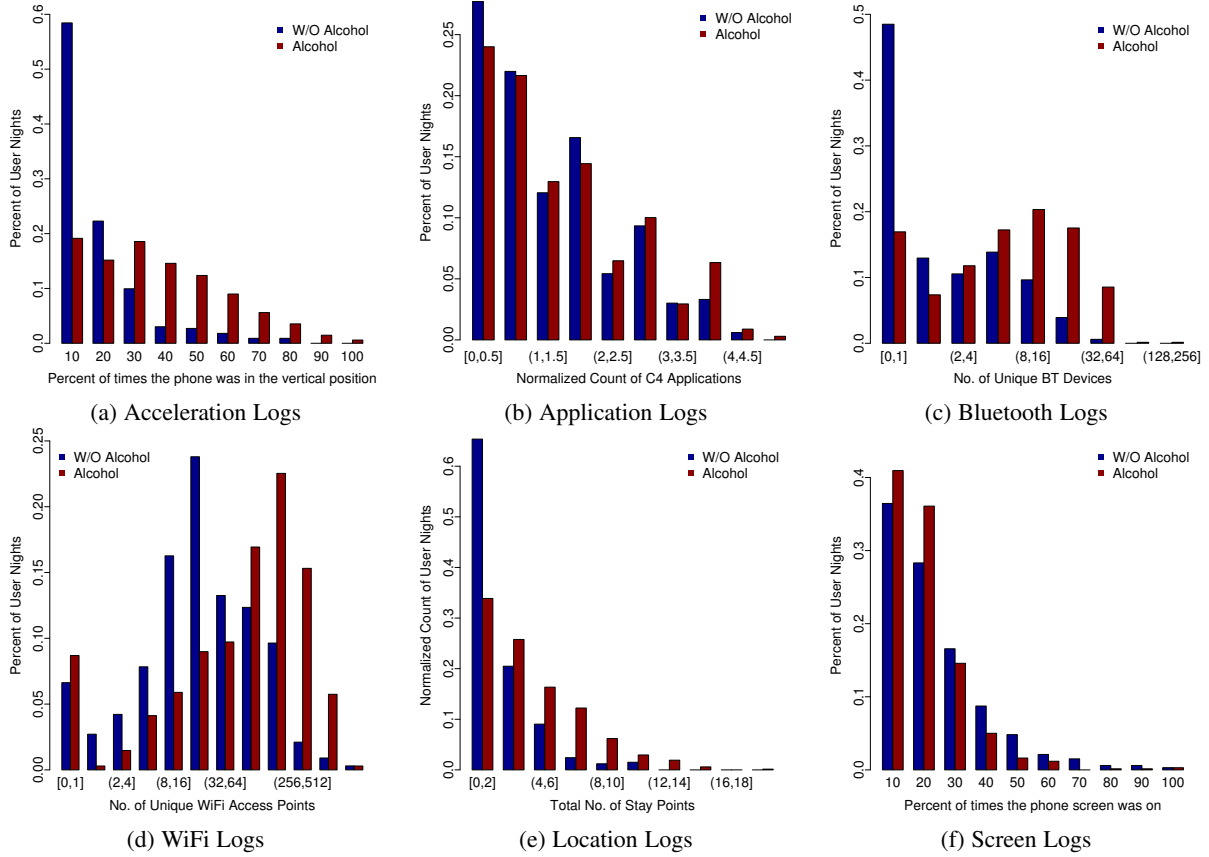


Fig. 5: Plots showing the histograms for a) Percent of times the phone was in the vertical position inferred using accelerometer logs, b) Normalized *Count 1* of “Social” applications (C4 category), c) Number of unique Bluetooth devices in range, d) Number of unique WiFi access points, e) Total number of stay-points, and f) Percent of times the phone screen was on inferred using screen logs for both alcoholic and non-alcoholic user-nights.

category accounts for applications that were not frequently used by the participants in the study, while $Y@N$ denotes applications designed for our study. Note that the $Y@N$ category contains only the *drink logger* application; the *sensor logger* app, being a daemon service, did not appear in the list of running applications. Applications falling under $Y@N$ category also belongs to the Top 100 applications. Note that each application was assigned to the closest possible category, though an application in principle could belong to multiple categories.

For each user-night, in addition to the basic features (e.g., number of records or scans, count of unique applications, duration), we generated a set of three additional count features using the categorized applications. First, for all the applications obtained per scan, we increment their respective category count (*Count 1*). Second, we examined the top-3 applications (based on their recency order) in a scan and increment their respective category count (*Count 2*). Note that while returning the list of running apps, the Android API returns the list in order of recency, i.e., the most recently app used is returned first, and so on. Using the Android API, we have no means to examine for how long an app has been used, thus we used the recency order as a proxy to understand temporal app usage. For the last and third count feature (*Count 3*), we count the number of times an app category has improved its rank (based on the recency order), with respect to the previous scan. For *Count 3*, we look at only the top-3 applications and their respective categories. We increment the count only if there is an improvement in the rank with respect to the previous rank.

For each count feature, we also added their normalized count version as features. Count normalization was performed based on the number of scans in a user-night. In total, we generated a set of 52 features using application logs (Table 3).

For all eight categories, we do not observe any significant differences for all count features between alcohol and non-alcohol user-nights. To illustrate it, in Figure 5b, we show the distribution of normalized count (*Count 1*) for “Social” applications (C4 category) across several bins for both types of user-nights.

