UNDERSTANDING EATING AND DRINKING IN CONTEXT FROM CROWDSOURCED DATA

Thèse n. 7674 2020
présenté le 7 Mai 2020
à la Faculté des sciences et techniques de l'ingénieur Laboratoire de l'IDIAP
Programme doctoral en génie électrique
École Polytechnique Fédérale de Lausanne

pour l'obtention du grade de Docteur ès Sciences
par

Thanh-Trung Phan

THÈSE ACCEPTÉE SUR PROPOSITION DU JURY:

Prof. J.-Ph. Thiran, président du jury
Prof. D. Gatica-Perez, directeur de thèse
Dr I. Kompatsiaris, rapporteur
Dr K. Farrahi, rapporteuse
Dr D. Gillet, rapporteur

Lausanne, EPFL, 2020

Everything Is Obvious:
Once You Know the Answer.
— Duncan J. Watts

To my parents…

Acknowledgements

I have countless memories with Prof. Daniel Gatica-Perez during lunch meetings with him. He is not only my supervisor, but also a boss, an inspiration, a role model, and especially a witness of my research's growth. In my 5-year journey towards this degree, he has been providing his patience and altruism, heartfelt support and invaluable guidance, inspiration and suggestions, in my paper work. Without his guidance, this thesis would not have been possible and I am indebted to him for his assistance.

I would like to thank my jury members: Prof. J.-Ph. Thiran (Jury President), Dr I. Kompatsiaris (Reporter), Dr K. Farrahi (Reporter), Dr D. Gillet (Reporter) for reading my thesis and providing insightful comments.

During these five years, I had the opportunity to work with my colleagues and fellow research scholars at Idiap Research Institute: Dr. Skanda Muralidhar, and Florian Labhart. Their support in the discussion and also in collaboration with writing for my research have been great contributors to the completion of this thesis. I would also like to thank the other members (including alumni) of the Social computing group (Dr. Trinh Minh Tri Do, Dr. Darshan Santani, Dr. Laurent Nguyen, Dr. Benoit Masse, Jessica Pidoux, Lakmal Meegahapola, and Sina Sajadmanesh), and my friends at Idiap (Dr. Subhadeep Dey, Dr. Dhananjay Ram, and Lesly Miculicich) for making my stay at Martigny an unforgettable experience. I am particularly grateful to all the people at Idiap, the Idiap secretaries, and the system administrators for assisting me with technical and admin process.

This work is dedicated to my parents, who did not finish primary school and could only write enough but they tried to raise and support me incredibly so I can acquire a high education like this. They always feel proud of me whenever they talk about me to their cousins or neighbors. However, I am the one who is proud of them when, luckily, I am their child. They are less educated farmers but they dare to dream of their child excelling in education. Had it not been for their loving upbringing and nurturing, I would not have been where I am today and what I am today. Especially, this work is to my late father for his care and support. I know he is always up there watching over me, and protecting me constantly. I thank my parents with all my heart and they will always be my heroes.

Lausanne, June 4, 2020 Thanh-Trung Phan

Abstract

With the increasing rate of urbanization, understanding food and beverage consumption, including alcohol drinking behaviour with its consequences, is relevant in such a megacity in the future. Especially, investigation of alcohol drinking is necessary for uncovering more drinking activities of young people in context, and providing more scientific references to authorities to improve public policies, in particular in the domain of public health. In previous research, data were collected a posteriori in face-to-face interviews or by using questionnaires. Thanks to the development of feature phones, an evolution of more traditional methods in ubiquitous food and alcohol research could have collected survey data via SMS. However, these methods have limitations, including a low recall and an expensive scaling up.

On the other hand, the adoption of smartphones and social media is opening new channels for investigating behaviours by collecting fine-grained in-situ data, following methodologies from social sciences, and using advanced technologies from computer sciences. Recently, crowdsourcing is a new paradigm that consists in using the inputs from a great number of people to facilitate and accelerate large scale data collection from broad samples, compared with traditional methods. In addition, mobile crowdsourcing, a form of crowdsourcing, has enormous potential in collecting in-situ data by taking advantage of embedded sensors, cameras, and being equipped with Internet connection.

In this dissertation, we investigate drinking and eating behaviour of young people in Switzerland, based on crowdsourcing data including records and metadata from mobile sensors (mobile crowdsensing) and data shared on social networks. Our contributions are three-fold, relative to either of two data sources or their combination. First, we conduct data analyses that uncover generic food and drink consumption on Instagram and reveal two types of drinking practices (casual and heavy drinking) on social media. This analysis provides an initial snapshot of the food consumption and alcohol drinking practices based on the way they appear online, and creates a preliminary alcohol consumption model that is developped in the rest of the dissertation. Second, we use mobile crowdsensing data, annotated a posteriori, both to identify heavy drinking and to understand the characteristics of private spaces (including ambiances) and activities (including drinking activities) of young people in the weekend nights in Switzerland. These results show how mobile crowdsensing data can be used to better understand and predict alcohol drinking practices and ambiances in private spaces.

Third, we combine mobile crowdsensing data with social media to retrieve the multi-factorial characteristics of drinking events depending on the type of beverage (multiple alcoholic and non-alcoholic categories) based on images features and contextual cues from individual and joint data sources. This result indicates the feasibility of using, individually or combined, data from mobile crowdsensing and social networks in discriminating drinking behaviour. This is a promising sign towards the development of a system that use machine learning for self-monitoring of alcohol consumption.
This dissertation not only contributes towards understanding urban drinking and eating activities, but also towards the computational inference of food/alcohol categories, drinking practices, and ambiances of private drinking places at a large scale. This dissertation, by combining the advanced machine learning of computer science and literature of social science, demonstrates the relevance of using a multidisciplinary approach to investigate social behaviours in urban areas.

Key words: Mobile Crowdsourcing, Mobile Crowdsensing, Social Media, Food Consumption, Alcohol Consumption, Casual Drinking, Heavy Drinking, Home Spaces, Ambiance, Youth, Nightlife, Alcohol, Ubiquitous Computing, Urban Computing, Social Computing, Instagram, Foursquare

Résumé

Avec l'augmentation du taux d'urbanisation, il est pertinent de comprendre les habitudes actuelles et futures de consommation de nourriture et de boisson dans les grandes villes, en incluant la consommation d'alcool et ses conséquences. Les enquêtes sur la consommation d'alcool sont nécessaires pour comprendre et analyser la manière dont les jeunes consomment de l'alcool et dans quels contextes. Cela permet aussi de fournir davantage de références scientifiques aux autorités afin d'améliorer les politiques menées, notamment en matière de santé publique. Traditionnellement, les données utilisées sont collectées a posteriori, en posant des questions aux participants pendant un entretien en face à face ou par l'intermédiaire d'un questionnaire. Avec la démocratisation des téléphones portables, ces même méthodes ont pu être appliquée à plus grande échelle, en envoyant des questionnaires par SMS, pour collecter des données sur la consommation de nourriture et d'alcool. Cependant, ces méthodes restent limitées, notamment par une mise à l'échelle coûteuse et un faible rappel.

D'autre part, l'adoption massive des smartphones et des réseaux sociaux permet aujourd'hui d'étudier les comportements à partir de données plus fines, collectées in situ d'après des méthodes d'analyses issues des sciences sociales, et en utilisant des technologies informatiques récents. Apparu depuis peu, le *crowdsourcing*, ou production participative, est un nouveau paradigme qui consiste à utiliser le savoir-faire d'un grand nombre de personnes pour faciliter et accélérer la collecte de données à grande échelle. En outre, le *mobile crowdsourcing*, ou production participative de données de smartphone, est une forme de *crowdsourcing* très prometteuse, qui consiste à collecter des données in situ en tirant parti de la caméra et autres capteurs intégrés dans le smartphone, et à les partager sur Internet.

Dans cette thèse, nous analysons les comportements de consommation d'alcool et de nourriture chez les jeunes en Suisses, à partir de données de *crowdsourcing*, obtenues soit par *mobile crowdsensing* (données et méta-données extraits par les capteurs du téléphone), soit sur les réseaux sociaux. Cette étude apporte une triple contribution, relative à l'une ou l'autre des deux sources de données considérées, ou à la combinaison des deux. Tout d'abord, nous analysons des données issues du réseau Instagram, qui révèle qu'il existe plusieurs type de comportements en fonction de ce que les gens mangent et boivent. En particulier, nous mettons en évidence deux types de pratiques en terme de consommation d'alcool (consommation occasionnelle ou excessive) sur les réseaux sociaux. Cette analyse donne un premier aperçu des pratiques de consommation de nourriture et d'alcool à travers la manière dont elles sont partagées en ligne. Ces résultats préliminaires permettent une modélisation de la consommation d'alcool qui est développée plus en profondeur dans le reste de cette thèse. Deuxièmement, nous utilisons des données de *mobile crowdsensing*, manuellement annotées a posteriori, pour identifier les cas de consommation excessive d'alcool et pour comprendre ce qui caractérise les lieux privés (notamment en terme d'ambiance) et les activités (dont la consommation d'alcool) des jeunes lors de soirées en Suisse. Ces résultats montrent comment les données de *mobile crowdsensing* peuvent être utilisées pour comprendre et prévoir la consommation d'alcool ou encore l'ambiance d'un lieu privé. Troisièmement, nous combinons les données issues de *mobile crowdsensing* avec celles des réseaux sociaux pour expliciter les caractéristiques des occasions ou l'utilisateur boit, en fonction du type de boisson (plusieurs catégories de boissons, alcoolisées ou non) ; lesdites caractéristiques sont extraites des images et des données contextuelles, provenant de l'une ou l'autre des sources de données, ou bien des deux conjointement. Ce résultat montre qu'il est possible d'utiliser, seuls ou combinés, les données de *mobile crowdsensing* ou des réseaux sociaux pour étudier les comportements en terme de consommation d'alcool. Ces résultats sont prometteurs et laissent envisager le développement d'un outil d'autocontrôle de la consommation d'alcool, basé sur l'apprentissage automatique. Chacune de ces contributions permet, d'une part, une meilleure compréhension des activités urbaines de consommation d'alcool et de nourriture, et d'autre part, de fournir des méthodes pour déterminer automatiquement des catégories d'aliments/alcool, des pratiques de consommation d'alcool ou encore des ambiances des lieux de consommation privés, à partir de données à grande échelle. Cette thèse, en combinant des techniques avancées d'apprentissage automatique et des résultats récents des sciences sociales, démontre la pertinence d'utiliser une approche pluridisciplinaire pour étudier les comportements sociaux dans les zones urbaines.

Mots-clés : Mobile Crowdsourcing, Mobile Crowdsensing, Réseaux sociaux, Consommation de nourriture, Consommation d'alcool, Consommation occasionnelle, Consommation excessive, Lieux privés, Ambiance, Jeunesse, Vie nocturne, Alcool, Informatique ubiquitaire, Informatique urbaine, Informatique sociale, Instagram, Foursquare

Contents

List of Figures

List of Tables

1 Introduction

1.1 Context and Motivation

Food and beverage consumption, including alcohol drinking, is an important research topic in context of increasing urban areas. According to United Nations Organization (of Economic & Affairs, 2018), 68% of the total world population, or 6.6 billion people, will live in urban areas by 2050, compared to 55% in 2018. This reflects that studying urban areas is one of the pressing and necessary requirements for understanding the potential issues in such mega-urban areas. In line with researching those potential problems, eating and drinking activity is an interesting proxy channel. Food and drink consumption not only is a hand-to-mouth activity, but also relates to many research perspectives, from spatio-temporal configuration to social context. For instance, alcohol consumption can be conceptualized as a central pillar to investigate drinking behaviors of people in relation to various topics, such as co-occurring activities (chatting, eating, or watching TV), social context (friends, colleagues, or partners), drinking places (restaurants or parks), or even ambiance of drinking venues. Some examples of work investigating food and drink consumption are, for instance, drinking motives (E. Kuntsche, Knibbe, Gmel, & Engels, 2005), and food consumption ("Food and Agriculture Pocket Statistic 2016", 2016). However, most existing evidence was collected using paper-and-pencil questionnaires, face-to-face interviews or survey data via SMS (E. Kuntsche & Labhart, 2013), which are methods that potentially have recall and scale limitations. Mobile crowdsourcing appears therefore as a promising alternative to investigate eating and drinking in context.

Mobile crowdsourcing is a term that describes "a powerful approach incorporating human wisdom into mobile computations to solve problems while exploiting the advantages of mobility and context-awareness" (Phuttharak & Loke, 2018). More specifically, in-situ data, geographically collected from mobile crowdsourcing, and reflecting the collective and aggregated experience of targeted people, can be used to tackle specific problems (Santani, Biel, et al., 2016). Based on the definition by (Phuttharak & Loke, 2018), the term of *mobile crowdsourcing* mainly includes two elements: *mobile crowdsensing* and *mobile social network*.

Mobile crowdsensing is defined as "where individuals with sensing and computing devices collectively share data and extract information to measure and map phenomena of common interest" (Ganti, Ye, & Lei, 2011). For instance, researchers (Biel, Martin, Labbe, & Gatica-Perez, 2017; Santani, Biel, et al., 2016) have used mobile crowdsensing and geolocalized human-generated input to collect data about drinking and eating habits. Mobile crowdsensing allows gathering different types of in-situ data, such as photos, sensors, time, and locations. Compared to traditional survey methods (e.g., manual interviews or questionnaires), this leverages complementary or new insights on food and drink consumption. Yet, despite bringing new opportunities, it also comes with possible limitations, including the small number of participants, high incentive cost, and data privacy issues.

The understanding of drinking-eating activities could also rely on the collection of data from users on social networks. In detail, large-scale data created by tens or hundreds of thousands of individuals will possibly reflect aspects of cities, in ways that mobile crowdsensing or traditional methods cannot. Social media data could be curated from social networks, e.g., Twitter or Instagram, where people have freedom to self-present their mood or activities in real-time, by means of posts, pictures, hashtags, geolocalized content, etc. For example, users can share pictures of food and beverage consumption with their friends on Instagram with captions, venue names, and locations. By aggregating the posts' contents, social media research might reveal various patterns at the city and nation levels. However, social media also have some data limitations, such as sparse spatio-temporal resolution, and population bias as well as important privacy implications.

In this dissertation, we aim to use *mobile crowdsourcing*, including both *mobile crowdsensing* with smartphone data collection and *social media* with large-scale data collection, to investigate drinking-eating activities and discover other related contexts, including eating/drinking patterns, drinking practices, and ambiances of private drinking spaces. By combining knowledge from social science (e.g., social psychology, drinking motives, etc.) and computer science (e.g., machine learning, data science, deep learning, etc.), we contribute to the picture of digital cities with drinking-eating activities in context as a use case.

1.2 Goals and Scenarios

The objective of this dissertation is to understand food and drink consumption in context by developing computational inference methods, based on social media and mobile crowdsensing data, to uncover hidden patterns and characteristics of food and drinking habits in Switzerland. For this dissertation, we selected Switzerland, a developed country with multicultural and multilingual influences of its nearby European countries ("Languages Used in Switzerland", 2020; Santani & Gatica-Perez, 2013a) because there is open government data on food and drink consumption in Switzerland ("Alcohol In Figures 2017", 2017; "menuCH: Resultats concernant la consommation alimentaire. Consommation des differents groupes d'aliments", 2017; Office, n.d.; "Results on Food Consumption Survey", 2017). In addition, social media applications, e.g., Instagram, Foursquare, are (or were) also popularly used in Switzerland. Our dissertation adds diversity of data sources to the studied country for both mobile crowdsourcing research (social media and mobile crowdsensing) and food/-drink consumption research. This work also can open scenarios towards systems to give recommendations of to city dwellers and provide insights to city governments.

The outline of this dissertation is the following:

1. **Social Media.**
    - Understand how social media, i.e., Instagram posts, can be used to characterize food and drink consumption compared to available open public statistical data, and to examine different eating events (e.g., breakfast, brunch, lunch, and dinner).
    - Study two aspects of drinking practices, namely, “casual” drinking (identified by the *#drink* hashtag or its variations) and possibly “heavy” drinking (*#drunk* hashtag or its variations), in Switzerland, by using Instagram users’ posts collected over five years.
2. **Mobile Crowdsensing.**
    - Identify characteristics of “heavy” versus “non-heavy” drinking on weekend nights, using data from mobile crowdsensing, including mobile sensors, visual content, and context.
    - Understand the characteristics of private spaces in youth nightlife on weekend nights by investigating physical environment features (including drinking activities) and then using a combination of human annotation and machine learning (computer vision & audio processing) to infer ambiances of home.
3. **Social Media and Mobile Crowdsensing.**
    - Investigate how crowdsensing and social media can be integrated as complementary data sources to improve the understanding of alcohol consumption and to develop their complementarities in machine learning approaches to classify drinking-related attributes.

1.3 Summary of Contributions

This dissertation is primarily based on curated social media data, i.e., Instagram and Foursquare, and mobile crowdsensing data, collected as part of the Youth at Night (Y@N) project (Santani, Biel, et al., 2016), to obtain multivariate data about food and drink consumption. To enrich both sources, we have carried out additional annotations using external annotators, coming from an online crowdsourcing website for annotating social media data and an offline cohort for annotating Y@N data. Figure 1.1 shows the schema of contributions of this dissertation.

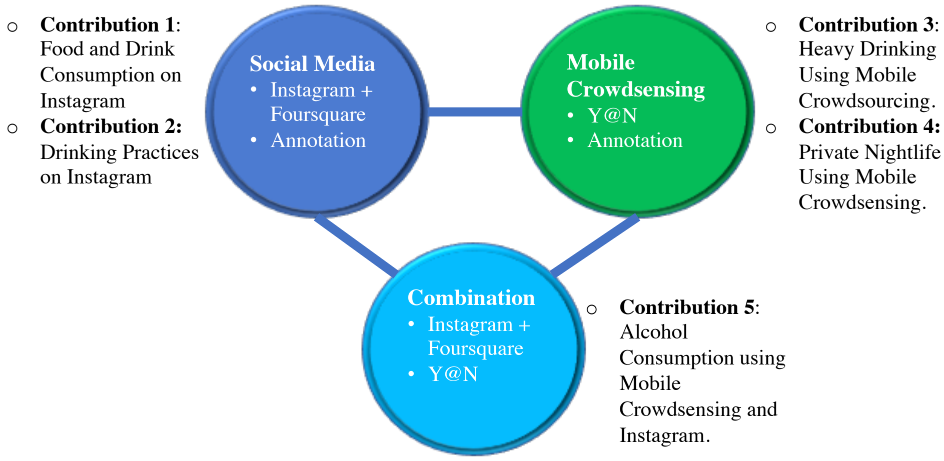


Figure 1.1 – Schema of contributions in each mobile crowdsourcing data used in the dissertation. Each contribution is equivalent to a chapter and a published paper.

Each contribution is equivalent to a chapter and a published paper which will be summarized in the next subsections.

The contributions of this dissertation are the following:

1. **Food and Drink Consumption on Instagram.** We analyzed and inferred food and drink (F&D) consumption posts on Instagram filtered from a pool of 2.8 million posts in Switzerland over a period of five years. Towards this, we define a hashtag-based dictionary of F&D (both food items and food-related concepts), based on their specific popularity within the country of interest. We first extracted F&D posting patterns, in terms of time, location, and types of eating events. In addition, we also mapped and compared the F&D consumption reported on Instagram with open available data (the Swiss Federal Food Safety and Veterinary Office, the Swiss Federal Statistical Office and with menuCH), including consumption comparisons based on food categories. We then used a random forest classifier to infer eating events, namely a two-class task (healthy vs. unhealthy eating) and a six-class task (breakfastt/brunch/lunch/dinner/-coffeetime/teatime) by using users' self-reported labels (hashtags). The results show that healthy vs. unhealthy eating posts can be inferred with 85.8% accuracy (with content features as most relevant), while the six daily eating occasions can be correctly inferred with 61.7% accuracy (with context features as most relevant). The material of this contribution was originally published in (Phan & Gatica-Perez, 2017).

2. **Drinking Practices on Instagram.** We analyzed multimodal signals and drinking practices (casual and heavy drinking) on Instagram by using a 1.7M Instagram dataset existing at least one hashtag, filtered from the same above dataset in Switzerland. Towards this, we manually defined dictionaries of hashtags for casual drinking (e.g., #drink), heavy drinking (e.g., #drunk), drinking occasion, location, social context, and alco-

hol categories. We found that #drunk posts occur more often for nightlife and events locations, and for party occasions, while #drink posts often occurs in food locations like restaurants. Drink-related objects (like beverages, glasses, or tables) are also more present in #drink posts, while #drunk posts have a higher presence of people. In addition, the result of manual annotation on human perception of drinking posts shows that drinking motives (social, coping, enhancement, and conformity) are perceived on #drunk posts with higher scores than #drink posts. Furthermore, 19% of #drunk posts and 6% of #drink posts were labeled as potentially problematic. We then infer #drink and #drunk by using random forest on textual and visual cues. Textual cues results at a classification accuracy of 82.3%, while visual cues achieve an accuracy of 75.0%. The material of this contribution was originally published in (Phan, Muralidhar, & Gatica-Perez, 2019a).

3. **Heavy Drinking Using Mobile Crowdsourcing.** We study heavy drinking nights through smartphone sensing versus active human engagement. Towards this, we used two data sources from mobile sensors, and manual human inputs (in-situ questionnaires, and pictures) collected from 200+ young people, aged 16 - 25 years old (Santani, Biel, et al., 2016). First, to identify participant mobility patterns during weekend nights, we segmented the nights by using location data (GPS) and timestamps. We found that participants tend to change locations more frequently and attend more crowded locations on heavy drinking nights, compared to non-heavy drinking nights. In addition, the time period between the first and the last drink is longer, and the number of drinks reported outside of private places are higher on heavy drinking nights. We then combined sensor data and manual human inputs (context information and images) in a binary classification task of heavy and non-heavy drinking nights. The results showed that phone sensors results in an accuracy of 71% while manual human inputs results in a max accuracy of 72%. These results showed potential considered data features for building up a system for self-monitoring of alcohol consumption among youth. The material of this contribution was originally published in (Phan, Florian, Muralidhar, & Gatica-Perez, 2020).

4. **Private Nightlife Using Mobile Crowdsensing.** We analyze characteristics of personal spaces of young people in the weekend nights. To address to this, we use annotations of physical attributes, social attributes, and ambiance on a 301-video dataset of home spaces collected by Swiss young people on weekend nights. We first found that the most popular activities of young people are eating, drinking, and entertainment. While, in a substantial number of nights, young people were alone, in other occassions, they tend to socialize with friends in noisier atmospheres. We then conducted a correlation analysis among the ambiance attributes that showed two main opposite dimensions, namely places perceived as large, colorful, comfortable, festive, stylish, and unique; and a second category of places perceived as confined, simple, and boring. Dark and bright ambiances did not show significant correlation with the rest of the ambiance attributes. In addition, we also found that using deep learning models applied on the audio and

video tracks to extract automatic features to represent private spaces at the level of objects (1000 objects), scenes (365 scenes), and sounds (527 sounds) is feasible in order to describe general home environments. Finally, we use sounds, objects, and scenes as features to infer ambiances of private spaces (as a regression task). The results showed that object and scene classes can predict six ambiances with R^2 between 0.21 and 0.69: space capacity (*large/spacious vs. cramped/confined*), brightness (*bright/well-lit vs. dark/badly-lit*), *comfortable/cozy*, and *dull/simple*. The material of this contribution was originally published in (Phan, Labhart, & Gatica-Perez, 2019).

5. **Alcohol Consumption using Mobile Crowdsensing and Instagram.** We characterize alcohol consumption through mobile crowdsensing and social media. The two data sources are the Youth@Night crowdsensing data and the Instagram data described earlier. We analyze temporal, spatial, and social contextual patterns of alcohol consumption on weekend nights as captured by both data sources. We classified alcohol categories and alcohol/non-alcohol, using contextual cues and image features from individual and joint data sources. The classification accuracies of alcohol categories and alcohol/non-alcohol were up to 82.3% and 90.0%, respectively, with contextual features being less discriminant than visual features. We found that the combination of two datasets produces no improvement in classification performance, possibly due to the noisier nature of Instagram data. The material of this contribution was originally published in (Phan, Muralidhar, & Gatica-Perez, 2019b).

1.4 Dissertation Outline

The rest of this dissertation is structured as follows.

In Chapter 2, we present the work on understading food and drink (FD) consumption posts on Instagram. In Chapter 3, we investigate the drinking practices namely casual and heavy drinking on Instagram. In Chapter 4, we investigate heavy drinking by using mobile crowdsensing. In Chapter 5, we present our study of nightlife activities of young people at home from mobile crowdsensing. In Chapter 6, we study how mobile crowdsensing and Instagram are used to investigate alcohol consumption. Chapter 7 concludes this dissertation by discussing the limitations of our work and potential directions for future work.

1.5 Publications

This dissertation is a compilation of works published in two international journals and three conference proceedings. Two of these papers were written in collaboration with Dr. S. Muralidhar (Idiap), one was written in collaboration with F. Labhart (Idiap and Addiction Switzerland), and one was written in collaboration with both of them.

Journal Papers

Phan, T.-T., Labhart, F., and Gatica-Perez, D. (2019). My own private nightlife: Understanding youth personal spaces from crowdsourced video. In Proceedings of the ACM on Human-Computer Interaction, CSCW (Austin, TX Nov 9-13, 2019). PACM-HCI 2019. Interact, 3(189).

Phan, T.-T., Muralidhar, S., and Gatica-Perez, D. (2019b). Drinks & crowds: Characterizing alcohol consumption through crowdsensing and social media. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(2), 59.

Conference Papers

Phan, T.-T., and Gatica-Perez, D. (2017). #Healthy #fondue #dinner: Analysis and inference of food and drink consumption patterns on instagram. In Proceedings of the 16th international conference on mobile and ubiquitous multimedia (pp. 327–338). ACM.

Phan, T.-T., Muralidhar, S., and Gatica-Perez, D. (2019a). #Drink or #Drunk: Multimodal signals and drinking practices on Instagram. In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare (Trento, Italy, May 20-23, 2019), (pp. 71–80). ACM. Pervasive Health 2019. Honorable Mention Paper Award

Phan, T.-T., Labhart, F., Muralidhar, S., and Gatica-Perez, D. (2020). Understanding Heavy Drinking at Night through Smartphone Sensing and Active Human Engagement. In Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare (Atlanta, USA, May 20-23, 2020). ACM. Pervasive Health 2020.

Funding Acknowledgments: This thesis was funded by the Swiss National Science Foundation (SNSF) through the Dusk2Dawn Sinergia project and by the EPFL Integrative Food Science and Nutrition Center through the Social Food and People project. We also acknowledge the financial support of a Swiss Government Excellence Scholarship.

2 Food and Drink Consumption on Instagram

Social media generate large-scale data to study food and drink consumption in everyday life. Using Instagram posts in Switzerland over five years, our goal is two-fold. First, we extract key food & drink consumption patterns, through the lenses of a data-driven dictionary of popular items extracted from hashtags, and of a food categorization system used by the Swiss Federal government for national statistics purposes. Patterns related to spatial and temporal distributions of food & drink consumption, and eating events are extracted and compared to official statistics. Second, using the insights from this analysis, we define two eating event classification tasks, including a two-class task (healthy vs. unhealthy) and a six-class task (the three main meals breakfast/lunch/dinner/ plus brunch/coffee/tea). Both tasks use hashtags as labels for supervised learning. We study how content (hashtags and food categories), context (time and location), and social features (likes) can discriminate these eating events. A random forest and a combination of content and context features can classify healthy vs. unhealthy eating posts with 85.8% accuracy, and the six daily eating occasions with 61.7% accuracy. The material of this chapter was originally published in (Phan & Gatica-Perez, 2017).

2.1 Introduction

Studying patterns of food and drink (F&D) consumption has been a research subject in academia, government, and the food industry for years. In the past, researchers and governments largely relied on data collected offline ("Food and Agriculture Pocket Statistic 2016", 2016), like retrospective surveys and phone interviews, which are not easy to obtain and are subject to recall biases and other issues. The ubiquitous use of smartphones and social media has generated new large-scale data sources through which food consumption in everyday life can be studied. This is the case for photo and location sharing services like Instagram (300 million monthly-active users worldwide) and Foursquare (8 billion check-ins worldwide), which opens new, data-driven questions for mobile and ubiquitous multimedia research.

As people share their real-time activities, mood, photos, and location at particular venues, certain aspects of the experience of eating and drinking in daily life can be examined, e.g.

how people get benefits from sharing food information on social media and what challenges they encounter (Chung, Agapie, Schroeder, Mishra, Fogarty, & Munson, 2017a). Instagram data (photos and video with captions, hashtags, user mentions, likes, and comments) can be enriched with Foursquare venue data (venues name, category, and additional metadata) through check-ins, which results in a rich description of the context in which F&D information is posted.

To investigate food-related phenomena on mobile social media, extracting and categorizing food and non-food content from posts is a first necessary step. Clearly, users posting photos at a restaurant do not necessarily imply that such photos depict food; they could be photos of people or views of the venue (Santani & Gatica-Perez, 2015). In addition, food posts often contain text content like hashtags and captions that describe the photo content but also the user location and activities. Hashtags are an essential data source to study the characteristics of food as well as the context and interests of users related to food (Abbar, Mejova, & Weber, 2015; Fried, Surdeanu, Kobourov, Hingle, & Bell, 2014; Kershaw, Rowe, & Stacey, 2014b; Rich, Haddadi, & Hospedales, 2016; Sharma & De Choudhury, 2015).

In public health and nutrition science studies, the national level is often the fundamental target (Berg et al., 2009; Camilleri et al., 2015; Duffey & Popkin, 2011). This is important because, despite globalization, eating still depends on the local context, i.e., the available food items and brands, food stores, and cooking styles are often local or regional. Furthermore, government statistics, which are essential to ground some of the findings from social media analytics, are most often available nationally. Much of the work on social media and food has focused on English-speaking countries (US and UK) (Abbar et al., 2015; Fried et al., 2014; Kershaw et al., 2014b; Sharma & De Choudhury, 2015), or has discussed global trends only scratching the surface with respect to the existing scientific knowledge about eating patterns in specific countries (Rich et al., 2016). We focus our analysis on Switzerland, a multilingual European country for which there is government data on food and eating ("Food and Agriculture Pocket Statistic 2016", 2016; "Results on Food Consumption Survey", 2017; "Swiss Food Composition Database", 2017), and in which Instagram is popular among youth. This chapter thus adds diversity to the countries studied under the social media and food angle.

In this chapter, we investigate two research questions:

RQ1: What food and drink consumption patterns can be extracted from geo-localized Instagram/Foursquare Swiss data, and how do these patterns compare to other sources of national statistical data?

RQ2: How do content and context-related features of Instagram food posts can be used to automatically classify everyday self-reported events, like eating healthy or having lunch?

To obtain answers to these questions, we filter all F&D posts from a pool of 2.8 million Instagram posts in Switzerland over a period of five years. Posts contain images, captions, hashtags, comments, timestamp, venue information, and other metadata. For our analysis, we first define a hashtag-based dictionary of F&D (both food items and food-related concepts), based on their specific popularity of the country under study. We then extract general patterns of F&D posting in terms of time, location, and types of eating events. We deepen the analysis by mapping the F&D dictionary into categories defined by the Swiss Federal Food Safety and Veterinary Office (FFSVO) ("Swiss Food Composition Database", 2017). This allows for a more systematic analysis of the F&D consumption reported on Instagram. Whenever possible, the patterns extracted from Instagram/Foursquare are compared with figures from the Swiss Federal Statistical Office (FSO) ("Food and Agriculture Pocket Statistic 2016", 2016) and with menuCH, the first national survey on food consumption in Switzerland conducted over 2014-2015, which used a combination of pencil-and-paper questionnaires with face-to-face interviews with trained dieticians in ten centers throughout the country ("Results on Food Consumption Survey", 2017). Some of the biases of Instagram data appear evident through these comparisons with traditional instruments for collection of food consumption data.

Finally, using the insights from the descriptive analysis, we define automatic inference tasks for two ways of conceptualizing eating events, namely a two-class task (healthy vs. unhealthy eating) and a six-class task (breakfast/brunch/lunch/dinner/coffeetime/teatime). Both tasks use self-reported labels (in the form of hashtags) for supervised learning. Our goal is to understand how content (hashtags and food categories), context (time and location), and social features (likes) can be informative of different eating events. A random forest approach shows that healthy vs. unhealthy eating posts can be inferred with 85.8% accuracy (with content features as most relevant), while the six daily eating occasions can be correctly inferred with 61.7% accuracy (with context features as most relevant).

The rest of the chapter is structured as follows. Section 2.2 discusses related work. Section 2.3 describes the datasets used in our study. Section 2.4 describes the generation of the hashtag-based F&D dictionary. Section 2.5 presents the analysis of spatio-temporal patterns and eating events. Section 2.6 presents the analysis based on FFSVO food categories. Section 2.7 presents the automatic inference tasks and discusses the experimental results. Section 2.8 concludes this chapter.

2.2 Related Work

In this section, we review work related to food post recognition and analysis of food consumption patterns in social media.

Food Post Recognition. Methods to recognize specific food items in social media involve text and images. On Twitter, several works have analyzed text content like hashtags and key terms to recognize tweets containing F&D (Abbar et al., 2015; Fried et al., 2014; Kershaw et al., 2014b). In other text-based research, works have investigated food items by examining the textual content and distinguishing the presentation of high-calorie, low-nutrient food items vs. fruits and vegetables (Holmberg, Chaplin, Hillman, & Berg, 2016), and by combining text topics with nutritional fact (Kusmierczyk & Nørvåg, 2016). Other works have processed Instagram hashtags to detect canonical names and retrieve nutrition information from online sources (Sharma & De Choudhury, 2015). Hashtags have the advantages of simplicity and direct semantics, but also have limitations due to polysemy, so it might be hard to identify some food items: for example, *orange* can be both food and a color.

The photos available in Instagram can be of great help to complement hashtags. Recent work (Rich et al., 2016) has used a hashtag-driven approach to discover the most popular food categories in a given Instagram dataset, and to learn visual recognizers of food images via supervised learning. This chapter inspires us to extract a data-driven dictionary of F&D items for the specific country under study, which will reflect national trends for food items.

Deep learning has become the preferred choice for learning visual food item classifiers (Akbari Fard, Hadadi, & Tavakoli Targhi, 2016; Kawano & Yanai, 2014), and has been applied on Flickr data for food and non-food classification (Ragusa, Tomaselli, Furnari, Battiato, & Farinella, 2016), Instagram (Rich et al., 2016), and other sources of online data connected to cooking recipes (Chen & Ngo, 2016), and restaurant menus (Meyers et al., 2015), often with the ultimate interest of counting calories by recognizing the contents of a single image and then extracting nutritional content such as calories. In this chapter, we do not aim to recognize food images automatically, but rather to use all other available information (hashtags, place and temporal context, and social features) to discriminate types of eating occasions, as opposed to identify food items or caloric content.

Analysis of Food Consumption Patterns. In the social media literature, various F&D consumption patterns have been studied. Alcohol drinking tweets throughout regions across the UK were tracked in (Kershaw et al., 2014b). A set of 27 health-related statistics of Twitter data at the US county level, including a few eating-related patterns (limited healthy food, fast food, diabetes, obesity) was studied in (Culotta, 2014). Another study of food-related tweets discovered correlations with obesity and diabetes rates at the US county level (Abbar et al., 2015). Using data from Instagram, further connections between food-related hashtags (#foodporn and others) and geographic aggregates of obesity indicators in the US context were studied in (Mejova, Haddadi, Noulas, & Weber, 2015). As we discussed in the next section, data-driven, food-related dictionaries built from popular hashtags (Abbar et al., 2015; Mejova et al., 2015) can have large variations across countries. This highlights the need to understand (and be sensitive to) these national differences.

Other work (Silva, de Melo, Almeida, Musolesi, & Loureiro, 2014) has looked at differences of reported F&D habits through check-ins in Foursquare (Lindqvist, Cranshaw, Wiese, Hong, & Zimmerman, 2011a), filtering check-ins at Food and Nightlife venues and allocating them to three classes: drink, fast food, and slow food. In a larger context, all this work is related to the interest on discovering links between geo-localized social media posts and socioeconomic characteristics of local people (Li, Goodchild, & Xu, 2013). This chapter focuses on extracting patterns of F&D consumption within a particular country as reflected on Instagram, which is not representative of the full population yet corresponds to a young population as we discussed later in this chapter.

Finally, a recent study with Instagram users (Chung, Agapie, Schroeder, Mishra, Fogarty, & Munson, 2017a) interviewed 16 women who posted about food to support themselves and others to maintain healthy eating behaviors, through the use of hashtags used for food tracking like #fooddiary, #foodjournal, and #caloriecounting. Inspired by this qualitative research, in this chapter we studied the feasibility of recognizing self-reported healthy or otherwise eating occasions, indicated by the use of hashtags like #healthyfood, through the use of text, context, and social features.

2.3 Datasets

At the time of data collection, we defined a spatial grid covering Switzerland, and use the Instagram public API to scan venues. From these venues, we downloaded posts. In this chapter, we focus on 1.7M posts with one or more hashtags dated between November 1, 2010 and March 31, 2016. In the rest of this chapter, we call this dataset the *Instagram 1.7M dataset*.

At the time of data collection, Instagram supported matching to Foursquare (4sq) venues, which contain venue category information. After matching, each venue belongs to a category tree declared by 4sq. This tree has many levels, and each level has a list of category nodes. We only focus on the ten top categories declared by 4sq[I]: None (-1), Arts & Entertainment (0), College & University (1), Events (2), Food (3), Nightlife Spots (4), Outdoors & Recreation (5), Professional & Other Places (6), Residence (7), Shop & Services (8), and Travel & Transport (9).

In order to define a hashtag-based Food & Drink vocabulary for the specific Swiss case, we started by examining all Instagram posts generated at venues that specifically match 4sq food venues, i.e., category (3) above. The assumption is that posts generated in food venues probably contain more hashtags involving Food & Drink than other venues. This step was practically important as F&D is only one of the hundreds of topics talked about on Instagram. Following this step, we obtained 3,745 matched food venues between Instagram and 4sq. From these food venues, we harvested a set of 65K Instagram posts. As described in detail in the next Section, this data set was used to define our data-driven, hashtag-based dictionary of F&D items, which consists of 184 items. We call this dataset the *Instagram 65K dataset*.

With the 184-item F&D dictionary, we revisit the larger *Instagram 1.7M dataset* to extract all posts containing at least one F&D item in the dictionary, i.e., harvesting as many images as possible that use our data-driven F&D dictionary. As a result, we obtained 95K posts. We call this dataset the *Instagram 95K dataset*.

Finally, from the *Instagram 95K dataset*, we filtered out those posts with Instagram-4sq matched venues, obtaining 55,342 posts with at least one hashtag in our F&D dictionary. We call this dataset the *Instagram 55K dataset*.

[I] https://developer.foursquare.com/categorytree

Table 2.1 – Instagram datasets used in the rest of this chapter.

Properties	95K Dataset	55K Dataset
# of images	95K	55K
# of total hashtags	1M	576K
# of unique hashtags	136K	80K
# of users	42K	26K
# of venues	22K	10K

In summary, our data is rich in terms of covered period (5 years) and detailed associations between individual posts and venues where they were created. Table 6.1 summarizes the filtered F&D datasets used in the rest of this chapter. Depending on the specific analysis, we will use the corresponding dataset.

2.4 Food & Drink Item Dictionary Design

2.4.1 Data-Driven Dictionary Creation

Hashtags describe photos and their context, and in the case of eating-related posts they are often used to name food elements in pictures. For instance, a photo posted in Gruyere, Switzerland can have as caption: *"Feel #happy in #Gruyere. Have lunch with #cheese, #rosti at #fancy restaurant with #friends"*. In this example, #rosti and #cheese are food hashtags, while references to the location, its social context, and the user's mood are also provided.

We defined a data-driven dictionary of food and drink items as follows. As mentioned in the previous section, we started with the 65K dataset, which contains posts at 4sq food venues, so their hashtags potentially contain names of food and drink items. From this dataset, we extracted the 2,500 most frequent hashtags. Second, we defined a coding system with five hashtag categories: 0 (non-food-or-drink items, such as #geneva, #picoftheday, etc.); 1 (definite food items, such as #fondue, #cheese, etc.); 2 (definite drink items, such as #espresso, #cappucino, etc.); 3 (food-related items, such as #dinner, #lunch, etc.); 4 (drink-related items, such as #coffeeholic, #drunk, etc.). In a third step, the author of this dissertation manually labeled all 2,500 hashtags according to this coding system.

