Understanding Heavy Drinking at Night through Smartphone Sensing and Active Human Engagement

Thanh-Trung Phan*
Idiap Research Institute and EPFL
Switzerland
tphan@idiap.ch

Florian Labhart
Idiap Research Institute and La Trobe University
Switzerland and Australia
flabhart@idiap.ch

Skanda Muralidhar
Idiap Research Institute and EPFL
Switzerland
smuralidhar@idiap.ch

Daniel Gatica-Perez
Idiap Research Institute and EPFL
Switzerland
gatica@idiap.ch

ABSTRACT

Heavy alcohol consumption can lead to many severe consequences. In this paper, we study the phenomenon of heavy drinking at night (4+ drinks for women or 5+ for men on a single evening), using a smartphone sensing dataset depicting about nightlife and drinking behaviors for 240 young adult participants. Our work has three contributions. First, we segment nights into moving and static episodes as anchors to aggregate mobile sensing features. Second, we show that young adults tend to be more mobile, have more activities, and attend more crowded areas outside home on heavy drinking nights compared to other nights. Third, we develop a machine learning framework to classify a given weekend night as involving heavy or non-heavy drinking, comparing automatically captured sensor features versus manually contributed contextual cues and images provided over the course of the night. Results show that a fully automatic approach with phone sensors results in an accuracy of 71%. In contrast, manual input of context of drinking events results in an accuracy of 70%; and visual features of manually contributed images produce an accuracy of 72%. This suggests that automatic sensing is a competitive approach.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; *Ubiquitous and mobile computing design and evaluation methods.*

KEYWORDS

Alcohol consumption, heavy drinking, mobile crowdsensing, deep learning

---

*This is the corresponding author.

1 INTRODUCTION

The consumption of alcohol is an important rite of passage for young adults in western societies [26]. In late adolescence and early adulthood, heavy drinking is reported to be more common than in any other period of life [24] and the amount of alcohol consumed is found to peak on Friday and Saturday nights [33, 51]. Alcohol is consumed mainly due to its disinhibiting and enhancing effects [31]. The consumption of several drinks per night is often the norm, considering that some studies show that about 70% of all alcohol consumed by young adults per week occurs within sessions of six or more drinks in a row [22]. Such a high intake of alcohol can lead to many adverse consequences not only for the person drinking (e.g., unprotected sex, injury, accidents or blackouts [37]) but also for whole society (e.g. violence, drunk driving). Hence, determining indicators of heavy drinking behavior is critical for healthcare professionals. In this work, we investigate patterns of heavy drinking behavior and compare the utility of automatically captured smartphone sensor data and manually contributed information including context data and photos.

A large body of epidemiological research shows that different characteristics of the drinking context are associated with increased drinking at the event level [58, 59]. These characteristics include motives or intentions for drinking, the time of day when drinking occurs, the type and number of people involved, the venue (e.g. pub or at home), the situation (characteristics of the venue) [59], and the interactions between these characteristics [46]. However, while drinking is a dynamic behavior, i.e., drinks can be consumed in different locations throughout the extended drinking occasion, most studies have focused on single drinking settings and overlooked the influence of sequences of drinking episodes (e.g. pub crawl, or drinking before going out) [58].

The ubiquity of smartphones has opened new opportunities for investigating alcohol consumption patterns using information from

Thanh-Trung Phan, Florian Labhart, Skanda Muralidhar,
Daniel Gatica-Perez

varied sensors. For example, previous work has shown the feasibility of using sensor data from a variety of wearable devices (e.g. breathalyzers, wristbands, smartwatches) to detect heavy drinking [73]. Similarly, crowdsensing methods, which collect and combine smartphone sensing with human-generated inputs in real-life situations, have shown the ability to capture various drinking patterns of interest [4, 6, 55]. Bae et. al [6] investigated the use of sensor data to infer heavy drinking among 30 participants and 71 short 30-minutes heavy drinking episodes using only passive smartphone sensors, and reported a classification accuracy of 96.6% (against a majority class baseline of 90.7%).

In this paper, we investigate the comparative value of automatic smartphone sensors and manually contributed information including the context of drinking events and pictures taken in-situ, in classifying a given night as being one in which heavy drinking occurred or not. Towards this, we use a previously collected crowdsensing dataset consisting of 241 participants and 847 user-nights. The classification at night-level rather than at event-level (i.e classifying each drinking event), is motivated by the relevance of providing insights to users over the course of an entire drinking occasion. We believe this to be an important insight for users of such a framework as an alternative to real-time feedback. Alcohol consumption research has shown that in-the-event preventive messaging (e.g., by sending SMS to the individuals who reached a certain number of drinks over the course of a drinking night) present multiple challenges [69]. Specifically, we address the following research questions:

RQ1: Using a mobile crowdsensing dataset generated on weekend nights, how do mobile sensor data, manual context and image content differ between heavy and non-heavy drinking nights?

RQ2: If such differences exist, how can the different data sources be used to classify the two types of drinking nights?

The contributions of the paper are the following:

(1) We investigate multiple aspects of real-life nightlife behaviors, collected by means of mobile sensors, in-situ questionnaires, and pictures, to characterize non-heavy drinking and heavy drinking nights. Using a large crowdsourced dataset collected from 200+ young people, aged 16 - 25 years old, from our previous work [54], nights were labeled based on the total number of "standard drinks" consumed over the course of the night. Utilizing location data and timestamps, we segment nights into moving and episodes as anchors to aggregate mobile sensing features.

(2) We showed that some derived features of time, place, multiple mobile sensors, and image features are statistically significantly different for heavy and non-heavy drinking nights. Our statistical analysis provides several interesting insights; (a) Participants tend to stay for shorter duration at a given location and change locations more frequently during heavy drinking nights; (b) they are more likely to drink heavily when attending crowded locations; (c) the time intervals between the first and the last drink on heavy nights are longer than on non-heavy nights; and (d) the number of drinks reported was greater outside private places for heavy drinking nights, while a greater number of drinks was reported on private places for non-heavy drinking nights.

(3) We use a machine learning framework to classify heavy drinking nights and compare the classification power of automatically captured sensor data vs. manually contributed context information and image content. Our binary classification task shows that a fully automatic approach with phone sensors results in an accuracy of 71%. The same task using manual input of context of drinking events, and visual features of manually contributed images results in an accuracy of 70% and 72% respectively. These results can be seen as an initial step towards developing an automatic system for self-monitoring of alcohol consumption.

2 RELATED WORK

2.1 Identifying Heavy Drinking Occasions

In alcohol research and epidemiology, the term 'heavy drinking' (also commonly called 'binge drinking' or 'risky drinking') characterizes an amount of alcohol that carries a significantly higher risk of experiencing adverse alcohol-related consequences, such as hangover, injury, or blackouts [32, 37, 64]. The most common definition of heavy drinking is the consumption of 4 or more standard drinks (containing 10-12 grams of pure alcohol per drink) for women and 5 or more drinks for men on a single drinking occasion [15]. For decades, research on heavy drinking has been conducted mainly using retrospective questionnaires with the aim of identifying faction that might result in drinking heavily [10, 68]. With the development of ecological momentary assessment (EMA), attention has been given to occasion-level factors associated with exceeding heavy drinking thresholds [34, 36, 37]. For example, heavier drinking was found to be associated with drinking in multiple locations, the size of the drinking group, as well as the duration of the drinking occasion [40, 62]. Such evidence are important to develop in-the-event prevention interventions [28] that are able to identify heavy drinking situations independently of the person's own consumption level. However, existing evidence is mostly based on participants' active self-reports, which are subject to response bias, and rarely considers cues from the extended social and physical context. This highlights the need to develop methods to capture and identify patterns of heavy drinking occasions that are less dependent on participants' reports.

This was for example done by investigating mobility (i.e., types of locations, geographical moves) and activity (i.e., body activity) as patterns of heavy drinking occasions. By combining participants' self reports of alcohol use with Global Positioning System (GPS) data from their smartphone and reading from a transdermal alcohol sensor (i.e., an ankle bracelet able to detect ethanol concentration in vapors formed above the skin), Clapp and al. [13] showed the ability of sensors to capture characteristics of real-life drinking occasions, including across multiple locations (e.g. pub crawls). More recently, other research groups have used a combination of smartphone sensors to discriminate between non-drinking, drinking and heavy drinking single episodes [5]. Regarding cues from the extended social and physical context, our previous work [50] demonstrated the possibility to distinguish between heavy and casual drinking by using visual and textual content on Instagram pictures and posts with an accuracy of 82% and 75%, respectively.