5.3 Battery Logs

The battery sensor returns a set of values whenever a change in the battery status is detected. Via this sensor, we obtained a set of features including battery status, level, temperature, and whether the phone was plugged to a power source. The battery can be in one of five states (charging, discharging, full, not charging, and unknown), depending on its power status [42]. Using the battery logs, we generated 10 features as summarized in Table 3.

5.4 Bluetooth and Wifi Logs

From the Bluetooth and WiFi sensors, we obtain the list of the nearby Bluetooth (BT) devices and WiFi access points (APs) respectively. For these sensors, the scan was performed every 5 minutes. For every scan, we compute the list of features as described in Table 3. All the generated features are self-explanatory and are computed based on prior work [35], and correspond to a total of 6 features (Table 3).

Feature Group	Features (Dimensions)	β
Accelerometer	Number of records (1)	.
	Hist. of angle between g vector and X, Y, and Z-axis (6x3)	$+5.86 \times 10^{-1}$
	Hist. of std. dev. of acceleration along X, Y and Z-axis (6x3)	+1.20
	Hist. of mean, median and max resultant acceleration (6x3)	+4.59
	Hist. of mean, median and std. of relative resultant acceleration (6x3)	$+8.12 \times 10^{-1}$
Application	Hist. of signal magnitude area (SMA) (6)	.
	Number of records (1)	.
	Count of apps in different categories (8x2)	$+4.22 \times 10^{-2}$
	Count of Top-3 apps per category (8x2)	$+8.45 \times 10^{-1}$
	Relative ranking of Top-3 apps per category (8x2)	+1.84
Battery	Duration between the first and last scan (1)	-3.51×10^{-5}
	Count of unique and total apps (2)	.
	Number of records (1)	-2.12×10^{-4}
	Count of various battery status (5)	$+3.91 \times 10^{-2}$
	Min. and Max. level of battery (2)	-2.49×10^{-3}
Bluetooth (BT)	Difference between min. and max. battery levels (1)	-5.17×10^{-3}
	Plug time (1)	-1.49×10^{-1}
	Number of records (1)	-1.98×10^{-3}
	Number of unique BT IDs in range (1)	$+1.26 \times 10^{-1}$
	Number of BT scans (1)	-6.07×10^{-3}
Location	Percent of empty BT scans (1)	$+2.74 \times 10^{-1}$
	Number of stay-points (1)	$+1.81 \times 10^{-1}$
	Sum of duration at stay-points (1)	.
	Sum of travel distance between consecutive stay-points (1)	$+1.17 \times 10^{-7}$
	Sum of travel time between consecutive stay-points (1)	.
Screen	Hist. of computed speed (8)	+7.93
	Number of records (1)	$+5.87 \times 10^{-3}$
	Percent of time screen was <i>on</i> (1)	-2.49
	Screen count after midnight (1)	.
WiFi	Duration of screen events (<i>on/off</i>) (2)	+1.64
	Number of records (1)	-8.7×10^{-4}
	Number of unique visible WiFi hotspots (1)	$+9.02 \times 10^{-3}$

TABLE 3: List of all features extracted (organized by feature group) for each user-night. For each feature, β refers to the coefficient estimate of the penalized logistic regression (PLR) model.

Bluetooth device density has often been used as a proxy to the social context of the user’s environment e.g., being alone or in the vicinity of a nearby group of people [40], [43]. When a large number of BT devices are detected, it likely signals the presence of people in the surrounding. Analysing BT logs becomes relevant to our work as we are interested in understanding the social context of participants and how it might potentially impact their drinking behavior. In our study, participants encountered an average of 9.6 unique BT devices and 128 WiFi APs in a typical user-night. For alcohol user-nights, 12.3 BT devices and 162 WiFi APs were encountered, while for non-alcohol user-nights only 3.95 devices and 59.3 WiFi APs were detected on average. Figure 5c and Figure 5d shows the distribution of BT devices and WiFi APs, respectively for both alcoholic and non-alcoholic user-nights.

5.5 Location Logs

The location sensor returns the set of location estimates using either GPS sensor or network-based (estimated using a lookup of cell tower and WiFi access points), depending on which sensor was available at the time of sampling. Location data was collected continuously for one minute, every two minutes. For every data point, we collect the location coordinates (longitude and latitude) and provider information (GSM-based or network-based). Additionally, the sensor logger also listened “passively” for any location requests made by other mobile applications and services (such as Maps) and recorded the location without actually initiating a location fix and consequently not draining the battery further. In our analysis, we have estimated the location using both the “active” and “passive” data sources. As a result, some of the location updates might have resulted in coarse grain estimates.

For each user-night, we extracted the sequence of stay-points using location pairs (longitude and latitude). A stay-point is a region (radius of d meters) where a user has stayed for a given duration of time (t minutes). It is a standard unit of analysis for place extraction [31]. In our analysis, we have used $d = 200$ meters and $t = 5$ minutes, which is similar to the ones reported in prior work [36]. Stay-points are extracted independently for each user. Using stay-points, we extracted a series of features, such as the duration of stay at stay-points, travel distance, and time between consecutive stay-points (Table 3). Additionally, we also computed user speed using the location data. The total number of extracted location-related features is 12.