Table 6.2 shows that only 338 (13.5%) of the top 2,500 hashtags from photos taken at 4sq food venues indeed correspond to F&D items. Furthermore, an additional 353 hashtags (14.1%) correspond to F&D-related concepts. Those F&D-related hashtags play an important role as semantic indicators of F&D events, e.g. breakfast, lunch, or healthyeating. In other words, they represent self-reported labels that indicate specific eating events. The remaining 72.4% of hashtags are about other topics. The manual coding process shows that several of these extra hashtags correspond to venues names, feelings of the users, current locations, etc.

The 338 F&D hashtags reflect common social media trends like frequent grammar variations (e.g. singular vs. plural), and the multilingual nature of Switzerland, a country with four

Table 2.2 – Result of manual coding of top 2,500 hashtags at 3,245 food venues in the 65K dataset.

Category	Non-food	Food	Drink	Food-related	Drink-related
# of hashtags	1805	255	83	297	56
Total	1805	338		353	
Percent	72.4%	13.5%		14.1%	

official languages (German, French, Italian, Romansh) and English as lingua franca Santani and Gatica-Perez, 2013b. Hence, we group the original 338 hashtags into 184 F&D items. The 184 items correspond to 142 food items (F), 20 non-alcohol items (NA), and 22 alcohol items (A). The 20 non-alcohol items include beverages such as coffee, tea, and soft drinks, while the 22 alcohol items include beer, wine, cocktail, and so on. The list can be seen in Table 2.4. To assess differences with previous work, we compare our data-driven F&D dictionary with the ones that are publicly available Abbar et al., 2015; Mejova et al., 2015. Interestingly, we found that they match only in 30.0% and 73.7% of elements, respectively, which highlights the interest to study world regions other than the US, which have certain globalized trends but also their own culinary variations.

2.4.2 Visual Validation

Previous work has shown that social media images at F&D venues cover more than just food Santani and Gatica-Perez, 2015. For our dataset, we perform a validation of a sample of the F&D item dictionary to understand how much the corresponding images indeed depict such food items. First, 30 of the 184 items are randomly chosen. Second, we randomly sampled 50 pictures for each of these 30 items. Third, we defined a three-value coding system to indicate F&D item-to-image correspondence: *true* if items definitely correspond to the image content; *false* if items do not correspond at all to any image content; and *unclear* if there is an apparent connection between item and image content but we cannot be sure about it, e.g., sugar is likely part of a cake but we do not see sugar explicitly. The author of this dissertation manually labeled all 1500 images (30 items x 50 pictures) according to this coding system. The results show 1,066 true cases (71.1%), 256 false cases (17.1%) and 178 unclear ones (11.8%). Unsurprisingly, food items normally used as ingredients in prepared dishes (e.g. sugar, pistachio, and mango) have the lowest visual correspondence, as sometimes they can be hardly recognizable as a separate food item. On the other hand, some items have higher visual correspondence, such as tiramisu, sashimi, or tart. In summary, this validation step highlights that our F&D item dictionary, while clearly useful, has a built-in level of uncertainty due to the way in which hashtags are created in Instagram, compared to custom-made methods to collect food labels and eating events, where people are specifically asked to label what they eat Cordeiro, Bales, Cherry, and Fogarty, 2015a; Zepeda and Deal, 2008.

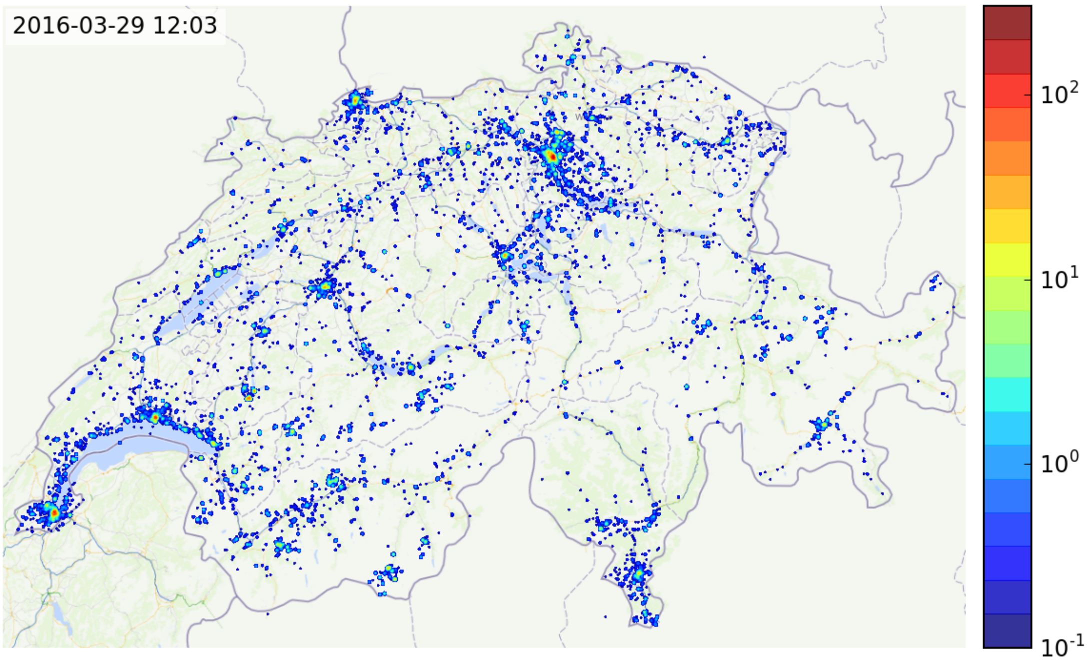


Figure 2.1 – Spatial distribution of the Instagram 95K F&D dataset.

2.5 Food & Drink Pattern Analysis (RQ1)

In this section, we examine the 184 F&D items on the Instagram 95K dataset. In some parts, we will mention if we use the Instagram 55K dataset.

2.5.1 Spatio-Temporal

Spatial Patterns

Figure 2.1 shows the spatial distribution of F&D posts in Switzerland. Unsurprisingly, most posts come from the largest cities (Zurich, Geneva, Basel, Lausanne, Bern) and across the various linguistic regions.

We are also interested in the mean distribution of F&D posts per individual at the ten 4sq venues categories. This accounts for the bias due to frequent contributors. We examine the Instagram 55K dataset. Figure 2.2 shows the corresponding percentages at each venue category. Alcohol is on average most often reported at nightlife venues, which are bars, pubs, and clubs. In 8 categories, alcohol is reported with a percentage above 10% of all posts at such venues. However, the absolute number of posts at each venue category is not evenly distributed. The top 4 categories are food (21,005), travel & transport (9,687), outdoor & recreation (9,575) and nightlife spots (6,026).

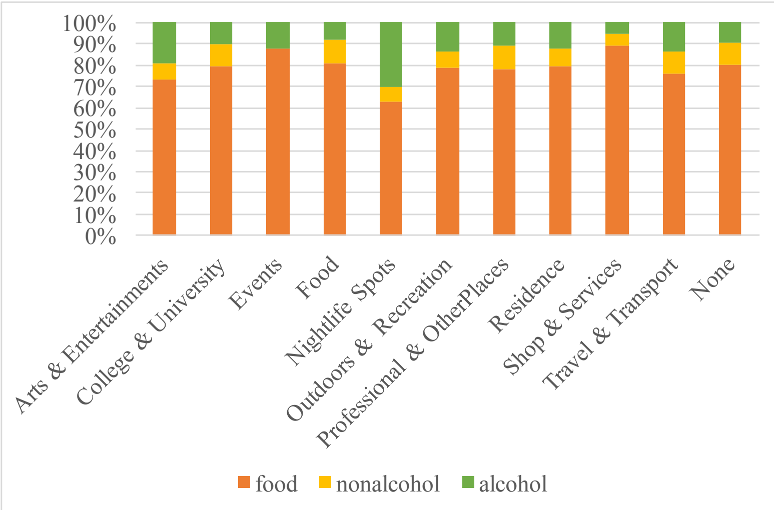


Figure 2.2 – Percentage of food (F), non-alcohol (NA), and alcohol (A) at ten 4sq venue categories in the Instagram 55K dataset.

Temporal Patterns

Based on the GMT time of posts, we estimate the posts' timestamp in Swiss local time. Then, we calculate the hourly distribution of food, alcohol, and nonalcohol for each user. Next, we obtain the mean for all users. Figure 2.3 shows that the distribution of food consumption increases until the evening along with alcohol, and both of them decrease late at night. In the meanwhile, non-alcohol keeps stable during daytime from morning to afternoon. Generally, F&D pictures posted by Instagram users peak around 20:00 - 21:00 and have a local peak around 13:00 - 14:00. This means that the number of posts around dinner time are the highest followed by other eating occasions. Our results are somewhat similar to those reported in (Silva et al., 2014) using 4sq checkins in terms of main peaks, although an exact comparison is not possible. Other papers have not reported temporal patterns (Abbar et al., 2015; Mejova et al., 2015; Rich et al., 2016).

2.5.2 Eating Event Patterns

Literature in nutrition science has investigated how people define meals (Leech, Worsley, Timperio, & McNaughton, 2015). It is known that people label eating events employing situational factors (e.g. where and when eating takes place) (Marshall & Bell, 2003; Wansink, Payne, & Shimizu, 2010), and that how people label their meals affects what they actually eat (Pliner & Zec, 2007). In our case, meals are defined by the users themselves through the use of hashtags. We investigate eating events in this section.

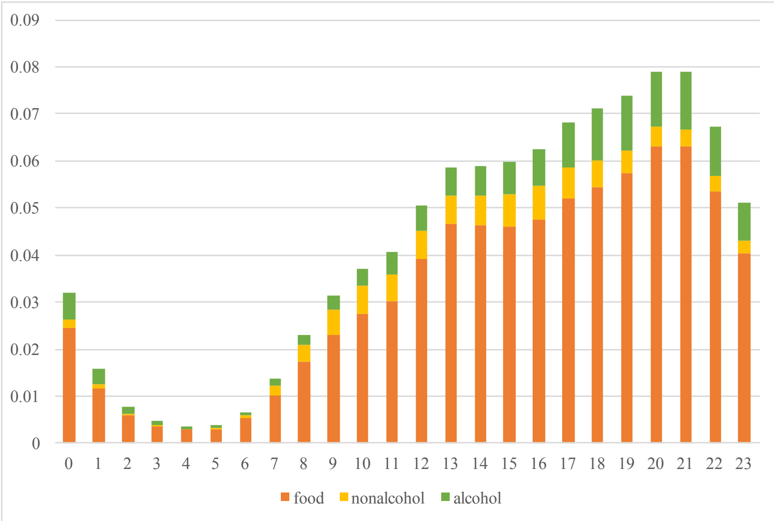


Figure 2.3 – Hourly distribution of F&D items in the 95K dataset.

Daily Meal Analysis

Hourly distribution of daily meals. We turn our attention towards daily meals: breakfast, lunch, dinner, brunch, as well as tea time and coffee time, which are often talked about (i.e., self-reported) on Instagram. From the pool of 353 F&D-related hashtags discussed in Section 4 (table 6.2), several of them refer to daily meals explicitly. The first author manually categorized these hashtags. They are shown in Table 2.3, and correspond to 11,168 posts by 6,125 users. If we only take into account the 3 main meals (9,298 posts), breakfast corresponds to 22.7% of posts, lunch to 30.2%, and dinner to 46.9%. There is uncertainty in a few hashtags (e.g. #diner can correspond to a type of restaurant or a misspelling of dinner). Breakfast posts are 7.5% (absolute) less frequent than lunch posts, and 24.9% (absolute) below dinner posts. As a relative point of comparison, the menuCH study (Bochud, Chatelan, Blanco, & Beer-Borst, 2017) involving over 2,000 individuals, found through surveys that 5.2% of the population never have breakfast, followed by 2.2% who never have lunch, and 0.6% who never have dinner. Regarding the temporal patterns, Figure 2.4.a shows the hourly distribution of the six meals, with peaks for breakfast between 09:00-10:00, lunch around 13:00, brunch between 12:00-13:00, and dinner around 21:00. Tea time peaks around 16:00, and coffee time has two peaks in the mid-afternoon and morning. The hourly patterns for meals are intuitive based on observation of everyday life in Switzerland, although leaning towards the later side of what one could expect. However, note that hashtags related to daily meals can be used outside their expected time (e.g. breakfast can be mentioned in the late afternoon). Users post in this way for a variety of reasons, e.g. lack of internet connection, which makes them post at a later time.

Distribution of daily meals over the week. Figure 2.4.b shows the distribution of self-reported meals over the week. As patterns, breakfast and brunch have an increase on weekends (highest on Sunday). For dinner, Saturday is the day with most posts; while for lunch the most popular day is Friday.

Table 2.3 – Daily meals: defining hashtags and frequency.

Meal	F&D Related Hashtags	Post
Breakfast	breakfast, petitdejeuner, frühstück	2,119
Lunch	lunchtime, lunch, lunchbreak, lunchwithaview, lunchdate, businesslunch, pranzo	2,811
Dinner	dinnertime, diner, dinner, finedining, birthdaydinner, dining, dinnerfortwo, helvtidiner, americandiner, abendessen,	4,368
Brunch	sundaybrunch, brunch, brunchtime	924
Coffee Time	coffeetime, coffetime, coffeebreak	570
Tea Time	teatime, afternoontea	376

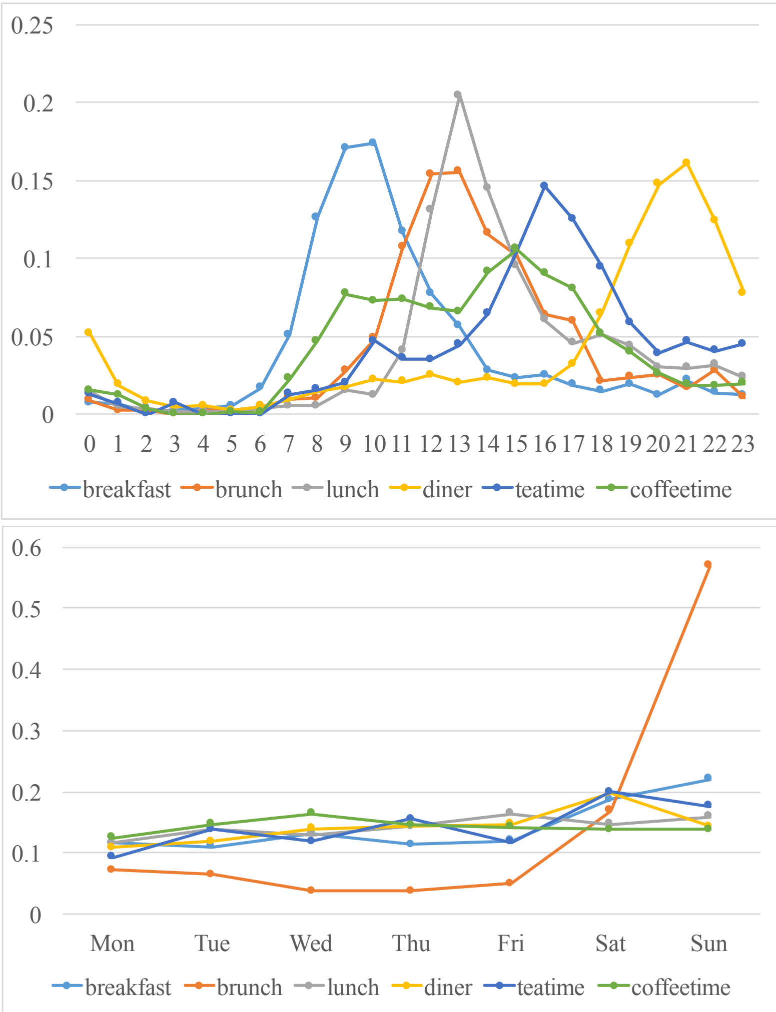
a)
b)


Figure 2.4 – Temporal distribution of daily meals over (a) 24 hours; (b) days of the week.

Co-occurrence of F&D Items in Daily Meals. Based on the co-occurrence (within the same Instagram posts) of the 184 F&D dictionary items and the six daily meals, we plot wordclouds to reveal popular items used in each meal in Figure 2.5. People use coffee, egg, fruits, and croissant for breakfast, while coffee, latte, crepes, and meat for brunch. Lunch has salad, vegetables, coffee, pasta, and burger, while dinner has wine, beef, cheese (including the

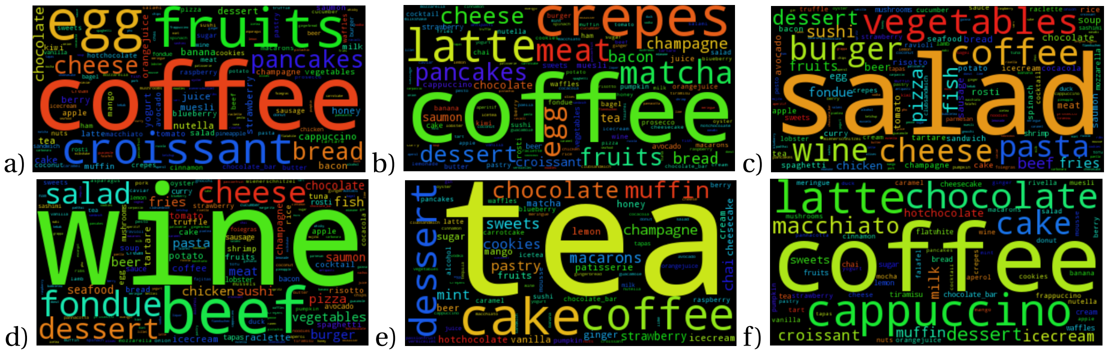


Figure 2.5 – Wordcloud of F&D dictionary items in a) Breakfast. b) Brunch. c) Lunch. d) Dinner. e) Tea Time. f) Coffee Time.

fondue Swiss traditional dish), and dessert. Coffee time and tea time are characterized by the corresponding beverages and cake. Note that while the use of specific F&D items serves as illustration, we will rely on food categories in the next section as a more parsimonious description of consumed food and drinks.

Healthy And Unhealthy Food Post Analysis

As discussed in Section 2, Instagram users comment on their own eating practices, labeling their posts as healthy (or otherwise) and using the app to keep their health goals (Chung, Agapie, Schroeder, Mishra, Fogarty, & Munson, 2017a). We now investigate this phenomenon in our dataset. From the manual coding results in Table 6.2, we use hashtags related to both healthy food and unhealthy food to collect posts that use them. Posts are marked as healthy food if they contain at least one of the following hashtags: healthy, healthyfood, goodfood, eatclean, cleaneating, healthyeating, eathealthy, healthychoices, eatwell, fitfood, and gesund. Posts are marked as unhealthy food if they contain at least one of these hashtags: junkfood, burgerlove, burgerporn, instaburger, fastfood, mcdonalds, burgerking, and holycow (the last three being fast food chains). As a result, we obtain 3,450 healthy food posts and 666 unhealthy food posts. Figure 2.6 shows the occurrences of our F&D dictionary within these posts. Healthy food relates to vegetables, salad, and fruits, while (unsurprisingly) unhealthy food relates to burger, fries, and cocacola. Recent work on Instagram has shown connections between Instagram hashtags and public statistics of obesity in the US (Mejova et al., 2015). Our results suggest that there could be space to study this kind of connection for the Swiss case. As a first step, in Section 7, we investigate an approach to automate the classification of these types of posts.

a) b)

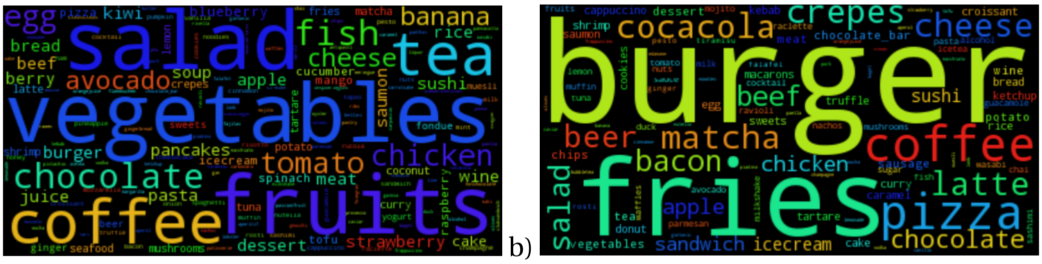


Figure 2.6 – Wordcloud for a) F&D dictionary items for Healty food-related hashtags. b) F&D dictionary items for Unhealthy food-related hashtags.

2.6 Food & Drink FFSVO Category Analysis (RQ1)

2.6.1 Mapping F&D Items to Categories

The 184 F&D item dictionary (FDI) was manually mapped into the 19 F&D categories defined by the Swiss Federal Food Safety and Veterinary Office (FFSVO) ("Swiss Food Composition Database", 2017) (Federal Department of Home Affairs) by the first author, using both local knowledge and web search (e.g. Wikipedia). Table 2.4 shows the distribution of Instagram F&D items over the FFSVO categories. Prepared dishes, sweets, vegetables, and milk & dairy products are the top food categories. Alcohol and non-alcohol drinks are also well represented. In contrast, the special foods FFSVO category (that includes food for gym or stimulants) does not exist in the dictionary. In the rest of this section, we focus on the other 18 F&D categories.

2.6.2 Food & Drink Category Patterns

Overall Distribution

We are interested in how the FFSVO categories are represented on Instagram, and how this compares with existing government data, provided among others by the Swiss Federal Statistical Office (FSO). Similarly to the previous charts, we first calculate the distribution over FFSVO categories for each user in the Insagram 95K dataset. Then, we obtain the mean distribution over all users. Note that these numbers will differ from those shown in Table 2.4 as they are generated from raw post counts. As a separate data source, the Swiss FSO ("Food and Agriculture Pocket Statistic 2016", 2016) provides data on food consumption in kilograms of raw products per head per year. We group food into FFSVO categories, add the consumed amount for each category (in kilograms) and then estimate the distribution over all categories. Table 2.5 shows the distributions for Instagram and government data. Note that an exact comparison is not possible, given the different sources of data (hashtags counts in one case, kilograms of consumed food in the other). However, they allow for a comparison of general trends. Unsurprisingly, Instagram is biased towards certain categories, including alcohol, sweets, and prepared dishes. In contrast, certain categories are not directly represented in the official stats of FSO like prepared dishes and non-alcoholic drinks.

Table 2.4 – Distribution of 184 F&D dictionary items (FDI) (non-normalized over users) over the FFSVO food categories for the Instagram 95K dataset.

Category	% FDI	Samples of food drink items (FDI)
Bread, Flakes & Breakfast Cereals	2.2	bagel, bread, croissant, muesli
Cereal Products, Pulses & Potatoes	6.0	vermicelles, noodles, pasta, penne, spaghetti, carbonara, ramen, gnocchi, ravioli, rice, risotto
Egg	1.0	egg, meringue
Fats & Oils	0.5	butter
Fish Crustaceans	4.9	saumon, tuna, shrimp, lobster, oyster, mussels, caviar, fish, seafood
Fruit	7.1	fruits, avocado, pineapple, strawberry, raspberry, apple, mango, berry, kiwi, blueberry, coconut, banana, passionfruit
Meat & Offal	3.8	beef, ribs, lamb, pork, chicken, duck, meat
Meat Substitute	0.5	tofu
Milk & Dairy Products	8.2	yogurt, cheese, parmesan, mozzarella, raclette, fondue, cream, latte, milk, chai, milkshake, cappuccino, macchiato, frappuccino, flatwhite
Nuts, Seeds & Oleaginous Fruit	1.1	nuts, pistachio
Prepared Dishes	16.3	sandwich, clubsandwich, soup, salad, wienerschnitzel, pho, padthai, paella, pizza, sashimi, sushi, maki, crepes, tapas, carpaccio, nachos, fajitas, guacamole, tartare, curry, escargot, flammkuchen, couscous, antipasti, kebab, falafel, dessert, matcha, piadina, burger
Sausages & Cold Meats	2.7	ham, sausage, bacon, salami, foiegras
Savoury Snacks	1.1	fries, chips
Sweets	13.0	carrotcake, tiramisu, macarons, cake, pancakes, cheesecake, muffin, donut, patisserie, tart, pastry, cookies, waffles, gingerbread, honey, sugar, sweets, chocolate, nutella, caramel, chocolate bars (Lindt, Toblerone, etc.), icecream, mousse, pannacotta
Various	1.6	ketchup, sauce, wasabi
Vegetables	11.4	mushrooms, asparagus, tomato, pumpkin, vegetables, onion, cucumber, spinach, truffle, rucola, mint, edamame, lemon, lime, vanilla, ginger, olives, cinnamon, pesto, potato, rosti
Alcohol Drink	12.0	alcohol, wine, prosecco, beer, tequila, lillet, grappa, aperol, spritz, cocktail, margarita, gintonic, gin, bellini, mojito, champagne,cognac, whisky, liquor, aperitif, vodka, sake
Non-alcohol Drink	6.5	coffee, tea, bubbletea, mocha, hotchocolate, cocacola, rivella, lemonade, gazosa, icetea, juice, orangejuice

Table 2.5 – Distribution of F&D FFSVO categories in the Instagram 95K dataset and official statistics (FSO). For some categories, data is not available (N/A).

FFSVO Category	Instagram (%)	FSO (%)
Bread, Flakes & Breakfast Cereals	0.76	N/A
Cereal Products, Pulses & Potatoes	1.59	16.85
Egg	0.56	1.28
Fats & Oils	0.08	2.91
Fish Crustaceans	2.42	0.93
Fruit	4.48	13.36
Meat & Offal	3.43	5.81
Meat Substitute	0.04	N/A
Milk & Dairy Products	13.47	29.05
Nuts, Seeds & Oleaginous Fruit	0.13	1.05
Prepared Dishes	10.54	N/A
Sausages & Cold Meats	1.23	N/A
Savoury Snacks	0.55	N/A
Special food or Stimulants	N/A	1.45
Sweets	19.20	4.65
Various	0.14	N/A
Vegetables	4.51	11.97
Alcohol Drink	23.85	10.69
Non-alcohol Drink	12.99	N/A

We see that the top 5 categories in government data are milk & diary products, fruit, vegetables, alcohol, and cereal & potatoes & starch. On the other hand, the top 5 categories on Instagram are alcohol, sweets, milk & dairy products, non-alcohol, and prepared food. It seems that what people consume in everyday life is different than what people share on Instagram. The top 5 FFSVO categories on Instagram data account for 80.0% of the probability mass, while for the official stats the top 5 categories account for 81.9%. In daily life, people consume 16.8%, 13.3%, 11.9% of cereal products, fruit, and vegetables, respectively. At the same time, Instagram users post 1.5%, 4.4%, 4.5% for the same three categories. This trend echoes media reports that state that what people post does not accurately reflect what people actually consume. There is a performative aspect to this practice, where users post what they want themselves or others to see (Chung, Agapie, Schroeder, Mishra, Fogarty, & Munson, 2017a). In the Swiss case, it seems to be often sweets (19.2%) and alcohol (23.8%). The relative overabundance of these categories could partly explain why previous studies looking at connections between Instagram posts and health problems like obesity at county levels in the US have been successful at finding significant correlations (Mejova et al., 2015), even though Instagram does not appear to accurately reflect true consumption patterns.

Main Meals and FFSVO categories

We use the Bhattacharyya distributional distance to compute the distance between the hourly distribution of each of the main meals (breakfast, brunch, lunch, and dinner) and the hourly distribution of each of the FFSVO categories (see Figure 2.7). For each meal type and FFSVO category, the lower the distance value, the more similar the temporal pattern is. Based on this

Figure 2.7 – Bhattacharyya distributional distance between 18 F&D categories and four daily meals. The lower the distance value (violet), the more the category is used in a specific meal.

computation, breakfast has close distance to bread, flakes, & breakfast cereal, eggs, fruit, and does not have close distance to the rest of the categories. In contrast, dinner is close to almost all categories except non-alcohol and bread, flakes, & breakfast cereal. Lunch stays somewhat in the middle with respect to many food categories. Brunch has closer distance to fruit, eggs, and non-alcohol.

2.7 Classification of Eating Events (RQ2)

The previous sections showed how Instagram users employ hashtags to mention eating events (e.g. lunch) and what they think about them (e.g. unhealthy). This form of self-report is interesting for two reasons. First, users make use of their own internal definitions to choose the hashtags they attach to their posts; in other words, they decide on their own what they call breakfast or healthy. Second, this bottom-up practice results in labeling eating events in a useful way for supervised learning. In this section, we follow such approach for two eating event classification tasks: a six-class daily meal classifier (breakfast, lunch, brunch, dinner, tea time, coffee time), and a binary healthy vs. unhealthy food classifier. In both cases, we study *content* features (184 F&D items and 17 food categories), *context* features (time of day, day of week, and 4sq venue category), and *social* features (likes and comments), with the goal of understanding their individual and combined discriminative power.

Table 2.6 – Features for classification of eating events.

Feature	Description	Type	Group Feature
hour	Time of the day (in minutes) when the picture is posted	numeric	Context (C)
day	week days when the picture is posted	numeric	Context (C)
vennuecat	4sq venue category where the picture is posted	categorical (10)	Context (C)
likes	number of likes of the picture	numeric	Social (S)
comments	number of comments of the picture	numeric	Social (S)
userInPhotos	number of userInPhotos in the picture	numeric	Social (S)
filter	filter user uses for the picture	categorical (44)	Social (S)
tags	number of tags in the picture	numeric	Picture Caption (P)
captions	number of words of captions in the picture	numeric	Picture Caption (P)
foodtags	number of F&D hashtags mentioned in the picture	numeric	Picture Caption (P)
foodrelatedtags	number of F&D related hashtags mentioned in the picture	numeric	Picture Caption (P)
checkinsCount	number of checkins of 4sq users at venue	numeric	Foursquare (4sq)
usersCount	number of users did check-ins at venue	numeric	Foursquare (4sq)
tipCount	number of tips posted by 4sq users at venue	numeric	Foursquare (4sq)
F	184 F&D binary vector	categorical (184)	F&D items (F)
FC	18 F&D category vector	categorical (18)	F&D Categories (FC)

2.7.1 Classification Method

Random Forest (RF) is a well-known supervised learning method for classification (Liaw, Wiener, et al., 2002). It builds up multiple decision trees, and the output of classification is the mode of the results over all individual trees. RF is able to deal with numerical data and categorical data (typically handled by using factors or one-hot encoding). In the reported experiments, we use one-hot encoding, and set parameters as ntree = 500 and mtry as recommended by (Liaw, Wiener, et al., 2002). We use repeated 10-fold cross validation over 5 times for accuracy evaluation, i.e., 9 data folds are used for training and 1 data fold is used for testing. This procedure is repeated 5 times.

2.7.2 Feature Extraction

Features are extracted from textual data of Instagram posts and 4sq venues. We group them into six groups: F&D items (F), F&D categories (FC), context (C), social (S), picture caption statistics (P), and Foursquare statistics (4sq). Note that we treat the picture caption statistics separate from the actual hashtag content (F), and the venue category (part of context C) separate from the specific venue statistics, so as to have a cleaner representation of content and context. We summarize all features used for classification in Table 6.3.

2.7.3 Classification Results and Discussion

For the two classifications tasks, the datasets are imbalanced. In case of daily meals, discussed in Section 5.2.1, there are 6 classes spreading from 4,368 dinner posts to 376 teatime posts. For

Table 2.7 – Classification results for healthy and unhealthy (N= 1332).

Feature	Acc(%)
Baseline	50.0
F	83.2
FC	79.3
C	71.1
4sq	62.8
P	61.7
S	58.0
F + FC	82.6
F + FC + P	84.7
F + FC + P + C	**85.8**
F + FC + P + C + 4sq	84.9
F + FC + P + C + 4sq + S	85.3

experiments, we decide to keep the original number of posts for the six classes. In the case of healthy and unhealthy posts, discussed in Section 5.2.2, we decide to balance the dataset. We randomly chose 666 healthy posts from the 3,450 available posts such that healthy posts and unhealthy posts are equally represented.

Healthy vs. Unhealthy Classification. In term of individual features, the F&D item feature (F) is the best feature, with 83.2% accuracy. Then, F&D category (FC) is the second best individual feature with 79.3% accuracy. This result is expected as specific food items are related to the corresponding posts being labeled as healthy or not (recall Figure 2.6). The combination of F and FC decreases slightly to 82.6%. In term of feature group combinations, the combination of content and context (F+FC+C+P) provides the highest accuracy with 85.8%. This suggests that time and venue category provide additional discriminative power. The rest of the features do not contribute to further improve classification performance.

Six Daily Meal Classification. We have 11,168 daily meal posts with the following distribution: breakfast (19.0%), brunch (8.3%), dinner (39.1%), lunch (25.2%), coffee time (5.1%), and tea time (3.4%). A majority class baseline (labeling everything as dinner) thus represents an accuracy of 39.1%. In terms of individual feature groups, context (C) is the best feature with 60.7% accuracy. In principle, time is intuitively a good cue to discriminate among some daily meals, although the problem is not trivial given the overlap in time, place, and menu items that many of these meals can have. Furthermore, the F&D item feature (F) is the second best feature with 56.6% accuracy. In term of combinations, a group that integrates content and context (F+C+P) provides the highest accuracy with 61.7%. It is interesting to see that the use of food items can indeed complement the context information albeit slightly. The results also show that the classification task remains open for future performance improvements.

In terms of the most relevant features from the RF for each of the two tasks, the top 10 sub-features for healthy vs. unhealthy are: burger (F&D item), the number of food related tags, prepared dishes (F&D category), salad (F&D item), the number of hashtags, savoury snack (F&D category), pizza (F&D item), food venue category, fries (F&D item), and unknown (venue

Table 2.8 – Classification results for six daily meals (N= 11,168).

Feature	Acc(%)
Baseline on majority class	39.1
F	56.6
FC	54.7
C	60.7
4sq	43.5
P	34.9
S	39.1
C + F	61.6
C + F + P	**61.7**
C + F + P + 4sq	61.3

category). In contrast, the top 10 sub-features for classifying daily meals are: time of the day, tea (F&D item), coffee (F&D item), cappuccino (F&D item), pancakes (F&D item), wine (F&D item), croissant (F&D item), day of the week, fruits (F&D item), and eggs (F&D item). For both classification tasks, the lists of most relevant features seem meaningful.

In summary, we have shown that F&D content and context are indeed informative features for eating event classification in the two tasks we studied. In this sense, by studying the social media setting, this chapter adds to recent work in ubiquitous computing that is examining how to automatically identify eating events from mobile and wearable sensors (Mirtchouk, Merck, & Kleinberg, 2016; Thomaz, Essa, & Abowd, 2015).

2.8 Final Discussion and Conclusion

In this chapter, we set out to study Instagram food and drink posting in a particular national context. We close the chapter by summarizing the answers we found to the two research questions we posed, and by discussing limitations and future directions.

Our first question (RQ1) inquired about the types of food and drink consumption patterns that could be mined from Instagram data generated in Switzerland, and about how these patterns compared to national statistics. We have shown that, starting from a large and longitudinal dataset of Instagram posts and the definition of a data-driven F&D item dictionary, several patterns related to spatial distribution, temporal distributions, food categories, and eating events can be extracted. We found that the F&D item dictionary, around which the whole study was conducted, plays a key role, and is not identical to dictionaries created in other western countries in previous work. This highlights the importance of understanding the national context under which social media studies on food and drink are conducted. We will publish the list of F&D items and categories we collected in our dataset. We also found that broad comparisons with national statistics on the subject are possible but not exact. Despite this limitation, some of the biases of Instagram data appear evident through these comparisons. This points out towards caution when investigating social media data as a proxy for everyday life. At the same time, this does not remove the value of understanding food and drink consumption on Instagram as a specific social media practice.

Our second question (RQ2) inquired whether content and context features could be used to automatically classify eating events. We have shown that a number of features could be defined from the insights obtained from the descriptive analysis, and a random forest approach was able to classify healthy vs. unhealthy posts with 85.8% accuracy, and could also classify six daily eating occasions with 61.7% accuracy, both with a combination of content and context features.

In this chapter, we understand food and drink patterns on social media in Switzerland by mainly using manual-annotated hashtag dictionaries of food and drinks. In the next chapter, we also focus on using hashtag dictionaries but aim to drinking practices.

3 Understanding Drinking Practices on Instagram

The understanding of alcohol consumption patterns, especially those indicating negative drinking behavior, is an important issue to researchers and health policymakers. On social media, people share daily activities, including alcohol consumption, representing these moments through images and text. This chapter, using a five-year dataset from Instagram, analyzes what machine-extracted textual and visual cues reveal about trends of casual drinking (concepts gathered around #drink) and possible negative drinking (concepts gathered around #drunk). Our analysis reveals that #drunk posts occur more frequently in party occasions and nightlife locations, with a higher presence of people, while #drink posts occur at food locations, with a higher presence of drink containers. Manual coding further shows that #drunk posts have a higher chance of being perceived as potentially objectionable. A random forest classifier shows that #drink and #drunk posts can be discriminated with accuracy up to 82.3%. These results have important implications for alcohol research among youth. The material of this chapter was originally published in (Phan, Muralidhar, & Gatica-Perez, 2019a).

3.1 Introduction

Alcohol plays an important role in many cultures including social interaction and health (Shrikhande, 2000). However, people also abuse alcohol, leading to negative social and health outcomes including injuries, violence, accidents, and fatalities (Gmel, Rehm, Kuntsche, et al., 2003; Rehm et al., 2003; Taylor et al., 2010). In this chapter, we examine the drinking behavior of youth using self-representation on social media (Instagram), with the aim to uncover patterns of positive drinking and potential negative drinking.

Drinking culture "refers to the customs and practices associated with the consumption of alcoholic beverages" ("Drinking culture", 2018) and is prevalent in many societies. As part of this, social drinking or responsible drinking, practices correspond to "casual drinking of alcoholic beverages in a social setting without an intent to become intoxicated" ("Drinking culture", 2018). This drinking pattern is in contrast with negative drinking, which involves alcohol intake "far beyond that which is socially acceptable" ("Drinking culture", 2018). In practice, alcohol consumption often begins as a casual, social activity (Johnston, 2010) until excessive consumption leads to negative effects (Gmel et al., 2003). Due to the above mentioned negative consequences, it is important to limit excessive drinking by setting up prevention efforts (Hawks et al., 2002), and to understand the possible transitions between casual drinking and negative drinking.

Researchers in psychology and alcohol consumption have studied drinking behavior from the perspective of drinking motives (Cooper, 1994; Cooper, Russell, Skinner, & Windle, 1992; E. Kuntsche & Kuntsche, 2009) and consequences of drinking (Gmel et al., 2003; Rehm et al., 2003; Taylor et al., 2010). Most of these works collected alcohol consumption data using face-to-face interviews or paper-and-pencil questionnaires (Cooper, 1994; Cooper et al., 1992), which are known to have limitations due to incorrect reporting, e.g. limited recall (Ekholm, 2004; E. Kuntsche & Labhart, 2012). The advent of ubiquitous sensors and smartphones aided researchers to collect larger amounts of data including in-situ responses via SMS on features phone (E. Kuntsche & Labhart, 2013), wearable sensor data (Arnold, Larose, & Agu, 2015; Bae et al., 2017; Kao, Ho, Lin, & Chu, 2012), and hybrid data (including sensor data, and human-generated data like photos, captions, and location) (Y. Chon, Lane, Li, Cha, & Zhao, 2012; Santani, Biel, et al., 2016; Santani et al., 2017; T. Yan, Marzilli, Holmes, Ganesan, & Corner, 2009). However, these methods of data collection have the disadvantage of being intrusive because participants are asked to intentionally report their alcohol intake.

As a complement to such data collection methods, social media from Twitter and Instagram also provide in-situ data related to drinking events, including photos, captions, locations, and other semantic information. By taking advantage of the large scale of users and media posts, researchers can study aggregate alcohol patterns in countries or cities (Hossain et al., 2016; Kershaw, Rowe, & Stacey, 2014a; Pang, Baretto, Kautz, & Luo, 2015; Phan & Gatica-Perez, 2017). It must be noted that social media data also has some limitations, in terms of population bias and the lack of continuous temporal data for individual users.

In this chapter, using data from Instagram collected over five years, we study two aspects of drinking culture (social drinking and negative drinking) inside Switzerland (as a European case study) using users' posts (both textual and visual). We hypothesize that #drunk (or its variations) might be indicative of **possibly negative drinking**; vs. #drink (or its variations) as **casual drinking** (i.e., with a more positive connotation). Specially, we investigate two research questions:

RQ1: Are there significant differences depending on the specific representation used to characterize the act of drinking? More specifically, what are the textual content, visual content, and perceived differences between possibly negative drinking (e.g. posts containing #drunk or related hashtags) and casual drinking (e.g. posts containing #drink)?

RQ2: If such differences exist, how can textual and visual features of Instagram posts be used to automatically classify #drink and #drunk drinking episodes?