2.2 Sensing Behaviors With Smartphones

Most modern mobile phones are equipped with physical sensors including GPS, accelerometer, WiFi and Bluetooth. GPS has been used for individual mobility recognition [49, 60], points of interest detection [11, 48], travelling behavior [48, 66], as well as links between behavior patterns at certain location and particular demographics [29]. Accelerometer has been used for physical activity recognition [3, 7, 45, 52], including classifying human movement in real-time [27], detecting daily movements [43], static/dynamic activities [44], and monitoring long term human movements at home [44]. Bluetooth and WiFi sensors have been used to extract social and urban context, e.g., estimating human density around the user [18] by considering number of Bluetooth device as a proxy for social context [19], or to infer human movement by using WiFi as a proxy. Finally, other sensors like proximity, battery, and screen time have been exploited to understand users' behavior in relationship with sleep [12, 47].

In the context of alcohol consumption, Santani et al. [54] designed a smartphone application to collect sensor data from youth on weekend nights. Using sensor features, a random forest is able to automatically classify alcohol and non-alcohol drinking nights with an accuracy of 76.6% [55]. More recently, Bae et al. [6] defined a classification task to identify non-drinking, drinking, and heavy drinking single episodes, using various smartphone sensors among a sample of 30 young adults and reported an accuracy of 96.6%. The majority class consisting of non-alcoholic drinking constituted 90.7% of the data. This study showed the potential of sensors to identify heavy drinking patterns.

Our work proposes a framework for detecting heavy drinking occasions presenting a comparative study of three different data sources, namely (1) passive sensor data (location, accelerometer, WiFi, Bluetooth); and (2) actively contributed contextual cues (time and place of drinks) and images.

3 DATASET AND TASK DEFINITION

The dataset for the present work comes from our previous work [54], as part of a project aimed at understanding young peoples' nightlife behavior in Switzerland using a custom-developed smartphone application [39, 54]. This section presents an overview of the study design, the data collection procedure, and the definition of the target task.

3.1 Study Design

With other collaborators, we designed a study of nightlife behavior of young adults on weekend nights [54] . Towards this, we developed an Android-based smartphone application that required participants to actively report details of their alcohol use and the corresponding context using event-level questionnaires and pictures from the early evening until the end of the night; and that passively collected sensor data (accelerometer, WiFi, etc. as well as app logs) from 8PM to 4AM the next morning. Participation criteria were being aged between 16 and 25, owning an Android phone, having been out in the city at least once in the last month, and having consumed alcohol at least once in the last month (see [38] for a detailed description of the recruitment procedure). The study was approved by the Ethical Review Boards of the Swiss cantons of

Vaud and Zurich . It should be noted that the legal drinking age for beer and wine is 16 in Switzerland. After signing the online consent form, 241 participants installed the app on their own smartphone and were expected to use it on 10 Friday and Saturday nights over 7 consecutive weekends [39].

3.2 Data Collection

The app contained different types of questionnaires to record the drinks consumed (alcoholic and non-alcoholic) and the characteristics of the locations attended. At 8PM, participants were requested to report the number of drinks consumed since 5PM. Then, from 8PM to the end of the night, whenever participants had a new drink, they had to take a picture of it, and to describe its content and the characteristics of the surrounding place in a short survey. Participants were expected to take a picture that clearly captured a container with liquid (with or without alcohol). In addition, in case they had forgotten to document a drink at the time of its consumption, participants could report it in a separate questionnaire (see [39] for a full description of the sequence and content of the questionnaires). Regarding mobile sensors, the sensor logger of the application, collected a variety of sensor and log data, including GPS coordinates, accelerometer, activity, using a background running process without any user interaction (see [54] for a full description of the sensors collection process). However, location sensors not being activated by few participants made information inconsistent in some nights. For this work, we selected the nights containing consistent information on drinking time, drink location, sensor data, and drink pictures, resulting in a final sample of 241 participants, and 847 user-nights as opposed to episodes. The participant pool is balanced in terms of sex (53% men), is young (mean=19.4 years old, SD=2.5), and is distributed across main activities (29.9% high-school, 38.9% university, and 31.1% employed people: apprenticeship, part-time or full-time job). Regarding participants' drinking history, they drink an average of 6.8 occasions per month (SD = 5.4), and an average of 3.8 drinks per occasion (SD = 2.1). Regarding nightlife habits, participants go out in the city an average of 5.9 nights per month (SD = 2.9).

3.3 Heavy Drinking Definition

The most common definition of heavy drinking in alcohol research, according to [15, 32], is 4+ drinks for women and 5+ drinks for men in a single drinking occasion. Note that 4+/5+ drinks correspond to a blood alcohol concentration of about 0.08 percent [32]. This percentage corresponds to the legal limit in the US to drive a vehicle [2]. We use this definition and define a heavy drinking night classification task for those nights in which participants drink 4+ or 5+ drinks for women and men, respectively, in a single night. Note that other authors [6] have defined a different classification task for non-drinking, drinking, and heavy drinking episodes. In other words, our task involves full night periods whereas the other one is about single episodes.

4 FEATURES AND LABELS

Table 1 outlines the various features that are extracted from the mobile sensor logs, questionnaires, and pictures recorded by the
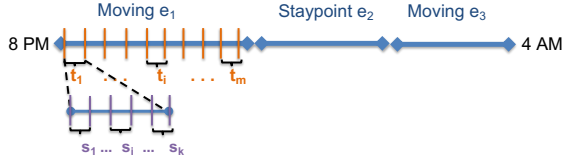
Thanh-Trung Phan, Florian Labhart, Skanda Muralidhar,
Daniel Gatica-Perez

Figure 1: Aggregation process of single data points into time blocks and into episodes (*staypoint* or *moving*). e = episode, t_i = time block within an episode during which data are collected with m = total number of blocks, s_i = single data point with k = total number of data points per block.

app. In this section, we describe the features and the methodology used to extract them.

4.1 Mobile Sensor Features

4.1.1 Location & Mobility: GPS was used to derive two different types of features: location, which includes a number of features such as location accuracy, speed, and GPS network percentage; and mobility, which includes distinct segments of user-night mobility patterns (*staypoints* vs. *moving episodes*) identified based on the location data and corresponding timestamps.

We compute mobility of participants using an algorithm proposed by Montoliu et al. [48] to distinguish between periods when a user stays at a specific location and periods of transition between locations. Figure 1 provides a hypothetical example of a young person's weekend night mobility pattern between 8PM and 4AM: the person goes out at 8PM to a pub for a couple of hours, and comes back home. Staying at home and staying at the pub are defined as "*staypoints*", which correspond to a spatio-temporal location where the participant has spent some time. The other segments, namely moving from home to the pub and coming back home, are defined as "*moving episodes*". This distinction is important as people's drinking behavior have been shown to be connected to whether they stay at a certain place for some time or are moving [8, 48].

We used GPS data to identify *staypoints* using a duration range limit between 5 minutes and 4 hours (Tmin = 5 minutes, and Tmax = 4 hours) and a maximum radius of 100 meters as recommended by Do et al. [20]. Time sequences between two *staypoints* were considered as *moving episodes*. In total, we obtained 2792 *staypoints* (1423 on heavy drinking nights on 1369 on non-heavy drinking nights) and 2762 *moving episodes* (1417 on heavy drinking nights and 1345 on non-heavy drinking nights) in the 847 user-nights. For each night, the other sensors' data (described below) were then aggregated within the corresponding episode (Figure 1).

4.1.2 Accelerometer. To recognize patterns of activities in a given time window, previous studies have calculated various statistics using the raw readings for the X, Y and Z-axes [7, 9, 45, 52]. Others have aggregated the values of three axes into one value, e.g., signal-magnitude area (SMA) [27, 43, 44, 65], or average resulting acceleration [3] (average of the square roots of the sum of the values of each axis squared) over the sampling duration. In the dataset, accelerometer data was collected for 10 seconds every 60 seconds, with a sampling frequency of 50Hz. We utilize the raw readings

of individual axes, the angles with g vectors, and the aggregated values of three axes. As for all other sensors, accelerometer features were then aggregated within the corresponding episode (namely *moving* or *staypoint*).