In this paper, we have not extracted any semantic features from location data i.e., to estimate if the participant were hanging out at a bar or sitting at home. Location data would play a dominant role if the semantic label of places were known. However, automatically extracting semantic place labels from raw location traces involves multiple research challenges including noisy and missing location data (see Figure 6), and collection of manual place annotations [44], [45]. To avoid collection of such labels, an alternative would be to use location services like Google Places API, to extract meta information about places. In urban areas (where the Youth@Night study took place), reverse geocoding the location data would result in a list of possible candidate places, and disambiguating them to locate actual places participants had been to constitutes a research challenge [46]. Errors in place disambiguation mechanisms could potentially lead to wrong places being discovered e.g., a bar next to a grocery store. Due to these research challenges, we plan to extract semantic place attributes as part of future work.

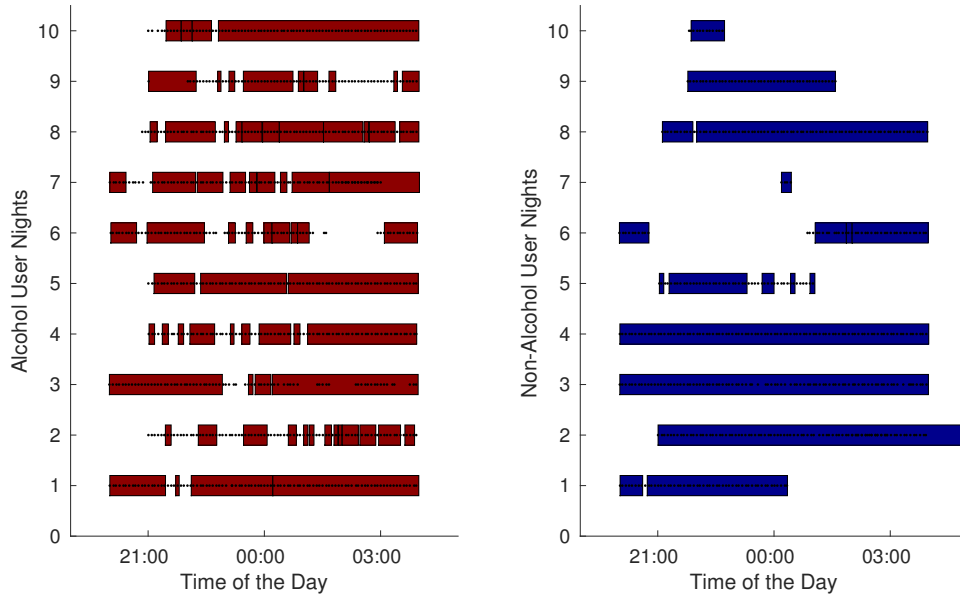


Fig. 6: Stay-points for a random selection of 10 alcoholic and non-alcoholic user-nights. A rectangle bar indicates a stay-point, and black dots indicate the raw location traces.

Mobility using Location Logs: In a user-night, on average a participant had stayed on 3.8 stay-points, with the mean stay duration of 2.1 hours in a given stay-point. For alcohol user-nights, a participant spent an average of 1.6 hours per stay-point, while for nights where alcohol was not consumed, a user spent an average of 3 hours per stay-point. We also observe significant differences with respect to the number of stay-points between nights spent consuming alcohol (4.5) and otherwise (2.5). Figure 5e shows the distribution of total number of stay-points across several bins for both alcoholic and non-alcoholic user-nights, which highlights the role of mobility on alcohol consumption. To contextualize these results and further understand the role of mobility, we examined the distance travelled between consecutive stay-points. We found that on average a participant travelled a distance of 3.7kms (resp. 1.4kms) in an alcoholic user-night (resp. non-alcoholic nights). These findings suggest potential links between mobility and drinking behavior. Besides, these results complement our understanding of the physical mobility of participants with respect to our previous work, where we conducted a first analysis of visited place categories using self-reported check-in data [27].

In Figure 6, we show the stay-points for a random selection of 10 alcoholic and non-alcoholic user-nights. In the figure, a rectangle bar indicates a stay-point, and black dots indicate the raw location traces. It was not always the case that we continuously obtained location traces for the entire user-night, as evident in Figure 6. For most user-nights, we observe missing location data for parts of the night. For some user-nights, the data was received only for few minutes (see non-alcoholic user-nights 7 and 10). These findings point towards the inherent missing data in our collected dataset and the subsequent challenges involved in using the data for an inference task.

5.6 Screen Logs

Using the screen logs, we measured any change in the state of the screen i.e., whether the screen was *on* or *off*. Using these logs, we computed a set of five features following [10] (Table 3). The Android API only provides the information when the state of the

mobile screen changes from *on* to *off* and vice-versa. Using this information, we computed the percent of time the screen was *on* (or *off*) by normalizing it based on the time difference between the first and last screen events in a given user-night.

We found that on average, participants had their mobile screen *on* for 15.6% of the time, indicating some form of interaction with their mobile devices. When examining differences between user-nights, we found that for alcohol user-nights (resp. non-alcoholic nights), a participant kept their screen on for an average of 14.4% (resp. 18.3%) of times (see Figure 5f for a visual comparison).

5.7 Ground-truth Alcohol Consumption

Finally, in order to obtain the ground-truth data on alcohol consumption i.e., whether a participant had drunk alcohol during a user-night, we rely on a combination of “in-situ” *drink survey* and “retrospective” *forgotten drink survey* data as opposed to a purely retrospective survey method (Section 3.3.1). The retrospective survey was designed to help participants report forgotten drinks during a specific user-night only, as opposed to longer recall periods as is the case with most alcohol epidemiology studies. In [6], considering 1 day as the reference recall period, the study showed a significant decrease in reported alcohol intake with increasing recall period (between 2-7 days). Furthermore, in our data 73% of drink survey responses were in-situ (Section 4.1).