The specific contributions of this chapter are the following: (1) We curate a 1.7M Instagram dataset collected over 5 years inside Switzerland to obtain a corpus for #drink and #drunk. This is done using a dictionary of hashtags defined manually for #drink, #drunk, drinking occasion, location, social context, and alcohol categories. (2) We automatically extract textual and visual features. Textual features include drinking occasion, location social contexts, and alcohol categories. Visual features consist of visual autotags, descriptions, and categories. In addition, we obtain crowdsourced perceptions of drinking motivations from #drink and #drunk posts using a popular online platform. (3) An analysis of textual features indicates that #drunk posts occur more often for nightlife and events locations, and for party occasions, while #drink occurs often in food locations like restaurants. (4) Visual analysis shows that photos in #drink posts contain higher presence of drink-related artifacts (like beverages, glasses, or tables), while, #drunk posts have a higher presence of people. (5) The analysis of manually annotated perceived drinking attributes in posts shows that #drunk posts have higher scores for drinking motives (social, coping, enhancement, and conformity, as defined by a validated short questionnaire) compared to #drink. Furthermore, 19% of #drunk posts and 6% of #drink posts were labeled as potentially problematic. (6) We show that using machine learning algorithms, textual and visual cues are able to discriminate #drink and #drunk. Specifically, textual cues achieve a classification accuracy of 82.3%, while visual cues achieve an accuracy of 75.0%.

This chapter has implications not just for alcohol consumption research but also for automatic classification of potential negative drinking social media posts in health tracking applications. This chapter could also be useful for understanding consequences related to mental and physical health through self-representation in social media.

3.2 Related Work

3.2.1 Self-Presentation and Social Media

Self-presentation refers to how people try to manage impressions of themselves with respect to other people or entities (Schlenker, 2012). Goffman explained it as strategic activities of the individual to “convey an impression to others which it is in his interests to convey” (Goffman, 1959). In detail, self-presentation relates to how people try to make up attitudes and reactions of audiences through the presentation of self-relevant information. In the age of the Internet, people have developed strategies for self-presentation in personal web spaces (Jensen Schau & Gilly, 2003) or online dating environments (Ellison, Heino, & Gibbs, 2006). The boom of social networks in the last decade, e.g. Instagram, has opened more channels for people to self-present.

The epitome of self-presentation on Instagram is the selfie (Souza et al., 2015) which is an efficient medium to occupy audiences’ attention. In previous work, (Bakhshi, Shamma, & Gilbert, 2014) shows that photos with faces are 38% more likely to receive likes and 32% more likely to receive comments. Likes and comments are means of responding to other Instagram users' posts. Based on a study of 27K teens and adults in Instagram (Jang, Han, Shih, & Lee, 2015), teens tend to post fewer photos but more hashtags and to get more likes/comments. In addition, teens show a higher sense of self-presentation than adults through posting more selfies and self-expression photos. Interestingly, self-presentation on Instagram can reveal aspects of user's mental health. For instance, (Reece & Danforth, 2017) uses color analysis, metadata components (i.e. image filters), and face detection to diagnose rate of depression. In addition, the relationship between self-presentation behavior in Instagram and narcissism is investigated in (Moon, Lee, Lee, Choi, & Sung, 2016), finding that the higher a user scores in narcissism, the more often they post selfies or update their profile pictures. Hence, self-presentation on Instagram has potential to be informative of other user behavioural patterns. In this chapter, we consider alcohol-tagged posts as one of kind of self-presentation that could reveal drinking practices of Instagram users.

3.2.2 Alcohol Consumption and Social Media

Alcohol consumption is a topic of great interest to researchers and policymakers. The literature has shown that drinking motivations can be broadly categorized into four types: social, enhancement, coping, and conformity (Cooper, 1994; Cox, 1990; Cox & Klinger, 1988). Specifically in the context of young adults and adolescents, it has been shown that alcohol is often used as a stimulant for initiating conversations (social) and/or due to peer-pressure (conformity) (Petraitis, Flay, & Miller, 1995). Previous works also indicate that some young adults indulge in heavy drinking with the motivation of enhancement (E. Kuntsche et al., 2005), leading to alcohol abuse (Kairouz, Gliksman, Demers, & Adlaf, 2002), (McCabe, 2002).

Traditionally, alcohol research has relied on paper-and-pencil questionnaires and/or face-to-face interviews for data collection (Cooper, 1994; Cox, 1990; Cox & Klinger, 1988; Yarosh, 2013). With the advances in ubiquitous computing and the widespread availability of smartphones, the use of mobile technology for data collection in alcohol research has increased. Specifically, literature has shown the validity of data collected using mobile phone applications or wearable devices (Arnold et al., 2015; Bae et al., 2017; Biel et al., 2017; Kao et al., 2012; A. Min, Lee, & Shih, 2018). Participants are asked to use smartphones to report their drinking events (E. Kuntsche & Labhart, 2012; Santani, Biel, et al., 2016). Researchers collect data on drinking through reported drink images, social context, ambiance context, etc. In addition, participants are asked to answer questionnaires or drinking habits, which are used as validation data. For example, (Santani et al., 2017) uses sensor and log data to classify drinking nights with 76.6% accuracy.

As mentioned above, alcohol-tagged posts could reveal drinking practices of Instagram users. (Boyle, Earle, LaBrie, & Ballou, 2017) stated that Instagram was the most likely destination for posts glamorizing college drinking. By asking volunteers to answer surveys about themselves and give access right of their social network accounts, researchers can collect participants' ground truth and their social network data. (Hendriks, Van den Putte, Gebhardt, & Moreno, 2018) asked 192 participants (mean age 20.64, 132 women and 54 men) to give access to their Facebook and/or Instagram profiles and their timelines which are analyzed to discover users' behaviour w.r.t drinking. Similarly, through group interviews with 37 young women aged 16-21, (Atkinson & Sumnall, 2016) explored experiences of drinking and intoxication, the use of social networks in their drinking cultures and the display of drinking practices on social media. All this work has involved manual work, e.g. no machine learning has been used.

In another direction, researchers collected available posts from social networks in a larger scale. In Chapter 2 (Phan & Gatica-Perez, 2017), we investigated the food and drink patterns in Switzerland by using a Instagram corpus. We created a vocabulary of food and drinks by manually annotating the top hashtags on Instagram and classified six daily meals at 61.7% and healthy/unhealthy posts at 85.8%. Similarly, (Abbar et al., 2015) and (Mejova et al., 2015) used hashtags to analyze food/drink patterns including alcohol consumption. These works did not investigate particular drinking patterns like casual drinking or possibly negative drinking, as we do in this chapter. The work in (Pang et al., 2015) investigated posts containing alcohol-related hashtags (textual cues) and Instagram user demographics (visual cues). This work obtained users' demographics from photos via Face++ (an online face processing platform). This work found patterns of alcohol consumption w.r.t time, age, location, and exposure to specific alcohol types, e.g. beer or vodka. We also use hashtags, captions, meta data, and picture contents on Instagram posts as users' self-presentation features, to examine the novel angle of distinguishing between casual and possibly negative drinking behaviours through two categories: *#drink and #drunk*.

3.3 #Drink and #Drunk Datasets

We reuse the dataset collected in Chapter 2. By using the alcohol-related vocabulary from the food & drink dictionary defined in (Phan & Gatica-Perez, 2017), we harvested 34K alcohol drink posts. We call this dataset the *34K dataset*. We hypothesize that posts containing #drink and similar hashtags, co-occurring with explicit alcohol hashtags (e.g. #wine, #beer, etc.), could be evidence for casual drinking, while those posts containing #drunk and similar hashtags could represent possibly negative drinking. In addition, we hypothesize that users posting about these two types of drinking practices also use hashtags related to social relationships (#friend, #family), occasions (#party, #festival), and locations (#bar, #restaurant). From the 34K dataset, we extracted and manually annotated the top 2000 hashtags ranked by their frequency. Based on the meaning of the 2000 hashtags, we manually annotated, grouped, and defined a #drink/#drunk hashtag dictionary, a location hashtag dictionary with ten venue categories, an occasion hashtag dictionary with six categories (*travel_vacation, holiday, sport, party, festival, other*), and a social hashtag dictionary with five categories (*friend, alone, partner_spouse, family, other*). In detail, the #drink/#drunk hashtag dictionary has 20 hashtags for casual drinking (*drinkup, alcoholinfused, alcoholdrink, alcoholicdrink, alcoholicdrinks, alcoholsucks, drink, drinks, drinking, instadrink, drinkingcraft, drinklocal, boire, drank, drinkin, instadrinks,* *saufen, slurp, trinken, drinkporn*) and 9 hashtags for possibly negative drinking (*drunk, wasted, getdrunk, ivresse, boozing, instadrunk, tipsy, drunken, getdrunk*) for a total of 29.

The hashtags are in multiple languages (German, French, English, and Russian) reflecting that Switzerland is a multilingual country and a tourist destination. With the location dictionary, we borrowed the definition from Foursquare venue categories, namely, Arts & Entertainment, College & University, Events, Food, Nightlife Spots, Outdoors & Recreation, Professional & Other Places, Residence, Shop & Services, and Travel & Transport, and Other or None ("Foursquare Venue Categories", 2018). Then, for the #drunk construct, we used the 1.7M dataset to filter out all posts which mentioned at least one of the hashtags in the #drunk hashtag dictionary. For the #drink construct, in order to avoid posts with non-alcoholic drinking, we selected posts with both an explicit alcohol hashtag (e.g. #wine) and at least one of the hashtags in the #drink dictionary. This resulted in a corpus of 2046 #drink and 1323 #drunk posts. We call this corpus *the #Drink/#Drunk corpus.* The corpus is composed of 1451 and 952, users respectively. Similarly to Chapter 2, we also link our original corpus to Foursquare venue categories. In the end, the corpus filtered by occasion, social, context location, alcohol hashtag dictionaries, and 4sq categories is shown in Table 3.1.

In order to validate the above 29 drink/drunk hashtags as being related to positive and negative drinking, we asked 10 trusted annotators to think about the meaning of each hashtag and search it on Instagram first and Google Images later. Then, we asked three questions. Specifically, we asked "Please think about this hashtag and the pictures you have searched and rate it according to its positive connotation". The second question was worded similarly asking for negative connotation. The annotator had to answer this on a 5-point Likert scale (1=strongly disagree and 5=strongly agree). In the last question, we asked "Do you think that if users posted those pictures they could have bad consequences?" with two answers: 1 (Yes) or 0 (No). To assess agreement of annotators, we used Intraclass Correlation Coefficients (ICC(2,k)) as recommended in (Koo & Li, 2016). ICC(2,k) of #drink and #drunk on positive/negative connotations is good (0.60-0.80) as shown in Table 3.1. We observe a difference between the mean for positive (4.06 vs. 2.52) and negative (1.66 vs. 3.22) connotation for #drink and #drunk hashtags. In addition, the mean of bad consequence of #drink hashtags is 0.16 while the value of #drunk hashtags is 0.54. These results suggest that our hypothesis of #drink hashtags as signaling casual, positive drinking, and of #drunk hashtags as signalling more negative drinking is reasonable.

3.4 #Drink and #Drunk Analysis (RQ1)

In this section,we analyze patterns of drinking (expressed by #drink and #drunk) using textual content (hashtags) and visual content (image) from the Instagram posts, as well as other attributes obtained from human perception.

Table 3.1 – Instagram datasets used in this chapter.

Corpus and its derives	#drink	#drunk
#Drink/#Drunk corpus	2,046	1323
#posts containing alcohol hashtags	2,046	453
#posts containing occasion hashtags	881	682
#posts containing social hashtags	594	585
#posts containing location hashtags	608	266
#posts linked to 4sq venues	1351	859
ICC(2,k) for negative connotation	0.80	0.60
Mean of negative connotation	1.66	3.22
ICC(2,k) for positive connotation	0.72	0.66
Mean of positive connotation	4.06	2.52
ICC(2,k) for bad consequence	0.68	0.56
Mean of bad consequence	0.16	0.54

3.4.1 Textual Content

As a first step, we manually extract all hashtags from the posts in our corpus (1,323 #drunk and 2,046 #drink) and cluster them according to (a) alcohol type (wine, beer, spirit & cocktails, and others), (b) occasion (holidays, events, party, travel, sports and festival), (c) social context (friend, family, partner_spouse, alone and other), and (d) location categories (ten 4sq venues) as described in Section 4.3.

The descriptive statistics for each group of hashtags are presented in Table 3.2. We use unpaired T-tests to compare the two groups for each variable. As p-values are known to be not sufficiently informative (Lee, 2016; Yatani, 2016a), we additionally use effect size, namely Cohen's d (CD), and 95% confidence interval (95% CI) to expand our understanding of statistical significance (Lee, 2016; Yatani, 2016a). In Table 3.2, the cases that are statistically significant (by combining the effect size and the CI not including zero) are the number of alcohol hashtags (medium effect size), social hashtags (small effect size) and location hashtags (small effect size). In other words, #drunk posts use fewer alcohol-related hashtags (e.g. wine, beer) and location hashtags but slightly more hashtags related to social interaction. As a basic point of reference to these numbers, we compute descriptive statistics on a disjoint, random sample of 2046 posts from the general Instagram 1.7M-post dataset containing at least one hashtag. Compared to our #drink/#drunk dataset, this random sample of general Instagram posts differs for all variables.

As the next step, we study the distribution of #drink and #drunk posts for the various categories (alcohol type, occasion, and location), see Figure 3.1. We also study the distribution of the social context but it is not shown for space reason. Unsurprisingly, friends are the dominant social context, with most posts that contain a social hashtag refer to friends (0.84 – 0.91). Figure 3.1a shows the distribution of *alcohol category* hashtags. We observe that #drunk posts have a similar frequency for wine, beer, and spirit (0.25 – 0.30), while #drink posts have a substantially higher frequency for *spirit & cocktails* (0.45), indicating that cocktail & spirit are

Table 3.2 – Descriptive statistics of hashtags for random posts (N=2046), #drink posts (N=2046) and #drunk posts (N=1323). (CD denotes Cohen's d, MD denotes mean difference, and CI denotes confidence interval)

# hashtag per post	Random Posts		#Drink		#Drunk		#Drink vs. #Drunk			
	avg	sd	avg	sd	avg	sd	MD	CD [95% CI]	T-Test	p-value
# hashtags	8.40	7.98	15.45	7.51	14.06	7.96	1.39	0.18 [0.11, 0.25]	5.1	1.0e-6
# words	5.42	12.06	4.17	8.76	3.35	5.81	0.82	0.11 [0.04, 0.18]	3.27	0.001
#comments	2.29	6.30	1.43	2.35	1.62	2.35	-0.19	-0.08 [-0.15,-0.01]	-2.3	0.021
# Alcohol hashtags	0.02	0.23	2.08	2.32	0.67	1.61	1.41	0.68 [0.61, 0.75]	20.88	1.0e-90
# Social hashtags	0.08	0.36	0.43	0.86	0.64	0.89	-0.22	-0.25 [-0.32, -0.18]	-7.0	1.0e-11
# Occasion hashtags	0.40	1.11	0.70	1.05	0.73	0.88	-0.03	-0.03 [-0.10, 0.04]	-0.98	0.327
# Location hashtags	0.17	0.45	0.37	0.65	0.25	0.55	0.12	0.20 [0.13, 0.27]	5.9	1.0e-8

popular representations of the #drink concept. Figure 3.1b shows the distribution of *occasion* hashtags. We observe that #drunk posts are often mentioned at parties (0.53), while #drink posts spread over other occasions. Figures 3.1c and 3.1d show the distribution of location categories for #drink and #drunk posts based on 4sq-linked venues and location hashtags, respectively. While the two methods that associate the location with posts produce different distributions for #drink and #drunk, we observe three consistent trends. Specifically, the distribution of #drink is higher than #drunk at food locations and travel & transport, while #drunk is higher at nightlife spots. On the other hand, private places (residence) are poorly represented, even though it is known that alcohol drinking at home is common among young people (Santani, Biel, et al., 2016). This quantitative analysis complements previous results in the literature reporting about the drinking location of young people using qualitative methods (Chatterton & Hollands, 2003; Demant & Landolt, 2014).

In summary, #drink and #drunk posts include references to friends, parties/events, and non-private venues, with #drunk posts being more often associated with parties and nightlife, and #drink posts with food and travel & transport. We also showed some small to medium effects regarding the frequency of use of alcohol, social, and location hashtags.

3.4.2 Visual Content

In this section, we examine the photos of #drink and #drunk posts through visual cues (categories, autotags, and descriptions) using the Computer Vision API from Microsoft Azure (Fang et al., 2015; Guo, Zhang, Hu, He, & Gao, 2016; He, Zhang, Ren, & Sun, 2016).

Visual Categories. Each image has at least one category name from the top 15 parent/child hereditary hierarchies ("Computer Vision API on Microsoft Azure", 2018). Figure 3.2a shows

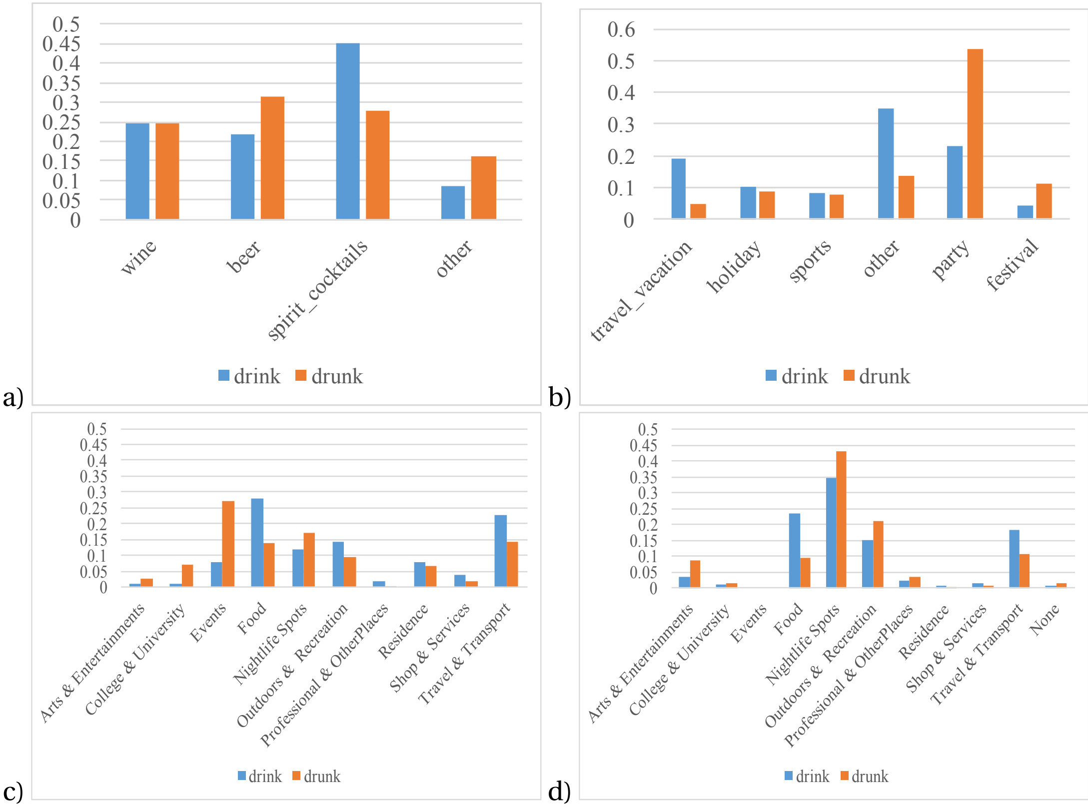


Figure 3.1 – Distribution of #drink and #drunk posts based on: (a) Alcohol categories (b) Occasion (c) Location-related Hashtags (d) Foursquare venues

the distribution of #drink and #drunk posts for these 15 visual categories. We observe that considerably more photos from #drunk posts than from #drink posts have people in them (0.55 vs. 0.21), while photos from #drink posts include drinks, indoor, and food. This differentiated trend suggests that #drunk photos are more often about people, while #drink photos have a wider variety of content, adding drinks and food content to human presence.

Visual Autotags and Descriptions. Autotags correspond to over 2,000 recognizable objects, living beings, scene hierarchy, and actions. These autotags are returned along with a confidence value. We use these autotags for our dataset, with 153 autotags having confidence value higher than 90%. Figure 3.2a shows the distribution of the top vision autotags with confidence above 90%. We again observe that #drunk posts have a higher frequency than #drink posts for autotags related to people, while #drink has a higher frequency for indoor scenes and objects including tables, cup, and bottle. Additionally, visual descriptions correspond to full semantic descriptions like "glass of beer on the table" for each image. Figure 3.2c shows that #drunk photos contain photos with people posing, while #drink photos contain objects related to eating/drink: (*"a glass of beer on a table, a glass of wine"*, etc.). This confirms that #drunk photos are quite often about people, while #drink photos also depict content related to the drinks themselves.

In summary, by using state-of-art computer vision algorithms to characterize the visual content of images, we found that that #drunk posts significantly depict more people in the corresponding photos, while #drink posts relate to both people and eating/drinking activities.

3.4.3 Human Perception of Drinking Posts

In this section, we study a final issue: are #drink and #drunk posts perceived differently by human observers? For this, we analyze how #drink and #drunk posts are perceived based on three issues: drinking motives, context, and problematic issues. We randomly chose a small sample of 200 posts from each category, and manually coded them for three dimensions: (1) 12 questions of an adapted questionnaire on four drinking motives (social, coping, enhancement, conformity) using a 1-to-5 Likert scale (E. Kuntsche & Kuntsche, 2009); (2) two questions about the context of the posts (place and social context); and (3) a binary flag of possibly problematic issues with the post. Annotations were conducted on Crowdflower ("Crowdflower Website", 2018), with each post annotated by 5 raters.

Perceived Drinking Motives. For the four dimensions of drinking motives, the descriptive statistics are shown in Table 3.3. All means are higher for the #drunk group. For #drunk posts, social and conformity motives are the two motives and have a mean equal or higher than 4.0 (i.e., one point above the middle of the Likert scale). Social and enhancement motives are the top two motives, for the #drink group. Similarly to the previous analysis in this section, we calculate effect size and 95% CI for the differences between #drink and #drunk for each drinking motive, as shown in Table 3.3. The differences for the four drinking motives appear to be significant, with one medium effect size (enhancement), three large effect size (social,

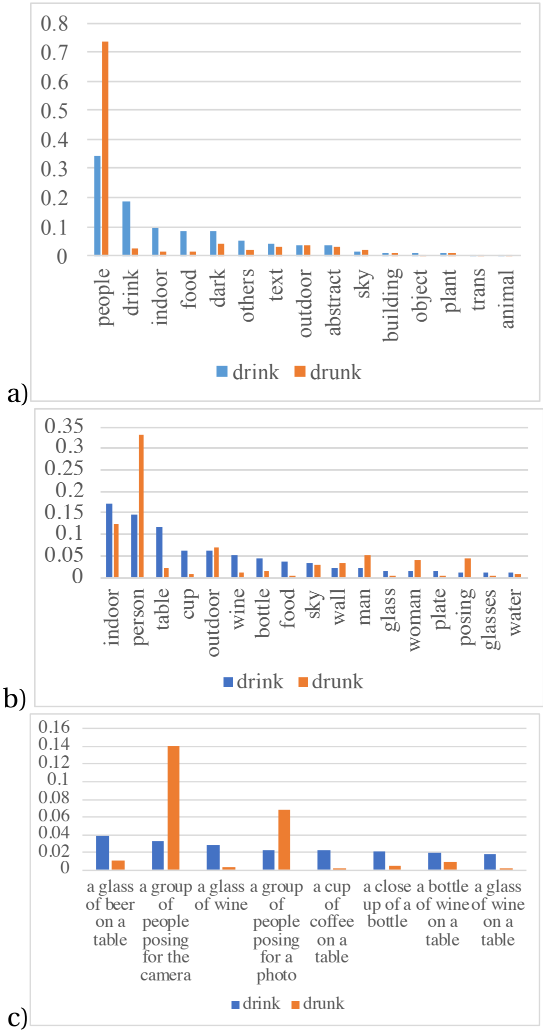


Figure 3.2 – Distribution of (a) visual categories, (b) visual autotags, and (c) descriptions for #drink and #drunk posts.

coping, conformity) and all CIs not including zero.

Perceived Problematic Posts in Context. Among the annotated posts, there are 11 #drink posts and 37 #drunk posts perceived as potentially problematic (for this difference, Fisher's Exact Test score = 3.9, p=8.5e-5). Manual coding of these posts indicated three reasons, related to people apparent behavior; alcoholic drinks combined with one another or with energy drinks; and assumptions that other events like driving could follow. Regarding place context, around 90% of images correspond to either bars, pubs, and nightclubs. Finally, regarding social context, for half of the photos this information was not possible to infer, while for the other half the categories included both partner and friends.

In summary, we obtain three results related to perception of Instagram drinking practices: perceived drinking motives are scored higher for #drunk posts, for which social and conformity

Table 3.3 – Drinking motives descriptive stats and significance for #drink posts (N=200) and #drunk posts (N=200). (CD denotes Cohen's d, MD denotes mean difference, and CI denotes confidence interval)

Motives	#Drink		#Drunk		#Drink vs. #Drunk			
	mean	SD	mean	SD	MD	CD[95% CI]	T-Test	p-value
Social	3.77	0.56	4.29	0.43	-0.52	-1.04[-1.25,-0.83]	-10.38	2.2e-16
Coping	3.0	0.56	3.8	0.70	-0.80	-1.28[-1.49,-1.06]	-12.79	2.2e-16
Enhancement	3.63	0.32	3.82	0.35	-0.18	-0.55[-0.75,-0.35]	-5.46	8.3e-08
Conformity	3.4	0.51	4.0	0.56	-0.65	-1.23[-1.44,-1.01]	-12.25	2.2e-16

motives are the top ones; the perceived drinking context generally agrees with the trends obtained with visual and textual content results; and more #drunk posts were perceived as problematic (19%) compared to #drink posts (6%).

3.5 Classifying #Drink and #Drunk Posts (RQ2)

We now investigate how the textual and visual cues analyzed in the previous section could be used to automatically discriminate between #drink and #drunk posts, defining a binary classification task.

3.5.1 Feature Extraction

In the first step, we extract a number of features from the visual and textual modalities (shown in Table 6.3 with name, description, type and group). Note that all hashtags used to define the #drink and #drunk clusters (see Section 4.3) have been removed so they are not part of the features. Picture caption (PC) contains the numerical count of hashtags in the caption (general hashtags, words, alcohol hashtags, occasion hashtags, social hashtags, alcohol category hashtags), and the categories of social context and occasions. Time (T) corresponds to the hour and weekday when photos are posted. Attention (A) includes the count of comments and likes on posts.

3.5.2 Classification Task

To classify #drink vs. #drunk posts, we use a Random forest (RF) algorithm. In parameter setting, we set ntree=500, mtry as recommended by (Liaw, Wiener, et al., 2002) and GridSearch supported by ("Sklearn with GridSearchCV", 2018). Then, we use 10-fold cross validation over 10 times for accuracy evaluation. For the classification experiments, we use 1,242 data points for the #drunk class, and randomly sample the same number of posts from the #drink class to have a balanced dataset (total: 2,484 data points) with a random baseline of 50%.

Table 3.4 – Features for classification of #drink and #drunk posts with the following group features Picture Caption (PC), Vision Autotag (VA), Fine Vision Autotags (VCA), Vision Colors (VCO), Attention (A), Time (T).

Feature	Description	Type	Group
tagCount	Total number of hashtags	numeric (1)	PC
wordCount	Total number of words in caption	numeric (1)	PC
alcoholTag	Total number of alcohol hashtags	numeric (1)	PC
alcoholCategory	Total number of wine/ beer/ spirit&cocktails hashtags	numeric (3)	PC
socialCount	Total number of social-related hashtags	numeric (1)	PC
occasionCount	Total number of occasion-related hashtags	numeric (1)	PC
socialCategory	5 social hashtags	categories (5)	PC
occasionCategory	6 occasion hashtags	categories (6)	PC
Visual Autotags	Generated autotags for auto description from Azure Vision	categories (520)	VA
Fine Visual Autotags	VA with confidence values higher 90%	categories (153)	VAC
Visual Categories	Generated categories of image from Azure Vision	categories (67)	VCA
commentCount	Total number of comments of picture	numeric (1)	A
likeCount	Total number of likes of picture	numeric (1)	A
hours	Hour when picture is posted (in minutes)	numeric (1)	T
day	Day when picture is posted (weekday)	numeric (1)	T

Table 3.5 – Classification accuracy of #drink/#drunk classifier for visual and textual.

Modality	Feature	Accuracy
Baseline		**50.0**
Visual	Visual Autotags (VA)	**75.0**
	Fine Visual Autotags (VAC)	68.9
	Visual Categories (VCA)	67.4
Textual	Picture Caption (PC)	**82.3**
	Attention (A)	52.5
	Time (T)	60.8

3.5.3 Classification Results and Comparison

Classification results are shown in Table 3.5. With visual features, we see that the visual autotags produce the highest accuracy (75%), followed by fine visual autotags (68.9%), and visual categories (67.4%). With textual features, we see that the picture captions (PC) produce the highest accuracy (82.3%), followed by time (60.8%) and attention (52.5%).

In summary, basic visual and textual features are useful to classify #drink and #drunk. We see that the textual content not only plays a better role in classification (82.3%) but also take less effort as they are directly available in the posts.

3.6 Discussion and implications

In this section, we discuss the results presented in the last two sections and their implications.

Social signals and Instagram drinking posts. The higher use of social-related hashtags (Table 3.1), combined with the machine recognition results (Figure 3.2) and the social motives for drinking (Table 3.3), suggest that #drunk posts might carry a stronger social connotation. In contrast, the higher visual presence of drink-related artifacts (glasses, bottles, etc.) in #drink posts suggests that the conveyed signal, while clearly social, could also serve other objectives related to documenting specific drink items, or moments when people are alone. The finding adds to the literature investigating social media practices related to personal tracking of food and drink (Chung, Agapie, Schroeder, Mishra, Fogarty, & Munson, 2017b; De Choudhury, Kumar, & Weber, 2017). Future qualitative research interviewing Instagram users could complement and deepen these results from the perspective of health-related issues.

Places and Instagram drinking posts. Our analysis also showed some differences with respect to the use of places (Fig. 3.1c,d). We observe more cases of #drunk posts in bars, pubs, and nightclub. This result, though not surprising, has implications for public health and security in and around these places. These also have implications for police authorities and policymakers. In this dataset, private places are underrepresented (see category “residence” in Fig. 3.1c). This is important as research has shown that alcohol is often consumed by youth at home (Andersson et al., 2007; Santani, Biel, et al., 2016). Further research could specifically investigate the issue of alcohol drinking and social media practices in the home context.

Human perception of Instagram drinking posts. Social motives, corresponding to external orientation and positive reward (Cooper, 1994), were perceived as the strongest motive for both #drink and #drunk posts. On the other hand, #drunk posts have higher scores for all motives compared to #drink posts, with large effect sizes (Table 3.3). To our knowledge, this issue has not been previously studied. Finally, on a small random sample, 19% of #drunk posts and 6% of #drink posts were labeled as potentially problematic. While previous survey-based research has found that youth are generally aware of the potential negative consequences of excessive drinking (33.7% of young Europeans believe that over drinking might make them do something they would regret, and 10% believe they might get in trouble with the police) (Andersson et al., 2007), our analysis shows that these kind of photos are clearly in circulation within Instagram. Furthermore, the risks of sharing posts about alcohol drinking take on a new dimension with the current advances in machine recognition and the possible ethical implications.

Machine recognition of Instagram drinking posts and health-drinking tracking applications. Our results showed that #drink and #drunk posts can be discriminated up to 82%. Given the current trends in deep learning, we anticipate that performance could increase if larger datasets were available and more advanced models were applied. Our results suggest that the recognition of different forms of drinking could be automated to some degree in the future. Possible health applications include privacy-sensitive tools for self-reflection and self-management of personal habits; and anonymized contributions towards public health studies. There are, however, important risks associated with the inference of such kind of personal information, especially if used for other purposes that are not directly meant to support or benefit users.

3.7 Conclusion

In this chapter, we investigated the patterns in drinking behavior (both social drinking and potential negative drinking) in Switzerland using Instagram posts. We conclude this chapter by summarizing the findings of the two research questions we posed.

Regarding RQ1, the following patterns were observed. (a) Textual features indicated that a majority of the #drink and #drunk posts include references to friends (over 84% of posts), parties/events, and non-private venues (nightlife spots, outdoors, food, and travel & transport). In particular, #drink posts occurred more often with food and travel & transport, while #drunk posts occurred more often at parties and nightlife. (b) Visual features indicated that #drink posts contain the higher presence of drink-related artifacts (like beverages, glasses, tables). (c) Manual coding indicated #drunk was rated higher for all drinking motivations, with social and conformity being the top ones. The perceived drinking context is in line with trends obtained using the visual and textual content. Furthermore, significantly greater drunk posts were perceived as being problematic (19%) compared to drink posts (6%).

Regarding RQ2, we observed that textual and visual cues in posts are able to discriminate #drink and #drunk, with textual cues showing improved classification accuracy (82.3%) but representing less computational effort due to their direct availability in captions as compared to visual cues. We believe that this chapter has implications not just for alcohol consumption research but also for future self-reflection applications that use social data analysis.

We see that social network has noisy data due to hashtag semantics, e.g., #rose (flower or wine). In addition, users on Instagram could be locals or visitors. This could have an effect on the reliability of the results. We will see how these limitations could be treated by using mobile crowdsensing in the next chapter.

4 Understanding Heavy Drinking Using Mobile Crowdsourcing

Heavy alcohol consumption can lead to many severe consequences. In this chapter, 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. The material of this chapter was originally published in (Phan et al., 2020).

4.1 Introduction

The consumption of alcohol is an important rite of passage for young adults in western societies (Jayne, Holloway, & Valentine, 2006). In late adolescence and early adulthood, heavy drinking is reported to be more common than in any other period of life (Gmel et al., 2003) and the amount of alcohol consumed is found to peak on Friday and Saturday nights (E. Kuntsche & Labhart, 2012; Phan, Muralidhar, & Gatica-Perez, 2019b). Alcohol is consumed mainly due to its disinhibiting and enhancing effects (E. Kuntsche et al., 2005). 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 (Gmel, Gaume, Faouzi, Kulling, & Daeppen, 2008). 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 (Labhart, Livingston, Engels, & Kuntsche, 2018)) 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 (Stanesby, Labhart, Dietze, Wright, & Kuntsche, 2019; Stevely, Holmes, & Meier, 2019). 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) (Stevely et al., 2019), and the interactions between these characteristics (McCarty, 1985). 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) (Stanesby et al., 2019).

The ubiquity of smartphones has opened new opportunities for investigating alcohol consumption patterns using information from 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 (You et al., 2015). 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 (Arnold et al., 2015; Bae et al., 2017; Santani et al., 2017). Bae et. al (Bae et al., 2017) 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 chapter, 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 (Wright et al., 2018). 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 the Youth@Night study, 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.

4.2 Related Work

4.2.1 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 (E. Kuntsche, Kuntsche, Thrul, & Gmel, 2017; Labhart, Livingston, et al., 2018; Van Liempt, Van Aalst, & Schwanen, 2015). 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 (Courtney & Polich, 2009). 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 (Carter, Brandon, & Goldman, 2010; Wechsler & Nelson, 2001). With the development of ecological momentary assessment (EMA), attention has been given to occasion-level factors associated with exceeding heavy drinking thresholds (E. Kuntsche, Otten, & Labhart, 2015; Labhart, Graham, Wells, & Kuntsche, 2013; Labhart, Livingston, et al., 2018). 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 (Labhart, Wells, Graham, & Kuntsche, 2014; Thrul & Kuntsche, 2015). Such evidence are important to develop in-the-event prevention interventions (Kazemi et al., 2017) 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. (Clapp, Madden, Mooney, & Dahlquist, 2017) 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 (Bae, Chung, Ferreira, Dey, & Suffoletto, 2018). Regarding cues from the extended social and physical context, our previous work (Phan, Muralidhar, & Gatica-Perez, 2019a) 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.

4.2.2 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 (Papliatseyeu & Mayora, 2009; Su, Tong, & Ji, 2014), points of interest detection (J. Chon & Cha, 2011; Montoliu, Blom, & Gatica-Perez, 2013), travelling behavior (Montoliu et al., 2013; Vich, Marquet, & Miralles-Guasch, 2017), as well as links between behavior patterns at certain location and particular demographics (Kelly, Smyth, & Caulfield, 2013). Accelerometer has been used for physical activity recognition (Anderson et al., 2007; Bao & Intille, 2004; Maurer, Smailagic, Siewiorek, & Deisher, 2006; Ravi, Dandekar, Mysore, & Littman, 2005), including classifying human movement in real-time (Karantonis, Narayanan, Mathie, Lovell, & Celler, 2006), detecting daily movements (M. Mathie, Celler, Lovell, & Coster, 2004), static/dynamic activities (Merryn J Mathie et al., 2004), and monitoring long term human movements at home (Merryn J

Mathie et al., 2004). Bluetooth and WiFi sensors have been used to extract social and urban context, e.g., estimating human density around the user (Do, Blom, & Gatica-Perez, 2011) by considering number of Bluetooth device as a proxy for social context (Do & Gatica-Perez, 2011), 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 (Christensen et al., 2016; J.-K. Min et al., 2014).

In the context of alcohol consumption, Santani et al. (Santani, Biel, et al., 2016) 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% (Santani et al., 2017). More recently, Bae et al. (Bae et al., 2017) 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.

4.3 Dataset and Task Definition

The dataset for the present work comes from the Youth@Night study, as part of a project aimed at understanding young peoples' nightlife behavior in Switzerland using a custom-developed smartphone application (Labhart et al., 2019b; Santani, Biel, et al., 2016). This section presents an overview of the study design, the data collection procedure, and the definition of the target task.

4.3.1 Study Design

Santani et al. developed a study of nightlife behavior of young adults on weekend nights (Santani, Biel, et al., 2016) . Towards this, they 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 (Labhart et al., 2017) 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 (Labhart et al., 2019b).

4.3.2 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 (Labhart et al., 2019b) 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 (Santani, Biel, et al., 2016) 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).

4.3.3 Heavy Drinking Definition

The most common definition of heavy drinking in alcohol research, according to (Courtney & Polich, 2009; E. Kuntsche et al., 2017), 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 (E. Kuntsche et al., 2017). This percentage corresponds to the legal limit in the US to drive a vehicle ("Centers for Disease Control and Prevention", 2020). 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 (Bae et al., 2017) 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.

4.4 Features and Labels

Table 4.1 outlines the various features that are extracted from the mobile sensor logs, questionnaires, and pictures recorded by the app. In this section, we describe the features and the methodology used to extract them.

4.4.1 Mobile Sensor Features

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. (Montoliu et al., 2013) to distinguish between periods when a user stays at a specific location and periods of transition between locations. Figure 4.1 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 (Bermingham & Lee, 2019; Montoliu et al., 2013).

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. (Do & Gatica-Perez, 2013). 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 4.1).

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 (Bao & Intille, 2004; Brezmes, Gorricho, & Cotrina, 2009; Maurer et al., 2006; Ravi et al., 2005). Others have aggregated the values of three axes into one value, e.g., signal-magnitude area (SMA) (Karantonis et al., 2006; Merryn J Mathie et al., 2004; M. Mathie et al., 2004; Veltink, Bussmann, De Vries, Martens, Van Lummel, et al., 1996), or average resulting acceleration (Anderson et al., 2007) (average of

Table 4.1 – 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)

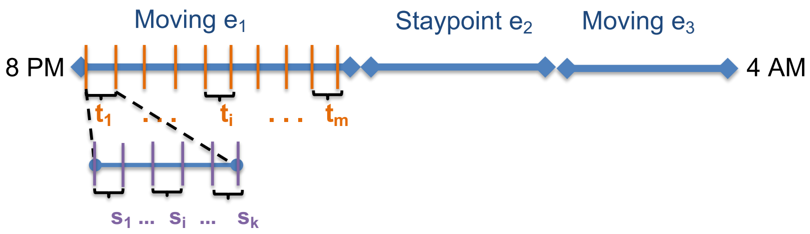


Figure 4.1 – Aggregation process of single data points into time blocks and into episodes (*staypoint* or *moving*). e = episode, ti = time block within an episode during which data are collected with m = total number of blocks, si = single data point with k = total number of data points per block.

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*).

Bluetooth and WiFi

Bluetooth and WiFi density can be used as a proxy to capture human mobility (Sapiezynski, Stopczynski, Gatej, & Lehmann, 2015), social context (Do et al., 2011; Eagle & Pentland, 2005; Kjærgaard & Nurmi, 2012; Z. Yan, Yang, & Tapia, 2013), and person-person proximity (Sapiezynski, Stopczynski, Wind, Leskovec, & Lehmann, 2017). For example, if there are a large number of distinct WiFi hotspots and Bluetooth devices at some points 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.