4.1.3 Bluetooth and WiFi. Bluetooth and WiFi density can be used as a proxy to capture human mobility [56], social context [18, 21, 30, 71], and person-person proximity [57]. For example, if there are a large number of distinct WiFi hotspots and Bluetooth devices at some point of the night, this likely indicates that the participants attended a busy or crowded area. In this work, we compute various features including signal strength and the number of devices visible via Bluetooth and WiFi.

4.1.4 Applications. Applications use might differ depending on the location or the activity [70]. For example, Likamwa et al. [42] developed a mood sensor using smartphone application usage, which could further be linked to drinking behaviors [53, 63]. We categorized each app name in the logs using Google Play Store 30 categories and based on this we computed app category counts within mobility episode.

4.1.5 Event-Triggered Sensors. Event-triggered sensor data includes information such as proximity to screen, battery status, and screen usage. These data have shown to be predictive of quality of sleep [47]. Screen proximity relates to particular behaviors like taking phone calls, putting the phone into one's pocket, or laying the phone upside-down on a table. Battery status records any change to battery including charging, discharging, full, not charging, the actual battery level, and if the phone is plugged in or not. Screen usage detects the status of the screen, e.g. ON or OFF. Counts are aggregated at the mobility episode level.

4.2 Context Features

From 8PM to 4AM, the participants were required to complete a series of small questionnaires describing the social, spatial, and temporal context for every new drink (alcohol or not) as well as a photo [39]. In this subsection, we describe the temporal and spatial context of drinking occasions.

4.2.1 Time Context. Based on the timestamp of the submitted questionnaires, we computed various statistics (min/ max/ mean/ median/ standard deviation) on the time interval between consecutive drinks (if possible), the time of the first drink, and the time difference between the first and the last drink. However, not all drinks had actual time because in some cases the participants forgot to report their drinks in-situ. In such situations, we set the features to 0.

4.2.2 Semantic Location Context. We obtain information of the place context of drinking activities in two different ways. First, while using the application, participants indicated in real time the type of semantic location they were drinking at from a list of 12 place categories (coffee, bar, restaurant, personal, travel, plaza, school, event, club, park or lake, on board, and other). Second, we designed a manual annotation task to obtain qualitative information on the location type. This task was done by five independent annotators, who viewed the entire corpus of photo (see next section) and answered the question "In what kind of place was the photo

Table 1: Features extracted from mobile sensors, survey data and photos.

Feature	Description	Type	Group
Location Count	Staypoint and moving count	Numeric (2)	Sensor (S)
Location Attributes	{Min., Max., Med., Avg., Std.} of Avg of duration of staypoint/moving, speed, acuuracy, in case of staypoint and moving	Numeric (30)	Sensor (S)
Location Signal	3 signal strength (GPS,network,unknown) in case of staypoint and moving	Numeric (6)	Sensor (S)
Accelerometer Raw	{Min., Max., Med., Avg., Std.} of Avg of xAxis, yAxis, zAxis of the accelerometer in case of moving, staypoints	Numeric (30)	Sensor (S)
Accelerometer Angle	{Min., Max., Med., Avg., Std.} of Angle of xAxis, yAxis, zAxis with g vector in case of moving, staypoints	Numeric (30)	Sensor (S)
Accelerometer Dynamic	{Min., Max., Med., Avg., Std.} of SMA, Dm, m, mNew in case of moving, staypoints	Numeric (40)	Sensor (S)
Bluetooth Count	The number of Bluetooth IDs surrounding devices, records, Bluetooth scan count, emtyScanCount in case of moving and staypoint	Numeric (8)	Sensor (S)
Bluetooth Strength	{Min., Max., Med., Avg., Std.} of The Bluetooth strength signal of surrounding devices in case of moving and staypoint	Numeric (10)	Sensor (S)
Wifi Count	wifiRecord , wifiIdSet in case of moving, staypoints	Numeric (4)	Sensor (S)
Wifi Attributes	{Min., Max., Med., Avg., Std.} of level, frequency, of wifi hostpot in case of moving, staypoints	Numeric (20)	Sensor (S)
Application Count	AppCount, AppRecord in case of moving, staypoints, general	Numeric (4)	Sensor (S)
Application Category	Normalized 33-bin histogram of 33 application categories in case of moving, staypoints	Numeric (66)	Sensor (S)
Proximity Count	proximityRecord in case of moving, staypoints	Numeric (2)	Sensor (S)
Proximity Distance	{Min., Max., Med., Avg., Std.} of Distance from phone to objects in case of moving, staypoints	Numeric (10)	Sensor (S)
Battery Status	5 Battery status in case of moving and staypoint, general	Numeric (15)	Sensor (S)
Battery Level	{Min., Max., Med., Avg., Std.} of BatLevel in case of moving and staypoint	Numeric (10)	Sensor (S)
Battery Count	Count of battery records and plugged times in case of moving and staypoint	Numeric (4)	Sensor (S)
Screen	screenRecord, percentage of time screen On in case of moving, staypoints	Numeric (2)	Sensor (S)
Drink Duration Time	Time (e.g. in minutes) between first and last drink in a user night.	Numeric (1)	Time (T)
Drink Start Time	Time of day in which the first drink was drunk.	Numeric (1)	Time (T)
Drink Gap Time	{Min., Max., Med., Avg., Std.} of time between consecutive drinks. If people only had one drink, {Min., Max., Med., Avg., Std.} should be 0.	Numeric (5)	Time (T)
Drink Time Histogram	Normalized 4-bin histogram (8-9pm,10-11pm, 0-1am, 2-3am) of time of night when drinks were taken.	Numeric (4)	Time (T)
Drink Day	Values of indicating drinking happens in Friday and Saturday night	Numeric (2)	Time (T)
Drink Semantic Place	Normalized 11-bin histogram of 11 locations (e.g. home, bar, etc) where a person visited in a user-night.	Numeric (11)	Place (P)
Image Objects	1000-object features on both heavy and non-heavy nights	Numeric (210)	Image (I)
Image Scenes	365-scene features on both heavy and non-heavy nights	Numeric (60)	Image (I)

taken?" using the same list of 12 place categories. This is used to create a histogram of the visited places throughout the night.

4.3 Visual Features

To understand the content of the 1733 drink pictures taken by participants, we use pre-trained deep neural network models. Specifically, we use two different models for object parsing and scene parsing. For object parsing, we use an Inception-v3 model [61] pre-trained on the ImageNET image corpus. This model outputs a 1000-dimensional vector consisting of class probabilities. To obtain the average probability for all objects recognized across the night,

we aggregate the output vectors at the night-level over all pictures taken by a user.

To obtain a scene-level description for each picture, we use a Resnet18 [74] model pre-trained on the Place-365 database. The output of this model is a 365-dimensional vector of class probabilities that the picture belongs to one of the 365 places. The place classes qualifies the semantic categories as well as functionality (e.g., beer hall for drinking beer). To represent the scene of all the pictures at the night level, we aggregate the output vectors of all pictures for each night by computing the average for each class.

Thanh-Trung Phan, Florian Labhart, Skanda Muralidhar,
Daniel Gatica-Perez

4.4 Labels

Two measures were used to quantify the quantity of alcohol consumed per night. First, for each drink documented during the night, the app proposed predefined lists of beverages (e.g. "beer", "wine", "spirits" and "premix/cocktails"), drink sizes (e.g. for beer: "small (25 cl)", "medium (33 cl)", or "large (50 cl)") and alcohol content (e.g. for beer: "alcohol–free", "light (2-4%)", "medium (4-6%)", and "strong (6% or more)"). Each drink was then converted into 'standard drinks' containing 10 grams of pure alcohol using the formula: standard drinks = size (in ml) * content (in %) * 0.793g/ml (density of alcohol) [23]. Second, the next morning, participants were asked to report the total number of standard alcoholic drinks consumed the previous night (answer categories ranging from 0 to 30+). For this work, the total night consumption was defined as the maximum number of drinks reported either as the sum of all 'standard drinks' consumed through the night or the next morning questionnaire. Following guidelines in alcohol research, heavy drinking nights were identified by splitting the total night consumption into two categories, namely "non-heavy" drinking nights (0-3 alcohol drinks for women/0-4 for men) and "heavy" drinking nights (4+ for women, and 5+ for men) [23]. In total, 39.8% (337) of the 847 user-nights were classified as heavy and 60.2% (510) as non-heavy.