Participants reported their drinks or forgotten drinks using the *drink logger* application. Participants reported both alcoholic and non-alcoholic beverages. If a participant reported consuming both types of drinks, we consider the night as alcoholic user-night. If a participant had not reported a consumed drink using the *drink survey* i.e., while consuming it in-situ, but reported it the next day using the *forgotten drink survey*, we considered those user-nights in our analysis as well.

In total, 67% of user-nights have reported alcohol consumption, while 33% of user-nights have no alcohol consumption. This imbalance signals potential biases of participants towards reporting mostly alcohol-related activities. After the in-situ data collection was over, we conducted a series of qualitative interviews

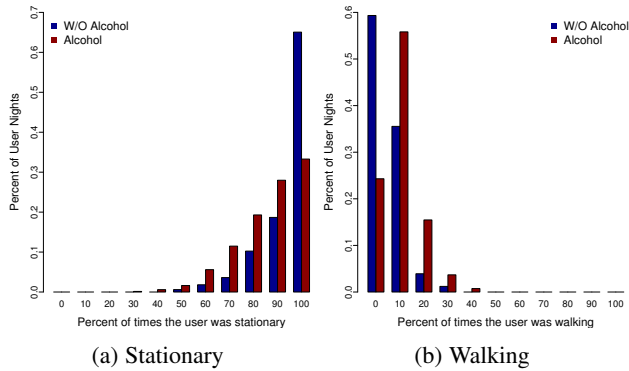


Fig. 7: Plots showing the histograms for a) Stationary, and b) Walking inferred using location logs for both night types.

with 40 participants to gain insights of their nightlife experience (Section 3.3). During the interviews, some of the participants mentioned perceiving our study as an “alcohol study” where they did not consider reporting consumption of non-alcoholic beverages (e.g., water, juice, etc.). This is interesting given the fact, that the instructions given to participants clearly stated them to report both alcoholic and non-alcoholic drinks (i.e., without any bias.) Few other interviewees concluded that they were only allowed to report if they had bought the drinks themselves. Such narrations provide insights into the ways a “drink” was contextualized by participants. Furthermore, they offer a starting point to reflect the observed reporting bias. Detailed analysis of interviews will be reported as part of social science research.

6 REGRESSION ANALYSIS

In this section, we perform regression analysis of different feature groups. To examine the relationship between alcohol consumption and feature groups, we employ penalized logistic regression (PLR) [47]. PLR measures the relationship of a binary response variable (consuming alcohol or not) as a function of the explanatory predictor variables (set of features derived from sensors and log data, as described in the previous section). In contrast with traditional logistic regression, PLR guards against collinearity amongst features via regularization. Regularization is implemented using the LASSO technique, which performs feature selection by forcing some of the coefficient estimates to zero, thereby increasing the interpretability of the model.

Precisely, we fit a generalized linear model with LASSO regularization ($\alpha = 1$) over a grid of 10,000 regularization parameter (λ) values, ranging from 10^{-5} to 10^5 . To choose the optimal λ that minimizes the error rate (λ_{MIN}), we used a 10-fold cross-validation technique for each PLR model. All predictor variables were standardized before fitting the model to increase model interpretability and compare the relative importance of (post-shrinkage) coefficient estimates. To conduct our regression analysis, we have used the `glmnet` library in R programming language. `glmnet` provides statistical routines to fit a generalized linear model for different penalties including L1 (LASSO) and the library is ideally suited for most data distributions, including sparse data [48].

6.1 Results

For each feature group, we fit the penalized logistic model involving all features of that group against the binary alcohol

consumption variable in a user-night. Table 3 shows the results of the PLR models, where we report the coefficient estimates (β) for each feature group. As λ varies, the number of coefficients in the fitted model also vary. In Table 3, we are reporting the β estimates at λ_{MIN} (λ_{MIN} is the value of λ that gives the minimum mean cross-validated error.). For feature sets involving multiple dimensions (e.g., histogram of mean relative acceleration), we report the maximum β value for that feature set. Due to regularization, features with zero coefficients are reported blank in Table 3.

We found that, while most features have either a positive or negative relationship with alcohol consumption, some features have zero β values (e.g., histogram of signal magnitude area, sum of duration at stay-points, screen events after midnight, etc.) Features with zero β values suggest that either these features do not possess any significant relationship with the alcohol consumption or these features are correlated with other significant predictive feature(s) within the same feature group. The accuracy of the PLR models ranges from 77% to 69% with accelerometer data obtaining the best predictive performance. While evaluating the accuracy of each PLR model, no cross validation was used, so the baseline accuracy of each model stands at 67% (Section 5.7).

Most of the regression results presented in Table 3 corroborates with the exploratory analysis reported in Section 5. For acceleration logs, the standard deviation of raw acceleration values (a_i) along the Y-axis has a positive relationship, and the maximum of the resultant acceleration values (a_i^r) also has a positive association with alcohol consumption. When examining the features extracted using location logs, we found that the number of stay-points (Figure 5e) and computed speeds are positively associated with alcohol consumption. The percent of time the mobile screen was *on* is negatively associated with alcohol consumption with β -value of -2.49 (Figure 5f).