Applications

Application use might differ depending on the location or the activity (Xu et al., 2011). For example, Likamwa et al. (LiKamWa, Liu, Lane, & Zhong, 2013) developed a mood sensor using smartphone application usage, which could further be linked to drinking behaviors (Rohsenow, 1982; Townshend & Duka, 2005). 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.

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 (J.-K. Min et al., 2014). 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.

4.4.2 Context Features

From 8PM to 4AM, the participants were required to complete series of small questionnaires describing the social, spatial, and temporal context for every new drink (alcohol or not) as well as a photo (Labhart et al., 2019b). In this subsection, we describe the temporal and spatial context of drinking occasions.

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.

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

4.4.3 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 (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016) 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 (Zhou, Lapedriza, Khosla, Oliva, & Torralba, 2018) 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.

4.4.4 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) (Gmel & Rehm, 2004). 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.

4.5 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 4.2). Cohen’s D and 95% confidence interval (95% CI) are reported to better understand the effect sizes (Lee, 2016; Yatani, 2016b) and qualifies the magnitude of the difference between the two groups as follows: 0.2 = “small”, 0.5 = “medium” and 0.8 = “large”(Cohen, 2013). As seen in Table 4.2, 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 4.3). 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 (Labhart et al., 2014). 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 4.3), 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 4.2). 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 4.3 and Figure 4.3). 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.

Table 4.2 – 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)

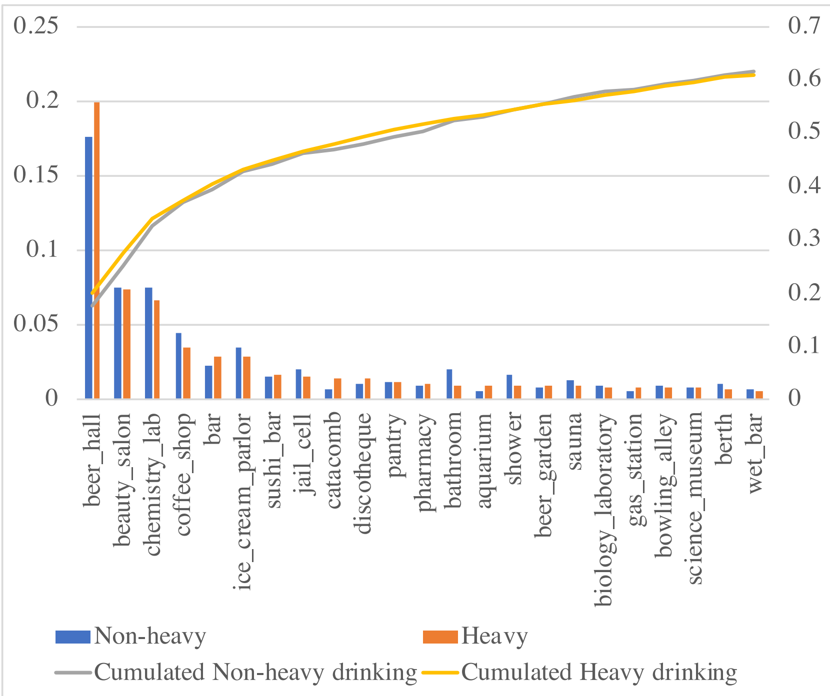


Figure 4.2 – The top 20 detected scenes of all images on heavy and non-heavy drinking nights. The unit of the y-axis is average scores which are ranked with respect to their value on the Y@N data.

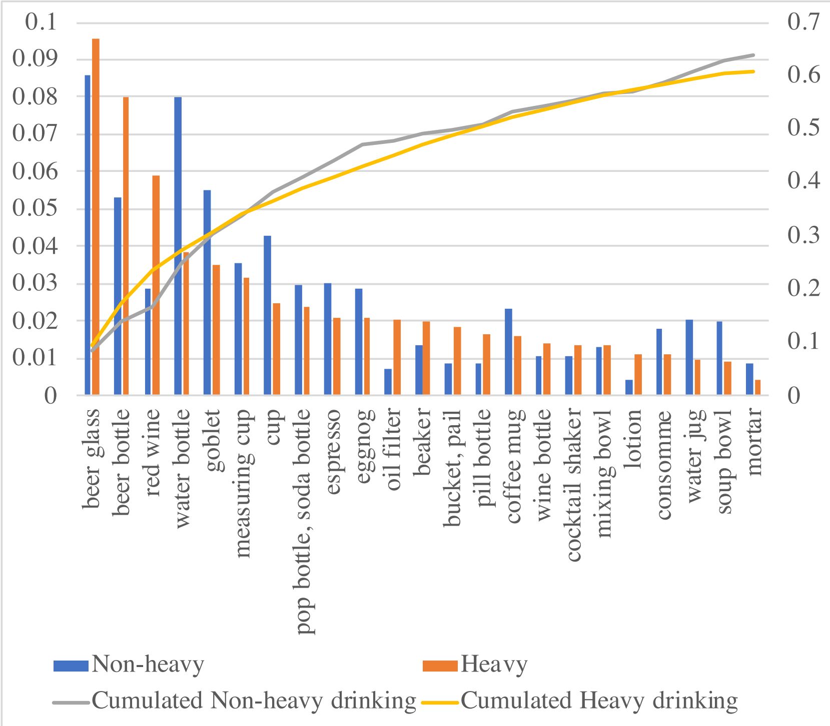


Figure 4.3 – The top 20 detected objects of all images on heavy and non-heavy drinking nights. The unit of the y-axis is average scores which are ranked with respect to their value on the Y@N data.

4.6 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 4.4 to train our models.

4.6.1 Models and Performance Measures

We use two classification methodologies: Random Forest (RF) and Support Vector Machines (SVM), using a Python package Scikit-learn ("Sklearn with GridSearchCV", 2018). 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).

4.6.2 Classification Results

The classification results are presented in Table 4.4. From the table, we observe that RF performs better than SVM for all feature groups.

Table 4.3 – 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)

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.

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 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 4.4 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

Table 4.4 – 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

human engagement (surveys or images).

4.7 Discussion and implications

In this work, using the Youth@Night dataset (Santani et al., 2017), 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 drinks consumed is a function of the duration of the drinking occasion (Labhart et al., 2014; Thrul & Kuntsche, 2015).
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

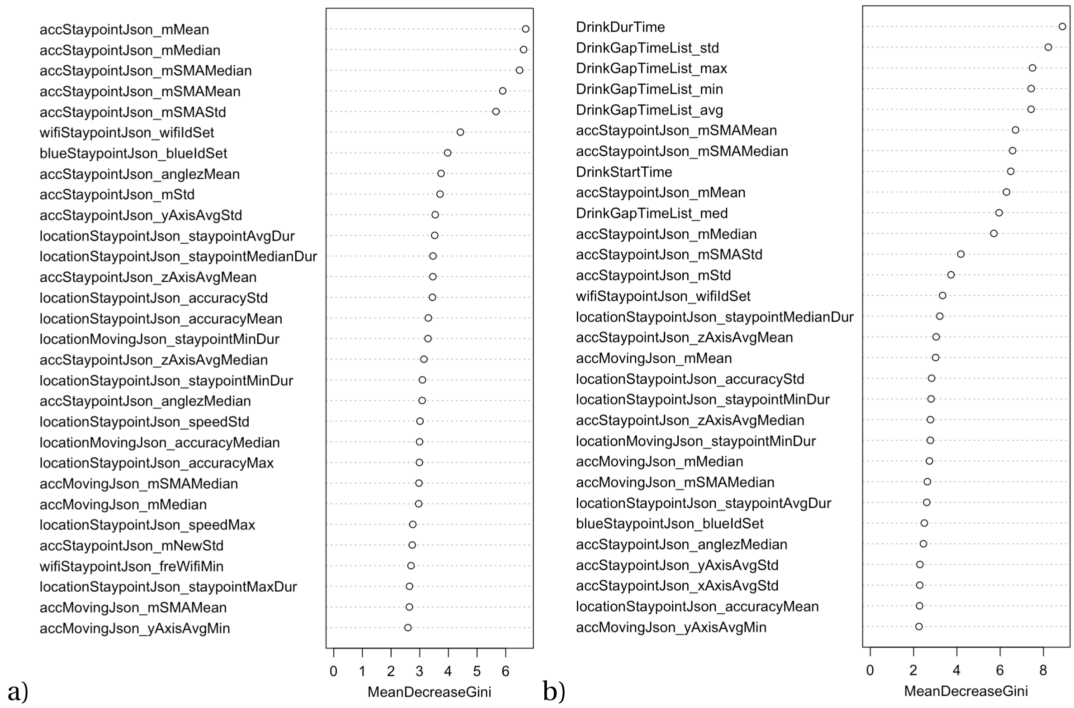


Figure 4.4 – Top important features detected by using a) Sensor (S), and b) Sensor (S) + Time (T) when using RF. A higher Mean Decrease in Gini indicates higher variable importance.

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 (Stanesby et al., 2019).

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 districts after starting drinking at home (also known as "predrinking" (Hughes, Anderson, Morleo, & Bellis, 2008; Labhart et al., 2013)). 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 (Dietze, Livingston, Callinan, & Room, 2014). It should be noted, though, that participants were recruited in the nightlife districts of two cities (Labhart et al., 2017), 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 (Curtis et al., 2019).

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 (e.g. (Wright et al., 2018)) 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 believe 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. (Bae et al., 2018), 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 (Bae et al., 2018), 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 (Wright et al., 2018). One reason for this is that heavy drinkers tend to show increased attention biases towards alcohol-related stimuli and are resistant to nudging (Weafer & Fillmore, 2013). 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 (Labhart, Engels, & Kuntsche, 2018) 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.

4.8 Conclusion

In this chapter, 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.

This chapter uses mobile crowdsensing data to understand and infer heavy and non-heavy drinking nights. Also in the context of using mobile crowdsensing, we will use videos recorded at home spaces by young people at night to understand physical and ambiance attributes of home spaces as well as participants' activities (including drinking activity) in the next chapter.

5 Understanding Nightlife at Home using Mobile Crowdsourcing

Private nightlife environments of young people are likely characterized by their physical attributes, particular ambiance, and activities, but relatively little is known about it from social media studies. For instance, recent work has documented ambiance and physical characteristics of homes using pictures from Airbnb, but questions remain on whether this kind of curated data reliably represents everyday life situations. To describe the physical and ambiance features of homes of youth using manual annotations and machine-extracted features, we used a unique dataset of 301 crowdsourced videos of home environments recorded in-situ by young people on weekend nights. Agreement among five independent annotators was high for most studied variables. Results of the annotation task revealed various patterns of youth home spaces, such as the type of room attended (e.g., living room and bedroom), the number and gender of friends present, and the type of ongoing activities (e.g., watching TV alone; or drinking, chatting and eating in the presence of others.) Then, object and scene visual features of places, extracted via deep learning, were found to correlate with ambiances, while sound features did not. Finally, the results of a regression task for inferring ambiances from those features showed that six of the ambiance categories can be inferred with R^2 in the [0.21, 0.69] range. The work in this chapter contributes to the understanding of home environments represented through digital media. The material of this chapter was originally published in (Phan, Labhart, & Gatica-Perez, 2019).

5.1 Introduction

The home environment is an important subject of study in several social sciences including psychology and geography, as well as architecture and design, and more recently computing (Abbott-Chapman & Robertson, 1999, 2001). Private spaces at home include common living spaces in households (living room and kitchen), but also individual personal rooms (bedrooms). It is known that young people appropriate their private spaces and manifest aspects of their personal beliefs and traits in this way (Lincoln, 2014, 2015; Lincoln & Robards, 2016).

One important feature of place (private and otherwise) is ambiance. This is defined as "the character of a place or the quality it seems to have" (Dictionary, 2018) and used for both indoor environments (Benkhedda, Santani, & Gatica-Perez, 2017; Can, Benkhedda, & Gatica-Perez, 2018; Nguyen, Ruiz-Correa, Mast, & Gatica-Perez, 2018; Redi, Quercia, Graham, & Gosling, 2015; Santani, Hu, & Gatica-Perez, 2016) and outdoor environments (Ordonez & Berg, 2014; Porzi, Rota Bulò, Lepri, & Ricci, 2015). In the context of commercial spaces, ambiance plays an important role in customer behavior related to shopping (Kotler, 1973), food choices (Bell, Meiselman, Pierson, & Reeve, 1994; Stroebele & De Castro, 2004) or hotel experiences (Countryman & Jang, 2006). Regarding the ambiance of personal spaces, previous work has shown that ambiance mediates other factors, like gender (Rheingold & Cook, 1975) and personality (Samuel D Gosling, Ko, Mannarelli, & Morris, 2002), on choices made on physical and environmental characteristics. Understanding home spaces is a relevant domain that has not been fully studied in social computing. A better understanding of physical and social attributes and ambiance of personal spaces could have various implications for social computing research as a part of an agenda on living spaces and well-being. For example, homes can be reconfigured by their inhabitants with respect to decoration, spatial organization of furniture, light, and music, thus inducing more appropriate ambiances for certain activities and social interactions at home (e.g. a romantic dinner vs. an end-of-year party). Designing systems that both recognize physical and social attributes and support users to reconfigure their home spaces based on their specific goals is a relevant application. This could integrate the many perspectives existing in psychology, architecture, human geography, and public health, with the availability of environmental and mobile sensors and social media, and is a particularly interesting angle to understand and support youth practices. Furthermore, many traditional studies have collected information of personal spaces by using paper-and-pencil questionnaires and interviews. The potential of collecting in-situ information of home spaces (physical and social attributes of the environment) through technological means could add to the existing set of research tools.

Recent work on recognition of indoor ambiance (Nguyen et al., 2018; Santani, Hu, & Gatica-Perez, 2016) has used still images from online social systems like Foursquare (Santani, Hu, & Gatica-Perez, 2016) or Airbnb (Nguyen et al., 2018). Social psychologists have also investigated impression formation on home environments (Samuel D Gosling et al., 2002; Graham, Gosling, & Travis, 2015). Yet, gaps in the existing body of work emerge as most previous work has been conducted using social media data that either (1) might lack diversity in the representations of private residences (Cramer, Rost, & Holmquist, 2011; Phan & Gatica-Perez, 2017), as they are naturally focused on outdoor and commercial spaces; or (2) might be beautified, e.g. on Airbnb and similar sites, due to the intrinsic motivations to create and share such images (Nguyen et al., 2018). To investigate home ambiance in a naturalistic setting, a different research direction could use crowdsourcing to collect in-situ videos of the personal environments inhabited by volunteers, which will yield vivid image and sound information, while reducing certain motivations of the video makers (e.g. performative or commercial) that could otherwise affect the generated content. Compared to previous work, this chapter uses crowdsourced video data of young people in their personal spaces during weekend nights. This provides a new view of youth nightlife activities in the private sphere, which enriches the kind of information that traditional methods in the social sciences can provide, in terms of temporal and scene granularity as well as scale.

The overall aim of this chapter is to understand the characteristics of private spaces in youth nightlife in the weekend by investigating physical environment features and ambiances of home. Specifically, using a combination of human annotation and machine learning (computer vision & audio processing), we address the following research questions:

RQ1: Given crowdsourced videos recorded at home spaces by young people at night, what patterns of physical and ambiance attributes of youth home spaces can be revealed by manual coding of videos using external annotators and machine-extracted features?

RQ2: Given machine-extracted features of videos at youth home spaces, can these features infer the perceived ambiance of such spaces?

To address these questions, we use a crowdsourced dataset about nightlife, involving the participation of 241 young people, aged 16 - 25 years old, in the two main Swiss hubs for nightlife (Zurich and Lausanne) (Santani, Biel, et al., 2016). To obtain detailed insights on the locations attended and their related ambiance, participants were requested to record panoramic 10-second video clips of their environment at the start of the night and whenever they changed locations. In total, 841 videos were collected on 10 weekend nights. In this dataset, a significant portion of locations documented were private places (Santani, Biel, et al., 2016), which provides a unique snapshot of how weekend nights are experienced by youth in their private environments. We design and implement an annotation task by asking external annotators to watch video clips. To build up this questionnaire, we have adopted several dimensions from related work (Samuel D Gosling, Craik, Martin, & Pryor, 2005; Graham & Gosling, 2011; Nguyen et al., 2018; Santani & Gatica-Perez, 2015). As a result, we generated a labeled dataset of 301 video clips in personal environments which contain richer in-situ information than what is often captured in questionnaires or surveys used in previous work, and manually annotated attributes of private spaces for our analysis.

This chapter has the following contributions:

(1) To address RQ1, we use a 301-video dataset of home spaces collected by Swiss young people on weekend nights. Our dataset contains video and audio files. A set of five independent raters annotated all videos with a rich set of questions, including physical attributes, social attributes, and ambiance. The results show that the video dataset can be consistently assessed by external raters, with at least moderate agreement, and in many cases with good or excellent agreement. Detailed analyses of the annotations produce several relevant results. First, we show that activities like eating, drinking, and entertainment (chatting, watching TV, and using digital portable devices) are all popular among young people, but with fluctuations over the night period. Second, we found a substantial number of cases where young people are alone and where home place loudness (chatter and music) is low. For those cases in which people

socialize, we observed a same-sex trend between study participants and their companions. Third, we performed a correlation analysis among the ambiance attributes that showed two main opposite dimensions, namely places perceived as large, colorful, comfortable, festive, stylish, and unique; and a second category of places perceived as confined, simple, and boring. Dark and bright ambiances did not show significant correlation with the rest of the ambiance attributes. Finally, we use deep learning models applied on the audio and video tracks to extract automatic features to represent private spaces at the level of objects, scenes, and sounds; the results indicate the feasibility of using deep learning to produce generic semantic descriptions of home environments, although in several cases interpretation remains an issue.

(2) To address RQ2, we find that several of the 1000-object, 365-scene, and 527-sound classes used in this chapter have a particular correlation with specific ambiances. Finally, we use a machine learning pipeline to automatically infer ambiances of private spaces (as a regression task) using features informative of sounds, objects, and scenes. The results show that object and scene classes can predict six ambiances with R^2 between 0.21 and 0.69: space capacity (*large/spacious vs. cramped/confined*), brightness (*bright/well-lit vs. dark/badly-lit*), *comfortable/cozy*, and *dull/simple*.

The chapter is organized as follows. Section 5.2 discusses related work. Section 5.3 presents the data collection and annotation process. Section 5.4 presents the in-depth analysis of private spaces based on the manual annotations. Section 5.5 presents the approach based on deep learning to extract visual and sound features of videos, examines the correlation between ambiance and the previously extracted cues, and presents experiments on automatic ambiance inference. Section 5.6 discusses the findings and limitations of this chapter from the perspectives of social computing. Section 5.7 concludes the chapter.

5.2 Related Work

This chapter is related to a body of work from various disciplines examining issues of urban nightlife and youth; characterization of private spaces; and ambiance modeling. Each of these themes is discussed in the next subsections.

5.2.1 Urban nightlife and youth

Work in geography has studied the urban night period, often with qualitative methods (Van Liempt et al., 2015). The authors in (Boyd, 2014; Hatuka & Toch, 2017; Toch & Levi, 2013) also studied the dynamics surrounding youth experiences and urban nightlife. There is other work that has investigated the phenomena of human mobility and space usage in urban areas (Chatterton & Hollands, 2002; Hubbard, 2008; Matthews, Limb, & Percy-Smith, 1998; Skelton & Valentine, 2005). From the perspective of alcohol consumption and urban youth, researchers investigated pubs and bars (Duff, 2012), house parties (Järvinen & Østergaard, 2009) and public spaces (Demant & Landolt, 2014). Especially, (Bellis & Hughes, 2011; Holloway, Jayne,

& Valentine, 2008; Van Liempt et al., 2015; S. Wilkinson & Wilkinson, 2018) studied alcohol consumption from "pre-loading" (drinking before going out for the night) to excessive drinking with risky consequences. In contrast with these works, this chapter aims to understand the characteristics and activities of the nightlife of youth in their home environments based on captured videos of the private spaces, contributed by the study participants in a crowdsensing setting.

5.2.2 Place characterization and private spaces

Regarding place characterization, the authors of (Wang, Lymberopoulos, & Liu, 2014) used mobile sensors, i.e. audio signals to infer occupancy, human chatter, music, and noise of places. Meanwhile, the authors of (Y. Chon et al., 2012) aim to categorize places by using audio signals and images. Chon et al. (Y. Chon, Lane, Kim, Zhao, & Cha, 2013) collected 48,000 place visits from 85 participants in Seoul to study the coverage and scaling properties of place-centric crowdsensing.

As a private space, the home is an environment where many social activities of young people unfold (Abbott-Chapman & Robertson, 1999, 2001). In geography, Abbott et al. (Abbott-Chapman & Robertson, 1999) investigated perceptions of young people about home as an idealized social construct and as a private space. Abbott et al. later investigated the social constructs of 'home' and 'neighborhood' as private and public spaces, in the context of leisure activities performed by young adolescents (Abbott-Chapman & Robertson, 2001). These studies used standard methods based on recall-based surveys. From a technical perspective, work in ubiquitous computing has developed approaches for place characterization, which use mobile sensors like microphones to extract audio signals through which certain features like human chatter and music can be inferred (Wang et al., 2014), or a combination of audio signals and still images that capture snapshots of everyday places (Y. Chon et al., 2013; Y. Chon et al., 2012). This body of work, however, has been largely focused on understanding outdoor spaces, often with goals of automatic place recommendation for urban users. In contrast to this work, we investigate how attributes of the home environment of young people are depicted on videos recording snapshots of weekend nights (a period of intense socialization among youth (Abbott-Chapman & Robertson, 1999, 2001)) using both human observers and machine-generated descriptors of the home environments.

5.2.3 Home spaces and activities

Home is conceived in different ways, including a physical space (house/apartment), someone's place of origin, or the place where people feel they belong ("Definition of home in English", 2019). Regarding the place of origin or where a person feels as belonging to, home is a site of 'shelter' (Hayden, 2004) or a 'meaningful' place with multiple experiences through which people feel belonging (Ahmet, 2013; Tuan, 1971). Home can not only be a fixed space but also an urban area, e.g., a street in town or a popular area in the city (Ahmet, 2013). Home can also be a material place where young people live with their family (Blunt, 2005; Schiano, Elliott, & Bellotti, 2007), or a student home or dormitory where students study or live away from their parents (Holton, 2016). In our research, we aim to understand home as a personal space where young people spend their time alone or with friends on weekend nights.

Home is one of the places where youth spend their leisure time e.g., watching TV, listening to music (Schiano et al., 2007), playing physical games (Sall & Grinter, 2007), or drinking before going out with their friends at public places in the city (S. Wilkinson & Wilkinson, 2018). Many people also socialize at their friends' or family's house (Holloway et al., 2008), which emphasizes the importance to understand these practices, as the use of rooms and spaces at home can be influenced by architectural constraints, culture, an individual's daily life (Alitajer & Nojoumi, 2016), or even mental distress (Tucker, 2010). Baillie et al. (Baillie & Benyon, 2008) studied four spaces in the home, including communication, work, leisure (private) and leisure (public) along with their utility to people living there. In this chapter, we investigate multiple dimensions of home of youth on weekend nights, including physical attributes (e.g. room types, brightness, music), social attributes (people present in the home environment), and ambiance (e.g. festive or fun).

5.2.4 Ambiance in architecture and psychology

The roles of interior architecture and design on human behavior have been studied in several disciplines, and provide background about the way humans interact with their living spaces. The characteristics of the places where we live, including space quality, interior design, and colors, affect how we feel, and reflect personal and social constructs (S. Gosling, Gifford, & McCunn, 2013). Three main factors discussed by (S. Gosling et al., 2013) influence living spaces: identity claims, thought/feeling regulators, and behavioral residues. Interior ambiance, i.e., "the character of a place or the quality it seems to have" (Dictionary, 2018) can have specific effects on people's behavior.

In the context of personal spaces, physical and environmental cues can reveal characteristics related to gender (Rheingold & Cook, 1975) and personality (Samuel D Gosling et al., 2002). A common method used in psychology (Samuel D Gosling et al., 2002; Rheingold & Cook, 1975) involves asking observers to manually rate physical spaces, which is an approach applicable to small-scale studies. This chapter uses this methodology, and expands it by using automatic analysis to characterize the content of videos using state-of-art deep learning methods.

Ambiance has also been studied in public and commercial spaces. Quercia et al. (Quercia, O'Hare, & Cramer, 2014) presented a crowdsourcing project related to ambiance-related constructs in the outdoor space, which studied how visual cues, color, and texture have effects on London neighborhoods along three perspectives: beautiful, quiet, and happy. In (Kotler, 1973), physical and decoration cues had effects on the shopping behaviors of customers, because people's emotions and behaviors can be affected by these places' ambiance. Ambiance cues like color, brightness, and style have an important impact on customer emotions at hotels

(Countryman & Jang, 2006), or on food intake and food choice at restaurants (Stroebele & De Castro, 2004). For instance, (Bell et al., 1994) showed that decorating the ambiance of a pasta restaurant with a distinctive Italian feeling can make customers order more food.

Specifically for home environments, the Personal Living Space Cue Inventory (PLSCI) (Samuel D Gosling et al., 2005) describes personal living spaces, including 42 physical attributes and the ambiances of the space along with a checklist of 100 individual items. PLSCI is used by (Carney, Jost, Gosling, & Potter, 2008; Samuel D Gosling et al., 2002) to study various questions in environmental psychology. We also adopt PLSCI for designing the video ambiance questionnaire for our study about home spaces of youth, which is discussed in Section 5.4.

5.2.5 Indoor ambiance inference in social computing

Several works have proposed methods to automatically recognize indoor ambiance from social media data. By observing the avatar pictures of Foursquare users, the work in (Graham & Gosling, 2011) showed that people can identify place ambiance, clientele, and their activities with some degree of reliability. The work in (Redi et al., 2015) also used 4sq profile pictures to infer place ambiance using aesthetics, colors, emotions, demographics (age and gender), and self-presentation. Although the number of data points used in this work was small (N=49), it showed promise for place ambiance inference. Using data from Foursquare, the work in (Santani & Gatica-Perez, 2015) generated crowdsourced annotations on an image corpus to study 13 ambiance dimensions. This dataset was later used to apply traditional visual features (color, GIST, HOG) and features extracted from a pre-trained CNN for ambiance inference (Santani, Hu, & Gatica-Perez, 2016). The work in (Benkhedda et al., 2017) further examined the problem of ambiance recognition through scene semantics, assuming that there are visual cues within scenes that can be extracted using a scene-centric semantic parser. We also adopt this assumption in this chapter for conducting annotation on ambiance by asking raters to watch videos. However, the datasets used in previous work are images from Foursquare places, thus covering restaurants, bars, cafes, etc.. In this chapter, we work with substantially different data, namely with videos capturing private spaces during 10 seconds, and through the combination of manual annotation results of ambiance based on the observation of the captured videos, and on semantic video cues extracted from deep learning models.

Airbnb is a social platform for hospitality that shows home environments to possible guests through photos. Ikkala et al. (Ikkala & Lampinen, 2015) conducted a qualitative research of hospitality exchanges on Airbnb. The study found that hosts on Airbnb have both financial and social reasons. In detail, money plays a role in supporting hosts in their efforts to manage social interaction, select guests consistent with their preferences, and control the volume and type of demand of visitors. In what constitutes the closest work to ours, (Nguyen et al., 2018) used a dataset of 1200 Airbnb venues represented by three images of each place to infer ambiances from visual features extracted from deep learning models. This work is an inspiration to this chapter, with one fundamental difference, namely that the visual data responds to very different motivations: crowdsensing for scientific research in our case, and illustrating home places for monetary purposes on Airbnb. This translates into rather different visual content: on Airbnb, images are curated to appear as appealing as possible to viewers; in this chapter, the videos produced by youth on weekend nights are unfiltered (except for reasons of sensitive situations and privacy) and non-beautified (as the study participants are sharing this data for research only and not for performative purposes as is often the case on Instagram and other social media).

To the best of our knowledge, this chapter extends the current understanding of private nightlife settings with respect to physical attributes at homes, activities of young people, and ambiances, building upon previous work in the social computing and ubicomp literature (Baillie & Benyon, 2008; Ikkala & Lampinen, 2015; Quercia et al., 2014; Redi, Aiello, Schifanella, & Quercia, 2018; Sall & Grinter, 2007; Schiano et al., 2007). In Table 5.1, we summarize the most closely related work and distinguish what we contribute to this domain.

5.3 Data Collection

This chapter uses data from the Youth@Night project (Santani, Biel, et al., 2016), which aimed at studying young people nightlife behavior in Switzerland using a smartphone application (Labhart et al., 2019a; Santani, Biel, et al., 2016). This section provides an overview of the study design, the data collection procedure, and the specific data we use in this chapter.

5.3.1 Study design

Study context

Participants were recruited in Zurich and Lausanne, two of the four largest Swiss cities (Labhart et al., 2017; Santani, Biel, et al., 2016) and the two main hubs of nightlife activities (Marquard, 2014; of Lausanne, 2010). They were approached by small groups of research assistants on the street between 8 PM and midnight in September 2014. In order to obtain a representative sample of nightlife goers, participants were recruited in popular areas (e.g., nightlife districts, public parks, streets), pro-rata of the area popularity at the city level. Quotas of people to recruit per area were determined using geo-localized venue data from Foursquare (Santani & Gatica-Perez, 2013b), and were validated with local experts (social workers and police). Eligibility criteria for participation were being aged between 16 and 25, owning an Android phone, having been out in the city at least once in the past month, and have consumed alcohol at least once in the past month (legal drinking in Switzerland, as in many other European countries, is 16 for beer and wine). The study protocol was approved by the ethical review boards of Vaud and Zurich cantons, and authorization to recruit on the street was obtained from the local authorities.

Table 5.1 – Comparison between previous work and this chapter.

Work	Goal	Data	Tasks	Finding
(Abbott-Chapman & Robertson, 2001) (2001)	Observe young people's favourite places and associated leisure activities at home and neighbourhood.	256 completed questionnaires and 58 interviews (28 girls, 30 boys from secondary school)	Quantitative and qualitative analyses	Young people want their homes to be friendly, spacious, modern, and quiet, they hangout with friends at home and friends' homes.
(Wang, Lymberopoulos, & Liu, 2014) (2014)	Infer the ambiance of business places from audio recordings	150 audio traces of indoor business and external surveys	Regression task for inferring the level of occupancy, human chatter, music, and noise levels using audio features	Classification performance of ambiance at 79% accuracy
(Redi, Quercia, Graham, & Gosling, 2015) (2015)	Determine which visual cues of profile pictures can predict places' ambiances	Ambiance surveys of 49 places, with 250 annotations on 25 profile pictures on each place.	Regression task for predicting place ambiance using profile pictures' features	Predict ambiance based on faces at 78%
(Santani & Gatica-Perez, 2015) (2015)	Investigate which types of social media images best convey indoor ambiance	50K images from 300 places on Foursquare, and 13 ambiance labels.	Interannotator agreement (ICC) analysis ans correlation analysis.	All 13 dimensions have ICC>0.5
(Santani, Hu, & Gatica-Perez, 2016) (2016)	Infer impressions of place ambiance, using generic and deep learning features	45,000 Foursquare images from 300 popular places in six cities	Regression task for inferring ambiance using machined-extracted features	Inferring place ambiance is feasible with a maximum R^2 of 0.53
(Benkhedda, Santani, & Gatica-Perez, 2017) (2017)	Examine correlation of visual cues with ambiance of Foursquare images to automatically infer place ambiance	50K Foursquare images and 20K scene centric image dataset	Regression task for inferring ambiance using deep learning features	Ten of the ambiances can be inferred using scene objects and demographic attributes
(Nguyen, Ruiz-Correa, Mast, & Gatica-Perez, 2018) (2018)	Predict ambiance from pictures of listings on Airbnb	1200 Airbnb listings and crowdsourced annotations of images	Regression task for inferring ambiance using deep learning features	Ambiance can be inferred with R^2 up to 0.42
This chapter	Describe youth personal spaces by means of crowdsourced videos recorded in-situ. Labels are different than all above work except (Nguyen, Ruiz-Correa, Mast, & Gatica-Perez, 2018). Infer ambiance at youth personal spaces from physical attributes	301 videos recorded in participants' home space on weekend nights. Manual annotations of the 301 videos by 5 independent annotators and CNN-based extraction of visual and audio descriptors.	Descriptive and correlation analyses of ambiances and physical features of home spaces. Regression task to infer ambiance using machined-extracted features	Living room, bed room, kitchen, and dining room are all represented at home on weekend nights. Top activities include drinking, chatting, watching TV, and eating. Home ambiance was often described as quiet and simply decorated. Regression for ambiance inference achieved R^2 between 0.21 – 0.69.

Data collection

The study took place on Friday and Saturday nights between September and December 2014. Participants were required to download and install Youth@Night applications. The survey logger application allowed participants to document, in real time, various aspects of their night, such as the locations attended (e.g. home, park, bar/pub), the type of drinks consumed, and 10-second video clips of their environment from 8 PM to 4 AM. Meanwhile, the sensor logger application, a background running app without any user interaction, collected many types of sensors and log data, such as GPS coordinates, accelerometer, and battery status (Labhart et al., 2017; Santani, Biel, et al., 2016). In this chapter, we will only use data from the survey logger application.

Questionnaires and sensor datasets were automatically uploaded to a back-end server when participants' smartphones had access to Wifi. Whenever the data was successfully uploaded, it was removed from the device. The participants could choose to manually upload data in case there was a problem with the automatic upload. At the end of the study, participants were paid 100 CHF if they documented at least 10 weekend nights. Participants completing less than 10 evenings with a minimum of three nights were paid on a pro-rata basis.

After the app-based data collection fieldwork, 40 qualitative interviews were conducted with study participants and focused on their experiences with the smartphone application, their experiences of nights out, and the ways in which mobile technologies shape contemporary nightlife (Truong, 2018a, 2018b).

5.3.2 Measures

Video clips of environments.

The survey logger application contained different questionnaires and media to capture participants nightlife behaviors, the locations attended, and the characteristics of their surrounding environment (see (Labhart et al., 2019a) for an overview of the different kinds and sequences of questionnaires). Participants were instructed to document any weekend night, including those during which they did not drink or did not go out, in order to have an overall representation of the different activities and events taking place on weekends. In the present chapter, we use the short video clips collected with the application at specific times of the night: whenever participants had their first drink (alcoholic or non-alcoholic) after 8 PM, and whenever they had a new drink (alcoholic or non-alcoholic) in a new location, they were required to indicate the type of location they attended (e.g. bar/pub, parks, home) and to record a 10-second video clip, which captured a panorama of their environment by slowly turning from left to right in landscape format. Participants thus recorded videos in varied environments, including pubs, clubs, public parks, means of transportation, and homes (Santani, Biel, et al., 2016). In case they were not able to record video (e.g., forbidden, felt uncomfortable), participants were told to skip the task and specify the reasons for it. Overall, participants recorded videos in 68% for

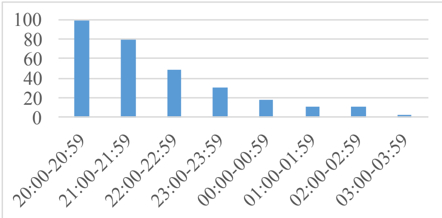


Figure 5.1 – Number of videos of private places (N=301) per hour.

the cases, while reasons for not recording were mostly because they did not it feel it as appropriate or safe (Labhart et al., 2019a). Each video file was stamped with its time of submission. In total, 843 videos were collected from 204 participants on 646 participant-nights.

Annotation of home environments.

After the fieldwork, we designed an annotation task to get qualitative information on the type of location, ambiance, physical attributes, and people shown on the 843 video clips recorded by the participants. Five independent annotators were hired and trained to watch the entire corpus of videos and answer 17 single and multiple choice questions on the type of location, the ambiance of the place, and characteristics of the social and physical environment. The exact questions and response options are presented in Section 5.4.

Identification of home environments.

Based on the annotators' answers to the question *"In what kind of place was the video clip taken?"*, places were considered as *"homes"* in case all five annotators agreed on this label. In total, this procedure retained 301 videos representing home places. In these environments, participants recorded videos in 64% of the cases. Reasons for not recording a video were: *"I was asked by someone not to do it"* (36%), *"it is not appropriate"* (25%), *"I don't feel safe"* (24%) and *"other"* (21%). Given that participants recorded videos of the environment whenever they had their first drink after 8 PM or moved to another location, (i.e., change of home, or come back home, in the present case), the 301 videos illustrate home environments throughout the night, although with a larger proportion of those taken early in the night if the participants did not change location. Figure 5.1 shows the number of videos per hour. Because of the small number of observations per hour after midnight, environments documented after midnight will be aggregated in the rest of the analyses.

Due to privacy requirements requested by the Ethical Review Boards that reviewed and approved the project, we cannot make this dataset publicly available.

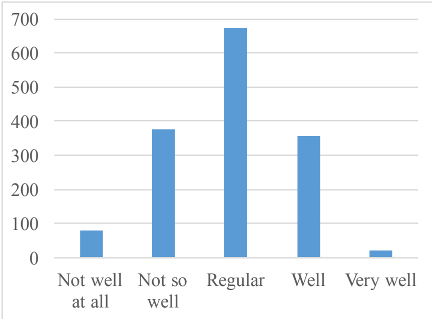


Figure 5.2 – How well physical spaces are captured in videos. The y-axis represents the total number of annotations.

5.4 Physical/Social Attributes and Ambiance of Home Spaces (RQ1)

In this section, we investigate how main patterns of physical attributes and ambiance can be extracted from videos recorded in private spaces using external annotators. In the following subsections, we first explain the measure (i.e. exact questions and response options), investigate the consistency of annotations across the five annotators using Intraclass Correlation analyses, and provide descriptive results. The Intraclass Correlation Coefficient (ICC) is a standard measure of reliability of raters (Shrout & Fleiss, 1979). As recommended by (Koo & Li, 2016), we used ICC(2,k) which is used for a fixed set of k judges rating each target (N=301) and reflects the absolute agreement. Following the guidelines from from Koo and Li (Koo & Li, 2016), ICC scores below 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.9 are indicative of poor, moderate, good, and excellent reliability, respectively. ICC(2,k) can only be computed on numerical variables, not on categorical ones, so there are few cases in this section that does not show ICC. We summarize results of ICC scores for each possible question in Table 5.2.

5.4.1 Overall representation of the space

In the annotation task, after carefully watching each video, several times if required, annotators were asked to indicate *"How well does the video capture the physical space (i.e. space layout, background scene, furniture, decoration, etc.)?"* with five single-item response options. *"[1]not well at all"*, *"[2]not so well""*, *"[3]regular"*, *"[4]well"*, and *"[5]very well"*. Results showed a good level of agreement on this question (ICC = 0.83). Figure 5.2 shows the histogram of all individual responses of all annotators to this question (301 x 5 =1505). As seen in Figure 5.2, most of the videos were rated as providing a *"regular"* representation of the space. The mean of this variable is 2.91 (SD=0.86), which is sightly lower than 3 ("regular"). In some cases, participants avoided recording directly physical spaces that could contain people. For instance, some participants recorded the ceiling or floor while panning the camera.

Table 5.2 – ICC of physical attributes at homes based on N(video)=301, N(raters)=5, with scale (1-5).

Physical attributes at homes	ICC(2,k)	mean	std	skew
Physical space (i.e. space layout, scene, decoration, etc)	0.83	2.91	0.86	-0.17
Amount of light	0.87	2.90	0.69	-0.30
How loud is the music	0.95	1.44	0.81	1.70
Level of overall chatter	0.94	1.71	0.94	0.80
Level of occupancy	0.97	1.82	1.09	1.12

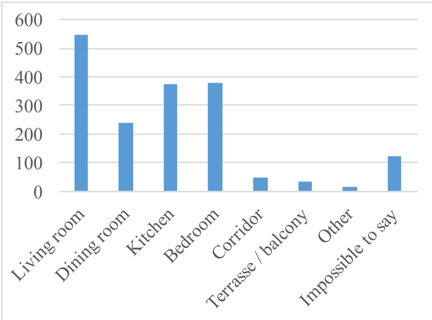


Figure 5.3 – Types of spaces at homes captured in videos. The y-axis represents the total number of annotations.

5.4.2 Physical and Social Attributes

Room of the home

In the annotation task, the room type within the home was labeled using the question: *"Where in the home was the video taken?"* and the following single-item response options: *"[a]living room"*, *"[b]dining room"*, *"[c]kitchen"*, *"[d]bedroom"*, *"[e]corridor"*, *"[f]terrace/balcony"*, *"[g]other"*, and *"[h]impossible to say"*.