5 NON-HEAVY AND HEAVY DRINKING ANALYSIS (RQ1)

This section presents insights into users' behavior on non-heavy and heavy drinking nights as captured through phone sensors, manual context, and image data.

Differences in mobility patterns and social contexts between heavy and non-heavy drinking nights were explored using independent t-tests and Cohen's D (Table 2). Cohen's D and 95% confidence interval (95% CI) are reported to better understand the effect sizes [41, 72] and qualifies the magnitude of the difference between the two groups as follows: 0.2 = "small", 0.5 = "medium" and 0.8 = "large"[14]. As seen in Table 2, the duration of moving and staypoints episodes are higher (MD= 0.10 and 0.91) on non-heavy drinking nights than heavy drinking nights. Conversely, the number of moving and staypoint episodes are lower (MD = -1.44 and -1.42). This suggests that participants tend to stay longer at a given location and change locations less frequently during non-heavy drinking nights compared to heavy drinking nights.

Concerning social context, a higher average number of Bluetooth devices are observed on heavy drinking nights than on non-heavy nights, for both *staypoints* and *moving episodes*. Similarly, the number of WiFi hotspots is also significantly higher on heavy drinking nights than on non-heavy nights. This indicates that participants tend to drink heavily in crowded locations rather than in empty ones.

We then analyzed the spatio-temporal context of heavy and non-heavy drinking nights (Table 3). Results show that the mean time intervals between the first and the last drink on heavy nights are longer than on non-heavy nights, echoing previous evidence that the number of drinks consumed is a function of the duration of the drinking occasion [40]. Furthermore, the number of places visited on heavy drinking nights is higher than on non-heavy drinking

nights, showing that people tend to move more on heavy drinking nights. Another observation is that the number of drinks at home (Place - Home) on non-heavy nights is higher than on heavy drinking nights, implying that heavy drinking nights tend to occur outside of private places.

Finally, regarding the scenes and objects recognizable on the pictures, to compare the most frequently identified scenes between heavy and non-heavy drinking nights (Table 3), we selected the top-5 output probabilities from each of the 365-dimensional vectors. Results show a few differences, notably in scenes labeled as coffee shop, as well as beauty salon and chemistry lab (also see in Figure 2). Although these labels appear to be irrelevant to the current investigation, manual inspection of randomly selected pictures showed visual similarities between these scenes and nightlife venues or home living rooms. An interesting observation was that non-heavy and heavy drinking nights could not be significantly discriminated by scenes labeled as bars. This finding might be related to the fact that, on heavier drinking night, people might also tend to drink in private places (i.e., before or after going out). On the other hand regarding detected objects, we observed that pictures that were significantly different between non-heavy and heavy nights contained "water bottle" (Table 3 and Figure 3). This can be explained by the fact that people are more likely to drink non-alcoholic beverages on non-heavy drinking nights.

To summarize, we used *moving* and *staypoint* episodes as anchors to segment all mobile sensor data, and aggregated them at the user-night level to compare heavy and non-heavy drinking nights. Additionally, we also studied the patterns of heavy and non-heavy drinking nights using derived features of mobility, social context, spatio-temporal context, and image context. Overall, we observe converging evidence, using activity captured by accelerometer, density of WiFi hotspots, number and duration of *moving episodes*, number of *staypoints*, that young adults tend to be more mobile, have more activities, and attend more crowded areas outside of homes on heavy drinking nights compared to non-heavy drinking nights.

6 CLASSIFYING HEAVY DRINKING VS. NON-HEAVY DRINKING (RQ2)

In this section, we investigate the feasibility of developing machine learning framework to classify a user-night as heavy or non-heavy drinking. Towards this, we use all the features described in Section 4 to train our models.

6.1 Models and Performance Measures

We use two classification methodologies: Random Forest (RF) and Support Vector Machines (SVM), using a Python package Scikit-learn [1]. The hyperparameters of the models were tuned using 5-fold cross validation (CV). For the final evaluation, we split the dataset into 80% (for training) and 20% (for testing) while ensuring that all the data points for a participant occur only in training or testing set. The baseline is the majority class i.e. all nights that are non-heavy drinking (accuracy: 0.60).

Table 2: Descriptive statistics of mobile sensor data of *moving* and *staypoint* for non-heavy drinking nights (N=510) and heavy drinking nights (N=337). (CD denotes Cohen's d, MD denotes mean difference, and CI denotes confidence interval). Entries in bold font correspond to |CD|≥ 0.2.

Sensor	Episode	Non-heavy		Heavy		Non-heavy vs. Heavy		
		mean	sd	mean	sd	MD	CD[95% CI]	T-Test (p-value)
Accelerometer - SMA	Moving	1.07	1.08	1.64	1.20	-0.57	**-0.50** [-0.53, -0.48]	-7.89 (9.4e-15)
	Staypoint	0.87	1.00	1.62	1.18	-0.75	**-0.70** [-0.73, -0.67]	-10.80 (2.0e-25)
Loc - Duration of episodes	Moving	0.81	1.05	0.71	0.99	0.10	0.09 [0.09, 0.10]	1.60 (0.10)
	Staypoint	2.35	2.20	1.43	1.39	0.91	**0.47** [0.45, 0.49]	8.35 (2.0e-16)
Loc - Count of episodes	Moving	2.64	2.31	4.08	3.07	-1.44	**-0.55** [-0.57, -0.52]	-8.20 (1.1e-15)
	Staypoint	2.69	2.32	4.11	3.09	-1.42	**-0.54** [-0.56, -0.52]	-8.05 (3.5e-15)
Loc - Mean Accuracy	Moving	172.00	365.90	214.25	383.75	-42.24	-0.113 [-0.118, -0.108]	-1.79 (0.07)
	Staypoint	85.43	157.72	107.00	189.07	-21.57	-0.126 [-0.132, -0.121]	-1.94 (0.052)
BT - Distinct Device Count	Moving	3.14	4.55	5.11	9.15	-1.97	**-0.29** [-0.31, -0.27]	-4.10 (5.1e-5)
	Staypoint	3.78	5.91	7.19	8.56	-3.42	**-0.48** [-0.50, -0.46]	-7.15 (2.2e-12)
WiFi - Distinct Hotspot Count	Moving	31.50	59.79	55.92	83.71	-24.42	**-0.35** [-0.36, -0.33]	-5.19 (2.8e-7)
	Staypoint	48.21	92.45	104.04	140.59	-55.82	**-0.49** [-0.51, -0.47]	-7.2 (1.8e-12)
Apps - Distinct App Count	Moving	12.06	9.08	13.50	9.67	-1.44	-0.155 [-0.161, -0.148]	-2.44 (0.015)
	Staypoint	12.78	9.04	13.69	9.79	-0.91	-0.097 [-0.101, -0.093]	-1.53 (0.127)
Proximity - Record Count	Moving	142.77	455.30	316.00	1770.06	-173.23	-0.150 [-0.156, -0.144]	-1.96 (0.05)
	Staypoint	256.97	752.70	747.06	6154.23	-490.09	-0.127 [-0.132, -0.121]	-1.62 (0.11)
Battery - Record Count	Moving	161.50	283.82	203.84	296.41	-42.33	-0.146 [-0.153, -0.140]	-2.32 (0.02)
	Staypoint	530.44	1229.70	376.59	369.51	153.85	0.155 [0.149, 0.162]	3.02 (0.003)
Screen - Record Count	Moving	20.99	58.36	28.94	39.90	-7.94	-0.152 [-0.159, -0.146]	-2.66 (0.008)
	Staypoint	34.33	59.35	56.37	100.72	-22.04	**-0.283** [-0.294, -0.271]	-4.05 (5.8e-05)

Table 3: Descriptive statistics of top five detected scenes in all pictures of non-heavy drinking nights (N=510) and heavy drinking nights (N=337). (CD denotes Cohen's d, MD denotes mean difference, and CI denotes confidence interval). Entries in bold font correspond to |CD|≥ 0.2.