Amongst all features within all feature groups, user speed computed using location logs has the highest absolute β value (7.93). To further understand the role of user speed, we examined the distribution of user speeds for both alcoholic and non-alcoholic user-nights. A participant was considered stationary if the computed speed was below 1 km/h, while a participant was assumed to be walking if the speed was in the range of 2–5 km/h [49]. In Figure 7, we show the percentage of time (as a fraction of a user night) when the user was stationary or walking for both night types. For more than 65% (resp. 33%) of non-alcoholic (resp. alcoholic) user-nights, a participant was stationary throughout the night, shown in Figure 7a. Note that we did not include the stationary or walking labels as features in the inference task (Section 7). In Figure 7, we have shown their respective distributions for illustrative purposes only.

7 INFERENCE OF DRINK STATUS

In this section, we investigate the problem of classifying whether a study participant consumes alcohol in a given user-night.

7.1 Classification Method

For automatically classifying the alcohol consumption status as a binary classification task, we use Random Forests. Random forest represents a tree-based supervised learning method for classification, which guards against overfitting of the model to the training data [50]. In our analysis, we set the number of trees to $N = 500$. From a practitioner point of view, it has been recommended to use at least 1,000 trees as a starting point [51], although at the

Feature Group	Accuracy	Precision	Recall	Feature Combination	Accuracy	Precision	Recall
Accelerometer	75.8%	74.0%	78.7%				
Location	68.5%	68.2%	68.0%	A+L	76.3%	74.5%	79.2%
WiFi	65.2%	65.1%	64.1%	A+L+W	76.4%	74.7%	79.1%
Bluetooth	64.2%	64.4%	62.0%	A+L+W+B	76.6%	74.8%	79.6%
Screen	61.1%	60.2%	63.2%	A+L+B+W+S	76.6%	74.9%	79.3%
Apps	61.1%	60.7%	60.5%	A+L+B+W+S+Ap	76.4%	74.7%	79.0%
Battery	58.8%	58.8%	56.3%	A+L+B+W+S+Ap+Ba	76.4%	74.7%	79.1%

TABLE 4: Classification accuracy, precision and recall metrics of alcohol consumption using each data type and using different combination of data types.

cost of increasing the computational time for the experiments. For our experimental setting, the computational time was not a bottleneck. We conducted our random forest classification using `randomforest-matlab` package in Matlab.

7.2 Performance Evaluation

Each user-night is represented by a feature vector from one or multiple feature groups. Since the user-nights are imbalanced (67% of user-nights have reported alcohol consumption, and 33% of user-nights have no alcohol consumption), we randomly sub-sample the majority class (i.e., alcohol consumption user-nights) to build 10 different balanced datasets. By training and evaluating the classifiers on balanced datasets, we target a classifier predicting binary alcohol consumption. For each balanced set, we use the 10-fold cross-validation approach. To guard against the case where the data of one user was distributed between both the training and the test sets, the 10-folds were created based on disjoint users (as opposed to the user-night). The final classification accuracy and its associated precision and recall metrics are averaged over 10 balanced sets. Using the above experimental setting, the baseline accuracy for our classification task is 50% (i.e., random guess).

7.3 Results

To highlight the contribution of each feature group, we first report the classification accuracy using each feature group separately in Table 4. Table 4 also reports their respective precision and recall scores. As can be seen, accelerometer data is the most informative feature among all features types, with 75.8% accuracy. Features extracted using location logs are the second best feature with an accuracy of 68.5%. Next, we see that Wifi and Bluetooth logs are also discriminant with the number of unique WiFi access points and nearby Bluetooth devices observed during a user-night selected as the most important features (see Figure 5d and Figure 5c). Furthermore, features extracted using screen, application, and battery logs are relatively less good features.

Table 4 also reports the classification accuracies along with precision and recall of alcohol consumption using a combination of different feature groups. Using all the feature groups, we obtained an overall accuracy of 76.4%, which is a modest increase over the accuracy achieved using just accelerometer features (75.8%). Accelerometer features when combined with location features achieve similar accuracy compared to all the features combined. To conclude, we have investigated alcohol consumption status as a binary classification task. Using a similar methodology, operationally it might be feasible to infer the amount of alcohol intake, although the performance of such a model is an open-ended research question, but worth exploring as an extension of the present work.

8 ROLE OF GOING OUT AND GENDER ON DRINKING EVENTS

To contextualize and reflect on the results, in this section we discuss the role of two potential confounding variables – going out and gender – on the reported alcohol consumption.

8.1 Role of Going Out on Drinking Behavior

For many people, consuming alcohol is inherently a social activity [12], [52], which typically happens outside the home environment in pubs, bars, nightclubs, etc. In the previous sections (Sections 6 and 7), it has been shown that accelerometer, location, bluetooth, and wifi are the top feature groups predictive of alcohol consumption. Accelerometer features are associated with users’ physical activities (e.g., standing, sitting, etc.); location features inform the physical mobility of users; Bluetooth and WiFi features can inform the social context of users’ environment (e.g., alone or in a group). Given the discriminative nature of these feature groups to characterize alcohol consumption, it can be argued that the list of features extracted for the inference model are in fact capturing participants’ social and physical activity, and only indirectly informing drinking behavior. In this subsection, we examine the role of going out behavior (i.e., going away from home) together with alcohol consumption.

As part of the study protocol when the participants documented their drinks, we also asked them to report the place (e.g., home, bar, restaurant, etc.) where they were having their drink [27]. Using this self-reported information, we know the type of place where the participants were at the time of reporting drinks. A participant could be at multiple venues during a night. To handle this case, if participants had reported at least one place outside their home during a user-night, they were considered as gone out for the night; and if participants had reported staying at home through out the night, they were considered as staying at home.