Figure 5.3 shows the frequency with which individual annotators identified specific rooms of the homes in the 301 videos. Living room, bedroom, kitchen, and dining room are the most attended spaces within homes. This result echoes previous work that using traditional methods reported that living rooms and bedrooms are the most used places in small and large homes by occupants (Khajehzadeh & Vale, 2015), and extends this previous finding by showing that for the specific case of young people on weekend nights, kitchens and dining rooms are also frequently used indoor spaces. As mentioned previously, a few videos avoid capturing directly the physical spaces by turning the camera to the ceiling and floor. This is one of reasons why *"Impossible to say"* appears in Figure 5.3.

Brightness

The annotators were asked to answer a single choice question *"Describe the amount of light in the place"* with five choices *"[1]It is very dark"*, *"[2]It is quite dark"*, *"[3]Normal"*, *"[4]It has a*

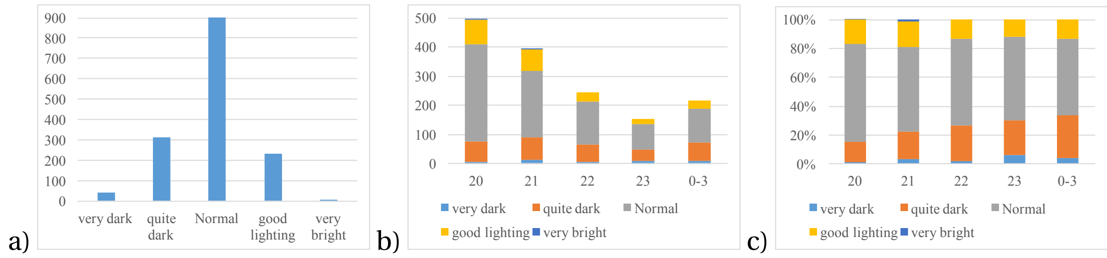


Figure 5.4 – (a) Brightness, (b) Brightness per hour (8:00-8:59 PM, etc), (c) Brightness per hour expressed as a percentage within that timeslot. The order of levels of brightness (very dark, quite dark, normal, etc.) is left-to-right in graph a, and top-to-bottom in graphs b and c. The x-axis on graph b and c is hour on Friday and Saturday nights from 20:00 to 3:00. The y-axis on graphs a and b represents the total number of annotations, while the y-axis on graph c represents the percentage normalized on each hour.

good lighting" to *"[5]Is is very bright"*. The ICC(2,k) of brightness is high (0.87). The brightness variable has a mean slightly below the middle of the scale (2.9, SD=0.86). Figure 5.4 shows the histogram of annotated brightness, brightness per hour (8:00-8:59 PM, etc.), and brightness per hour expressed as a percentage within that timeslot, respectively. The percentage of darkness (*quite dark* and *very dark*) increases from 18% (8PM) to 35% (0-3AM) in Figure 5.4c. Conceptual work in geography (Shaw, 2015) has recently discussed how individuals at home in the dark might be more willing to open themselves to others, and how adjusting the darkness of the home environment can be empowering. Our annotations suggest that as the weekend night goes on, young people at home indeed tend to be in conditions of lower illumination. As a reminder, note that given the season of the year when the data was collected (mid September through December), it was past sunset time at the beginning of each recorded night (8PM).

Music Loudness

Regarding music loudness at home places, the annotators were asked to answer *"Describe how loud is the music in the place"* with five choices *"[1]No music"*, *"[2]Low"*, *"[3]Medium"*, "*[4]Loud*", and *"[5]Very loud"*. The ICC(2,k) of music loudness is excellent (0.95). The mean value (1.44) is low (SD = 0.81). The skew is large (1.70) showing that the distribution has a tail. Figure 5.5 also shows the corresponding temporal trends. Overall, the present results on music and brightness levels at home are consistent with recent ethnographic research showing that young people tune their home by turning off lights and choosing slow paced music when they spend time drinking with their friends at night (S. Wilkinson, 2017). We found that no music was played in most of the recorded environments (frequency: 76%; see Figure 5.5a). When music was played, the loudness level was quite low throughout the night (Figure 5.5c), suggesting that the cohort of young people are relatively quiet in their private nightlife.

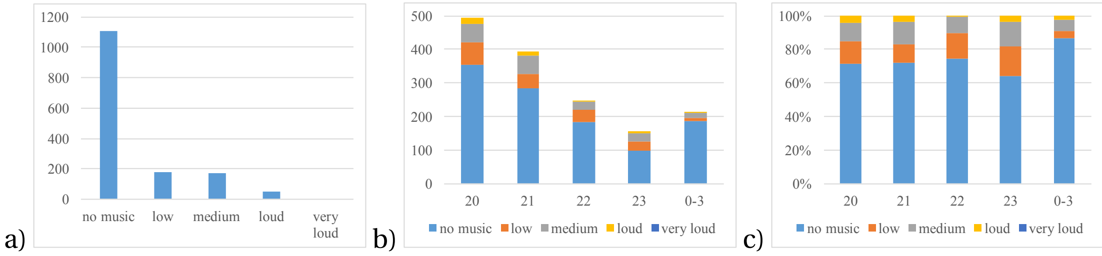


Figure 5.5 – (a) Music loudness, (b) Music loudness per hour, (c) Music loudness per hour expressed as a percentage within that timeslot. The x-axis on graph b and c is hour on Friday and Saturday nights from 20:00 to 3:00. The y-axis on graphs a and b represents the total number of annotations, while the y-axis on graph c represents percentage normalized on each hour.

Chatter Loudness

The annotators were asked to describe the level of chatter loudness at home space by answering the question *"Describe how loud is the chatter in the place"* with five single choices *"[1]No chatter"*, *"[2]Low"*, *"[3]Medium"*, *"[4]Loud"*, and *"[5]Very loud"*. Similarly to music loudness, the ICC agreement for chatting loudness is very high (0.94). The mean value is low (1.71, SD = 0.94). Figure 5.6a-c shows that there is not much loud talking in the recorded videos. Relative to each hourly slot, medium and loud chatting slightly increase from 8 PM to 11 PM (Figure 5.6c). This result is clearly connected to the results obtained for the occupancy of the physical space discussed next.

Occupancy

Annotators were also required to describe the level of occupancy of the place by using the following single choice question *"Describe the level of occupancy of the place based on what you hear or see"* with five choices *"[1]Empty"*, *"[2]There are few people for this space"*, *"[3]It's half empty/half full"*, *"[4]It's well attended, but there could still be more people"* to *"[5]It's highly crowded/packed"*. Annotator agreement of occupancy level was excellent (ICC = 0.97). The mean of level occupancy of the place is 1.82 (SD=1.09). While we anticipated that most young people meet with others at home on weekend nights, Figure 5.6d shows that empty is the most common category. Figure 5.6f also shows that young people slightly reduce gathering together from 8 PM to 10 PM; then, gathering increases at 11 PM, and decreases again after midnight.

Number of people present

As a complement to occupancy, we asked the annotators *"How many people appear on the video (in addition to the phone holder)"* with six choices *"[1]0 (the person seems to be alone)"*, *"[2]1"*, *"[3]2-4"*, *"[4]5-10"*, *"[5]More than 10"* to *"[6]Impossible to say"*. Figure 6g shows that around 40% of videos are labeled as containing no people, which is consistent with the labeling

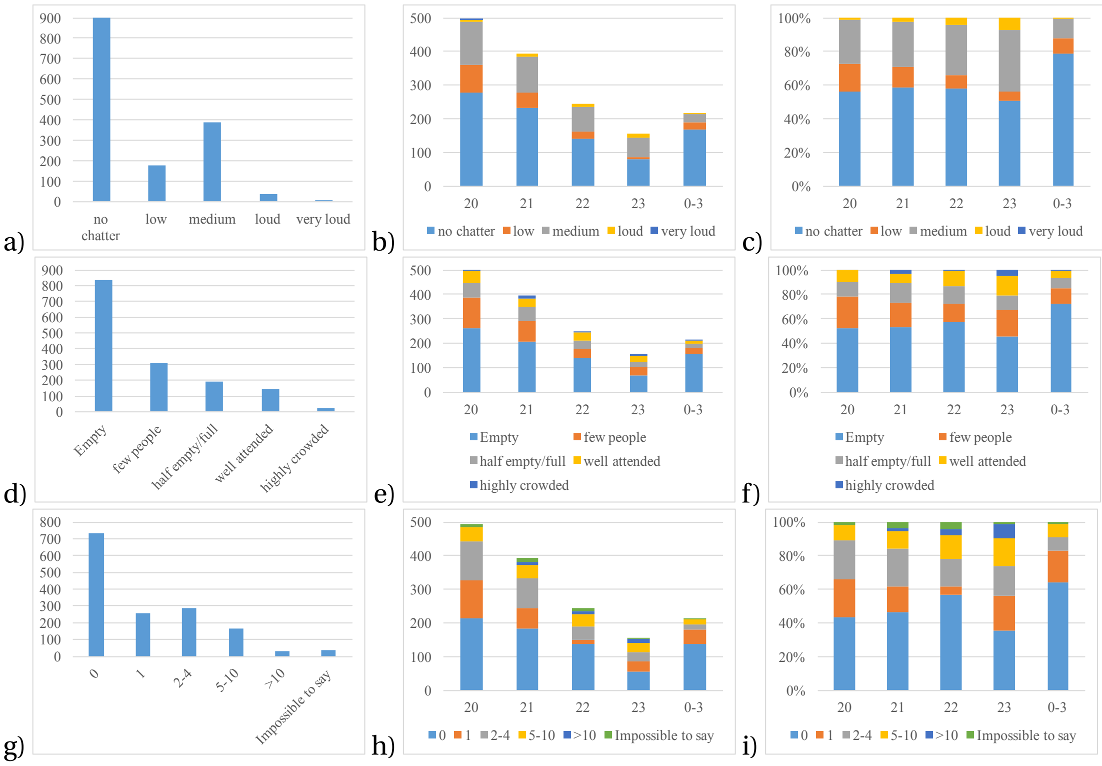


Figure 5.6 – Annotation of (a-c) chatter level, (d-f) occupancy level, (g-i) the number of people in the videos. The left column shows the overall trend, the middle column the trend per hour, and the right column the relative percentage for each timeslot. The order of values of all legends is left-to-right, top-to-bottom in all graphs. The x-axis on graphs b-c, e-f, h-i is hour on Friday and Saturday nights from 20:00 to 3:00. The y-axis on graphs a-b, d-e, g-h represents the total number of annotations while the y-axis on graph c, f, i represents percentage normalized on each hour.

of occupancy.

Gender of people present in videos

Among the people present in the videos, we examined their gender ratio by asking one single choice question: *"What is the gender ratio of the relatives, friends, or acquaintances appearing in the video?"* with 6 response options *"[1]Women only"*, *"[2]Mostly women"*, *"[3]Half-half"*, *"[4]Mostly men"*, *"[5]Men only"* to *"[6]Impossible to say"*. Figure 5.7a shows that "men only" is the most common situation, followed by "women only" and "half-half". The total number of situations with "men only" and "mostly men" is higher than those with "women only" and "mostly women", suggesting that men appeared more often in the videos than women. Surprisingly, the 301 videos were fairly evenly distributed per gender, with 144 videos recorded by 52 male participants and 157 videos recorded by 50 female participants. Figures 5.7b and 5.7c show the gender repartition of the people present in the videos recorded by male and

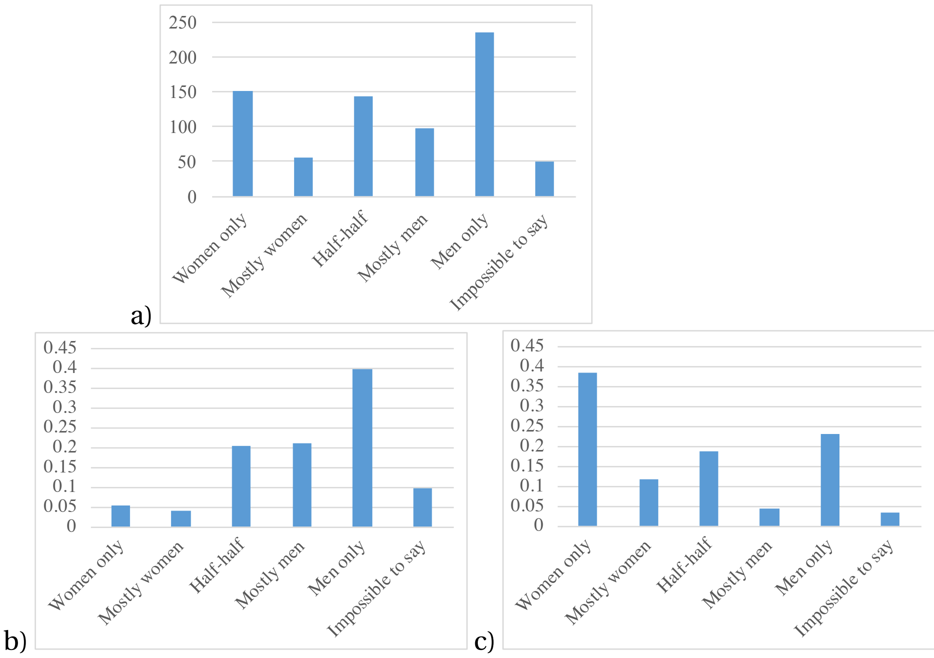


Figure 5.7 – (a) Frequency of gender of people appearing in 301 videos recorded by 102 male and female participants, (b) percentage of gender of people appearing in 144 videos recorded by 52 male participants, and (c) percentage of gender of people appearing in 157 videos recorded by 50 female participants. The y-axis on graph a represents the total number of annotations, while the y-axis on graphs b and c represents percentage normalized on each possible value on the x-axis.

female participants, respectively. Male participants mostly tend to spend their nights at home with other male friends and less so with women, while no clear preference could be observed for female participants. As a point of reference, work on a sample of 377 students (Watts, 2004) showed that young females tend to hang out at home with friends more than males do.

Activities of people

In order to assess the activities of young people at their home spaces, we asked annotators to indicate *"What things are people doing in the video?"* with 14 multiple choices items shown in Figure 5.8a. Results showed that activities are quite diverse, with drinking, chatting, watching TV, using smartphone/tablet/computer, and eating as the five most common activities. As seen in Figure 5.8c, these main activities are roughly constant from 8 PM to midnight. Drinking, as the most commonly annotated activity, takes 15-25 percent in relative terms across all hourly slots. The prevalence of this activity at home is not surprising given that participants were requested to document the environment when they had their first alcoholic or non-alcoholic drink there. Nevertheless, this finding also echoes to previous research from Valentine et al. (Valentine, Holloway, Knell, & Jayne, 2008) showing that 73% of young people report having consumed alcohol at their homes and 64% at their friends' houses over the last year.

Yet, our analysis brings a finer grained description of temporal trends. In addition, we also examine activities of young people depending on the level of occupancy and type of space at homes, as shown in Figure 5.9. In Figure 5.9a, when the place is empty (i.e. only the person recording the video is present), the most commonly annotated activities are watching TV, using a computer/tablet/smartphone and, to some extent, drinking. Conversely, in the presence of other people, the commonly annotated activities are chatting, drinking, and eating, whose proportions increase along with levels of occupancy. It might also be noticed that playing board games was the most frequently reported in "half empty/full" homes, and some dancing was reported in highly crowed homes. Figure 5.9b shows that there are four places at home spaces that co-occur with specific activities: terrace/balcony/corridor; kitchen/dining room; living room; and bedroom. In related work, Baillie et al. (Baillie & Benyon, 2008) study leisure (private) and leisure (public) places in terms of their utility to inhabitants of a house. We complement this by showing that chatting and drinking occur more (in distributional terms) in leisure public areas within homes (terrace/balcony/corridor, kitchen/dining room, living room), while activities like using computer/tablet/smartphone and watching TV occur around 60% in private leisure spaces (bedroom).

Reactions of people around in videos

To conclude our research on physical and social attributes at home spaces, we examined reactions of people around in videos by asking five annotators to answer a single choice question *"Can you see or hear one or more persons reacting to or being aware of the video being recorded?"* with two answers *"[1]Yes"* and *"[2]No"*. If the previous question gets answered *"Yes"*, we will ask five questions listed in Figure 5.10b with three single choices *"[1]Yes", "[2]No", "[3]Not sure"*

We are interested in how people in videos react to video recording in home spaces. As we mentioned, many videos did not get recorded by design, as participants were told not to do it if not appropriate. Regarding the 301 recorded videos at home spaces, in 25% of cases did people in the video react to the camera (shown in Figure 5.10a). Two of the main reactions were *having fun while the video is recorded* and *asking about or commenting on the purpose of the video*. It is important to note that participants in the study were explicitly instructed to record video only when it was socially acceptable and agreed and they were free to avoid recording (Santani, Biel, et al., 2016). The video dataset used here was recorded with such guidelines. There are just a few cases showing that people in the video were not comfortable about being recorded or to hide their face.

5.4.3 Ambiance attributes

To assess the ambiance of home environments , we used a modified version of the Personal Living Space Cue Inventory (PLSCI) (Samuel D Gosling et al., 2005). This instrument was originally designed to describe personal living spaces, e.g. rooms in family households, dormi-

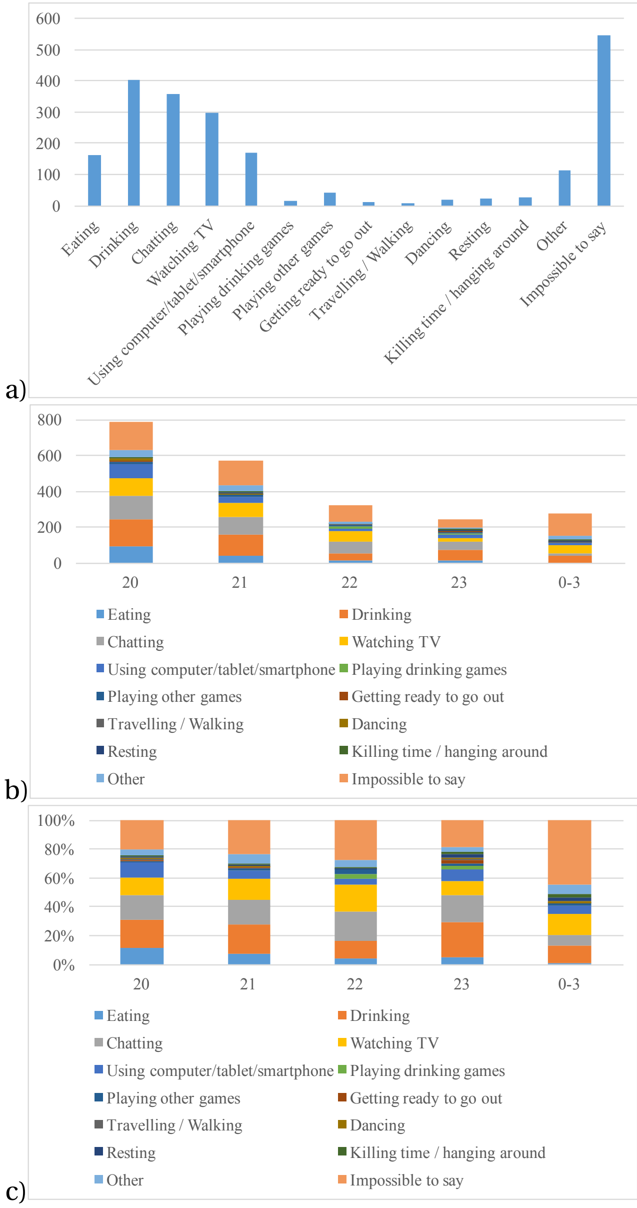
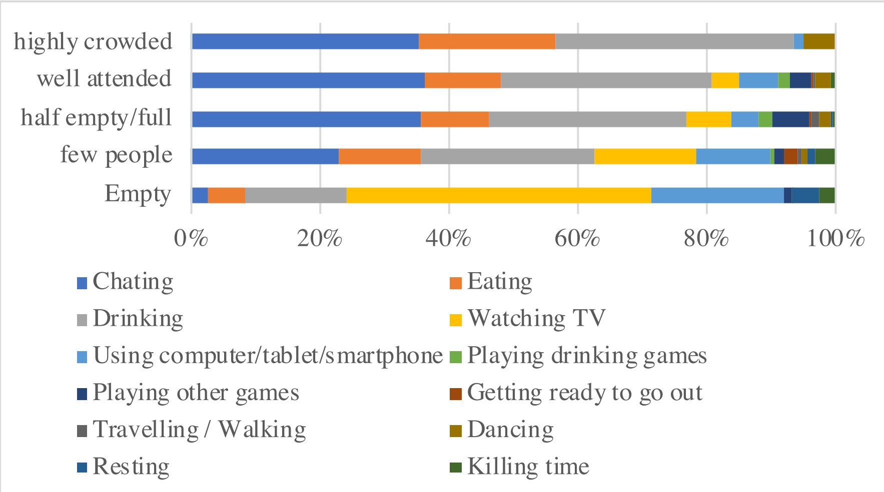
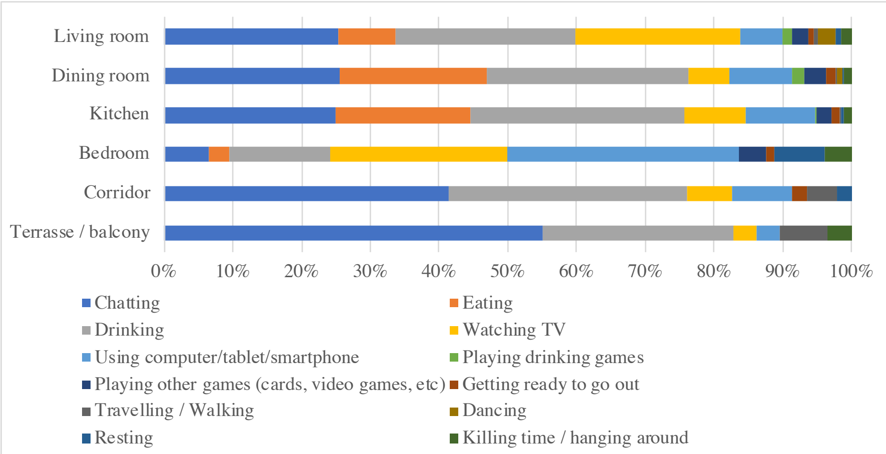


Figure 5.8 – Frequency of occurrence of (a) activities (b) activities per hour, and (c) percentage of activities within each timeslot. The order of activities (eating, drinking, chatting, watching TV, etc.) is left-to-right in graph a, and top-to-bottom in graphs b and c. The x-axis on graph b and c is hour on Friday and Saturday nights from 20:00 to 3:00. The y-axis on graphs a and b is the total number of annotations while the y-axis on graph c is percentage normalized on each hour.

tories, or residential places. In our case, we augmented the PLSCI with ambiance attributes from previous work (Graham & Gosling, 2011), (Santani & Gatica-Perez, 2015), (Nguyen et al., 2018), (Redi et al., 2018). As a result, we obtained a list of 11 ambiance word groups (e.g.

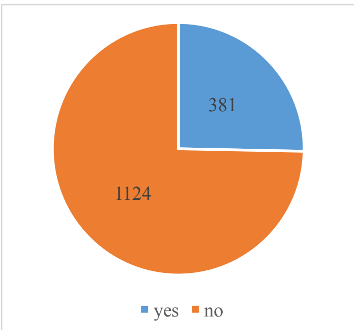


a)

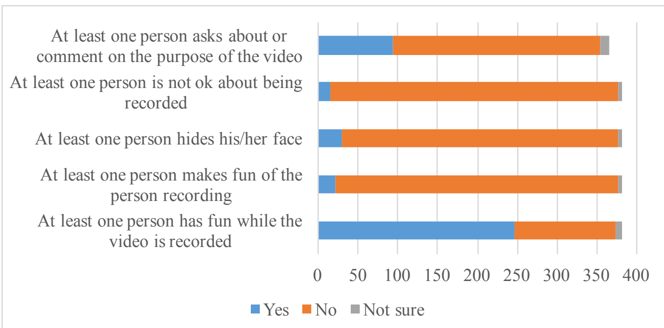


b)

Figure 5.9 – Percentage of activity based on a) level of occupancy, and b) type of space at homes. The order of activities (chatting, eating, drinking, watching TV, etc.) is left-to-right, top-to-bottom in graphs a and b.



a)



b)

Figure 5.10 – (a) Yes/No reactions of people in videos. (b) Description of reacrions to the videos.

Table 5.3 – ICC of ambiance categories at homes based on N(videos)=301, N(raters)=5 with scale (1-7).

Ambiances	ICC(2,k)	min	max	mean	std	skew
Large, spacious	0.81	2.0	6.4	3.9	1.3	-0.09
Dark, badly-lit	0.83	1.4	7.0	3.8	1.6	-0.26
Colorful, decorated	0.66	1.8	6.2	4.0	1.3	-0.40
Cramped, confined	0.67	1.4	6.2	3.7	1.6	-0.23
Bright, well-lit	0.81	1.0	6.2	3.7	1.5	-0.18
Comfortable, cozy	0.61	1.8	5.6	3.7	1.2	-0.33
Dull, simple	0.63	2.0	6.2	4.1	1.3	-0.40
Festive, fun	0.27	1.8	4.8	2.9	1.4	0.00
Sophisticated, stylish	0.65	1.0	6.0	2.8	1.6	0.52
Off-the-beaten-path, unique	0.35	1.4	6.2	2.8	1.5	0.18
Serious, boring	0.21	1.8	5.2	3.5	1.5	-0.42

large/spacious, cramped/confined; all items are listed in Table 5.3). A Likert scale, used in previous ambiance work but also as a reliable methodology to annotate image aesthetics (Siahaan, Hanjalic, & Redi, 2016) was used in this chapter. Annotators had to rate each ambiance by indicating, on a 7-point Likert scale ranging from *"[1]strongly disagree"* to *"[7]strongly agree"*, the degree to which they agreed with each of the ambiance attributes.

As seen in Table 5.3, moderate-to-good agreement levels were found for 8 out of the 11 ambiance characteristics (ICC greater than 0.5), but 3 items, namely *festive/fun*, *serious/boring*, and *off-the- beaten-path/unique* had ICC under 0.5. Attributes relating to physical characteristics of the place (*large/spacious*, *cramped/confined*) and its brightness (*dark/badly-lit*, *bright/well-lit*) have the highest agreement ranked as good (between 0.75 and 0.9). This indicates that the ambiances relating to physical attributes are easier to rate than attributes relating to the annotators' judgments on more subjective variables (*Serious, boring, Festive, fun* and *Off-the-beaten-path/unique*). This result is in concordance with the work of Nguyen et al. (Nguyen et al., 2018) on Airbnb personal homes, in that annotation on ambiance requires observers to make abstract impressions, which makes consistent annotation challenging for variables like *festive/fun*, *serious/boring* and *Off-the- beaten-path/unique*. Regardings descriptive statistics, the highest mean values are obtained for *dull/simple* (4.12), *colorful/decorated* (4.02), *large/spacious* (3.89), and *dark/badly-lit* (3.82).

Ambiance Correlation

Table 5.4 displays the Pearson correlation between the annotated ambiances for all home places (N=301). In Table 5.4, we only show correlation above 0.20 and p-value <0.001. From this analysis, we can identify opposing pairs, e.g. *large/spacious* vs. *cramped/confined*, and *dark/badly-lit* vs. *bright/well-lit.* but also observe other effects. All characteristics are associated with some others, with clearly identifiable patterns. First, characteristics related to brightness, namely *dark/badly-lit* and *bright/well-lit*, are uncorrelated to all other ambiance characteristics, suggesting that variations in lightings are independent of the general per-

Table 5.4 – Pearson correlation of ambiance (based on N(video)=301 having p-value <0.001). Entries marked with (*) correspond to correlation <0.20 and p-value >0.001.

Ambiance attributes	[a]	[b]	[c]	[d]	[e]	[f]	[g]	[h]	[i]	[j]	[k]
[a] Large, spacious	-	*	*	-0.92	*	0.42	-0.38	0.23	0.66	0.26	-0.29
[b] Dark, badly-lit		-	*	*	-0.94	*	*	*	*	*	*
[c] Colorful, decorated			-	*	*	0.56	-0.72	0.55	0.28	0.54	-0.56
[d] Cramped, confined				-	*	-0.41	0.35	*	-0.66	-0.21	0.24
[e] Bright, well-lit					-	*	*	*	*	*	*
[f] Comfortable, cozy						-	-0.61	0.48	0.49	0.45	-0.50
[g] Dull, simple							-	-0.67	-0.54	-0.67	0.72
[h] Festive, fun								-	0.29	0.54	-0.68
[i] Sophisticated, stylish									-	0.41	-0.31
[j] Off-the-beaten-path, unique										-	-0.64
[k] Serious, boring											-

ceived ambiance. Second, characteristics of *serious/boring*, *cramped/confined*, and *dull/simple* were all grouped together (i.e., positive correlations between all three characteristics), while characteristics of *large/spacious*, *colorful/decorated*, *comfortable/cozy*, *sophisticated/stylish*, *off-the-beaten-path/unique*, and *festive/fun* were also grouped together.

Co-occurrence of Ambiance and Activities

Figure 5.11 shows the relative distribution of activities for the different types of ambiances. For the figure, each ambiance was binarized, such that each place is associated to a given ambiance only if the average rating over all annotators is above the mean scale (4.0). Overall, 'Drinking', 'Chatting', 'Watching TV', 'Using computer/ tablet/ smartphone' and 'Eating' were the most prevalent activities, independently of the ambiance, although subtle variations can be observed. For example, chatting was more prevalent in unique, large, and sophisticated places, while the use of electronic devices seemed more prevalent in serious, dull, and confined places. The only ambiance that seemed largely different from the others is festive, fun, which showed a lower proportion of watching TV and using electronic devices than the other ambiances.

5.4.4 Automatic extraction of audio and visual descriptors of home environments

Video Preprocessing

We extract visual and audio descriptors of places from the 301 10-second video clips using deep learning. Table 5.5 summarizes the outcomes of the learning models presented below. Following the recommendation in (Keval & Sasse, 2008) to extract at least 8 frames per second using uniform sampling, we extract a total of 29K frames. Meanwhile, we also extract 301 audio files for all videos by using command line FFmpeg (Ffmpeg, 2010).

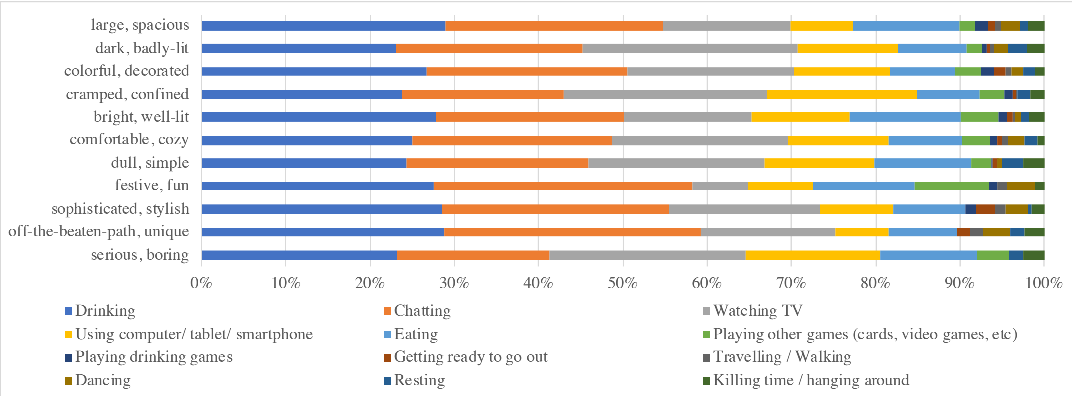


Figure 5.11 – Percentage of all activities co-occur with all ambiances. The order of activities (left-right and top-bottom on the legend), e.g., drinking, chatting, watching TV, using computer/tablet/smartphone, etc. are plotted from the left (0%) to the right (100%) on the stacked bar.

Object Parser

To obtain an object-level description for each video, we used a deep learning model to extract the probability of object appearance in each frame. We applied the Inception-v3 model (Szegedy et al., 2016) trained on the ImageNet Large Visual Recognition Challenge. This model classifies entire images into 1000 classes (e.g. dishwasher, refrigerator, etc.) where the output for each image at the last layer is the probability distribution over all object classes (i.e., the sum of the scores over the 1000 classes is 1.0). The work in (Szegedy et al., 2016) presented the fraction of test images for which the correct class label is not among the top five labels identified by the algorithm, namely "top-5 error rate", reported to be 3.46%. As a result, for each frame, we have a 1000-dimensional vector with each element as a probability. Then, we aggregate them at the video-level over all frames to include all the existing objects by computing, for each class, the maximum probability over the set of video frames.

Scene Parser

To obtain a scene-level description for each video, we extract 365 place classes (e.g. kitchen, living room, etc) using Resnet18 (Zhou et al., 2018) trained on the Place-365 database (Zhou, Lapedriza, Khosla, Oliva, & Torralba, 2017) for each frame. The semantic categories of the place classes are defined by their function, e.g., dressing room for dressing, locker room for storing, etc. As explained online, the database is meant to be used for "high-level visual understanding tasks, such as scene context, object recognition, and action and event prediction." The output of the last layer is a 365-dimensional vector in which the sum of all element values is 1. In order to represent the scene of the full video, we aggregate vectors over all frames of each video by computing the average for each class.

Table 5.5 – Visual and sound extracted features for the video dataset.

Feature Classes	Frame Level (28K frames)	Video Level (301 videos)
1000 classes	Probability distribution over 1000 object classes (Sum of 1000 classes is 1)	Class-specific aggregate for each video: maximum probability over the set of frames for each class. Purpose: obtain a representation of the objects present in the video.
365 scene categories	Probability distribution over the 365 scene classes (Sum of 365 classes is 1)	Class-specific aggregate for each video: average probability over the set of frames for each class. Purpose: obtain a representation of the most likely scene in the video.
527 sounds	Not available	Probability distribution over the 527 sound classes

Sound Parser

To get a scene-level representation of the sounds present in a video, we extract 527 audio classes using Vggish trained on the Audio Set dataset of generic audio events, which has 1.7 million human-labeled 10-second YouTube video soundtracks (Gemmeke et al., 2017). The output of the last layer is the probability of each individual sound detected by the model.

Figure 5.12a, b, c shows the top 30 descriptions extracted for 1000 objects, 365 places, and 527 sound classes, respectively. Overall, most of the identified top objects (e.g. TV, closet, sliding door, etc.), places (e.g. dorm, closet, etc.) and sound (e.g. speech, music, etc.) clearly correspond to home environments. This said, a few unexpected results are worth commenting. First, the first place obtained by category “jail cell” in Figure 5.12b seems strange. However, manual inspection of these images shows that studios with shelves or small rooms can indeed be mistaken with jail cells.

In summary, this section answers RQ1 (consistency of annotation and the main findings from the annotation results and machine-extracted features). In each section of physical/social attributes and ambiances, we present measures, ICC, and main findings . The ICC(2,k) shows that ambiance and physical/social attributes at home (e.g., presentation of home spaces, brightness, music loudness, chatter loudness) can be consistently annotated by external observers. The results also reveal that living room, dining room, kitchen, and bedroom are common places at home where nightlife activities like eating/drinking, entertainment (watching TV or using mobile devices) and chatting happen. Young people at home weekend nights seem to be mindful about the loudness of music and level of chatter. In addition, we found a surprisingly large proportion of videos with no people other than the volunteer, engaged in relatively quiet activities. Although the number of videos contains people do not take a large portion, they describe the gender ratio and their activities as well as their reactions to our participants. Moreover, although there are still unexpected results of extracted objects and scenes, many identified CNN-extracted classes from objects, scenes, and sounds are relevant to home environments. To our knowledge, this analysis of nightlife activities at home, which was enabled by the crowdsensing experience, has not been previously reported.

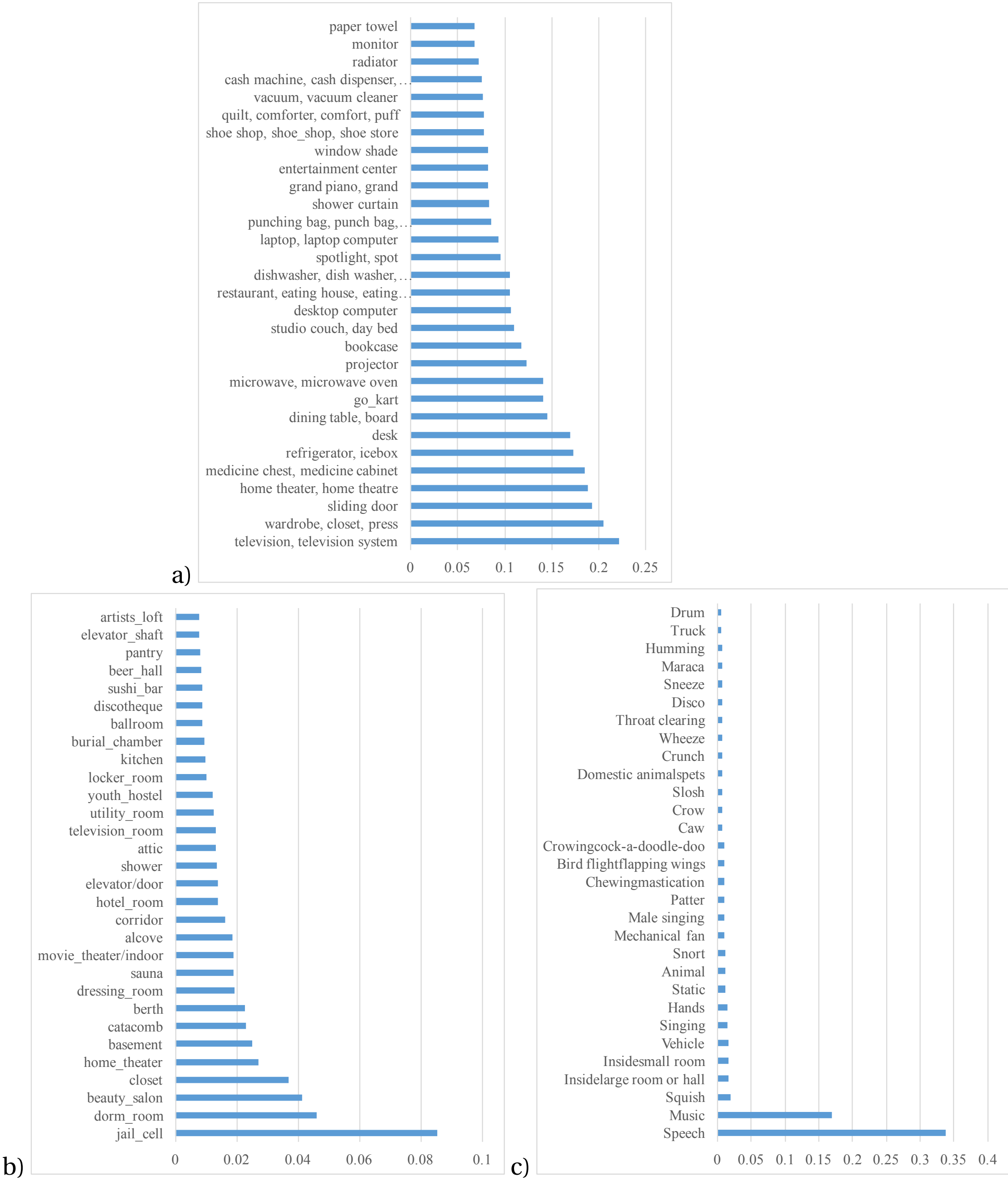


Figure 5.12 – Top 30 features of (a) 1000 object classes, (b) 365 place classes, and (c) 527 sound classes.

5.5 Machine-extracted Features and Ambiance Recognition of Home Spaces (RQ2)

This section describes how machine analysis of the audio-visual tracks of videos can be used to characterize and enrich the understanding of youth home spaces on weekend nights.

5.5.1 Correlation between Machine-extracted Features and Ambiance

This section aims to identify what machine-learning extracted features (1000-object classes, 365-place classes, and 527-sound classes) are correlated with the 11 ambiances categories assessed by the annotators.

Correlation between ambiances and object classes

Correlation results with ambiance are shown in Table 5.6. Only the largest correlations are shown, (i.e. those higher or equal to 0.25 and with p-value < 0.001). Places described as *comfortable/cozy* have couches and beds present in the videos, while *festive/fun* places were positively correlated with eating places and movie places. These results were confirmed by manual inspection of the videos. We also noted that, in a few cases, participants recorded the TV program they were watching as part of their home space videos. This might explain why dark ambiances are correlated with objects like cinema, but also with seemingly random objects like car mirror or grey fox. This is a known limitation of using CNN models trained on datasets which are not specifically designed for home environments (Szegedy et al., 2016). This could make some unexpected objects recognized and associated. Interestingly, object category *"restaurant, eating house, eating place, eatery"* has a positive association with *festive/fun* ambiance, while has a negative correlation with *dull/simple*, and *serious/boring* ambiance.