Top 5 Features	Non-heavy		Heavy		Non-heavy vs. Heavy		
	mean	sd	mean	sd	MD	CD[95% CI]	T-Test (p-value)
Time - Mean Gaps	2705.50	4279.88	3552.85	3631.47	-847.35	**-0.218** [-0.228, -0.208]	-3.03 (0.0025)
Place - Count	5.39	2.17	7.25	4.52	-1.86	**-0.534** [-0.567, -0.500]	-5.69 (2.87e-08)
Place - Home	2.92	2.54	2.06	2.83	0.858	**0.319** [0.299, 0.339]	3.50 (0.0005)
Object - Beer glass	0.078	0.0060	0.077	0.0163	0.00086	0.0055 [0.0052, 0.0057]	0.08 (0.94)
Object - Beer bottle	0.048	0.0007	0.064	0.0025	-0.016	-0.114 [-0.119, -0.109]	-1.60 (0.10)
Object - Red wine	0.03	0.0007	0.05	0.0014	-0.02	-0.176 [-0.184, -0.168]	-2.36 (0.018)
Object - Water bottle	0.070	0.001	0.030	0.002	0.04	**0.230** [0.220, 0.240]	3.70 (0.0002)
Object - Goblet	0.050	0.002	0.020	0.003	0.021	0.190 [0.180, 0.200]	3.00 (0.003)
Scene - Beer hall	0.138	0.203	0.103	0.158	0.035	0.188 [0.179, 0.197]	2.82 (0.005)
Scene - Beauty salon	0.058	0.075	0.039	0.063	0.019	**0.270** [0.257, 0.283]	3.98 (7.66e-05)
Scene - Chemistry lab	0.059	0.102	0.036	0.074	0.024	**0.258** [0.246, 0.270]	3.92 (9.61e-05)
Scene - Coffee shop	0.035	0.074	0.018	0.048	0.017	**0.264** [0.251, 0.276]	4.08 (4.9e-05)
Scene - Bar	0.018	0.029	0.016	0.034	0.001	0.040 [0.039, 0.042]	0.56 (0.58)

6.2 Classification Results

The classification results are presented in Table 4. From the table, we observe that RF performs better than SVM for all feature groups.

We observe that fully automatic sensing (S) gives 0.71 accuracy compared to a baseline of 0.60. Note that this feature set does not need any active user input. We now compare this result with features obtained from active user involvement - requiring human effort. Using context features, Time (T) and Place (P) obtain a classification accuracy of 0.70 and 0.67 respectively, while the fusion of these two features (T+P) shows a small improvement with 0.71. Similarly, the classification using Visual features (I) is 0.72. This indicates the feasibility of using mobile phone sensors unobtrusively to monitor heavy drinking behavior and potentially provide feedback/intervention.

Thanh-Trung Phan, Florian Labhart, Skanda Muralidhar,
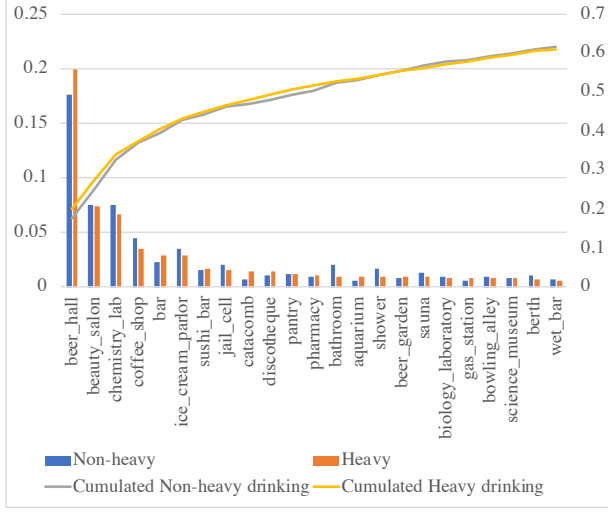Daniel Gatica-Perez

Figure 2: The top 20 detected scenes of all images on heavy and non-heavy drinking nights. The unit of the y-axis is average scores which are ranked with respect to their value on the Y@N data.
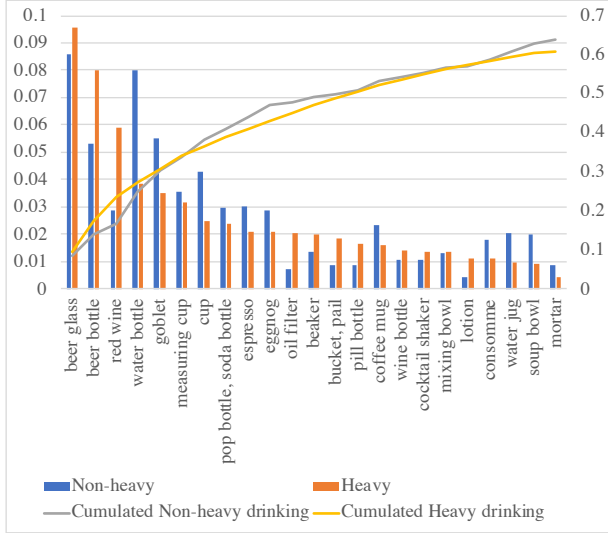

Figure 3: The top 20 detected objects of all images on heavy and non-heavy drinking nights. The unit of the y-axis is average scores which are ranked with respect to their value on the Y@N data.

For the sake of completeness, we present the results of feature fusion. Please note that we do not propose this as a very practical approach as it would involve active input from users, and it is known that user burden is one of the limitations of the current methodology of self-reported questionnaires. In the first step, we observe a marginal improvement in classification accuracy (0.76) by combining sensors and context (S, T, P). Addition of image features and contextual features do not improve over the classification

Table 4: Classification accuracy of heavy and non-heavy nights on 847 user-nights (337 heavy and 510 non-heavy). (A denotes accuracy, F1-score denotes weighted-average F1-score).

Feature Group	Features	RF		SVM	
		Acc	F1-score	Acc	F1-score
Baseline	Majority class	0.60	0.66	0.60	0.66
Sensor	**S**ensor (S)	0.71	0.69	0.49	0.46
Context	**T**ime (T)	0.70	0.68	0.70	0.68
	Place (P)	0.67	0.63	0.66	0.63
Visual	**I**mage (I)	0.72	0.70	0.67	0.65
Data Source Fusion	T + P	0.71	0.68	0.70	0.68
	S + T	**0.76**	**0.74**	0.70	0.68
	S + P	0.71	0.70	0.49	0.46
	S + I	0.76	0.73	0.49	0.46
	T + I	0.72	0.69	0.69	0.67
	S + T + I	0.76	0.74	0.70	0.68
	S + P + I	0.75	0.73	0.49	0.46
	S + P + T	0.76	0.75	0.70	0.68
	S + P + T + I	0.75	0.74	0.70	0.68

accuracy (0.75). Best classification accuracy of 0.76 is obtained by combining sensor (S) and time (T) features.

To understand the contributions of features towards classification, we compute the top 20 features from the random forest implementation in sci-kit learn. The metric for feature importance is Gini impurity. Figure 4 shows the top 20 features by using sensors (S), and sensor and time (S+T). Note that a higher mean decrease indicates higher variable importance. We observe that for features from sensors (S), the top important features are from accelerometer, Bluetooth/WiFi, and location. For the best classification model (S+T), time of drinking plays is considered more important than sensor features. This indicates the importance of data collected from the mobile sensors.

To summarize, our results show initial promise of developing a framework for mobile apps to automatic infer heavy alcohol drinking nights using in-built sensors, compared to active human engagement (surveys or images).

7 DISCUSSION AND IMPLICATIONS

In this work, using a dataset from our previous work [55], we studied user behavioral patterns on heavy and non-heavy drinking nights. Towards this we used sensor data from mobile phones as well as images taken and questionnaires answered by participants in-situ. Statistical analysis of features extracted (RQ1) from all these modalities showed the following trends.