Out of 1,011 user-nights, participants reported place information for 629 (62.2%) user-nights. In Table 5, we show the total number of user-nights of drinking contrasted with going out behavior. Out of 629 user-nights, participants had gone out for 441 (70%) user-nights, while they had stayed at home for 30% of nights. If a participant had gone out, it was highly likely that they had consumed alcohol (82.8% of user-nights). On the other hand, while staying at home, for 44.7% (84) of user-nights, participants consumed alcohol at home, supporting the increasing trend of home drinking [53]. In [53], the authors examined the potential reasons of why people drink at home. Using focus groups spanning different age groups, the authors reported that cost, convenience, safety, and social occasions are the key reasons of drinking at home. In the same study, for young people aged 16–25 years old (the age group of our study participants), social occasions was found to be the principal reason for home drinking. Given that the likelihood of consuming alcohol at home (44.7%) and not

consuming alcohol away from home (17.2%) is not negligible, it is fair to say that the extracted features might be capturing more nuanced notions of drinking behavior, and not just informing the social and physical context of users.

As stated earlier that the focus of this paper is on using the physical location as derived from location traces. Future work will investigate the role of semantic place location on alcohol consumption e.g., knowing that if a participant was in a bar, the likelihood of consuming alcohol would be higher, when compared to participant being in a shopping mall.

8.2 Role of Gender on Drinking Behavior

As stated in Section 2.1, it is known that gender plays a determinant role in alcohol use. In [15], using 16 population surveys across 10 countries, the authors found that males drank more frequently and consumed more alcohol when compared to females. Similar findings were reported in another large-scale study conducted amongst an adolescent population in Europe [54]. To inspect the potential gender differences on alcohol consumption, in this section we present results of reported alcohol consumption differentiated by gender as observed in our study.

From a total of 160 users studied in this paper (Section 4.2), gender information is available for 159 participants. In our study we observed a fairly balanced gender ratio with 85 males (53.45%) and 74 females (46.55%). Males contributed towards a total of 502 user-nights, and females provided slightly higher number of 507 user-nights. In Table 6, we report the total number of user-nights by each gender and their reported alcohol (and non-alcohol) consumption. Of all the user-nights reported by the female cohort, 62.5% (resp. 37.5%) of them were reported with alcohol (resp. non-alcohol) consumption. On the other hand, males reported 72% of nights with alcohol consumption, which is higher when compared to the female population. These results show that gender alone cannot explain the patterns observed in Sections 6 and 7.

9 DISCUSSION AND CONCLUSIONS

In this paper, we have examined the feasibility of collecting and using smartphone sensor data to automatically characterize basic drinking behavior of young adults (drinking/not drinking) in an urban, in-the-wild nightlife setting. To reflect on these findings, we get back to the research questions posed at the beginning of the paper.

RQ1: Is it possible to infer whether a person has consumed alcohol from smartphone sensor and logs data in an uncontrolled, real-life setting? Table 3 and 4 provide evidence to suggest the feasibility of characterizing and classifying drinking behavior (as a binary state) using smartphone data. From an alcohol epidemiology point of view, we believe our study produced promising results. In this domain, there is a recent push to study alcohol consumption patterns using mobile phones to increase research validity and reliability and to widen the understanding of contextual factors [8]. Smartphones, which can collect a wide range of sensor data (activity, location, proximity), event logs (including all phone applications), and media (photos, audio, video), offer a promising alternative to improve the way in which drinking-related phenomena can be studied. Furthermore, the recurrent use of retrospective surveys, which induces recall biases, can be potentially reduced using smartphones.

RQ2: Which features are more predictive of alcohol consumption? Table 3 and Figure 5 highlight the discriminative power of physical

		Alcohol Consumption	
		Negative	Positive
Going Out	Negative	104	84
	Positive	76	365

TABLE 5: Alcohol Consumption vs. Going Out Behavior.

		Alcohol Consumption	
		Negative	Positive
Gender	Female	190	317
	Male	141	361

TABLE 6: Alcohol Consumption vs. Gender

activity (accelerometer), physical mobility (location) and social context (WiFi and Bluetooth) features as partly predictive of alcohol consumption. Furthermore, we found that features derived from accelerometer data are the most informative data type, achieving an accuracy of over 75%. The best combination of features produced an accuracy of 76.6%.

With the increasing availability and acceptability of wrist-worn wearable devices (like Jawbone, Fitbit, Apple Watch), which embed accelerometers inside them, our findings points to potentially interesting ways to track drinking behavior and promote self-reflection based on behavioral data. As potential applications, near real-time information on consumption could potentially influence alcohol intake, assisting users in their decision-making to drink further or not. It has already been shown that self-monitor and real-time feedback can help reduce alcohol consumption and heavy drinking [26]. For real-time detection, the classification task would have to infer the likelihood of alcohol consumption upto the present time in a night, which would involve some form of dynamic time-based models. We plan to pursue this direction in future work. Finally, consumption information could be integrated into potential future systems of epidemiological research at the population level.

The data collected as part of the study provide numerous opportunities to explore different angles of drinking behavior. In this paper, we studied one aspect of drinking behavior. Alcohol use is context dependent, so from a scientific point of view, understanding the temporal aspects of consumption and how it affects intake is an area of future exploration. We are also interested in exploring how the social context and mobility cues affect drinking behavior. Does increased mobility or large social groups increases the likelihood of consumption? We believe our dataset would facilitate to empirically test these hypotheses, some of which are currently analysed using questionnaires [12], [13]. We plan to explore these angles of drinking activity as part of future work.