Correlation between ambiances and scene classes

Correlations between the 365-scene classes and the 11 ambiances categories are shown in Table 5.7. Overall, the results show similar associations to those identified in Table 5.6. For example, a bedroom and a living room associate positively with comfortable ambiance. A dining hall and dining room are positively linked to *large/spacious* ambiance, while pantry or closet do with *cramped/confined* and *dull/simple* ambiance. Results also show that dark and bright ambiances are correlated, negatively and positively, with a large number of scene classes. As mentioned above, participants have sometimes recorded videos of TV programs in dark places at homes, which made the model recognize some places types erroneously, i.e., the places depicted on the TV shows; recall that watching TV was a very popular activity (Figure 5.8).

Correlation between ambiances and sound classes

Finally, we examine the correlation between the 527 sound features with ambiances. Only three correlations were above 0.2 with p-value <0.001. In particular, *festive/fun* ambiance are positively associated with *chewing/mastication* (r = 0.21), which might be explained as the same ambiance categories were associated to people eating (see Tables 5.6 and 5.7). Also, the correlation of *female singing* (0.20) and *techno* (0.20) (both music-related sounds) with *off-the-beaten-path/unique* ambiance could help explain why that ambiance has positive

Table 5.6 – Pearson correlation between ambiance and 1000-object classes limited to classes with Pearson correlation score >=0.25 and p-value <0.001. Negative and positive correlation values are ranked in descending order by absolute correlation value and are shown in red and blue, respectively. Entries marked with (*) correspond to p-value >0.001 and are not discussed.

Ambiance Features	1000-object classes
[a] Large, spacious	medicine chest, medicine cabinet(-0.33), refrigerator, icebox(-0.31)
[b] Dark, badly-lit	cinema, movie theater, movie theatre, movie house, picture palace(0.37), grey fox, gray fox, Urocyon cinereoargenteus(0.32) , suspension bridge(0.30) , hyena, hyaena(0.28) , wing(0.25), badger(0.27) , miniature pinscher(0.26), jack-o-lantern(0.26), desktop computer(0.26), car mirror(0.25) whiptail, whiptail lizard(-0.28), microwave, microwave oven(-0.28)
[c] Colorful, decorated	tobacco shop, tobacconist shop, tobacconist(0.25) restaurant, eating house, eating place, eatery(-0.25)
[d]Cramped, confined	medicine chest, medicine cabinet(0.33), refrigerator, icebox(0.31) grand piano, grand(-0.27)
[e]Bright, well-lit	whiptail, whiptail lizard(0.27), microwave, microwave oven(0.25), dishwasher, dish washer, dishwashing machine(0.25) cinema, movie theater, movie theatre, movie house, picture palace(-0.35), grey fox, gray fox, Urocyon cinereoargenteus(-0.28), suspension bridge(-0.27), badger(-0.26), jack-o-lantern(-0.26), hyena, hyaena(-0.25), theater curtain, theatre curtain(-0.25), wing(-0.25)
[f]Comfortable, cozy	studio couch, day bed(0.29) dishwasher, dish washer, dishwashing machine(-0.28)
[g]Dull, simple	restaurant, eating house, eating place, eatery(-0.29)
[h]Festive, fun	restaurant, eating house, eating place, eatery(0.26), cinema, movie theater, movie theatre, movie house, picture palace(0.25)
[i]Sophisticated, stylish	*
[j]Off-the-beaten-path, unique	lumbermill, sawmill(0.31), dam, dike, dyke(0.27)
[k]Serious, boring	restaurant, eating house, eating place, eatery(-0.26)

correlations in home environments with features like bedchamber (0.42), throne room (0.37), or living room (0.33) in Table 5.7.

5.5.2 Ambiance Inference

This section presents the investigation of whether and how the ambiance of home places can be automatically inferred using machine-extracted features.

Inference task, method, and performance evaluation.

The goal is to infer (in the regression sense) the ambiance of home spaces as perceived by external observers. This inference task uses the aggregated annotations of ambiance discussed in previous sections and is run on the video, which is aggregated as described in the previous section. Random Forest (RF) (Breiman, 2001) is used as a regression model in our inference task. By using RF, multiple decision trees are built up to form various classification outputs. In this experiment, we set parameters ntrees = 500 as recommended by (Liaw, Wiener, et al.,

Table 5.7 – Pearson correlation between ambiance and 365-scene classes limited to classes with Pearson correlation score >=0.25 and p-value <0.001. Negative and positive correlation values are ranked in descending order by absolute correlation value and are shown in red and blue, respectively. Entry marked with (*) corresponds to p-value >0.001 and Pearson correlation score <0.25.

Ambiance Features	365-place classes
[a] Large, spacious	lobby(0.45) , living room(0.35) , restaurant patio(0.30) , dining room(0.28) , dining hall(0.26), waiting room(0.25) closet(-0.30), pantry(-0.26), clean room(-0.26), shower(-0.25)
[b] Dark, badly-lit	catacomb(0.58) , movie theater/indoor(0.52) , barn door(0.51) , alley(0.49), stage/indoor(0.48), ruin(0.47) , orchestra pit(0.46), auditorium(0.45), arena/performance(0.43) , castle(0.43), elevator shaft(0.43), grotto(0.43), mosque/outdoor(0.42), skyscraper(0.42), tower(0.41), , house(0.41) , courtyard(0.41), aquarium(0.41) , cockpit(0.41), downtown(0.39), music studio(0.38), mausoleum(0.38), tree house(0.37) , fountain(0.37), forest path(0.37), water tower(0.36), palace(0.36), temple/asia(0.36), hotel/outdoor(0.36), motel(0.35), office building(0.35), cottage(0.35), volcano(0.35), pagoda(0.35), plaza(0.35) , mansion(0.34) , throne room(0.34), viaduct(0.33) , canal/urban(0.33) , oast house(0.32), arch(0.32), building facade(0.32), church/outdoor(0.31) , aqueduct(0.31), oilrig(0.30), schoolhouse(0.30), waterfall(0.30), amphitheater(0.30) , cemetery(0.29),tree farm(0.29) , lock chamber(0.29), mountain(0.29) , creek(0.28), landing deck(0.28), formal garden(0.27), diner/outdoor(0.27), forest road(0.27) , village(0.27) , home theater(0.27), chalet(0.27), amusement park(0.27), burial chamber(0.27), harbor(0.26), hardware store(0.26), embassy(0.26) , bridge(0.26), parking lot(0.26), campsite(0.26), kasbah(0.26), windmill(0.26), jail cell(0.25), medina(0.25) laundromat(-0.40), kinder garden classroom(-0.35), art studio(-0.35) , pantry(-0.34), clean room(-0.33), nursery(-0.33), beauty salon(-0.32), hunting lodge/outdoor(0.32), playroom(-0.32), art school(-0.31), utility room(-0.30), art gallery(-0.28), storage room(-0.28), veterinarians office(-0.28), department store(-0.28), bathroom(-0.27), classroom(-0.26), office cubicles(-0.26), garage/indoor(-0.26), pet shop(-0.25), reception(-0.25), artists loft(-0.25)
[c] Colorful, decorated	bazaar/outdoor(0.30), throne room(0.28), bazaar/indoor(0.27), bedchamber(0.27), lobby(0.25)
[d]Cramped, confined	closet(0.30), pantry(0.29),clean room(0.27) living room(-0.38), dining room(-0.25), lobby(-0.41), restaurant patio(-0.26), waiting room(-0.29)
[e]Bright, well-lit	laundromat(0.41), clean room(0.36), kinder garden classroom(0.33), art studio(0.33), nursery(0.32), utility room(0.32), pantry(0.30), beauty salon(0.30), art gallery(0.29), playroom(0.29), veterinarians office(0.28), bathroom(0.27), biology laboratory(0.27), artists loft(0.27), department store(0.27), art school(0.27), physics laboratory(0.26), office cubicles(0.26), dressing room(0.25), garage/indoor(0.25) catacomb(-0.55), barn door(-0.48), movie theater/indoor(-0.48), stage/indoor(-0.46), alley(-0.45), auditorium(-0.44), orchestra pit(-0.43), ruin(-0.42), grotto(-0.41), elevator shaft(-0.40), arena/performance(-0.40), aquarium(-0.39), cockpit(-0.39), skyscraper(-0.38), castle(-0.38), music studio(-0.38), tower(-0.37), mosque/outdoor(-0.37), house(-0.36), courtyard(-0.35), throne room(-0.34), forest path(-0.34), mausoleum(-0.34), volcano(-0.34), tree house(-0.34), downtown(-0.34), fountain(-0.33), water tower(-0.33), temple/asia(-0.33), pagoda(-0.32), hotel/outdoor(-0.32), palace(-0.32), motel(-0.32), cottage(-0.31), plaza(-0.31), arch(-0.31), office building(-0.30), oast house(-0.30), canal/urban(-0.29), mansion(-0.29), waterfall(-0.28), aqueduct(-0.28), mountain(-0.28), cemetery(-0.28), viaduct(-0.28), hunting lodge/outdoor(-0.28), building facade(-0.28), oil rig(-0.27), burial chamber(-0.27), schoolhouse(-0.27), amphitheater(-0.27), landing deck(-0.26), church/outdoor(-0.26), lock chamber(-0.26), amusement park(-0.26), campsite(-0.26), tree farm(-0.26), forest road(-0.26), creek(-0.25), canyon(-0.25), home theater(-0.25)
[f]Comfortable, cozy	living room(0.35), bedroom(0.28), hotel room(0.26) pantry(-0.36), laundromat(-0.31), bedchamber(-0.29), clean room(-0.28)
[g]Dull, simple	alcove(0.26), closet(0.26) lobby(-0.30), throne room(-0.28), living room(-0.25)
[h]Festive, fun	discotheque(0.29), auditorium(0.28), stage/indoor(0.26)
[i]Sophisticated, stylish	lobby(0.37), roof garden(0.31), restaurant patio(0.30), living room(0.28) closet(-0.26)
[j]Off-the-beaten-path, unique	bedchamber(0.42), throne room(0.37), living room(0.33), bazaar/outdoor(0.32), bazaar/indoor(0.30), market/indoor(0.29), diner/outdoor(0.28), lobby(0.28), sandbox(0.28), junkyard(0.27), stable(0.26), pavilion(0.25)
[k]Serious, boring	*

Table 5.8 – Pearson correlation between ambiances and 527-sound classes with Pearson score >=0.20 and p-value <0.001. Entries marked with (*) correspond to p-value >0.001 and are not discussed.

Ambiance Features	527 sound classes
[a] Large, spacious	*
[b] Dark, badly-lit	*
[c] Colorful, decorated	*
[d]Cramped, confined	*
[e]Bright, well-lit	*
[f]Comfortable, cozy	*
[g]Dull, simple	*
[h]Festive, fun	Chewing mastication(0.21)
[i]Sophisticated, stylish	*
[j]Off-the-beaten-path, unique	Female singing(0.20), Techno(0.20)
[k]Serious, boring	*

2002). We ensure that the train and test set take 80% and 20%, respectively. We also apply 5-fold cross validation for training phase. After obtaining RF trained models, we quantify the performance by using Pearson's correlation coefficient (r), and the coefficient of determination (R^2). In the context of our RF model, R^2 measures how much variance in ambiance is explained by the RF model.

Experiment and results

We randomly divide the 301 videos into two subsets: 80% (241 videos) for training and 20% (60 videos) for testing. We apply RF on 241 videos for training with 5-fold cross validation. The evaluation of RF model is shown in Table 5.9. We observe that the audio features are not capable of improving over a simple prediction of the mean score ($R^2 \sim 0$). In contrast, using 1000 object classes can infer certain ambiances of home spaces with $R^2 > 0.2$, namely *large/spacious, dark/badly-lit, bright/well-lit, and dull/simple*. The highest R^2 obtained is 0.44 for *bright/well-lit*. Meanwhile, the rest of ambiance categories cannot be inferred by the object representation. Recall that three of these ambiance categories (*festive/fun, serious/boring, off-the-beaten-path/unique*) had not reach sufficient ICC agreement (Table 5.3), but we decided to include the results for purposes of completeness. Regarding the 365-scene classes, five of the eleven ambiance variables (*large/spacious, dark/badly-lit, bright/well-lit, cramped/confined, and comfortable/cozy*) are predicted by using 365-scene classes with $R^2 > 0.2$ ($R^2 = 0.69$ for dark/badly-lit). In Section 5.5, we discussed the correlation of ambiances and scenes. Clearly, certain scenes can predict those ambiances related to space capacity (*large/spacious vs cramped/confined*), and brightness (*bright/well-lit vs dark/badly-lit*). For *comfortable/cozy* ambiance, Section 5.5 also showed that living room with couch, and bedroom with bed, have positive correlation. Interestingly, two of the ambiance variables (*colorful/decorated* and *sophisticated/stylish*) could not be inferred by any of the visual representations, regardless of the fact that they achieved good inter-annotator reliability, (0.66 and 0.65, respectively, see Table 5.3).

Table 5.9 – Inference results including Pearson's correlation coefficient (r), coefficient of determination (R^2). All R^2 with score >= 0.20 are shown in bold font. Rows marked with (*) correspond to ambiance categories that did not reach sufficient annotator agreement (ICC).

Feature Groups	**127 sound classes**		**1000 object classes**		**365 scene classes**	
	r	R^2	r	R^2	r	R^2
[a] Large, spacious	0.07	0.005	0.47	**0.23**	0.52	**0.27**
[b] Dark, badly-lit	0.08	0.01	0.66	**0.43**	0.83	**0.69**
[c] Colorful, decorated	-0.13	0.02	0.24	0.06	0.31	0.10
[d] Cramped, confined	0.03	0.001	0.44	0.19	0.56	**0.31**
[e] Bright, well-lit	0.02	0.0005	0.67	**0.44**	0.79	**0.63**
[f] Comfortable, cozy	-0.03	0.0007	0.36	0.13	0.46	**0.21**
[g] Dull, simple	0.002	0.000006	0.48	**0.23**	0.44	0.19
[h] Festive, fun (*)	0.17	0.03	0.12	0.01	0.31	0.09
[i] Sophisticated, stylish	0.04	0.001	0.24	0.06	0.25	0.06
[j] Off-the-beaten-path, unique (*)	-0.09	0.008	0.12	0.02	0.28	0.08
[k] Serious, boring (*)	-0.02	0.0005	0.32	0.11	0.37	0.14

In summary, we use RF to train a regression model and use R^2 as the main measure to evaluate which features can predict the ambiance of a home space. Our findings show that six of the ambiance categories can be inferred with R^2 in the $[0.21, 0.69]$ range (four with object-based features, and five with scene-level features), and with higher R^2 values when a scene deep network is used. More specifically, space capacity (*large/spacious vs cramped/confined*), brightness (*bright/well-lit vs dark/badly-lit*), *comfortable/cozy*, and *dull/simple* can be predicted by object-level and scene-level description. In contrast, audio features were not effective at inferring ambiance.

5.6 Discussion and implications

Table 10 summarizes our main findings for RQ1 and RQ2. We now discuss the results and some of their implications.

5.6.1 Crowdsourcing as an Alternative to Collect Video

In terms of data source, collecting data on home environments via crowdsourced videos is novel in comparison to previous work using social media sources. This includes research on Foursquare, which showed that users underreported home presence by checking into homes considerably less frequently than into other places, given the logic of such social network (Cramer et al., 2011; Lindqvist, Cranshaw, Wiese, Hong, & Zimmerman, 2011b); and also includes recent work on Airbnb, which is known to feature photos of homes that are taken with the explicit purpose of attracting possible guests, in some cases taken by professional photographers (Nguyen et al., 2018). Our study used 301 ten-second video clips of young people's home spaces on weekend nights. To our knowledge, this is a unique dataset of real-life home environments that cannot be compared to any other publicly available dataset, in the sense that participants' showed their home spaces simply as they are (with no artistic filters or advertising intentions) on their weekend nights.

From the total set of 843 videos collected in the study, slightly less than one third were consistently identified as representing homes by the five annotators. Yet, this does not mean that only one-third of the nights were spent at home, but rather it can be seen as a consequence of the study design, which requested participants to provide only one video per night if they did not change location during the night. Given that about half of all drinks (non-alcoholic or alcoholic) in the Youth@Night dataset were documented in homes (Labhart et al., 2019a), this result suggests that participants were less likely to change locations when starting the night at home than when going out (Labhart et al., 2013), highlighting the relevance to research and understand what happens in this usually hidden or hard-to-reach kind of environment. In addition, the levels of inter-annotator agreement for most of the physical attributes at homes were globally good to excellent. This result echoes previous work in psychology (Samuel D Gosling et al., 2002) that found that personal environments elicit similar impressions from independent observers, while adding the novel angle of using short video as stimuli (rather than photos). This result also indicates that, despite being relatively short, 10-second videos are long enough to provide adequate cues of the physical and social environment, the ongoing activities, and the ambiance.

5.6.2 Home as a Nightlife Space

As mentioned above, about one third of the Y@N videos were recorded in homes, and participants were less likely to change locations when starting the night at home than when going out. This highlights the need to understand this particular environment. Qualitative feedback from the participants at the end of the fieldwork echoed previous research that has found that homes can serve both as ‘prequel night out spaces’, where young people meet, dress up, and get ready for the night out, as well as a standalone nightlife space where they hang out with friends or have parties (Lincoln, 2012). For one participant, home was his main nightlife destination: “Now that I study in Lausanne and live here, when I go out it’s really to other people’s place or at my place. Which still does not prevent me from going out [to pubs and clubs] now and then”. Another participant mainly conceived home as the starting point of the night: “Well, when I go out, I prefer drinking before going out, well, not before going out but, let’s say we meet with friends and we go to someone’s place to drink or just eat and we drink something, or in a park during the summer, yeah, let’s say I start drinking [in a residential neighborhood] and then we move on and continue the party downtown”. Finally, several participants considered the home as an alternative to commercial nightlife venues: “For me, there are two types of nights out: the dancing ones, when we go to clubs and the point is to dance [. . .] and then there are the quiet ones, when we just sit, at someone’s place or in a bar, and we talk and that’s it” or “There are different kinds of nights out. Sometimes, people want to go out to meet others and that’s it, it all depends on the mood we are in that night. It’s true that sometimes we enjoy staying with friends and have big parties in homes, or go out in the city, but as a small group.”

In order to better represent home environments, the annotation task developed for this study revealed detailed attributes of physical and social environment, including the types of rooms attended, levels of brightness, loudness and occupancy, the number and gender of people, and the ongoing activities. Altogether, this information provides a comprehensive picture of young people's nightlife environments. Specifically, we examined co-occurrence between activities and levels of occupancy and types of spaces at home. The authors of (Alitajer & Nojoumi, 2016; Baillie & Benyon, 2008; Tucker, 2010) studied usage of domestic spaces that were used in daily life activities, and specific psychological states (e.g., mental stress). In our research, home spaces were analyzed from the perspective of activities of young people on weekend nights. Through physical and social attributes, we have insights of activities in the context of Swiss young people (16-25 year-old), who present differences to other populations, e.g. in the US, where legal drinking age and norms about the use of the public space differ from those in Europe. We found that young people spent weekend night time watching TV, listening to music (Schiano et al., 2007), and playing games. Previous findings about pre-drinking before going out in (S. Wilkinson & Wilkinson, 2018) or drinking at friends' or family's homes (Holloway et al., 2008) were also partly shown in our work.

One particular instrument of the annotation task, the ambiance scale, aimed to capture the different dimensions of this construct. Dimensions related to the physical space (e.g. large, spacious), which could be rated rather objectively by the annotators, showed a high degree of agreement among them. Dimensions relating to the personal evaluation of the annotators (e.g. off-the-beaten-path) were indeed more subjective and showed a lower degree of agreement among the annotators. From the correlation analysis, three main groups of ambiances were identified: positively perceived characteristics (large, colorful, festive, stylish, and unique), negatively perceived characteristics (cramped, simple, and boring) and independent characteristics (dark and bright). In addition, the main types of ongoing activities were consistent across ambiance categories (drinking, chatting, watching TV, and computer device use), and small variations were found, e.g. less TV watching for the fun/festive ambiance.

While the aim of the annotation task was to describe home spaces from the perspective of human annotators, the aim of the machine learning task was to observe home spaces through automatic-extracted features using CNNs models, (Gemmeke et al., 2017; Szegedy et al., 2016; Zhou et al., 2018). Thus, without using external annotation of physical and social attributes, the latter task was able to automatically describe home spaces by observing the probability distribution of visual and audio labels. Correlation results between automatically extracted features based on the image frames of the videos and ambiance labels provided promising results for the visual cues (i.e. objects or scenes) from the videos. Yet, results also showed that the existing classes are made to recognize all kinds of objects or situations, even some that are not supposed to be in homes, such as jail cell, car parts, etc. Future research is clearly needed for the development of a specialized dictionary of classes focused on home environments.

Regarding automatic-extracted features based on the soundtrack of the videos, however, only two of the sounds dominated the dataset (speech and music), and thus only a few associations were found with ambiance features. These might be related to the way audio was recorded, but also because homes at night are generally quiet or because not enough information was found in the sound measure in (Gemmeke et al., 2017).

5.6.3 Feasibility of Ambiance Inference

We examined the use of machine-extracted features, i.e., 527-sound, 100-object, 365-scene features, for automatic inference of ambiance. As a result, *large, dark, bright, confined, comfortable*, and *simple* ambiances could be inferred by using object and scene classes. These ambiances could be also be perceived by people while unique, festive, and boring ambiances could be ambiguous when being annotated by humans. Comparing these inference results to those reported in (Nguyen et al., 2018) on Airbnb home photos, our results corroborate that ambiances closer to physical attributes reach better recognition performance, although the performance we obtained is lower than that obtained on Airbnb data for three variables (large, comfortable, and simple), similar for one variable (confined); and higher for two variables (bright, dark). Note that in addition to the datasets being different, the specific CNN models and the CNN outputs used as features are different too (last convolutional layer in (Nguyen et al., 2018) vs. final output equal to the number of objects or scenes in our work). Note also that we made this choice in order to interpret the CNN-derived features in the correlation analysis in Section 5.5.1. For future work, we believe that regression performance could be improved by CNN adaptation, i.e., by fine-tuning the last CNN layers to the ambiance target class as demonstrated in other visual tasks (Long, Cao, Wang, & Jordan, 2015). Home ambiance recognizers built around short duration mobile videos could be advantageous as they might in general contain more information than still images, and used in future applications, as discussed in the next subsection.

5.6.4 Implications for future research

We conclude this section by discussing some of the implications of our work for social computing research.

Understanding youth practices at home from mobile crowdsourced data. Using crowdsourced personal videos as input, we showed that a mixed methodology combining manual annotation and automatically extracted features enabled an in-depth study of youth personal spaces on weekend nights with respect to physical attributes, activities, and social attributes, including joint patterns of activities and places. Crowdsourced visual datasets like the one used here complement another common source of data used in research, namely social media. While early research showed that the home environment was infrequently reported or talked about by users (Cramer et al., 2011; Lindqvist et al., 2011b), future research could investigate whether certain sub-communities specifically depict nightlife in private spaces, and what specific practices are promoted or enacted around this theme, including ephemerality, self-representation, and sociality. This investigation would require the use of mixed methods of inquiry, combining machine analyses with user interviews and surveys. Furthermore, given that the concept of nightlife is broad and encompasses both the private and public spheres, a second promising line of future work could investigate the interplay between private and public spaces in urban nightlife, and how this is expressed digitally both in crowdsourced campaigns and social media. For instance, recent qualitative work showed that several participants in the Youth@Night campaign coordinated nightlife activities via Whatsapp (Truong, 2018a). This research could benefit from previous literature on coordination of action and social participation.

Applications of home ambiance recognition. Our work on recognition of ambiance at home also has potential implications for future work. First, it is evident from our study that state-of-art deep visual learning systems, while useful, still generate erroneous visual descriptors. We believe that it is important to make these limitations explicit to inform other researchers who plan to use deep learning as a toolbox for their future work. At the same time, in a fast-moving domain, it is not unreasonable to expect progress that could mitigate some of the current limitations, and thus to anticipate that the shown recognition performance will be improved (e.g., Facebook has published results on deep learning models trained on 1 billion Instagram images) (Mahajan et al., 2018). With this, one could envision applications in home supporting systems. Homes are reconfigurable spaces, in which certain elements can be readily changed (decoration, spatial organization of furniture, light, and music). A system able to recognize ambiance could also make recommendations of suitable ambiances at home for specific activities, e.g. to promote socialization. This kind of work would require human-centered approaches to design such prototypes, integrating perspectives of privacy, ethics, and transparency, all of whom are active topics of investigation in CSCW and social computing (Badillo-Urquiola et al., 2018; Lau, Zimmerman, & Schaub, 2018; Wong, Mulligan, Van Wyk, Pierce, & Chuang, 2017; Zong & Matias, 2018).

Future work on human factors in home research. Human factors play important roles requiring interdisciplinary researchers (social psychologists, sociologists, anthropologists, computer scientists) to find appropriate methods for an individual or group to adopt technology into their daily activities. In this case, technology plays a supporting role while the human factor plays a central one. In our work, we focused on home environments and inferring home ambiance from videos. Besides physical and social attributes, emotional states and nightlife behaviors and their links to ambiances could need the expertise of other researchers (Petrill, Pike, Price, & Plomin, 2004; Scherer, Zentner, et al., 2001). In future work, youth practices at home and ambiances, technologists could collaborate with specialists in interior decoration art, or psychologists, to build systems to support people to link their home ambiances to their current emotions as well as their behaviors. Beyond building this technology, users would increase their self-awareness about their home ambiances and their own behaviors to promote positive changes and share them with others.

Table 5.10 – Summary of findings related to our two RQs.

RQ	Factors	Message
RQ1 - Physical and Social Attributes	Home spaces	The most attended type of room is the living room; followed by bedroom; kitchen/dining room were also frequently attended rooms at night
	Brightness	It tends to reduce from early night to late night
	Music loudness	Videos contained no music on 76% of all situations
	Chatter loudness	Home are mostly quiet with slight increase from 8PM to 11PM
	Occupancy and number of people present	Around 60% of videos contained people gathering from 8 PM to 11 PM; then reducing after 11 PM
	Gender	A gender-matching pattern is evident: female participants tend to gather more with other women, and conversely for male participants. Mixed groups, however, also occur.
	Activities	Drinking, chatting, watching TV, using smartphones/computer, and eating are the most popular activities of young people on weekend nights.
RQ1 - Ambiances	Agreement on ambiance	8 of the 11 ambiance variables have ICCs above 0.5.
	Correlation between ambiances	Place ambiances are grouped on two main opposite dimensions, namely places seen as large, colorful, comfortable, festive, stylish, unique; versus places seen as confined, simple, boring. Dark and bright ambiances do not have a significant correlation with the rest of ambiances.
RQ1 - Machine-extracted Features	Automatic description	1000-object, 365-scene, 527-sound auto-extracted features can express ambiances but with a certain level of noise, because labels of these classes for CNN models are not specifically designed for homes.
RQ2 - Ambiance Regression	Correlation between ambiance and automatic descriptions	Although there are some limitations on the labels of CNNs model, automatic-extracted features have reasonable correlation with ambiances.
	Regression Performance	Six of the ambiance variables (large, dark, bright, confined, comfortable, simple) can be inferred by using object and scene features with coefficient of determination above 0.2. For the other five variables (including three with low ICC). regression performance is low.

5.7 Conclusion

In this chapter, we presented an original study of the characteristics of night personal spaces, including manual coding of places, machine extraction of acoustic and visual description of places, and inference of ambiance of homes of young people in the weekend night setting. We conclude by revisiting the research questions posed at the beginning of the chapter.

RQ1: *Given crowdsourced videos recorded at home spaces by young people at night, what patterns of physical and ambiance attributes of youth home spaces can be revealed by manual coding of videos using external annotators and machine-extracted features?* By describing measures, discussing ICC, and showing results, we sequentially analyzed the problem from physical/social attributes (home spaces, brightness, loudness, human presence, activities) to ambiances. We observed co-occurrence between activities and spaces at homes as well as ambiances. Then, we showed that ambiances could be grouped into two clusters: “unlike” characteristics with *serious/boring*, *cramped/confined*, *dull/simple*, and “like” characteristics with *Large/spacious*, *colorful/decorated*, *comfortable/cozy*, *sophisticated/stylish*, *off-the-beaten-path/unique*, *festive/fun*. Finally, we used state-of-the-art pre-trained deep learning models to extract automatic features to represent videos, namely objects, scenes, and sounds. Most machine-extracted classes relevantly characterize home environments, but there were some unexpected features.

RQ2: *What do machine-extracted features of videos reveal about physical attributes of youth home spaces? Can these machine-extracted features infer the perceived ambiance of such spaces?* Correlations between ambiance and automatic features potentially show the feasibility of using machine-extracted features to automatically describe home spaces, although there are certain limitations. Regarding the inference task, ambiances like space capacity (*large/spacious* vs. *cramped/confined*), brightness (*bright/well-lit* vs. *dark/badly-lit*), *comfortable/cozy*, and *dull/simple* can be inferred for private spaces in the weekend nights by using 1000 object classes and 365 scene classes. The total number of videos (N=301) could be a limitation for mo0del training in the automatic inference experiments. However, our results show that six of the ambiance categories can be inferred with R^2 in the $[0.21, 0.69]$ range, and with higher R^2 values when a scene deep network is used.

Chapter 3 uses Instagram to understand drinking practice with limitations on using social networks. Chapter 4 and this chapter use mobile crowdsensing to understand the heavy drinking and private nightlife of young people, respectively. The clear limitation of using crowdsensing is a small number of participants with the high paid expense, while social media has noisy data (e.g., semantic hashtags, or local/non-local users). In the next chapter, we will investigate how alcohol consumption is studied on a combination of mobile crowdsensing and social media.

6 Understanding Alcohol Consumption from Crowdsourced Data Sources

The design of computational methods to recognize alcohol intake is a relevant problem in ubiquitous computing. While mobile crowdsensing and social media analytics are two current approaches to characterize alcohol consumption in everyday life, the question of how they can be integrated, to examine their relative value as informative of the drinking phenomenon and to exploit their complementarity towards the classification of drinking-related attributes, remains as an open issue. In this chapter, we present a comparative study based on the two data sources studied in previous chapters. Our contributions are two-fold. First, we conduct data analyses that uncover temporal, spatial, and social contextual patterns of alcohol consumption on weekend nights as represented by both crowdsensing and social media. This comparative analysis provides a contextual snapshot of the alcohol drinking practices of urban youth dwellers. Second, we use a machine learning framework to classify individual drinking events according to alcohol and non-alcohol categories, using images features and contextual cues from individual and joint data sources. This chapter uncovers important patterns in drinking behaviour across these two datasets and the results of study are promising towards developing systems that use machine learning for self-monitoring of alcohol consumption. The material of this chapter was originally published in (Phan, Muralidhar, & Gatica-Perez, 2019b).

6.1 Introduction

Alcohol consumption in excess can lead to many adverse consequences, including violence and accidents (Gmel et al., 2003; Rehm et al., 2003; Taylor et al., 2010). These are serious problems in many countries, concerning both policymakers and the public (Ali et al., 2012; Measham & Brain, 2005). From the perspective of urban nightlife security, understanding patterns of drinking in urban areas is a relevant subject. Previous research (Chatterton & Hollands, 2003; Demant & Landolt, 2014) has shown that young people drink alcohol both indoors (at home, bars, nightclubs) and outdoors (in parks or open streets). In the past, scientists investigating alcohol consumption traditionally conducted face-to-face interviews or used paper-and-pencil questionnaires to collect data. These methods have limitations, including recall limitations and the issue of scale (Ekholm, 2004; E. Kuntsche & Labhart, 2012). On the other hand, the booming of smartphones and social media are opening new channels for investigating alcohol consumption trends, through the integration of methods from social science and computer science.

Ubiquitous computing research has studied alcohol consumption in the last years from two main perspectives. In the first one, machine learning methods involving sensor data from a variety of wearable devices (wristbands, smart watches, earbuds, glasses) have been developed to detect fluid intake moments (Rahman et al., 2014; Thomaz et al., 2015), including alcoholic beverages (Arnold et al., 2015; Bae et al., 2017; Kao et al., 2012). These studies, using accurate on-body sensors and fine-grain temporal resolution, have typically involved a couple of tens of users in experiments that have ranged from controlled (Mirtchouk et al., 2016; Thomaz et al., 2015) to in-the-wild settings (Cordeiro, Bales, Cherry, & Fogarty, 2015b; Zepeda & Deal, 2008). The second direction in ubicomp comes from crowdsensing (Y. Chon et al., 2012; T. Yan et al., 2009). These works have used a combination of smartphone sensing and geolocalized human-generated input to collect data about everyday life habits, which has been used to document alcohol intake for a couple of hundred volunteers in (Santani, Biel, et al., 2016) and (Santani et al., 2017). This line of work focuses on understanding the real-life context in which drinking occurs, as a first step towards automatic recognition of drinking events, and can be seen as an evolution of more traditional methods in alcohol research, which have collected survey data via SMS on feature phones (E. Kuntsche & Labhart, 2013), as well as ubiquitous health research on food diaries and diet monitoring (Biel et al., 2017; Mirtchouk et al., 2016; Thomaz et al., 2015).

Mobile crowdsensing shares many aspects with the way in which social media is generated on Twitter, Facebook, and Instagram, involving deliberate human actions to take photos and add hashtags, combined with phone sensing to geolocalize content. In particular, social media research has been conducted to reveal a few aspects of alcohol consumption, mainly with data from the US or the UK. By using the location of users or their tweets, various alcohol intake patterns in urban and suburban areas and local and national levels can be extracted and compared to available public data (Hossain et al., 2016; Kershaw et al., 2014a; Pang et al., 2015). Social media is generated at a global scale and potentially allows for studies with tens or hundreds of thousands of individuals. On the other hand, it is known that as a data source of behavioral traces, social media have limitations in terms of population bias and sparse temporal resolution.

To our knowledge, an important issue in ubicomp that has not been previously investigated is how these two threads of research (crowdsensing and social media) can be integrated, both to improve our understanding of their similarities as data sources informing the same phenomenon (alcohol consumption in everyday life), and to exploit their potential complementarities in machine learning approaches to classify drinking-related attributes. In this chapter, we investigate this problem through two research questions:

RQ1: Given comparable datasets (with respect to country and time) generated by mobile crowdsensing and social media, what alcohol consumption patterns can be extracted with respect to temporal, place, and social context? What are the similarities and differences between these patterns depending on the data source?

RQ2: Can such complementarity be used to improve the performance of classification tasks applied to individual alcohol-consumption episodes, namely to infer alcohol categories?

The contributions of this chapter are the following:

1. As a social media source, we curate new datasets of Instagram posts related to alcohol consumption in Switzerland . The resulting dataset includes alcohol-related hashtags, standardized alcohol categories (wine, beer, spirits), timestamps, venue information and other metadata, and images. As a mobile crowdsensing source, we use the Youth@Night dataset first presented in (Santani, Biel, et al., 2016), which comes from a large-scale mobile crowdsensing campaign that aimed at capturing nightlife activities (including drinking) of young people over three months in Switzerland.
2. We conduct new data analyses to reveal temporal, spatial, and social contextual patterns of alcohol consumption on weekend nights as captured by both crowdsensing and social media. Our analysis produces several relevant findings, including: (1) as a result of their respective population biases, both data sources show differences from general-population official statistics with respect to the distribution of consumed alcohol categories, with a significantly increased consumption of hard alcohol on weekend nights; (2) both sources show very similar temporal patterns, with a predominance of hard alcohol to be consumed in the late night; (3) crowdsensing can capture a significant amount of drinking in personal places like homes, while this practice is not common on Instagram; and (4) both sources show very similar social context patterns, with a strong predominance to drink with friends on weekend nights. These results paint a rich picture of young urban night dwellers consuming alcohol in the target country, and complement results from more traditional studies, which usually do not investigate time-specific trends.
3. Based on this analysis, we use a machine learning framework to classify individual drinking events according to alcohol categories, using contextual cues and image features from individual and joint data sources. We achieve classification accuracies in the range of 80% and 90% for the classification of alcohol categories and alcohol/non-alcohol, respectively, with contextual features being less discriminant than visual features. Combining the two datasets produces either marginal or no improvement in classification performance, which we hypothesize is due to the noisier nature of Instagram data. These results are encouraging for developing an automatic system towards self-monitoring of alcohol consumption.

This chapter is organized as follows. Section 6.2 discusses related work. Section 6.3 presents the crowdsensing and Instagram datasets. Section 6.4 introduces the alcohol drink categories defined in our study. Section 6.5 presents the in-depth analysis of alcohol categories with respect to temporal, spatial, social, and occasion-related patterns for both crowdsensed and social media data sources. Section 6.6 presents and discusses the work on classification of alcohol categories and alcohol/non-alcohol episodes. Section 6.7 discusses the results. Section 6.8 concludes the chapter.

6.2 Related Work

In this section, we review work related to alcohol consumption research in social media and ubiquitous computing.

6.2.1 Social Media and Alcohol Consumption

Methods for automatic estimation of alcohol volume sales from Twitter streams were studied in (Culotta, 2013). Drinks, as a broad category next to fast food and slow food, were studied in (Silva et al., 2014) in the context of differences in check-in habits in food/nightlife Foursquare venues. These two works did not analyze types of alcoholic drinks in any detail. Other works on Twitter and Instagram that analyze food patterns have included alcoholic drink hashtags as part of their analyses but did not investigate specific alcohol drinking patterns (Abbar et al., 2015; Mejova et al., 2015). Chapter 2 presented a descriptive analysis of patterns of food consumption on Instagram in the same target country, in which non-alcoholic and alcoholic drinks are treated as generic consumed items. This chapter significantly extends Chapter 2, both with an in-depth analysis of alcoholic drink types, and with an assessment of the objective utility of Instagram data to improve automatic inference of attributes from crowdsensed data.

In other work, (Marczinski et al., 2016) designed a 10-item alcohol-related Facebook activity (ARFA) questionnaire which asked 146 college students for past 30 day postings related to alcohol use and intoxication. By using regression analyses, they revealed that the ARFA scores were important predictors of recent drinking behaviors. In addition, they proposed that the ARFA scale can be a good tool for recognizing risky alcohol use. Boyle et al. (Boyle, LaBrie, Froidevaux, & Witkovic, 2016) attempted to predict possible drinking from manual coding of past alcohol-related content on social media sites (Facebook, Instagram, Snapchat) in a group of 408 first-year students during 6 weeks of college. This is different with respect to our work, which uses automatic feature extraction for classification of alcohol categories and alcohol/non-alcohol. This previous research also also identifies that drinking with close friends, and drinking with a group of people of different gender could be predictive of later alcohol use, in case of not having previous information about alcohol-related posts on social media sites.

On Twitter, several works have used predefined alcohol-related hashtags to filter tweets containing alcohol items and to extract basic patterns (Cavazos-Rehg, Krauss, Sowles, & Bierut, 2015; Hossain et al., 2016; Kershaw et al., 2014a). Several of these works (Hossain et al., 2016; Kershaw et al., 2014a) assume that all tweets containing alcohol-related hashtags are events of alcohol consumption or alcohol use. However, (Cavazos-Rehg et al., 2015) proposed a more nuanced approach by manually classifying the themes of 5000 alcohol-related tweets into pro-alcohol tweets (79%), anti-drinking tweets (7%), and neutral tweets (13%). In addition, by extracting the location of users or location of tweets, various patterns of alcohol consumption in urban/suburban areas and local/national levels are discovered and compared to available public data in (Hossain et al., 2016; Kershaw et al., 2014a).

Finally, Instagram photos can give visual cues to complement alcohol hashtags. The work in (Pang et al., 2015) extracted Instagram user demographics (age, gender, and race) by using Face++ (an online face processing toolkit) to automatically detect and analyze faces in selfie photos. The work then used a hashtag-driven approach to identify posts related to alcohol consumption in terms of time, under-age drinking patterns (using the demographics above), location-specific drinking patterns, and youth exposure to alcohol media. This chapter also uses visual analysis of Instagram photos to extract content cues, but instead uses this information to investigate how social media datasets can complement mobile crowdsensing datasets and to automatically classify alcohol categories in posts.

6.2.2 Ubicomp and Alcohol Consumption

As a kind of hand-to-mouth activity, alcohol drinking is related to ubicomp research on recognition of eating and smoking occasions. Work on recognition of eating occasions includes (Amft & Troster, 2009; Mirtchouk et al., 2016; Thomaz et al., 2015). First, Amft et al. (Amft & Troster, 2009) used multiple embedded sensors to collect data of an individual's food intake, which was used for building up personalized weight-loss plans for users. In the work of (Thomaz et al., 2015), Thomaz et al. implemented and evaluated methods to predict eating moments based on 3-axis accelerometer data from smartwatches. They collected data with 20 subjects of 2 groups and recognized eating moments in two conditions: 7 participants in 1 day (F-score: 76.1%) and 1 participant in 31 days (F-score: 71.3%). In the same line of research, Mirtchouk et al. (Mirtchouk et al., 2016) automatically estimated food type and amount of consumed food by using body-worn audio and motion sensors. Using 40 unique food types, the work reported a classification accuracy of 82.7% with a combination of sensors, compared to 67.8% for audio alone, and 76.2% for head and wrist motion.