(1) Participants tended to stay for shorter duration at a given location and change locations more frequently during heavy drinking nights.
(2) Participants were more likely to drink heavily when attending crowded locations rather than in empty ones.
(3) The mean time intervals between the first and the last drink on heavy nights were longer than on non-heavy nights. This is supported by literature that reports that the number of
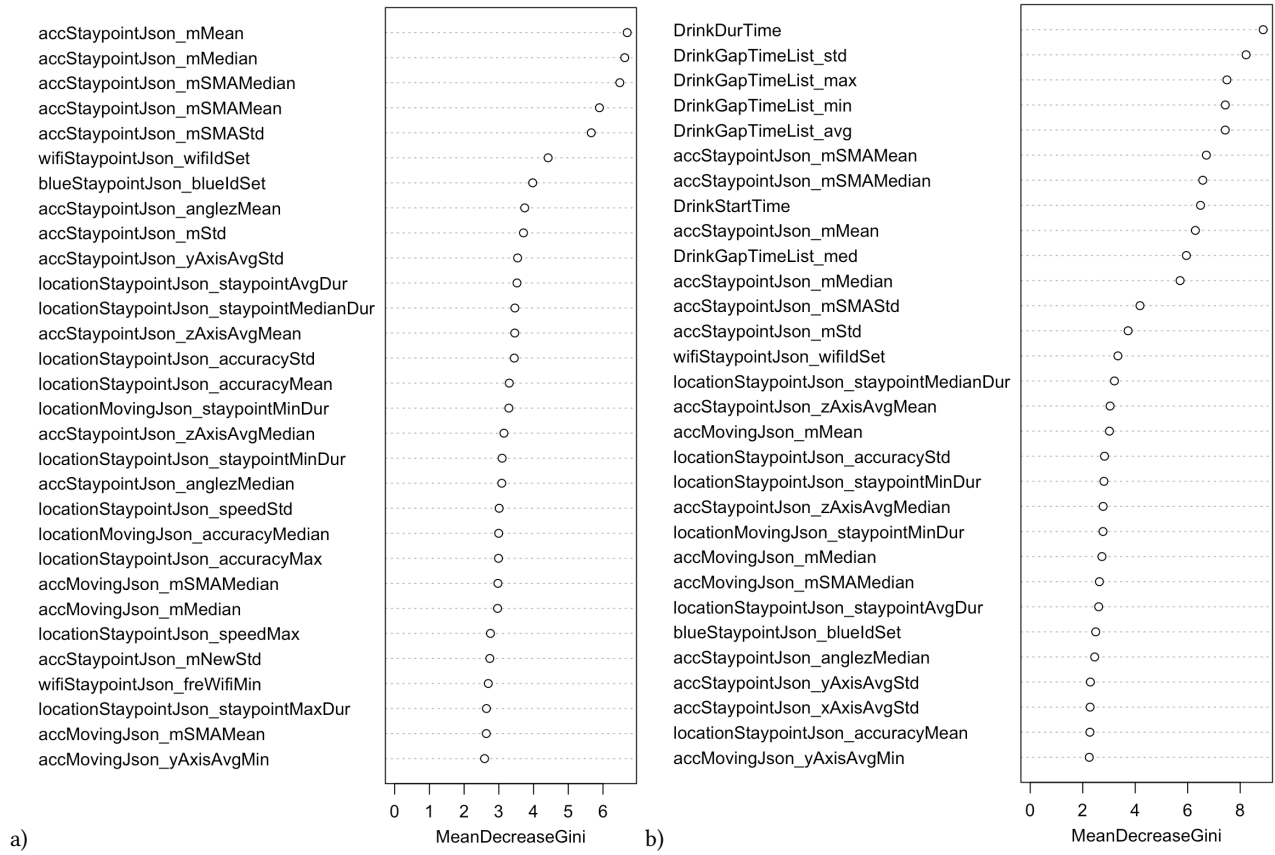
Figure 4: Top important features detected by using a) Sensor (S), and b) Sensor (S) + Time (T) when using RF. A higher Mean Decrease in Gini indicates higher variable importance.

drinks consumed is a function of the duration of the drinking occasion [40, 62].

(4) The number of drinks reported was greater outside private places for heavy drinking nights, while a greater number of drinks was reported on private places (homes) for non-heavy drinking nights.

These results underline the dynamic nature of nightlife-related drinking behaviors in the sense that multiple drinks can be consumed in different locations over the course of an extended drinking occasion. Future research on drinking behaviors at the event-level are thus recommended to take into account the contextual characteristics of each drink separately (types of location, drinking companions, drink prices, etc.) and the sequence of these drinking episodes [58].

The results also suggest that a significant proportion of participants might have moved from their home to licensed venues (e.g., pub, nightclub) in their cities' entertainment district after starting drinking at home (also known as "predrinking" [25, 36]). However, this was not the only possible trajectory: for example, some people might have moved to non-commercial places, such as a public park or someone else's home, while other might have moved several times, which is characteristic for heavy drinking nights [17]. It

should be noted, though, that participants were recruited in the nightlife districts of two cities [38], which might partly contribute to the finding that they tended to come back to nightlife venues during the field study. Yet, small rural cities might also have drinking places looking like a bar or a beer hall, where young people might go to on weekend evenings. We therefore believe that the tendency to move out from home to nightlife outdoor places does not only concern young people living within or close to a city downtown area, but might also be a normalized behavior for those living in the surrounding neighborhoods and further away.

With regard to public health, the finding that young people tend to move across locations on heavy drinking nights means that they might travel, drive, or seek to enter nightlife venues under the influence of alcohol. The development of targeted public transport systems, or the implementation of controls at nightlife venues entrance to detect inebriated patrons might therefore contribute to the reduction of the burden of heavy drinking and related harms at both the individual and the community levels [16].

The results of the classification task (RQ2) can be seen as an initial feasibility step towards developing applications for alcohol consumption self-awareness using automatic sensors in mobile phones. Current existing prevention and intervention programs

Thanh-Trung Phan, Florian Labhart, Skanda Muralidhar,
Daniel Gatica-Perez

(e.g. [69]) rely on participants self-reports of drinks consumed. If the participants do not accurately report details of their consumption, the intervention system fails. In contrast, our results suggest that smartphone sensors are able to unobtrusively identify the drinking pattern of their owners to some degrees (70% accuracy). In this respect, it should be noted that the accuracy of the classification task is the average accuracy for all participants, which was not tailored for each specific participant-situation. We believed that an adaptive model, which learns and updates based on each individual's behavioral patterns, might increase the accuracy.

Our work, although related to that of Bae et al. [5], differs significantly in the following ways. First, our dataset in 8 times bigger than the dataset used by Bae et al., who reported their findings on a pool of 30 participants. Our work used a dataset consisting of 241 participants with 847 user-nights. Second, the aim of our investigation is to be able to provide insights to the users over the course of an entire drinking occasion (i.e night-level). While in-the-event interventions methods might be seen as the panacea for preventing alcohol intoxication [5], previous attempts to deliver in-the-event preventive messages (e.g., by sending SMS to the participants when they reached a certain number of drinks over the course of a drinking night) has shown to be challenging [69]. One reason for this is that heavy drinkers tend to show increased attention biases towards alcohol-related stimuli and are resistant to nudging [67]. Our proposed approach of inferring heavy drinking status of the previous night of a user might not intervene heavy drinking in real-time, but might efficiently contribute to the literature researching heavy drinking [35] by reflecting about what elements of nightlife (e.g. the number of locations or the sequence of events) might play a role in this practice. Third, we used the images of drinks taken by participants to understand the main types of surrounding objects and partial scenes captured in them. The fact that photos are ultimately taken at the discretion of users (e.g, drink photos could be taken before being consumed or while being consumed) could result in a possible limitation. Finally, phone sensing involves sensitive data, and therefore any future public health phone applications related to alcohol consumption must be designed with ethics, privacy, and the well-being of users as fundamental values.

8 CONCLUSION

In this work, we investigated the comparative value of automatic smartphone sensors and manually contributed information including the context of drinking events and pictures taken in-situ towards the understanding of behavior during heavy drinking nights.

We first applied well-known definitions of heavy drinking in alcohol research to assign heavy or non-heavy drinking label for each night in the dataset. Next, we used location data with timestamps to segment participants' nights into *moving* and *staypoint* episodes. Then, we extracted features from sensors and images, as well as context data, to analyze the differences between heavy and non-heavy nights. We observed that during heavy drinking nights, participants tended to stay for shorter times at a given location, and change locations more frequently; they were more likely to be in crowded locations; the time intervals between the first and the last drink on heavy drinking nights were longer than on

non-heavy nights; and the number of drinks reported was greater outside private places.

We then developed a machine learning framework to classify heavy drinking nights and compared the classification power of automatically captured sensor data vs. manually contributed context information and images. The results showed that a fully automatic approach with phone sensors results in a accuracy of 71%. The performance on the same task, when using manual input of context of drinking events and visual features of manually contributed images, corresponds to an accuracy of 70% and 72%, respectively.

In the present study, we considered each night independently, although nights were clustered within participants. This approach has the disadvantage that part of the variance between the nights might relate to the participants habits rather than to the nights themselves. However, this approach has the advantage that participants serve as their own controls considering that we have assessed multiple night within the same individuals rather than random nights from unrelated individuals. We believe that our work has implications not only for alcohol research, but also for potential applications related to self-tracking and to health interventions.