10 ACKNOWLEDGEMENTS

This work was funded by the Swiss National Science Foundation Youth@Night project. We thank Flavio Tarsetti, Olivier Bornet and Hughes Salami for their work towards developing the mobile application; and Trung Phan Thanh for initial pre-processing of the data. We thank the research assistants (Anna Katz, Pauline Ndong, Lucie Chambeyron, Vera Bariswyl) and following students for their contribution to the recruitment of participants for the study: Ball Christian, Bazzi Laura, Becker Lilian, Bieri Cornelia, Bovet Stefanie, Cachelin Elodie, Christen Silvan, Eberle Stefanie, Hotzendorfer Peter, Lepdor Florence, Marfurt Josephine,

Meier Petra, Mhidi Mehdi, Nicoli Fanya, Sigam Bansa, Stojanovic Tamara, Tobler Hannes, Zollinger Larissa and Zust Lea. We also thank all the Y@N study participants.

REFERENCES

- [1] F. Measham and K. Brain, "Binge drinking, british alcohol policy and the new culture of intoxication," *Crime, media, culture*, 2005.
- [2] A. A. Ali, S. M. Hossain, K. Hovsepian, M. M. Rahman, K. Plarre, and S. Kumar, "mpuff: automated detection of cigarette smoking puffs from respiration measurements," in *Proceedings of the 11th international conference on Information Processing in Sensor Networks*. ACM, 2012.
- [3] G. Gmel, J. Rehm, and E. Kuntsche, "Binge drinking in europe: definitions, epidemiology, and consequences." *Sucht: Zeitschrift für Wissenschaft und Praxis*, 2003.
- [4] J. Demant and S. Landolt, "Youth drinking in public places: The production of drinking spaces in and outside nightlife areas," *Urban Studies*, 2013.
- [5] P. Chatterton and R. Hollands, *Urban nightscapes: Youth cultures, pleasure spaces and corporate power*. Psychology Press, 2003.
- [6] O. Ekholm, "Influence of the recall period on self-reported alcohol intake," *European Journal of Clinical Nutrition*, 2004.
- [7] E. Kuntsche and F. Labhart, "Investigating the drinking patterns of young people over the course of the evening at weekends," *Drug and alcohol dependence*, 2012.
- [8] E. Kuntsche and F. Labhart, "Icat: Development of an internet-based data collection method for ecological momentary assessment using personal cell phones," *European Journal of Psychological Assessment*, 2013.
- [9] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila, "Towards rich mobile phone datasets: Lausanne data collection campaign," *Proc. ICPS, Berlin*, 2010.
- [10] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 2013.
- [11] J. Petratis, B. R. Flay, and T. Q. Miller, "Reviewing theories of adolescent substance use: organizing pieces in the puzzle." *Psychological bulletin*, 1995.
- [12] H. Kuendig and E. Kuntsche, "Solitary versus social drinking: an experimental study on effects of social exposures on in situ alcohol consumption," *Alcoholism: Clinical and Experimental Research*, 2012.
- [13] F. Labhart, S. Wells, K. Graham, and E. Kuntsche, "Do individual and situational factors explain the link between predrinking and heavier alcohol consumption? an event-level study of types of beverage consumed and social context," *Alcohol and Alcoholism*, 2014.
- [14] J. Thrul and E. Kuntsche, "The impact of friends on young adults' drinking over the course of the evening - an event-level analysis," *Addiction*, 2015.
- [15] R. W. Wilsnack, N. D. Vogeltanz, S. C. Wilsnack, and T. R. Harris, "Gender differences in alcohol consumption and adverse drinking consequences: cross-cultural patterns," *Addiction*, 2000.
- [16] J. Ostergaard, "Mind the gender gap!" *Nordic Studies on Alcohol and Drugs*, 2007.
- [17] W. F. Wieczorek, B. A. Miller, and T. H. Nochajski, "Multiple and single location drinking among dwi offenders referred for alcoholism evaluation," *The American journal of drug and alcohol abuse*, 1992.
- [18] P. M. Dietze, M. Livingston, S. Callinan, and R. Room, "The big night out: What happens on the most recent heavy drinking occasion among young victorian risky drinkers?" *Drug and alcohol review*, 2014.
- [19] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *Proceedings of the National Academy of Sciences*, 2009.
- [20] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury, "Stresssense: Detecting stress in unconstrained acoustic environments using smartphones," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012.
- [21] K. Hovsepian, M. al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "cstress: towards a gold standard for continuous stress assessment in the mobile environment," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015.
- [22] Q. Tang, D. J. Vidrine, E. Crowder, and S. S. Intille, "Automated detection of puffing and smoking with wrist accelerometers," in *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, 2014.
- [23] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis, "Risque: recognizing smoking gestures with inertial sensors on a wrist-band," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM, 2014.
- [24] H.-L. C. Kao, B.-J. Ho, A. C. Lin, and H.-H. Chu, "Phone-based gait analysis to detect alcohol usage," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012.
- [25] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan, "Mobile phone based drunk driving detection," in *4th International Conference on Pervasive Computing Technologies for Healthcare*, 2010, pp. 1–8.
- [26] K.-C. Wang, M.-C. Huang, Y.-H. Hsieh, S.-Y. Lau, C.-H. Yen, H.-L. C. Kao, C.-W. You, H.-H. Chu, and Y.-C. Chen, "Soberdiary: a phone-based support system for assisting recovery from alcohol dependence," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014.
- [27] D. Santani, J.-I. Biel, F. Labhart, J. Truong, S. Landolt, E. Kuntsche, and D. Gatica-Perez, "The night is young: Urban crowdsourcing of nightlife patterns," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '16. ACM, 2016.
- [28] W. H. Organization et al., *Global status report on alcohol and health-2014*. World Health Organization, 2014.
- [29] E. Oliver, "The challenges in large-scale smartphone user studies," in *Proceedings of the 2nd ACM International Workshop on Hot Topics in Planet-scale Measurement*. ACM, 2010.
- [30] D. Ferreira, V. Kostakos, and A. K. Dey, "Lessons learned from large-scale user studies: Using android market as a source of data," *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 2012.
- [31] R. Montoliu, J. Blom, and D. Gatica-Perez, "Discovering places of interest in everyday life from smartphone data," *Multimedia tools and applications*, 2013.
- [32] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen, "Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2015.
- [33] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explorations Newsletter*, 2011.
- [34] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *AAAI*, 2005.
- [35] Z. Yan, J. Yang, and E. M. Tapia, "Smartphone bluetooth based social sensing," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. ACM, 2013.
- [36] T. M. T. Do and D. Gatica-Perez, "The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data," *Mobile Computing, IEEE Transactions on*, 2014.
- [37] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1166–1172, 2010.
- [38] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive computing*. Springer, 2004.
- [39] T.-M.-T. Do and D. Gatica-Perez, "By their apps you shall understand them: mining large-scale patterns of mobile phone usage," in *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 2010.
- [40] T. M. T. Do, J. Blom, and D. Gatica-Perez, "Smartphone usage in the wild: a large-scale analysis of applications and context," in *Proceedings of the 13th international conference on multimodal interfaces*. ACM, 2011.
- [41] M. Bohmer, B. Hecht, J. Schoning, A. Kruger, and G. Bauer, "Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage," in *Proceedings of the 13th international conference on Human computer interaction with mobile devices and services*. ACM, 2011.
- [42] "Android api," <http://developer.android.com/reference/android/os/BatteryManager.html>.
- [43] N. Eagle and A. Pentland, "Social serendipity: Mobilizing social software," *Pervasive Computing, IEEE*, 2005.
- [44] M. Ye, D. Shou, W.-C. Lee, P. Yin, and K. Janowicz, "On the semantic annotation of places in location-based social networks," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011.
- [45] T. M. T. Do and D. Gatica-Perez, "The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data," *IEEE Transactions on Mobile Computing*, 2014.