There are also specific works studying drinking occasions (Bedri et al., 2017; Rahman et al., 2014). Bedri et al. (Bedri et al., 2017) detected eating and drinking moments by using a wearable system. They discovered that inertial sensing is the best to detect eating events compared to two other features (optical and acoustic sensing features). The system can recognize chewing instances with the accuracy of 90.1% in a semi-controlled lab study, compared to 93% for outside-the-lab evaluation. Towards recognizing eating and drinking by using sound, Body-

Beat (Rahman et al., 2014) is a mobile sensing system capturing and detecting non-speech body sounds (food intake, breath, laughter, cough), that uses a microphone attached to a 3D printed neckpiece.

Other ubicomp works have focused on developing approaches to collect food and drink diaries, which have some similarities with photo-taking and tagging practices in social media (Biel et al., 2017; Cordeiro et al., 2015b; Zepeda & Deal, 2008). Zepeda et al. (Zepeda & Deal, 2008) study logs of food through photo and text as interventions to create awareness and change dietary habits with 43 participants. Their results reveal that photo food consumption diaries can alter attitudes and food choice behaviors more than written diaries. Cordeiro et al. (Cordeiro et al., 2015b) examine and discuss the benefits of lightweight photo-based capture in mobile food journals, compared to other existing journaling methods. Biel et al. recently studied a cohort of 128 college students who collected in-situ diaries of food, drinks, and context using a mobile application (Biel et al., 2017).

There are also approaches to detect smoking episodes (Chatterjee et al., 2016; Scholl, Kücükyildiz, & Laerhoven, 2013). Scholl et al. (Scholl et al., 2013) evaluated two different prototypes which detected smoking behaviors based on augmenting a lighter. They studied the deployment of the systems, with 11 frequently-smoking participants. Interestingly, smokers were not aware of their daily smoking patterns and often mis-estimated their consumption. Chatterjee et al. (Chatterjee et al., 2016) developed a model to estimate cigarette craving (during smoking abstinence) by using mobile sensor data. They found that stress and self-reported high craving are associated. By using this feature and other insights, they infer craving probabilities. In public health research, smoking can be associated with alcohol drinking (Ma, Betts, & Hampl, 2000). Ma et al. examined relationships between cigarette smoking, alcohol use, and dietary intake. They reported that the use of cigarettes and alcohol are closely related (Ma et al., 2000).

Note that unlike these works, which exploit sound, motion, or inertial sensors, in this chapter we are interested in contextual and visual cues for recognition like semantic location, social context, and visual features. Furthermore, in some of these works there is also an interest in automatically detecting the moment of the eating/drinking/smoking occasion, whereas in our case we use timestamps voluntarily provided by people at the moment of documenting a drink (in the case of crowdsensing) or posting a photo (in the case of Instagram).

In the ubicomp literature, various alcohol consumption patterns have been studied by mobile crowdsensing, which collect data from installed phone applications and wearable devices. Researchers have asked participants to report every single drink event via text using feature phones in (E. Kuntsche & Labhart, 2012). This approach shows that alcohol consumption increases on Saturday for males and females. Regarding using smartphones for daily logging of drinking behavior, (Attwood, Parke, Larsen, & Morton, 2017) used a mobile health application to allow users self-report types of alcohol beverages, amount of consumed alcohol, and drinking time. Also using a smartphone app, Santani et al. investigated urban nightlife patterns for over 200 young people, collecting smartphone sensor and log data, alcohol drinking surveys, place drinking surveys, and videos of drinking environments (Santani, Biel, et al., 2016). Santani et al. also used the sensor and log data to automatically classify nights in which alcohol was consumed, with a reported accuracy of 76.6% (Santani et al., 2017). Other studies ask participants to carry wearable devices or bring along their smartphones before and during alcohol use (Arnold et al., 2015; Bae et al., 2017; Kao et al., 2012). In detail, they use mobile phone sensors (wifi, accelerometers, etc.) to track drinking occasions, from non-drinking to heavy drinking (Arnold et al., 2015; Bae et al., 2017; You et al., 2015). Bae et al (Bae et al., 2017) defined a classification task for identifying non-drinking, drinking and heavy drinking episodes for 30 young adults aged 21-28 and reported an accuracy of 96.6%. The baseline was not explicitly mentioned, although it must be noted that the majority class, consisting of non-alcoholic drinking, constituted 90.7% of the data. You et al. proposed SoberDiary, a phone-based support system, which logged drinking through a Bluetooth breathalyzer connected to the mobile phone. According to quantitative and qualitative results, the system helped alcohol-dependent patients to self-track and self-manage their drinking behaviors (You et al., 2015). None of the ubicomp works described conduct any comparisons to social media-based approaches.

Finally, relatively few comparative studies between basic patterns from smartphone sensing campaigns and social media data have been conducted in the past. It was shown in (Malmi, Do, & Gatica-Perez, 2012) that the hourly distributions of generic Foursquare check-ins and of automatically inferred place visits in the Nokia Mobile Data Challenge were similar. This similarity was used to learn a temporal model from Foursquare data in the context of cold-start phone applications (Malmi, Do, & Gatica-Perez, 2013). These studies, however, did not focus on specific human activities as we do here for alcohol consumption. Furthermore, only time was studied as a variable, in contrast to this chapter that includes location, social context, alcohol categories, and gender in addition to temporal variations.

6.3 Datasets

In this chapter, we use two data sources: (a) alcohol-related posts from Instagram, (b) self-reported alcohol drinking from smartphone-based crowdsensing. This section outlines the data collection and curation process.

6.3.1 Instagram Datasets

Instagram allows users to share pictures and videos and tag such posts with hashtags to express mood, location, social context, etc. For instance, a photo posted in Lausanne, Switzerland can have a caption: *"#champagne with my family #summer #holiday #lausanne #sunshine #family #happiness"*. In this example, #champagne is an alcohol hashtag, while references to the location (#lausanne), its social context (#family), its occasion (#summer, #holiday), and the user's mood (#happiness) are also provided. In this chapter, we use the Instagram dataset first described in Chapter 2. This corpus will be referred to as *Instagram 1.7M dataset* in the

Table 6.1 – Instagram datasets used in the rest of this chapter.

Dataset	Instagram Alcohol dataset (34K)	Instagram Alcohol Spatial-driven dataset (19K)
# of images	34K	19K
# of users	19K	11K
# of venues	12K	5.3K
# of total hashtags	384K	205K
# of alcohol hashtags	49K	27K

rest of the chapter.

From this data, we curated a dataset consisting of alcohol posts. This was done by manually defining a hashtag-based alcohol vocabulary suited for Switzerland. The alcohol vocabulary was generated in a two-step process. In the first step, all Instagram posts in the *Instagram 1.7M dataset* that matched 4sq nightlife venues (like bar, pub, or club) were examined. We assume that posts generated on nightlife places are a reasonable starting point to identify alcohol-related hashtags, resulting in a set of 76 alcohol hashtags. This vocabulary was further enriched by using 41 alcohol hashtags from Chapter 2 and a set of 149 hashtags from Pang et al. (Pang et al., 2015). The three sets of alcohol hashtags contained some common elements that were filtered out, resulting in 199 unique hashtags. These 199 hashtags were then used to harvest a set of 30,192 posts We then re-examined the alcohol hashtags in this data manually, and further identified 249 alcohol hashtags not included in the 199 hashtag dictionary. We extracted all the posts which contained at least one of the 448 hashtags from the 1.7M dataset. This resulted in 34,856 posts that contain at least one alcohol-related hashtag, and is named *Instagram alcohol dataset (34K)*.

From the *Instagram alcohol dataset (34K)*, we filter the posts into the *Spatial-driven Dataset* based on location. This dataset consists of all Instagram posts that have matching 4sq venues. The motivation behind this is to enrich the spatial description of drinking behaviour of youth. Towards this objective, Instagram venues are not detailed enough, as they only show the coordinates and the name of the place (i.e., name of the restaurant or pub). This lacks the details provided by 4sq, which also include a higher level view for each venue like Arts & Entertainment (0), College & University (1), Events (2), Food (3), Nightlife Spots (4), Outdoors & Recreation (5), Professional & Other Places (6), Residence (7), Shop & Services (8), Travel & Transport (9) and Other or None (-1)[I]. Hence, we use the Instagram API with 4sq venue identities as inputs to match, and obtain a total of 84K Instagram-4sq matched venues. We keep all posts with Instagram-4sq matched venues, obtaining 19,409 posts, and name it the *Instagram alcohol spatial-driven dataset (19K)*.

Table 6.1 summarizes the details of the two datasets used in the rest of this chapter. Depending on the specific analysis, we will use the corresponding dataset.

[I]https://developer.foursquare.com/categorytree

6.3.2 Youth at Night Data

As discussed in previous chapters, the Youth at Night (Y@N) crowdsensing study investigated nightlife behavior (including drinking) of young people in two Swiss urban areas: Lausanne and Zurich (Santani, Biel, et al., 2016). The study used an Android-based application to collect data including places, social context, and nightlife activities for Friday and Saturday nights from 8:00 PM to 4:00 AM for a period of three months. This data also included mobile sensor data like accelerometer, wifi, etc. as well as app logs.

Drink surveys asked participants to record the drink name, drink size, and social connection to surrounding people (e.g. family, friends, partner, etc.) whenever a participant had a new drink (alcohol or non-alcohol). Participants were also requested to fill in a place survey if a new place was visited. Place surveys collected city name, place categories (e.g. restaurants, private homes, etc.), and ratings of place. The Y@N crowdsensing campaign collected 2532 drink surveys from 218 participants and 1394 place surveys from 206 participants. For this chapter, we mainly focus on alcohol consumption and hence use the alcohol-related drink surveys. As a result, we obtain 1247 alcohol drink surveys provided by 204 participants. For more details of the data collection process, the readers are referred to (Santani, Biel, et al., 2016).

6.3.3 Non-alcohol Datasets

In addition to investigating alcohol consumption, we examine non-alcohol consumption data in a classification setting in Section 6.6. For Instagram, we curated 16K non-alcohol posts by applying 51 non-alcohol hashtag vocabularies from Chapter 2 on the Instagram 1.7M dataset. In the context of classification of beverages consumed during the night, we only keep those pictures posted from 8 PM to 4 AM. As a result, we harvested *1431 Instagram non-alcohol posts* from the 16K non-alcohol dataset. Meanwhile, Y@N also collected non-alcohol drink data in the same setting as for alcohol consumption. There were 644 non-alcohol drink reports that include similar information to that of alcohol drinks, i.e., picture, time, places, social context, etc.

6.4 Alcohol Category Definition

There are four main categories of alcoholic drinks, as used by the World Health Organization (WHO) ("Global status report on alcohol and health from WHO", 2018): beer, wine, spirits, and cider. These categories reflect different methods to produce alcohol (brewing for beer; fermentation for wine and cider; distillation for spirits) and different taxing policies in many countries. However, the consumption of cider is marginal in Switzerland and the alcohol content is similar to beer, so the two categories can methodologically be combined. In the Y@N study, we had six different types of reported alcoholic drinks - the three main ones (beer/cider, wine, pure spirits) and three three additional declinations of distillate alcohol

Table 6.2 – Result of manual coding of 449 alcohol hashtags into four alcohol categories.

Category	Wine	Beer	Spirit & Cocktail	Other
# of hashtags	152	141	128	28
Percent (%)	33.9	31.4	28.5	6.2

that are associated with different risks among the target population: mixed drinks (with soda or energy drinks), aperitifs, and shots. However, at the coarser level, they all are spirit-based products. For our analysis, we use three main categories of alcoholic drinks: wine, beer, and spirits (including cocktails). Table 6.2 shows the distribution of the 449 alcohol hashtags (from the Instagram alcohol dataset 30K) over the three alcohol categories.

6.5 Alcohol Consumption Analysis (RQ1)

In this section, we extract and compare alcohol consumption patterns from the social media and crowdsensing datasets, and from public data (i.e., WHO ("Alcohol Statistic from WHO", 2014),("Global status report on alcohol and health from WHO", 2018) and Swiss Federal Statistical Department ("Alcohol In Figures 2017", 2017)). Specifically, we study (1) types and quantities of alcohol consumed; (2) temporal patterns of drinking; (3) geospatial patterns of drinking; (4) social patterns of drinking; and (5) connection between occasions and alcohol drinking. With the Instagram datasets, we compute a four-dimensional vector of each user representing the four alcohol categories (wine, beer, spirit, and other). This method accounts for the issue of some alcohol items being dominantly mentioned by some users and/or some posts containing more than one alcohol category. The methodology was also adopted to compute a personal distribution of alcohol consumption for the same categories in the Y@N dataset.

6.5.1 Analysis of Alcohol Categories and Quantity

In this section, we analyze the distribution of quantities consumed for each alcohol category.

Federal alcohol board data: Figure 6.1a shows the percentage of alcohol beverage consumed (in liters) in the three categories (wine, beer, spirits). We observe that beer is the most consumed alcoholic beverage (59%), followed by wine (37%) and spirits the least (4%). In contrast, in terms of the amount of pure alcohol consumed, as estimated by the Swiss federal statistics, wine is the highest (48%) followed by beer (34%) as shown in Figure 6.1b. This is due to the amount of alcohol in these drinks, i.e wine has 12-14% of alcohol, while for beer it is 4-7%. This also explains why the percentage of pure alcohol from spirits (18%) is greater.

Y@N dataset: Figure 6.1c illustrates the percentages of reported drinks over categories of alcohol consumed in the Y@N data, while Figure 6.1d shows the quantity of pure alcohol consumed for the same. The amount of pure alcohol consumed for this dataset was computed using the method proposed by (Labhart et al., 2013). Comparing these values with the data from the federal alcohol board, we observe several inconsistencies. Specifically, the federal data shows 48% of pure alcohol (liters) to be consumed from wine, but the Y@N data shows wine to contribute only 10%. Similarly, the percentage of pure alcohol consumed from spirits as estimated by the federal board is 18%, while this value in the Y@N data is 55%. We hypothesize that these differences are due to the differences in the sampled population and time period. The federal survey is conducted over a larger representative population, while the Y@N dataset is restricted to young people drinking between 8pm-4am on Fridays and Saturdays (Sept-Dec 2014).

Instagram: Figure 6.1(e-f) shows the percentage of alcohol alcohol posts per person per category for Instagram 34K (24-hour and seven days a week) and Instagram 34K (8PM-4AM on Friday and Saturday nights), respectively. For the Instagram dataset for 8PM-4AM on Friday and Saturday night, we observe a higher percentage of spirit (31%) as compared to Figure 6.1e, and a corresponding decrease for wine and beer.

Overall, the distributions of alcohol categories in the three data sources are different. It must be noted that only Y@N and Instagram datasets can be directly compared (Figure 6.1c and Figure 6.1f) as they correspond to the distribution of alcohol posts per person over the same time of day and day of the week, while the federal board data is collected as a traditional survey. Specifically, we observe that in Y@N 42% was spirits, 45% was beer, and 13% was wine. The corresponding percentage in the Instagram dataset is 31%, 26%, and 38% respectively. One reason for this difference is the data sample; participants from Y@N were young Swiss people mostly from Lausanne and Zurich, while the Instagram dataset contains posts from different demographics including wider age ranges from all parts of the country and tourists.

In summary, the main observation is that trends of alcohol consumption reported in federal statistics, Y@N and Instagram do not match with respect to types and quantities of alcohol consumed. Furthermore, both Y@N and Instagram contain considerably more reports of hard drinks. As the "other" alcohol category comprises only 4% in the Instagram dataset, we will not be discussing this further.

6.5.2 Drinking Time Analysis

In this section, we analyze the temporal patterns of Instagram drinking with respect to time of day and day of the week. In addition, we also compare alcohol consumption between Y@N and Instagram through Friday and Saturday nights from 8 PM to 4 AM.

Instagram 34K distribution of alcohol consumption across time of day: We calculate the time distribution of alcohol categories per user. Then, we aggregate these values to obtain the average of all users (Figure 6.2). Figure 6.2a shows that alcohol posting increases from morning until noon (12AM-1PM); then, it keeps growing and peaks at night (7PM-10PM). At night, beer, wine, and spirits peak at 7 PM, 9 PM, 10 PM respectively. The difference in the peak time of these alcohol categories may be affected by the legal rules of selling alcohol in Switzerland.

a) b) c) d) e) f)

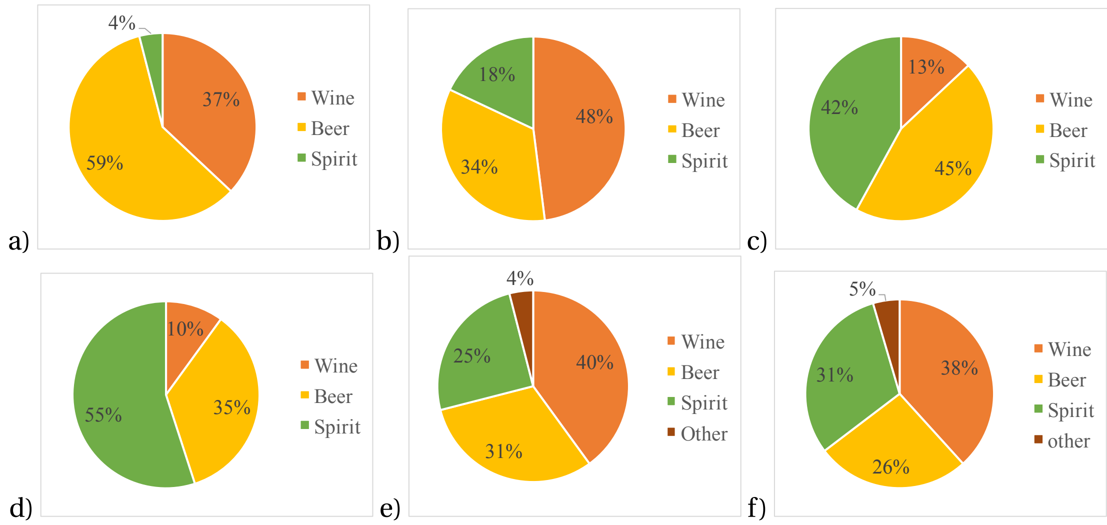
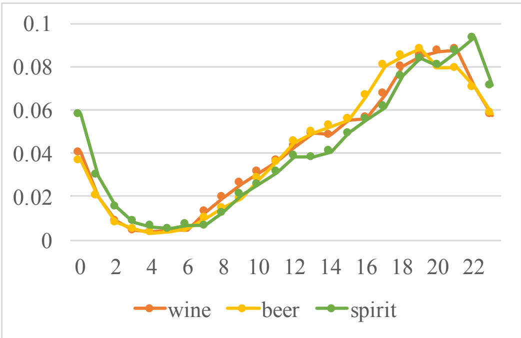


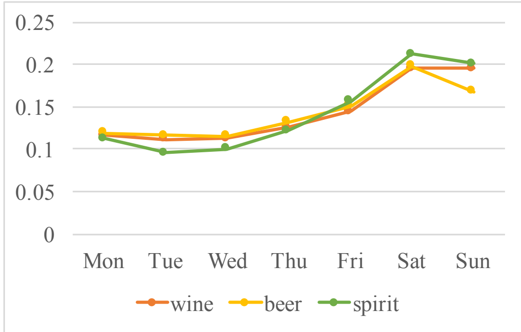
Figure 6.1 – Distribution of alcohol per person obtained from (a-b) Swiss federal statistics; (c-d) Y@N; and (e-f) Instagram. (a) Liters of effective alcohol estimated by the Swiss Federal Alcohol Board ("Alcohol In Figures 2017", 2017) (2010-2016). (b) Liters of pure alcohol estimated by the Swiss Federal Alcohol Board ("Alcohol In Figures 2017", 2017) (2010-2016). (c) Y@N alcohol drink survey 8PM-4AM on Friday and Saturday nights (Sep-Dec 2014). (d) Estimation of pure alcohol from Y@N drink survey on same period as (c). (e) Posts in the Instagram alcohol dataset (34K) (24-hour period and seven days a week) and seven days a week (2010-2016). (f) Posts in the Instagram alcohol dataset (34K) from 8PM-4AM on Friday and Saturday nights (2010-2016).

Specifically, the Swiss federal law (Federal Law, 2012), since January 2012, bans retail sales of alcohol from 10 PM to 6 AM. Particularly, spirits are observed to be consumed late at night as compared to beer and wine. Figure 6.2a shows that the spirits category is lower than wine and beer before 9PM, and dominates wine/beer after 9PM until early morning (also see Figures 6.3b and 6.3d discussed later in this section). These distributions follow the general trends of eating and drinking reported in the literature (Phan & Gatica-Perez, 2017), and also follow more generally the trends of Instagram posting.

Instagram 34K distribution of alcohol consumption over day of week:. We apply a method similar to the one described above for estimating the distribution of alcohol categories over the seven days of the week. Results are shown in Fig 6.2b. Our results are in line with those results reported by Kuntsche et al. (E. Kuntsche & Labhart, 2012) who used a traditional method based on questionnaires. These questionnaires were sent out as hypertext messages to all the participants hourly between 8 PM and midnight, and again at 11 AM the following day. The authors of (E. Kuntsche & Labhart, 2012) reported that consumption of alcohol increased slightly from Monday to Wednesday, and then dramatically increased, peaking on Saturday and decreasing on Sunday, as compared to the baseline percentage of drinks for men and women. In the Instagram alcohol dataset (34K), we observed similar patterns as shown in Figure 6.2b. Another interesting trend is that while wine and beer are slightly more common



a)



b)

Hours	Mon	Tue	Wed	Thu	Fri	Sat	Sun
00:00 - 03:59	0.009	0.007	0.009	0.009	0.013	0.015	0.022
04:00 - 07:59	0.003	0.003	0.004	0.003	0.003	0.003	0.005
08:00 - 11:59	0.013	0.011	0.012	0.012	0.014	0.019	0.021
12:00 - 15:59	0.022	0.019	0.019	0.022	0.024	0.038	0.047
16:00 - 19:59	0.033	0.032	0.032	0.037	0.045	0.065	0.052
20:00 - 23:59	0.036	0.037	0.035	0.043	0.051	0.059	0.043

c)

Figure 6.2 – Empirical time distribution of alcohol consumption on Instagram over a) time of day; b) day of week; c) joint time of day and day of week. The unit of the y-axis of a) and b) is probability.

at the beginning of the week, hard alcohol is comparatively more prevalent on the weekends.

Joint distribution of alcohol consumption over time of day and day of week: This distribution is shown as a matrix of 7 days and 6 four-hour time slots in Figure 6.2c. Each entry corresponds to the relative frequency of posts and the color of the cell indicate the density of posts from lowest (green) to highest (red). Figure 6.2c shows that the density of alcohol posts is highest on the evenings of Friday/Saturday and late afternoon on Sunday. The time slot between $00:00-03:59$ on Sunday is yellow as compared to the same time slot across the week, indicating that users drink at a comparatively higher density on Saturday night. On Sundays, people start drinking earlier than other days.

We are not aware of official Swiss statistics that report on the time consumption of alcohol across drink types. This result in itself shows that data from Instagram could be potentially useful to describe the population of social media users. In the next step, we compare the temporal trends of alcohol consumption in Y@N and Instagram. To make a fair comparison, we use Instagram data from Friday and Saturday night from 8 PM to 4 AM.

Comparison of Alcohol Consumption in Y@N and Instagram: Figure 6.3a and Figure 6.3c show the time distribution of alcohol consumption in Y@N, while Figure 6.3b and Figure 6.3d show the corresponding curves for Instagram alcohol dataset (34K). Figure 6.3a and Figure 6.3b show similar patterns for wine, beer, and spirits. In both datasets, the initial ranking early in the night (wine, beer, spirits) gets reversed in the 22:00 - 23:00 time slots, to (spirits, beer, wine). Figure 6.3c and Figure 6.3d show alcohol consumption patterns for Friday and Saturday night. In both Y@N and Instagram, wine, beer, spirits have overall lower probabilities on Friday night compared to Saturday night. Especially, spirits on Saturday nights have higher

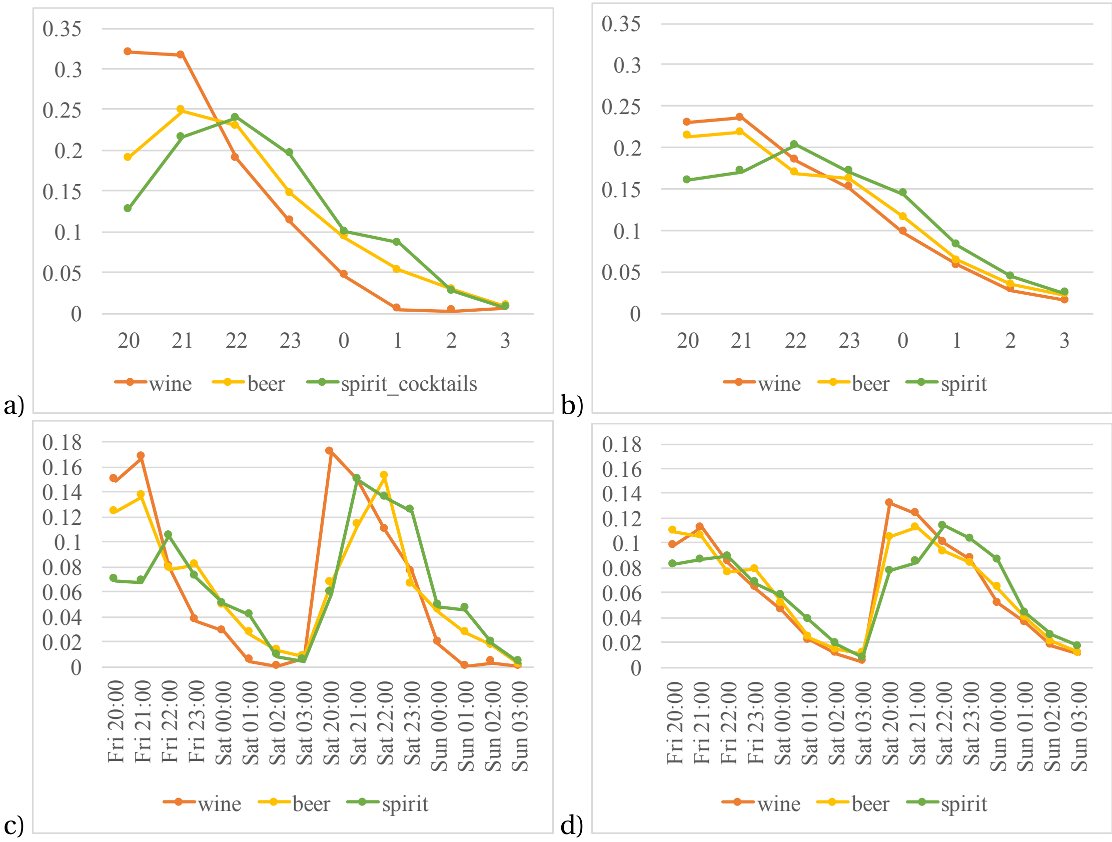


Figure 6.3 – Empirical time distribution of alcohol consumption over a) Combined Friday night and Saturday Night in Y@N. b) Combined Friday night and Saturday Night in the Instagram alcohol dataset (34K). c) Separate Friday night and Saturday Night in Y@N. d) Seperate Friday night and Saturday Night in the Instagram alcohol dataset (34K). The unit of the y-axis of all figures is probability.

values than on Friday nights. Although wine, beer, and spirits peak at 10PM and 11PM on Friday nights (Figure 6.3c and Figure 6.3d), they have different peaks on Saturday. Wine peaks at 8 PM in both Y@N and Instagram, while beers and spirits swap the peak at 9 PM and 10 PM on Saturday nights in Y@N and Instagram.

In summary, we observe that (1) there is a higher relative number of posts about spirits as the night progresses on weekends as compared to wine and beer for which there is a higher relative number of posts earlier in the evening; (2) a comparison between Y@N reports and Instagram posts on Friday and Saturday nights (8 PM to 4 AM) share similar trends, with a greater amount of wine, beer, and spirits being posted/consumed on Saturdays than on Fridays.

6.5.3 Place Analysis

In this section, we compare patterns of drinking venues between Y@N and Instagram using the alcohol spatial-driven dataset (19K).

Y@N Places: study participants were asked to complete a place survey whenever they had a new drink in a new place. The place survey contains three kinds of values: city name, place category, and ambiance. In this chapter, we focus on the place category. Participants could choose the following place categories: *coffee, bar, restaurant, personal, travel, plaza, school, events, club, park or lake, on board, other, and public*. In case participants chose "other", they had to manually enter the name of the place in a text field. As we are interested in comparing place categories across datasets, we use the place hierarchy defined in Foursquare to find the equivalence across the two place category systems. After examining the hierarchy of ten Foursquare Venue Categories[II], we match "coffee" and "restaurant" to Food, "bar" and "club" to Nightlife, "personal" to Residence, "travel" and "on board" to Travel & Transport, "plaza" and "park/lake" to Outdoor & Recreation, "school" to University & College, and "event" to Event. "Public" can correspond to ambiguous 4sq categories so it is matched to Unknown. Finally, for "other" Y@N category, we collected 47 free text inputs and manually allocated then into one of the ten 4sq categories. Some cases are vague, so we allocated them into "unknown". For each alcohol type, Figure 6.4a shows the average distribution per user for each place category. Wine is predominantly consumed at home (58%) followed by food venues (12%). In contrast, beer and spirits are mainly consumed at nightlife spots and home (65 to 70% of the probability mass). All other venues, individually, account for 10% or less.

Instagram Places: Figure 6.4b,c shows the distribution over places for each alcohol category with and without time filtering (8PM-4AM in Friday and Saturday nights) in the Instagram alcohol spatio-driven dataset (19K). We observe that four venue categories, namely Food, Nightlife Spots, Outdoor & Recreation, and Travel & Transport are the most predominant ones. Drinking categories in these four venues follow similar trends in the night-only and full-day conditions. Specifically on Friday and Saturday nights, wine has the highest probability mass at Food venues (0.29) followed by Outdoor & Recreation. Both beer and spirits are consumed in Nightlife Spots (0.47 for spirits, 0.36 for beer) and Food (around 0.25 for both alcohol categories). Importantly, drinking at Residence places is almost absent. This observation matches previous research in location-based social networks which found low rates of check-in at home (Lindqvist et al., 2011b).

Comparison between Y@N and Instagram: Both datasets share a high occurrence of spirits, beer, and wine at nightlife spots (bar, club, etc.). In contrast, Residence in Y@N is one of the top venues for the three alcohol categories, while there are few posts on Instagram. If we compare Figure 6.4a and Figure 6.4c, beer and spirit both have over 30% probability mass at Nightlife Spots in both Y@N and Instagram, the next ones being Outdoor & Recreation, and Travel & Transport.

[II]https://developer.foursquare.com/categorytree

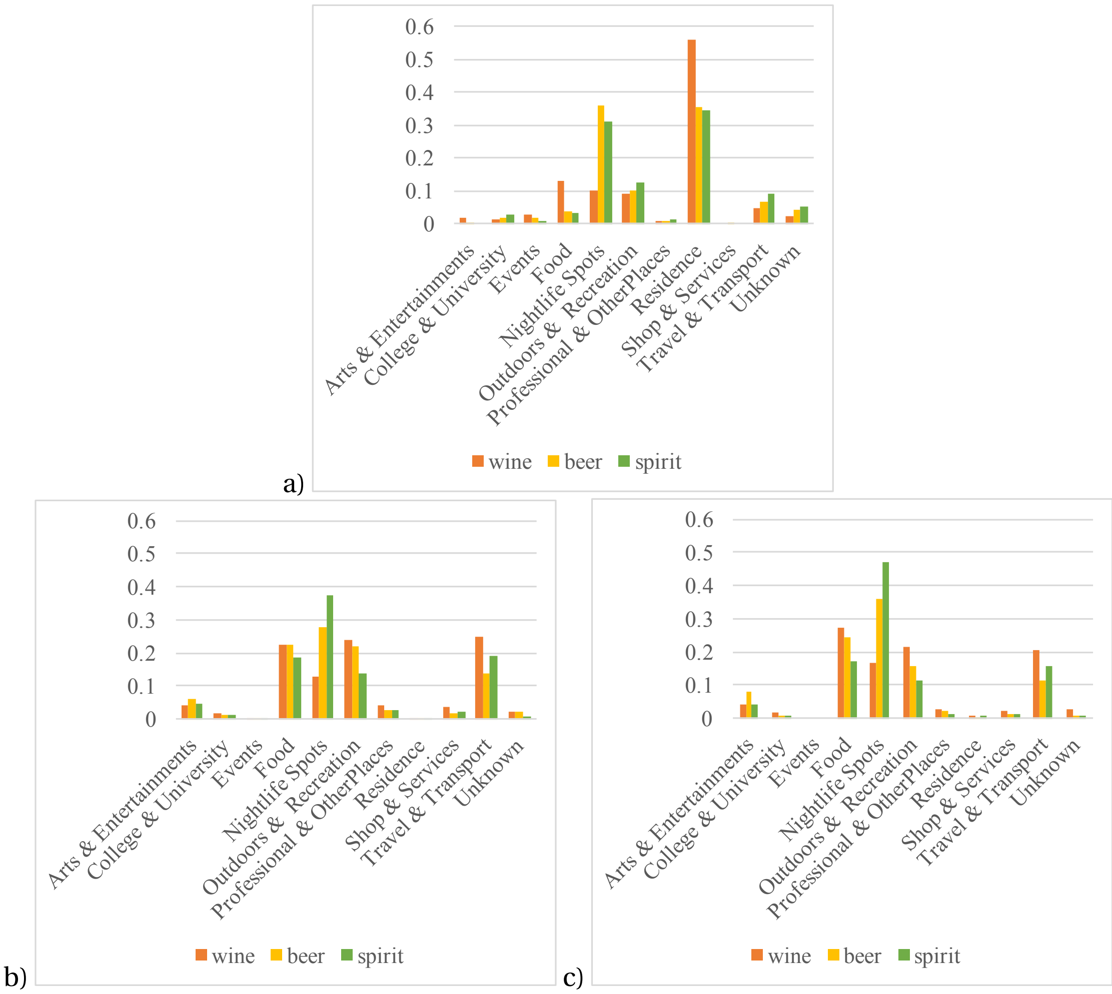


Figure 6.4 – Empirical distribution of alcohol consumption per place categories for each alcohol categories in a) Y@N b) The Instagram alcohol spatio-driven dataset (19K) c) The Instagram alcohol spatio-driven dataset (19K) from 8PM to 4AM in Friday nights and Saturday nights. The unit of the y-axis of all figures is probability.

In summary, we observe that (1) Nightlife Spots and Food venues appear as important venues in Y@N and Instagram datasets for alcohol consumption, although in different relative order. (2) In contrast, Residence venues are essentially absent from Instagram, while they are the top venue for wine and spirits consumption and top-2 for beer. This illustrates the key value of crowdsensing as a potential way to "fill in" for the absence of social media reports made at home (which is the result of a number of factors and practices associated to the Instagram logic).

6.5.4 Social Context Analysis

In this section, we analyze patterns of drinking according to social context for the Y@N and Instagram alcohol datasets (34K).

Y@N: As mentioned in the previous sections, Y@N participants completed a survey whenever they had a new drink. In this survey, participants indicated their social context when the drink was consumed with five item choices: partner and spouse, female friends or colleagues, male friends or colleagues, family members, and other people. All items were asked for the specific number of people in each case. Based on the five items, we grouped female and male friends or colleagues into "friend_colleagues". We also created "alone" when the counter values of all five choices are zero. In the end, we defined five types of social context, namely partner_spouse, friend_colleagues, family, alone, and other. As usual, we normalize at the level of the user and compute average distribution over all users. The distribution is shown in Figure 6.5a for each alcohol category. Friends and colleagues are the most common social context (55-65% of prob. mass). In particular, we observe that wine is consumed more than beer and spirits with family and partner & spouse (12-20% of prob. mass). In contrast, beer and spirits are predominantly consumed with friends & colleagues.

Instagram: As described in Section 6.4, we collected 151 social context hashtags. Due to issues of multiple forms of nouns, or multiple languages, we clustered these social hashtags into 39 social items. Based on the meaning of the items, we mapped these 39 social items into the five social context categories of Y@N. For example, items such as family, mom, parents, daughter, cousin are clustered into context "family". The resulting average distribution is shown in Figure 6.5b (full day and all days of the week) and Figure 6.5c (Friday and Saturday nights). Overall, we observe a similar trend to what was observed in Y@N, with one main difference between full day and night timeslots, namely that during nights, alcohol is predominantly consumed with friends & colleagues with almost total absence of family. The predominance of friends & colleagues as drinking actors has some backing in the literature. Thrul et al. (Thrul & Kuntsche, 2015) concluded that the number of friends present is associated positively with hourly drinking frequency during the course of weekend evenings among young adults in Switzerland.

In summary, we observe two main trends: (1) friends and colleagues are the most common social context for alcohol drinking at night for all drink categories, and (2) this is the case for both datasets with a more pronounced trend for Instagram. This is not surprising given the overall young population captured in both datasets.

6.5.5 Occasion Analysis

Here we present the analysis of the occasion context when users share alcohol consumption posts on Instagram. In Section 6.4, we obtained 152 occasion hashtags (showing the events when people drink) with a similar method of clustering hashtags as described in Section

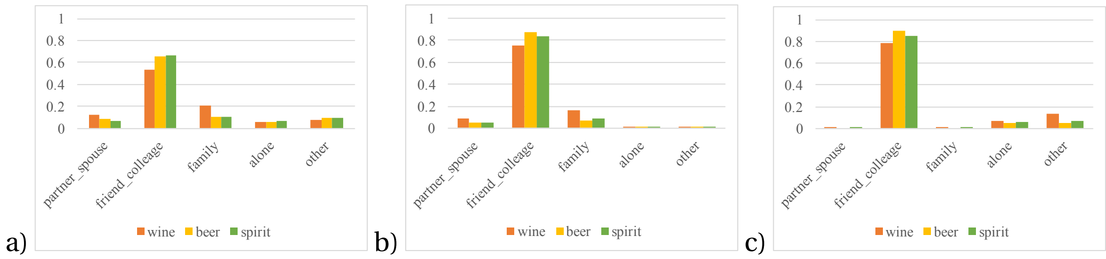


Figure 6.5 – Distribution of alcohol consumption per person with social relationship in a) Y@N b) The Instagram alcohol dataset (34K) c) The Instagram alcohol dataset (34K) from 8PM to 4AM on Friday and Saturday nights. The unit of the y-axis of all figures is probability.

6.5.4. We cluster these hashtags into 36 occasion items. Based on the meaning of each item, we manually group items into four categories, namely *party, holidays, events, festival*. As before, we first compute a user-level distribution over occasion categories. Then, we aggregate and estimate the average distribution over all users. Figure 6.6a and Figure 6.6b show how wine, beer and spirits are distributed in the full dataset and on Friday and Saturday nights. Furthermore, Figure 6.6c shows a comparison of the distributions for the more detailed occasion items. We observe that people drink at parties, during holidays, events, and festivals. On Friday and Saturday night, this party trend in more pronounced, especially with spirits. A party is not a location. It is an activity that takes place in any type of location (e.g. home, dorm, park, pub, club). According to the definition of Cambridge Dictionary ("Definition of party in English", 2018), a party is a social event a social event where people meet to talk, eat, drink, dance, etc., often in order to celebrate a special occasion. For example, (Bellis et al., 2015) showed that parties related to alcohol consumption included birthdays, weddings, and engagements. Based on Figure 6.6a, b, people drink spirits at parties, wine at parties and holidays, and beer at parties, festivals and events. Figure 6.6c shows the distribution for the detailed list of items for the full Instagram dataset, and for Friday and Saturday nights. From this, we first see that a few items dominate the kind of occasions, while the rest are rather marginal.

In summary, these results show that (1) partying is the predominant occasion context in which Instagram alcohol posts are produced, and (2) this trend is more pronounced for spirits, and slightly less so for wine and beer. As this information is not available for the Y@N study, a comparative analysis is not possible.