ACKNOWLEDGMENTS

REFERENCES

[1] 2019. *Machine Learning in Python.* Retrieved October 28, 2019 from https://scikit-learn.org/stable/
[2] 2020. Centers for Disease Control and Prevention. https://www.cdc.gov/alcohol/faqs.htm Accessed: 2020-01-13.
[3] Ian Anderson, Julie Maitland, Scott Sherwood, Louise Barkhuus, Matthew Chalmers, Malcolm Hall, Barry Brown, and Henk Muller. 2007. Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. *Mobile networks and applications* 12, 2-3 (2007), 185–199.
[4] Zachary Arnold, Danielle Larose, and Emmanuel Agu. 2015. Smartphone inference of alcohol consumption levels from gait. In *Healthcare Informatics (ICHI), 2015 International Conference on.* IEEE, 417–426.
[5] Sangwon Bae, Tammy Chung, Denzil Ferreira, Anind K Dey, and Brian Suffoletto. 2018. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive behaviors* 83 (2018), 42–47.
[6] Sangwon Bae, Denzil Ferreira, Brian Suffoletto, Juan C Puyana, Ryan Kurtz, Tammy Chung, and Anind K Dey. 2017. Detecting Drinking Episodes in Young Adults Using Smartphone-based Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 5.
[7] Ling Bao and Stephen S Intille. 2004. Activity recognition from user-annotated acceleration data. In *International conference on pervasive computing.* Springer, 1–17.
[8] Luke Bermingham and Ickjai Lee. 2019. Mining place-matching patterns from spatio-temporal trajectories using complex real-world places. *Expert Systems with Applications* 122 (2019), 334–350.
[9] Tomas Brezmes, Juan-Luis Gorricho, and Josep Cotrina. 2009. Activity recognition from accelerometer data on a mobile phone. In *International Work-Conference on Artificial Neural Networks.* Springer, 796–799.
[10] Ashlee C Carter, Karen Obremski Brandon, and Mark S Goldman. 2010. The college and noncollege experience: a review of the factors that influence drinking behavior in young adulthood. *Journal of studies on alcohol and drugs* 71, 5 (2010), 742–750.
[11] John Chon and Hojung Cha. 2011. Lifemap: A smartphone-based context provider for location-based services. *IEEE Pervasive Computing* 10, 2 (2011), 58–67.
[12] Matthew A Christensen, Laura Bettencourt, Leanne Kaye, Sai T Moturu, Kaylin T Nguyen, Jeffrey E Olgin, Mark J Pletcher, and Gregory M Marcus. 2016. Direct measurements of smartphone screen-time: relationships with demographics and sleep. *PloS one* 11, 11 (2016), e0165331.

[13] John D Clapp, Danielle R Madden, Douglas D Mooney, and Kristin E Dahlquist. 2017. Examining the social ecology of a bar-crawl: An exploratory pilot study. *PLoS one* 12, 9 (2017), e0185238.

[14] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences.* Routledge.

[15] Kelly E Courtney and John Polich. 2009. Binge drinking in young adults: Data, definitions, and determinants. *Psychological bulletin* 135, 1 (2009), 142.

[16] Ashlee Curtis, Nicolas Droste, Kerri Coomber, Belinda Guadagno, Richelle Mayshak, Shannon Hyder, Alexa Hayley, and Peter Miller. 2019. Off the rails—Evaluating the nightlife impact of Melbourne, Australia's 24-h public transport trial. *International Journal of Drug Policy* 63 (2019), 39–46.

[17] Paul M Dietze, Michael Livingston, Sarah Callinan, and Robin Room. 2014. The big night out: What happens on the most recent heavy drinking occasion among young V ictorian risky drinkers? *Drug and alcohol review* 33, 4 (2014), 346–353.

[18] Trinh Minh Tri Do, Jan Blom, and Daniel Gatica-Perez. 2011. Smartphone usage in the wild: a large-scale analysis of applications and context. In *Proceedings of the 13th international conference on multimodal interfaces.* ACM, 353–360.

[19] Trinh Minh Tri Do and Daniel Gatica-Perez. 2011. Groupus: Smartphone proximity data and human interaction type mining. In *2011 15th Annual International Symposium on Wearable Computers.* IEEE, 21–28.

[20] Trinh Minh Tri Do and Daniel Gatica-Perez. 2013. The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. *IEEE Transactions on Mobile Computing* 13, 3 (2013), 638–648.

[21] Nathan Eagle and Alex Pentland. 2005. Social serendipity: Mobilizing social software. *IEEE Pervasive Computing* 4, 2 (2005), 28–34.

[22] Gerhard Gmel, Jacques Gaume, Mohamed Faouzi, Jean-Pierre Kulling, and Jean-Bernard Daeppen. 2008. Who drinks most of the total alcohol in young men—risky single occasion drinking as normative behaviour. *Alcohol & Alcoholism* 43, 6 (2008), 692–697.

[23] Gerhard Gmel and Jürgen Rehm. 2004. Measuring alcohol consumption. *Contemporary Drug Problems* 31, 3 (2004), 467–540.

[24] Gerhard Gmel, Jürgen Rehm, Emmanuel Kuntsche, et al. 2003. Binge drinking in Europe: definitions, epidemiology, and consequences. *Sucht* 49, 2 (2003), 105–116.

[25] Karen Hughes, Zara Anderson, Michela Morleo, and Mark A Bellis. 2008. Alcohol, nightlife and violence: the relative contributions of drinking before and during nights out to negative health and criminal justice outcomes. *Addiction* 103, 1 (2008), 60–65.

[26] Mark Jayne, Sarah L Holloway, and Gill Valentine. 2006. Drunk and disorderly: alcohol, urban life and public space. *Progress in human geography* 30, 4 (2006), 451–468.

[27] Dean M Karantonis, Michael R Narayanan, Merryn Mathie, Nigel H Lovell, and Branko G Celler. 2006. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE transactions on information technology in biomedicine* 10, 1 (2006), 156–167.

[28] Donna M Kazemi, Brian Borsari, Maureen J Levine, Shaoyu Li, Katie A Lamberson, and Laura A Matta. 2017. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. *Journal of health communication* 22, 5 (2017), 413–432.

[29] Daniel Kelly, Barry Smyth, and Brian Caulfield. 2013. Uncovering measurements of social and demographic behavior from smartphone location data. *IEEE Transactions on Human-Machine Systems* 43, 2 (2013), 188–198.

[30] Mikkel Baun Kjærgaard and Petteri Nurmi. 2012. Challenges for social sensing using wifi signals. In *Proceedings of the 1st ACM workshop on Mobile systems for computational social science.* ACM, 17–21.

[31] Emmanuel Kuntsche, Ronald Knibbe, Gerhard Gmel, and Rutger Engels. 2005. Why do young people drink? A review of drinking motives. *Clinical psychology review* 25, 7 (2005), 841–861.

[32] Emmanuel Kuntsche, Sandra Kuntsche, Johannes Thrul, and Gerhard Gmel. 2017. Binge drinking: Health impact, prevalence, correlates and interventions. *Psychology & health* 32, 8 (2017), 976–1017.

[33] Emmanuel Kuntsche and Florian Labhart. 2012. Investigating the drinking patterns of young people over the course of the evening at weekends. *Drug and alcohol dependence* 124, 3 (2012), 319–324.

[34] Emmanuel Kuntsche, Roy Otten, and Florian Labhart. 2015. Identifying risky drinking patterns over the course of Saturday evenings: An event-level study. *Psychology of addictive behaviors* 29, 3 (2015), 744.

[35] Florian Labhart, Rutger Engels, and Emmanuel Kuntsche. 2018. What reminds young people that they drank more than intended on weekend nights: an event-level study. *Journal of studies on alcohol and drugs* 79, 4 (2018), 644–648.

[36] Florian Labhart, Kathryn Graham, Samantha Wells, and Emmanuel Kuntsche. 2013. Drinking before going to licensed premises: An event-level analysis of predrinking, alcohol consumption, and adverse outcomes. *Alcoholism: Clinical and Experimental Research* 37, 2 (2013), 284–291.