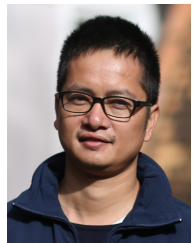
- [46] R. J. Agrawal and J. G. Shanahan, "Location disambiguation in local searches using gradient boosted decision trees," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. ACM, 2010.
- [47] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, 2010.
- [48] J. Friedman, T. Hastie, and R. Tibshirani, "glmnet: Lasso and elastic-net regularized generalized linear models," *R package version*, vol. 1, no. 4, 2009.
- [49] K. Fitzpatrick, M. Brewer, and S. Turner, "Another look at pedestrian walking speed," *Transportation Research Record: Journal of the Transportation Research Board*, 2006.
- [50] L. Breiman, "Random forests," *Machine learning*, 2001.
- [51] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013, vol. 26.
- [52] E. Kuntsche, R. Knibbe, G. Gmel, and R. Engels, "Why do young people drink? a review of drinking motives," *Clinical psychology review*, 2005.
- [53] J. Foster, D. Read, S. Karunanithi, and V. Woodward, "Why do people drink at home?" *Journal of public health*, 2010.
- [54] E. Kuntsche, M. Wicki, B. Windlin, C. Roberts, S. N. Gabhainn, W. van der Sluijs, K. Aasvee, M. G. de Matos, Z. Dankulinova, A. Hublet *et al.*, "Drinking motives mediate cultural differences but not gender differences in adolescent alcohol use," *Journal of Adolescent Health*, 2015.



Emmanuel Kuntsche is the Director of the Centre for Alcohol Policy Research and Professor of Public Health at La Trobe University, Melbourne, Australia; Associate Professor of Developmental Psychopathology at Radboud University Nijmegen, the Netherlands.



Darshan Santani received his PhD in Electrical Engineering from EPFL and the Social Computing Group at Idiap Research Institute in 2016. His research interests lie at the intersection of mobile, social and urban computing.



Trinh-Minh-Tri Do is a Data Scientist at Trusting Social. He received his PhD degree in Computer Science from Pierre and Marie Curie University, France in 2010. His research interests include machine learning, optimization, and ubiquitous computing.



Daniel Gatica-Perez is the Head of the Social Computing Group at Idiap and Professor at EPFL in Switzerland. His research interests include social computing, social media, ubiquitous computing, and crowdsourcing. He is a member of the IEEE.



Florian Labhart is a research associate at the Addiction Switzerland and Idiap Research Institute in Switzerland. His research mainly focuses on social and environmental determinants of risky alcohol use among young people.



Sara Landolt is the head of the research group "youth, migration, urban nightlife" at the Department of Geography, University of Zurich. Her research interests include geographies of youth and migration, urban studies and feminist methodology.