[37] Florian Labhart, Michael Livingston, Rutger Engels, and Emmanuel Kuntsche. 2018. After how many drinks does someone experience acute consequences—determining thresholds for binge drinking based on two event-level studies. *Addiction* 113, 12 (2018), 2235–2244.

[38] Florian Labhart, Darshan Santani, Jasmine Truong, Flavio Tarsetti, Olivier Bornet, Sara Landolt, Daniel Gatica-Perez, and Emmanuel Kuntsche. 2017. Development of the Geographical Proportional-to-size Street-Intercept Sampling (GPSIS) method for recruiting urban nightlife-goers in an entire city. *International Journal of Social Research Methodology* 20, 6 (2017), 721–736.

[39] Florian Labhart, Flavio Tarsetti, Olivier Bornet, Darshan Santani, Jasmine Truong, Sara Landolt, Daniel Gatica-Perez, and Emmanuel Kuntsche. 2019. Capturing drinking and nightlife behaviours and their social and physical context with a smartphone application–investigation of users' experience and reactivity. *Addiction Research & Theory* (2019), 1–14.

[40] Florian Labhart, Samantha Wells, Kathryn Graham, and Emmanuel Kuntsche. 2014. 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* 49, 3 (2014), 327–335.

[41] Dong Kyu Lee. 2016. Alternatives to P value: confidence interval and effect size. *Korean journal of anesthesiology* 69, 6 (2016), 555–562.

[42] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. 2013. Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services.* ACM, 389–402.

[43] MJ Mathie, Branko G Celler, Nigel H Lovell, and ACF Coster. 2004. Classification of basic daily movements using a triaxial accelerometer. *Medical and Biological Engineering and Computing* 42, 5 (2004), 679–687.

[44] Merryn J Mathie, Adelle CF Coster, Nigel H Lovell, Branko G Celler, Stephen R Lord, and Anne Tiedemann. 2004. A pilot study of long-term monitoring of human movements in the home using accelerometry. *Journal of telemedicine and telecare* 10, 3 (2004), 144–151.

[45] Uwe Maurer, Asim Smailagic, Daniel P Siewiorek, and Michael Deisher. 2006. *Activity recognition and monitoring using multiple sensors on different body positions.* Technical Report. CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE.

[46] Dennis McCarty. 1985. Environmental factors in substance abuse. In *Determinants of substance abuse.* Springer, 247–281.

[47] Jun-Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I Hong. 2014. Toss'n'turn: smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI conference on human factors in computing systems.* ACM, 477–486.

[48] Raul Montoliu, Jan Blom, and Daniel Gatica-Perez. 2013. Discovering places of interest in everyday life from smartphone data. *Multimedia tools and applications* 62, 1 (2013), 179–207.

[49] Andrei Papliatseyeu and Oscar Mayora. 2009. Mobile habits: Inferring and predicting user activities with a location-aware smartphone. In *3rd Symposium of Ubiquitous Computing and Ambient Intelligence 2008.* Springer, 343–352.

[50] Thanh-Trung Phan, Skanda Muralidhar, and Daniel Gatica-Perez. 2019. # Drink Or# Drunk: Multimodal Signals and Drinking Practices on Instagram. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare.* ACM, 71–80.

[51] Thanh-Trung Phan, Skanda Muralidhar, and Daniel Gatica-Perez. 2019. Drinks & Crowds: Characterizing Alcohol Consumption through Crowdsensing and Social Media. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 59.

[52] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. 2005. Activity recognition from accelerometer data. In *Aaai,* Vol. 5. 1541–1546.

[53] Damaris J Rohsenow. 1982. Social anxiety, daily moods, and alcohol use over time among heavy social drinking men. *Addictive Behaviors* 7, 3 (1982), 311–315.

[54] Darshan Santani, Joan-Isaac Biel, Florian Labhart, Jasmine Truong, Sara Landolt, Emmanuel Kuntsche, and Daniel Gatica-Perez. 2016. The night is young: urban crowdsourcing of nightlife patterns. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 427–438.

[55] Darshan Santani, Florian Labhart, Sara Landolt, Emmanuel Kuntsche, Daniel Gatica-Perez, et al. 2018. DrinkSense: Characterizing youth drinking behavior using smartphones. *IEEE Transactions on Mobile Computing* 17, 10 (2018), 2279–2292.

[56] Piotr Sapiezynski, Arkadiusz Stopczynski, Radu Gatej, and Sune Lehmann. 2015. Tracking human mobility using wifi signals. *PloS one* 10, 7 (2015), e0130824.

[57] Piotr Sapiezynski, Arkadiusz Stopczynski, David Kofoed Wind, Jure Leskovec, and Sune Lehmann. 2017. Inferring person-to-person proximity using WiFi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 24.

[58] Oliver Stanesby, Florian Labhart, Paul Dietze, Cassandra JC Wright, and Emmanuel Kuntsche. 2019. The contexts of heavy drinking: A systematic review of the combinations of context-related factors associated with heavy drinking occasions. *PloS one* 14, 7 (2019).

[59] Abigail K Stevely, John Holmes, and Petra S Meier. 2019. Contextual characteristics of adults' drinking occasions and their association with levels of alcohol consumption and acute alcohol-related harm: A mapping review. *Addiction* (2019).

Thanh-Trung Phan, Florian Labhart, Skanda Muralidhar,
Daniel Gatica-Perez

[60] Xing Su, Hanghang Tong, and Ping Ji. 2014. Activity recognition with smartphone sensors. *Tsinghua science and technology* 19, 3 (2014), 235–249.

[61] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2818–2826.

[62] Johannes Thrul and Emmanuel Kuntsche. 2015. The impact of friends on young adults' drinking over the course of the evening—an event-level analysis. *Addiction* 110, 4 (2015), 619–626.

[63] Julia M Townshend and Theodora Duka. 2005. Binge drinking, cognitive performance and mood in a population of young social drinkers. *Alcoholism: Clinical and Experimental Research* 29, 3 (2005), 317–325.

[64] Ilse Van Liempt, Irina Van Aalst, and Tim Schwanen. 2015. Introduction: Geographies of the urban night.

[65] Peter H Veltink, HansB J Bussmann, Wiebe De Vries, WimL J Martens, Rob C Van Lummel, et al. 1996. Detection of static and dynamic activities using uniaxial accelerometers. *IEEE Transactions on Rehabilitation Engineering* 4, 4 (1996), 375–385.

[66] Guillem Vich, Oriol Marquet, and Carme Miralles-Guasch. 2017. Suburban commuting and activity spaces: using smartphone tracking data to understand the spatial extent of travel behaviour. *The Geographical Journal* 183, 4 (2017), 426–439.

[67] Jessica Weafer and Mark T Fillmore. 2013. Acute alcohol effects on attentional bias in heavy and moderate drinkers. *Psychology of addictive behaviors* 27, 1 (2013), 32.

[68] Henry Wechsler and Toben F Nelson. 2001. Binge drinking and the American college students: What's five drinks? *Psychology of Addictive Behaviors* 15, 4 (2001), 287.

[69] Cassandra Wright, Paul M Dietze, Paul A Agius, Emmanuel Kuntsche, Michael Livingston, Oliver C Black, Robin Room, Margaret Hellard, and Megan SC Lim. 2018. Mobile phone-based ecological momentary intervention to reduce young adults' alcohol use in the event: a three-armed randomized controlled trial. *JMIR mHealth and uHealth* 6, 7 (2018), e149.

[70] Qiang Xu, Jeffrey Erman, Alexandre Gerber, Zhuoqing Mao, Jeffrey Pang, and Shobha Venkataraman. 2011. Identifying diverse usage behaviors of smartphone apps. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 329–344.

[71] Zhixian Yan, Jun Yang, and Emmanuel Munguia Tapia. 2013. Smartphone bluetooth based social sensing. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. ACM, 95–98.

[72] Koji Yatani. 2016. Effect sizes and power analysis in hci. In *Modern Statistical Methods for HCI*. Springer, 87–110.

[73] Chuang-wen You, Kuo-Cheng Wang, Ming-Chyi Huang, Yen-Chang Chen, Cheng-Lin Lin, Po-Shiun Ho, Hao-Chuan Wang, Polly Huang, and Hao-Hua Chu. 2015. SoberDiary: A Phone-based Support System for Assisting Recovery from Alcohol Dependence. In *Proceedings of CHI (CHI '15)*. ACM, New York, NY, USA, 3839–3848.

[74] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2018. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence* 40, 6 (2018), 1452–1464.