6.6 Classifying Individual Events Into Alcohol Categories (RQ2)

In the previous sections, we presented an analysis of temporal, spatial, and social context patterns of alcohol consumption on weekend nights as captured by both crowdsensing and social media. We now utilize these contextual cues, features extracted from images as input to a three-class classifier with the objective of classifying single reports/posts into their alcohol type (wine, beer, spirit). We examine the value of combining data from different sources to

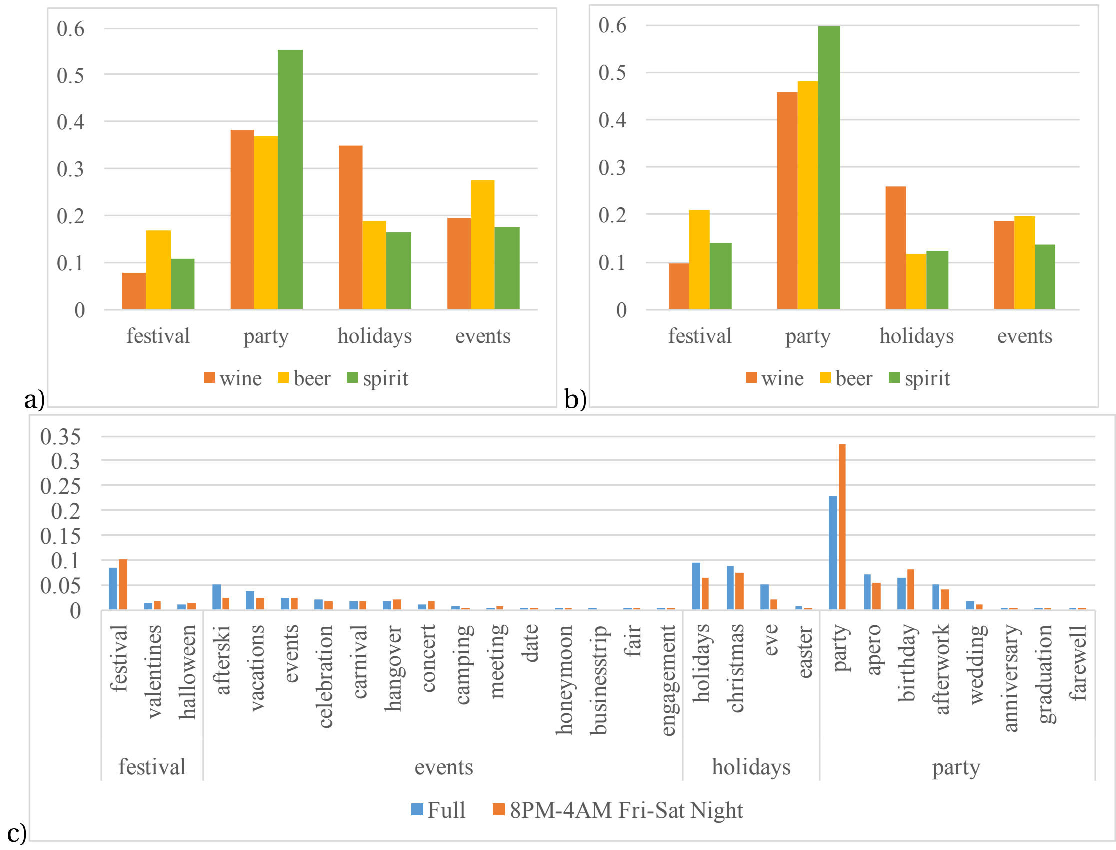


Figure 6.6 – Distribution of alcohol consumption with the occasion in a) The Instagram 34K. b) The Instagram 34K on Friday and Saturday nights. c) Detailed comparison of full, and Friday and Saturday nights for each detailed occasion. The unit of the y-axis of all figures is the probability.

improve classification accuracy. Finally, we also implement and evaluate a binary classification task to discriminate alcohol and non-alcohol consumption from single events.

6.6.1 Classification Method

Towards the objective stated above, we investigate two classifiers: random forest (RF) (Breiman, 2001) and Support Vector Machines (SVM) (Cortes & Vapnik, 1995), using the scikit-learn package for Python (“Sklearn with GridSearchCV”, 2018). The classification tasks are conducted on individual and combined datasets. Specifically, we conduct classification of drink type (alcohol and non-alcohol) and alcohol categories using Instagram, Y@N and combined (Instagram and Y@N) datasets. In both classification tasks, when using the individual datasets, the hyper-parameters for the two machine learning algorithms were optimized using 10-fold inner cross-validation (CV), and the final performance was assessed using 5 independent runs of the 10-fold CV procedure and computing the average over the 5 runs. To study the effects of combining the datasets, we first partitioned each of the datasets into training and testing data. Specifically, we used 80% of data from Instagram to create the training set (Insta_Train), while the 20% was used as test set (Insta_Test). Similarly, we create Y@N_Train and Y@N_Test. We then fused the training sets Insta_Train and Y@N_Train to create a combined training set. Using this combined training dataset, we learn a model using 10-fold inner cross-validation (CV), while final performance was assessed using the average of 5 independent runs of 10-fold CV. We report accuracy, precision, and recall (weighted precision and recall for the three-class task).

6.6.2 Feature Extraction

For Instagram, we extracted a set of 1872 posts corresponding to Friday and Saturday nights from 8 PM to 4 AM. Of these 1872 posts, 781 mentioned wine, 556 mentioned beer and 535 posts mentioned spirits. For Y@N, we used the 1247 surveys described in Section 6.3.2. 606 reports mentioned beer, 479 mentioned spirits, and 162 mentioned wine as the alcohol type being consumed.

For both datasets, we first extracted a number of contextual features including time (Section 6.5.2), place (Section 6.5.3), social context (Section 6.5.4). We then grouped the features into 4 groups: context (C), social (S), and time (T) (Table 6.3).

In addition, we extracted various image features utilizing a convolutional neural network model (CNNs) as feature extractor. Specifically, using Keras library with TensorFlow backend, we investigated the following architectures:

- InceptionResNet_V2 (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017) is a competitive CNN that outperformed Inception_v3 in the ImageNet ILSVRC image classification benchmark and is considerably deeper. This model takes images of size $299x299$ as input.

- MobileNet is a family of convolutional neural networks which are fast, small, and accurate (Howard et al., 2017). The default input size for this model is 224x224.

- XceptionNet is based on depthwise separable convolution layers with residual connections (Chollet, 2017). This model takes images of size $299x299$ as input.

All these models were pre-trained using ImageNet, a large corpus corpus containing over 14 million images across 1,000 categories. After examining the above architectures, we choose InceptionResNet_V2 to extract the final features. The final features are obtained by applying max pooling to the output of the last convolutional layer and by removing the last fully-connected layer (this layer's outputs are the 1000-class cores) (Table 6.3). We used the default options in Keras for all the models investigated.

Table 6.3 – Features for classification of drinking posts.

Feature	Description	Type	Group Feature
hour	Timestamp (in minutes) of Instagram post or Y@N survey	numeric	Time (T)
day	Day of the week of Instagram post or Y@N survey	numeric	Time (T)
Venue categories	4sq venue category of Instagram post or Y@N place survey	categorical (10)	Place (P)
Social categories	Social categories of Instagram post or Y@N survey	categorical (5)	Social (S)
Image features	Features from a convolutional neural network (CNN)	categorical (2048)	Image (I)

6.6.3 Alcohol Category Classification

We first present the classification results when we train on individual datasets (Table 6.4). We observe that using contextual cues only, RF performs better than SVM for Instagram data. Specifically, the best classification accuracy (47.59%) using contextual cues is obtained using RF and the fusion of time and place (T+P) cues. This accuracy is moderately higher than the majority-class baseline. For Y@N, the performance of RF and SVM are very similar. We also observe that the best performance obtained for the Y@N data is higher than the one obtained for Instagram, for both SVM and RF.

The use of image features improves classification accuracy, with the best performance obtained using SVM. Specifically, we observe an accuracy of 81.6% for Y@N and 62.76% for Instagram. We believe this difference in performance is due to Y@N being a “cleaner” data source and images being more representative of the alcohol being consumed by design of the crowdsensing mechanism. In contrast, Instagram is noisier due to the images not always being representative of hashtags and captions used. To understand the difference in performance between the two data sources, we examine the difference of visual content between Y@N and Instagram by extracting 1000 object features using the Inception model pre-trained on ImageNet. The 1000 object features as the last fully-connected layer correspond to the 1000-class scores. Then, we calculated the average score over all images in both Y@N and Instagram and filtered the top 30 objects for each dataset, as shown in Figure 6.7. We observe that the top-30 detected objects between Y@N and Instagram have important differences. Y@N images clearly tend to depict alcohol containers (e.g. beer bottle, wine bottle, beer glass) while Instagram, in addition to this, has eating-related concepts (e.g. restaurant, plate, candle, etc.). The cumulative scores also show that Y@N images are better described by drinks than Instagram images. Thus, these results indicate that Y@N pictures are overall cleaner than Instagram photos. This potentially explains that using visual features to classify alcohol categories results in higher accuracies for Y@N data.

In the next step, we explore the impact of merging data from the two data sources on classification accuracy of alcohol type (see Table 6.5). Here, we combine 80% of data from both Instagram and Y@N for training the model and test individually on 20% of each of the datasets separately. Comparing these classification results with those obtained for individual datasets

Table 6.4 – Classification accuracy of alcohol categories for models trained/tested on individual datasets i.e., Instagram only, and Y@N only using RF and SVM classifiers. A denotes accuracy, P denotes precision, R denotes recall. Best performances for Y@N are higher than for Instagram for both contextual features and image features.

Features	Random Forest (RF)						SVM					
	1872 Insta			1247 Y@N			1872 Insta			1247 Y@N		
	A	P	R	A	P	R	A	P	R	A	P	R
Baseline	41.72			48.59			41.72			48.59		
T	36.16	0.36	0.36	49.50	0.42	0.49	41.72	0.42	0.42	48.6	0.38	0.47
P	46.63	0.45	46.6	48.00	0.43	0.48	46.52	0.44	0.46	48.59	0.36	0.48
T+P	47.59	0.47	0.47	49.25	0.47	0.50	40.51	0.39	0.41	50.63	0.42	0.49
S	41.72	0.17	0.42	50.78	0.54	0.51	41.73	0.17	0.42	50.60	0.53	0.51
T+P+S	46.91	0.45	0.47	49.54	0.45	0.46	40.46	0.40	0.41	49.89	0.42	0.49
I	58.30	0.58	0.59	75.66	0.77	0.72	62.76	0.63	0.61	81.60	0.79	0.83

Table 6.5 – Classification accuracy of alcohol categories for models trained on combined dataset (Instagram + Y@N) and tested on Instagram-only and Y@N-only data using RF and SVM classifiers. A denotes accuracy, P denotes precision, R denotes recall.

Features	Random Forest (RF)						SVM					
	1872 Insta			1247 Y@N			1872 Insta			1247 Y@N		
	A	P	R	A	P	R	A	P	R	A	P	R
Baseline	41.72			48.59			41.72			48.59		
T	37.10	0.37	0.37	50.00	0.43	0.5	42.12	0.42	0.42	48.9	0.40	0.49
P	47.23	0.45	0.47	49.00	0.43	0.49	46.92	0.45	0.47	49.19	0.42	0.49
S	38.33	0.39	0.38	47.67	0.48	0.47	39.00	0.40	0.38	49.22	0.49	0.49
T+P	45.80	0.45	0.46	46.00	0.45	0.47	44.90	0.43	0.45	47.55	0.46	0.48
T+P+S	45.10	0.45	0.45	48.40	0.48	0.49	45.50	0.46	0.44	50.12	0.51	0.50
I	57.80	0.54	0.62	77.00	0.77	0.77	62.10	0.60	0.64	81.00	0.82	0.80

Table 6.6 – Classification accuracy of alcohol and non-alcohol on individual datasets (Instagram dataset including 1872 alcohol posts and 1431 non-alcohol posts; and Y@N dataset including 1247 alcohol reports and 644 non-alcohol reports). (A denotes accuracy, P denotes precision, R denotes recall).

Features	Random Forest (RF)						SVM					
	3303 Insta			1891 Y@N			3303 Insta			1891 Y@N		
Metrics	A	P	R	A	P	R	A	P	R	A	P	R
Baseline	56.68			65.94			56.68			65.94		
T	61.03	0.61	0.61	65.94	0.46	0.66	61.30	0.62	0.61	65.94	0.44	0.66
P	62.36	0.65	0.62	68.53	0.67	0.69	62.36	0.65	0.62	68.53	0.67	0.68
T+P	63.76	0.63	0.64	68.3	0.67	0.68	64.72	0.65	0.65	68.50	0.67	0.68
S	56.43	0.46	0.56	76.00	0.76	0.76	56.67	0.45	0.57	76.00	0.76	0.76
T+P+S	63.03	0.63	0.63	76.80	0.76	0.77	65.03	0.65	0.65	76.00	0.76	0.76
I	74.65	0.76	0.75	80.00	0.82	0.80	78.42	0.78	0.78	84.20	0.85	0.83
I+T+P+S	74.62	0.76	0.75	80.11	0.83	0.80	79.74	0.80	0.80	90.00	0.90	0.90

Table 6.7 – Classification accuracy of alcohol and non-alcohol on combined datasets (Instagram + Y@N) using RF and SVM classifiers. (A denotes accuracy, P denotes precision, R denotes recall).

Features	Random Forest (RF)						SVM					
	3303 Insta			1891 Y@N			3303 Insta			1891 Y@N		
Metrics	A	P	R	A	P	R	A	P	R	A	P	R
Baseline	56.68			65.94			56.68			65.94		
T	63.78	0.65	0.63	66.04	0.66	0.66	63.63	0.64	0.63	66.04	0.66	0.66
P	62.72	0.64	0.62	64.45	0.64	0.64	60.90	0.60	0.60	65.51	0.57	0.65
S	56.36	0.49	0.56	73.74	0.72	0.73	56.36	0.49	0.56	72.94	0.73	0.72
T+P	65.45	0.65	0.65	66.31	0.64	0.66	60.75	0.60	0.60	66.31	0.65	0.66
T+P+S	65.45	0.65	0.65	76.39	0.76	0.76	65.30	0.65	0.65	76.92	0.76	0.76
I	74.54	0.79	0.74	79.04	0.8	0.79	77.12	0.77	0.77	86.47	0.86	0.86
I+T+P+S	72.57	0.76	0.72	79.31	0.80	0.79	81.06	0.81	0.81	87.26	0.87	0.87

(Table 6.4), we observe no improvement using SVM for Instagram and Y@N. The performance improvement using RF is only marginal (0.5% for Instagram, 1.5% for Y@N for I or T+P+S). This is slightly lower than those obtained for SVM when learned only on individual datasets.

In summary, we have shown that (1) Instagram alcohol posts are more challenging as a data source for classification compared to Y@N due to Instagram's unconstrained nature, (2) that image features are significantly more effective than context for this task (yet context itself has some discriminative power), and (3) that the direct combination of datasets for training does not result in significant performance gains.

6.6.4 Alcohol vs. Non-Alcohol Classification

For a practical application, alcohol drinking events first have to be distinguished from non-alcohol ones. In this section, we expand the chapter to study a binary alcohol/non-alcohol classification task. Toward this, we use the same 1872 Instagram posts and 1247 responses from

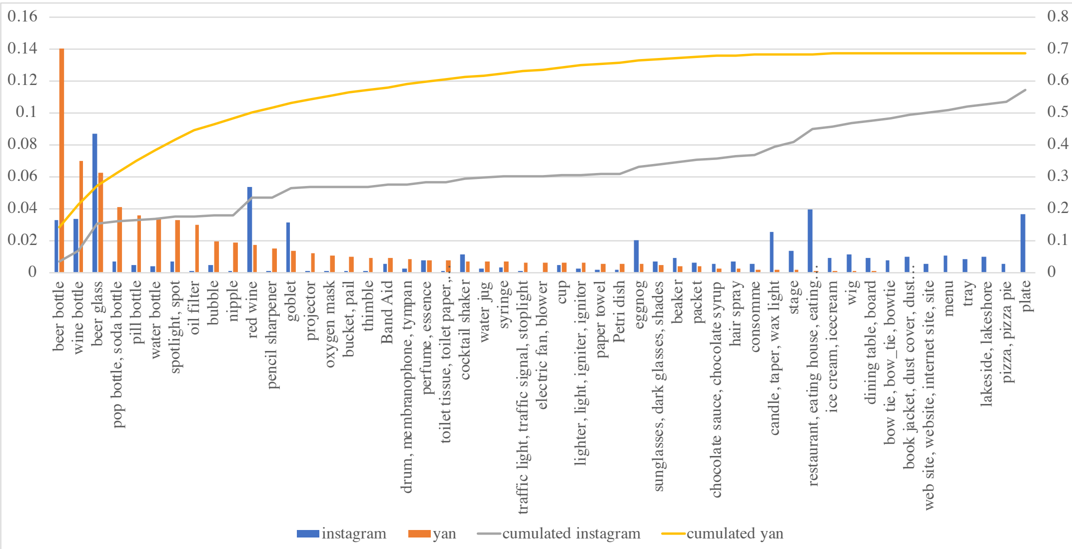


Figure 6.7 – Comparison between top 30 objects detected on Y@N and Instagram pictures. Average scores are ranked with respect to their value on the Y@N data. Continuous lines show the cumulative scores for both datasets.

Y@N for the alcohol class. For the non-alcohol class, we extract 1431 posts from Instagram and 644 reports from Y@N as mentioned in Section 4.3. We follow the same methodology as described in Section 6.6.1. The classification results using individual datasets are presented in Table 6.6 while those obtained by fusing the two datasets are presented in Table 6.7. The baseline method is again a majority-class classifier.

For the classification results based on individual datasets, Table 6.6 shows that contextual features alone can achieve 65% accuracy for Instagram and 76% for Y@N. Furthermore, the combination of contextual and image features can obtain the highest accuracy of classification (79.74% on Instagram and 90% on Y@N). The combination of context and visual information results in a better classifier than context-only and image-only cases. In general, the performance of SVM is better than RF.

For the classification results involving the combined datasets for training, Table 6.7 shows that the best accuracy (81.06% for Instagram or 87.26% for Y@N) is obtained using the fusion of image features and context features. Comparing the performance between individual and combined datasets, the performance on Instagram improves from 79.74% on individual in Table 6.6 to 81.06% on combination in Table 6.7, while Y@N decreases from 90.00% to 87.26%. In other words, the advantage of combining datasets for classification of alcohol vs. non-alcohol is marginal on Instagram (a noisier dataset), but decreases on Y@N (a cleaner dataset) as discussed earlier (Figure 6.7).

In summary, compared to the performance of alcohol category classification, alcohol vs. non-alcohol classification results in higher accuracy, and could be used as a first step towards an application that identifies consumed alcohol categories.

6.7 Discussion and implications

In this section we discuss the results previously presented and their implications.

For **RQ1**, we studied and compared the alcohol consumption patterns in data collected from social media (Instagram) and mobile crowdsensing (Y@N).We now discuss our main findings:

1. *Quantity:* The amount and type of alcohol consumed in Y@N and Instagram do not follow the same trends. In Y@N, we observed a predominant reporting of beer consumption, while the dominant post in Instagram for the same time period was wine. The observations made in these two datasets do not match those reported by the Swiss federal statistics ("Alcohol In Figures 2017", 2017; "Alcohol Statistic from WHO", 2014; Federal Law, 2012) with a disproportionate amount of spirits consumption in both crowd-generated datasets. While the specific time covered by our datasets (nights on weekends) could partly explain these differences, other sources of bias are likely present. This is an important observation from the ubicomp point of view as most literature has relied on either crowdsensed data like Y@N or social media data like Instagram, and relatively little work so far has been conducted on how bias affects machine based inference in urban ubicomp systems (Hecht & Stephens, 2014; Thebault-Spieker, Terveen, & Hecht, 2015).

2. *Time:* We observed similar trends in both datasets. Specifically, more spirits were reportedly consumed as the night progresses and on weekends. Beer and wine were consumed earlier in the evening and in the week. These results are backed by the results reported by Kuntsche et al., who found that people drink the most on weekends, especially from Saturday night to the first hours of Sunday morning (E. Kuntsche & Labhart, 2012). Furthermore, this observation could be affected by Swiss federal laws that ban the retail sales of alcohol from 10 PM to 6 AM (Federal Law, 2012). This illustrates the complexity of nightlife affected by personal choices, regulation, and social practices.

3. *Place:* In our datasets, we observed that alcohol was consumed with some differences. In Y@N, alcohol was consumed in private spaces, while on Instagram the top venue was nightlife locations. Food locations, outdoor & recreation, and travel & transport are common across both datasets. In line with our results, the results reported by (E. Kuntsche & Gmel, 2013), (Valentine et al., 2008) using traditional methods for data collection show that the top venue categories of alcohol consumption are bar/pub/nightclubs, private homes, events (festivals), public places (parks, streets), restaurants (food) and travel & transport. These observations have implications for ubicomp and alcohol research, as social media (possibly due to both performative aspects and privacy concerns) create a gap regarding the report of drinking at home. This could be mitigated by mobile

crowdsensing methods, where participants share their data neither with their friends nor with online audiences.

4. *Social Context:* In both Y@N and Instagram, we showed that friends and colleagues were the most common social context. These results are backed up by previous research which found that the most common social context of alcohol consumption are friends and colleagues (Beck et al., 2008; Valentine et al., 2008). More specifically, (Valentine et al., 2008) reported that 63.8% of people drinking alcohol do so in the home of friends and family.

5. *Occasion:* As the Instagram data was collected over a five-year period, this allowed us to observe various patterns of alcohol consumption during festivals and holiday occasions. For instance, in this dataset we observed a greater relative frequency of posts about spirits in parties, wine during holidays, and beer on festivals and events. In line with this chapter, the work in (Bellis et al., 2015) summarized the main occasions of alcohol drinking using a telephone survey . The list included birthdays, wedding parties, events, and festivals.

For **RQ2** we we developed and tested a method to classify alcohol categories using contextual features and visual features. The main results can be summarized as follows.

1. *Context Features.* We consider time, place, social as context features for classification. For classification of alcohol categories, context features improved performance compared to a majority-class baseline, as shown in Tables 6.4 and 6.5. For classification on alcohol and non-alcohol, context features generated a larger performance improvement compared compared to the majority-class baseline, as shown in Tables 6.6 and 6.7.

2. *Visual Features.* We consider image features and its combination with context features. Generally, the use of visual features to classify both alcohol categories and alcohol/non-alcohol substantially improves accuracy compared to the majority-class baseline (around 20% in both tasks), as shown in Tables 6.4 and 6.6. We also showed that the classification accuracy on Y@N is better than on Instagram.

3. *Dataset Combination.* Overall we found that Y@N represents less noisy data for classification compared to Instagram (Figure 6.7) resulting in consistently higher accuracies for both contextual and visual features. Likely due to these differences, the combination of datasets for training did not produce major classification improvements. In other words, the performance improvement by combining the datasets was null or minimal. There was a marginal improvement in classification accuracy in some cases, while in others, the combined dataset has lower accuracy than the individual dataset. We hypothesize that if the content of Instagram images used for dataset combination was filtered to only depict actual drinks and drink containers, the performance of a classifier trained on a combined dataset could increase. The use of machine learning to filter Instagram posts to obtain more focused content about drinking episodes could be part of future work.

6.8 Conclusion

In this chapter, we conducted a comparative study using data from a crowdsensing campaign and social media to examine the phenomenon of alcohol consumption on weekend nights and to devise and evaluate contextual classifiers of alcohol categories, which is relevant for both ubicomp and alcohol research. We conclude the chapter by reviewing the answers to our research questions and discussing directions for future research.

Our RQ1 inquired about alcohol consumption patterns on weekend nights that could be extracted from both data sources with respect to temporal, place, and social context, and about the similarities and differences revealed by examining the two sources jointly. This chapter has provided specific answers to these questions for the specific case of young urban dwellers, with potentially novel ideas for ubicomp research regarding (1) thinking about the consumption of hard alcohol as a relevant issue in and of itself, specially given recent literature that discusses how different emotions can be elicited by different types of alcoholic drinks, and (2) the importance of youth drinking in personal places. We believe that current ubicomp research on drinking episode recognition could make use of some of the findings from our contextual analysis.

Our RQ2 examined how image features and contextual cues could be used to classify individual drinking occasions according to alcohol category, and whether the complementarity of crowdsensing and social media sources could improve the performance of this classification task. This chapter has provided answers to these questions, with best classification accuracies in Y@N data of up to 80% for wine/beer/spirits, and up to 90% for alcohol/non-alcohol, where image features are stronger cues than contextual features.

Future work could investigate a number of open issues. First, regarding pattern analysis, we would like to understand in more detail the possible effects that the local regulations on alcohol retail sales can have on the emerging trends we observed. Second, it would be relevant to investigate possible regional differences; in our case, this refers to the French-speaking and German-speaking regions of Switzerland, both of which are represented in the crowdsensed and social media data sources. This could motivate the use of multiple models and more sophisticated transfer learning.

7 Conclusions and Future Work

In this chapter, we conclude this dissertation by summarizing the main contributions of each chapter as well as their implications in Section 7.1. Section 7.2 discusses the limitations of our work and suggests potential future work.

7.1 Contributions and Implications

In this dissertation, we developed a computational framework, based on mobile crowdsourcing and social media, for studying food and drink consumption activities of people and its related consequences. Towards this, we used the collections of mobile crowdsourcing data, collected from social datasets (Instagram and Foursquare) and mobile crowdsensing datasets (known as Youth at Night dataset (Y@N)).

Social Media. We curated datasets from Instagram in the country under study, enriched with Foursquare information. Using this data, our contributions were the following:

- In Chapter 2, we studied food and drink consumption on Instagram by using 95K posts filtered from 2.9M Instagram posts in Switzerland based on a manual hashtags dictionary. As a result, users tend to post #breakfast posts from 9:00 to 10:00, #lunch posts at 13:00, and #dinner posts from 20:00 to 21:00. Also, vegetables, salad, fruit, coffee are co-occurrence food in healthy posts vs. non-healthy posts with the burger, fries. We then use context, social, picture caption, 4sq, food/drink item, food/drink categories to classify healthy/non-healthy at 85.8% and 6 meals (breakfast/ brunch/ lunch/ dinner/ coffee/ tea) at 61.7%. Chapter 2 implies preliminary steps to understand food and drink on social networks and a window view of food in a country from multiple perspectives: context (time and location), and content (food and drink).

- In Chapter 3, we conducted research on drinking practices on Instagram in Switzerland: casual (#drink) and heavy (#drunk) drinkings. We used a dataset containing 2046 #drink and 1323 #drunk posts filtered from 2.9M Instagram posts in Switzerland based on manual hashtags dictionary of casual (#drink) and heavy (#drunk) drinking. Chapter 3 showed that the number of hashtags/words in captions in #drink posts is higher than #drunk posts. We then classify #drink and #drunk at an accuracy of 82.3% by using picture caption and at an accuracy of 75% by using visual features. The implication of Chapter 3 is to warn users about their picture content, especially to casual and possibly heavy drinking posts.

Mobile Crowdsensing. We used the existing Youth@Night smartphone sensing dataset about nightlife activities of young people in Switzerland. Based on this data, our contributions were the following:

- In Chapter 4, we studied heavy drinking at night (4+ drinks for women/5+ for men on a single evening) over the entire night (night-level) by using a smartphone-based crowdsensing dataset on nightlife and drinking behaviors involving 240 young adult participants. This chapter showed 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 other nights. We developed a task to infer heavy and non-heavy drinking nights, comparing automatically captured sensor features versus manually contributed contextual cues and images provided over the course of the night. As a result, using phone sensors results in an accuracy of 71% while visual features of manually contributed images produce an accuracy of 72%. This suggests that automatic sensing is a competitive approach.
- In Chapter 5, we studied private nightlife activities and ambiances at home spaces of young people by using 301 in-situ videos recorded at the home of 204 participants whenever they drink somethings as well as annotations on physical and social attributes. This chapter analyzed multiple patterns. Young people tended to stay in the living room, bedroom, kitchen, dining room with a higher possibility of low light, no music, fewer people, and less noise on weekend nights. Their common activities were drinking, chatting, watching TV, using the computer, and eating. Males tended to hang out with males while females tended to hang out with both males and females. Different ambiances went with corresponding activities. We then use automatically extracted features from deep learning (objects, scenes, audios) to infer ambiances. As a result, large/spacious, dark/badly-lit, bright/well-lit, comfortable/cozy, dull/simple could be inferred by using automatic features. Chapter 5 implies an automatic system to infer ambiances of private places.

Social Media and Mobile Crowdsensing. Regarding combination of social media and mobile crowdsening dataset, we firstly contribute to investigation of alcohol drinking by examining advantages of two data sources. We summarizes the contributions and implications of drinking behaviours in Switzerland by using multiple crowdsourced data sources in the following:

- In Chapter 6, we contribute to the research area by studying Alcohol consumption (wine, beer, and spirit) on crowdsensing (sensor data, images) and social networks (Instagram, Foursquare). In detail, we used 34K posts of alcohol dataset on Instagram including 19K users and 12K venues and 1247 alcohol drink surveys provided by 204 young participants. We analyzed the temporal patterns of drinking behavior on night hours on Instagram and mobile crowdsensing. Besides, Instagram alcohol pictures were highly posted at nightlife spots, outdoor recreation, food, travel & transport while drinking events reported on mobile crowdsensing are higher at residence and nightlife spots. We then classify alcohol categories (wine, beer, and spirit) on the individual dataset and combined datasets by using context (time, place, and social) and image. Chapter 6 is implied to develop systems in machine learning for self-monitoring of alcohol consumption.

7.2 Limitations and Future Work

In this section, we discuss limitations of our work and, possible future research directions.

Social media. An advantage of social media is its potential large scale, but also open the possibility of the following sources of bias:

(1) *Temporal bias.* Some users do not post right after the time when they take photos. We hypothesize that there several possible reasons, e.g., lack of Internet connection or being busy with social activities. Besides, there is also a lack of continuous data for individuals. Hence, longitudinal activity patterns cannot be observed in general.

(2) *Spatial bias.* In this dissertation, we showed that social media is biased towards commercial public places like bars and restaurants, while other non-commercial places cannot be represented in the same manner. Due to this limitation, researchers should carefully contextualize the spatial patterns when using location-based social networks.

(3) *Hashtags semantics.* The multi-meaning of hashtags could lead to label posts incorrectly. For instance, #apple could be used to describe a kind of fruit or could be a branch of electronic devices. However, its visual content in photo and other co-occurring words or hashtags in caption could be essential information to help verifying the relevant meaning of the hashtag.

(4) *Population noise.* Social media users are both local people and tourists. The diversity of both types of users can result in potentially different patterns. Besides, a social network like Instagram with young adults might not exhibit the behavior of people of all ages, e.g., seniors.

(5) *Textual context.* The contents of posts possibly contain text with a higher level of semantic meaning, emoticons, and other textual types (e.g., exclamation marks). Those could be potential sources for future work to investigate patterns of self-presentation of drinking on social media.

The above limitations of social media can also be regarded as one of the main limitations of our work. The issue of dealing with bias in social media remains an issue for future work in the domain.

Mobile Crowdsensing. The advantage of mobile crowdsensing is to capture detailed data of volunteering participants but also has limitations:

(1) *Small number of participants.* Our algorithms were trained on data collected from 200+ participants. While these numbers are considerable for ubicomp research, the achieved performance could be affected by the biases of a limited cohort. This bias could be mitigated with larger budgets to recruit more people to join the crowdsensing campaign, but also have intrinsic limitations related to managing large groups in an academic setting.

(2) *Incentives Mechanisms.* In this dissertation, it took human effort of external raters to annotate Y@N images and videos. As a direction for future work, we could propose engaging the participants themselves as annotators of their own data after the collection (e.g., via extra-pay mechanisms). This could be a chance for them to reflect about their own data, while helping researchers to verify certain questions that might require specific contextualization.

Machine learning and inference tasks. In this dissertation, we worked on several inference tasks. For several of them, we conclude that the performance can be improved. Besides obtaining higher performance, there are other limitations in our methods:

(1) *Visual and sound feature extraction.* Visual and audio features used in this dissertation are extracted by using CNN models trained on datasets that are not specifically designed for home environments or drinking areas in the specific context of the world region under study. This could result in unexpected objects or scenes recognized in photos or videos shown in the previous chapters. In future work, we build visual models by using a trained dataset from local images to obtain more relevant features relating to local drinking objects and scenes.

(2) *Methods for inference.* Most of the supervised learning methods we used are classic, e.g. random forests or support vector machine, which are used individually or in combination in classifications of each chapter. They are considered as a means to compare how multiple features are competitive to infer classes. We do not target to compare their performance, albeit random forests often produced the highest performance. However, the justification for choosing these algorithms could be targeted in future work. In addition, as a state-of-art technology in machine learning, deep learning needs to be considered in future work to potentially boost classification performance.

(3) *Generality.* Both data sources collected from social media and mobile crowdsensing are in Switzerland, a western developed country in the European continent. The findings in this work might have limitations of generality w.r.t. other locations, e.g., countries in Asia or Africa. An important topic of future work is to investigate how the methodology proposed in this dissertation can applied to other countries.

Ethical and privacy issues. The results of the methodologies in this dissertation warn users about posting content on public social networks or using privacy-sensitive tools for self-monitoring drinking. There is potentially a lot of harm that could be done if users' data is not used with ethics, privacy, and the well-being of users as fundamental values. For instance, our results of identifying heavy drinking by using mobile sensors or drinking practices on Instagram imply that different forms of drinking could be automatically recognized at high levels. This requires data owners to conduct anonymization and aggregation processes on sensitive inferences. This also exposes issues that who has access to these inferences (e.g., individual public health office, or companies). To deal with this privacy issue, GDPR is the general data protection regulation in Europe to address data protection, privacy, and also the transfer of personal data outside the European Union and European Economic Areas (Parliament & of the European Union, 2016).

Bibliography

Abbar, S., Mejova, Y., & Weber, I. (2015). You tweet what you eat: studying food consumption through twitter. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 3197–3206). ACM.

Abbott-Chapman, J., & Robertson, M. (1999). Home as a private space: some adolescent constructs. *Journal of youth studies*, *2*(1), 23–43.

Abbott-Chapman, J., & Robertson, M. (2001). Youth, leisure and home: space, place and identity. *Loisir et société/Society and Leisure*, *24*(2), 485–506.

Ahmet, A. (2013). Home sites: the location (s) of 'home'for young men. *Urban Studies*, *50*(3), 621–634.

Akbari Fard, M., Hadadi, H., & Tavakoli Targhi, A. (2016). Fruits and vegetables calorie counter using convolutional neural networks. In *Proceedings of the 6th international conference on digital health conference* (pp. 121–122). ACM.

Alcohol In Figures 2017. (2017). Accessed: 2017-11-20. Retrieved from https://www.eav.admin.ch/eav/fr/home/dokumentation/publikationen/zahlen-und-fakten.html

Alcohol Statistic from WHO. (2014). Accessed: 2017-11-20. Retrieved from http://www.who.int/substance_abuse/publications/global_alcohol_report/profiles/che.pdf

Ali, A. A., Hossain, S. M., Hovsepian, K., Rahman, M. M., Plarre, K., & Kumar, S. (2012). Mpuff: automated detection of cigarette smoking puffs from respiration measurements. In *Information processing in sensor networks (ipsn), 2012 acm/ieee 11th international conference on* (pp. 269–280). IEEE.

Alitajer, S., & Nojoumi, G. M. (2016). Privacy at home: analysis of behavioral patterns in the spatial configuration of traditional and modern houses in the city of hamedan based on the notion of space syntax. *Frontiers of Architectural Research*, *5*(3), 341–352.

Amft, O., & Troster, G. (2009). On-body sensing solutions for automatic dietary monitoring. *IEEE pervasive computing*, *8*(2).

Anderson, I., Maitland, J., Sherwood, S., Barkhuus, L., Chalmers, M., Hall, M., ... Muller, H. (2007). Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. *Mobile networks and applications*, *12*(2-3), 185–199.

Andersson, B., Hibell, B., Beck, F., Choquet, M., Kokkevi, A., Fotiou, A., ... Trapencieris, M. (2007). Alcohol and drug use among european 17–18 year old students. *Data from the ESPAD Project. The Swedish Council for Information on Alcohol and Other Drugs (CAN) and the Pompidou Group at the Council of Europe.*

Arnold, Z., Larose, D., & Agu, E. (2015). Smartphone inference of alcohol consumption levels from gait. In *Healthcare informatics (ichi), 2015 international conference on* (pp. 417–426). IEEE.

Atkinson, A. M., & Sumnall, H. R. (2016). 'if i don't look good, it just doesn't go up': a qualitative study of young women's drinking cultures and practices on social network sites. *International Journal of Drug Policy*, *38*, 50–62.

Attwood, S., Parke, H., Larsen, J., & Morton, K. L. (2017). Using a mobile health application to reduce alcohol consumption: a mixed-methods evaluation of the drinkaware track & calculate units application. *BMC public health*, *17*(1), 394.

Badillo-Urquiola, K., Yao, Y., Ayalon, O., Knijnenurg, B., Page, X., Toch, E., … Wisniewski, P. J. (2018). Privacy in context: critically engaging with theory to guide privacy research and design. In *Companion of the 2018 acm conference on computer supported cooperative work and social computing* (pp. 425–431). ACM.

Bae, S., Chung, T., Ferreira, D., Dey, A. K., & Suffoletto, B. (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*, 42–47.

Bae, S., Ferreira, D., Suffoletto, B., Puyana, J. C., Kurtz, R., Chung, T., & Dey, A. K. (2017). Detecting drinking episodes in young adults using smartphone-based sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(2), 5.

Baillie, L., & Benyon, D. (2008). Place and technology in the home. *Computer Supported Cooperative Work (CSCW)*, *17*(2-3), 227–256.

Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: photos with faces attract more likes and comments on instagram. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 965–974). ACM.

Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In *International conference on pervasive computing* (pp. 1–17). Springer.

Beck, K. H., Arria, A. M., Caldeira, K. M., Vincent, K. B., O'Grady, K. E., & Wish, E. D. (2008). Social context of drinking and alcohol problems among college students. *American Journal of Health Behavior*, *32*(4), 420–430.

Bedri, A., Li, R., Haynes, M., Kosaraju, R. P., Grover, I., Prioleau, T., … Abowd, G. (2017). Earbit: using wearable sensors to detect eating episodes in unconstrained environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* *1*(3), 37:1–37:20. doi:10.1145/3130902

Bell, R., Meiselman, H. L., Pierson, B. J., & Reeve, W. G. (1994). Effects of adding an italian theme to a restaurant on the perceived ethnicity, acceptability, and selection of foods. *Appetite*, *22*(1), 11–24.

Bellis, M. A., & Hughes, K. (2011). Getting drunk safely? night-life policy in the uk and its public health consequences. *Drug and alcohol review*, *30*(5), 536–545.

Bellis, M. A., Hughes, K., Jones, L., Morleo, M., Nicholls, J., McCoy, E., … Sumnall, H. (2015). Holidays, celebrations, and commiserations: measuring drinking during feasting and fasting to improve national and individual estimates of alcohol consumption. *BMC medicine*, *13*(1), 113.

Benkhedda, Y., Santani, D., & Gatica-Perez, D. (2017). Venues in social media: examining ambiance perception through scene semantics. In *Proceedings of the 2017 acm on multimedia conference* (pp. 1416–1424). ACM.

Berg, C., Lappas, G., Wolk, A., Strandhagen, E., Torén, K., Rosengren, A., … Lissner, L. (2009). Eating patterns and portion size associated with obesity in a swedish population. *Appetite*, *52*(1), 21–26.

Bermingham, L., & Lee, I. (2019). Mining place-matching patterns from spatio-temporal trajectories using complex real-world places. *Expert Systems with Applications*, *122*, 334–350.

Biel, J.-I., Martin, N., Labbe, D., & Gatica-Perez, D. (2017). Bitesnbits: inferring eating behavior from contextual mobile data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(4), 125.

Blunt, A. (2005). Cultural geography: cultural geographies of home. *Progress in human geography*, *29*(4), 505–515.

Bochud, M., Chatelan, A., Blanco, J.-M., & Beer-Borst, S. (2017). Anthropometric characteristics and indicators of eating and physical activity behaviors in the swiss adult population. In *Results from menuch 2014-2015 (2017)*.

Boyd, D. (2014). *It's complicated: the social lives of networked teens*. Yale University Press.

Boyle, S. C., Earle, A. M., LaBrie, J. W., & Ballou, K. (2017). Facebook dethroned: revealing the more likely social media destinations for college students' depictions of underage drinking. *Addictive behaviors*, *65*, 63–67.

Boyle, S. C., LaBrie, J. W., Froidevaux, N. M., & Witkovic, Y. D. (2016). Different digital paths to the keg? how exposure to peers' alcohol-related social media content influences drinking among male and female first-year college students. *Addictive behaviors*, *57*, 21–29.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Brezmes, T., Gorricho, J.-L., & Cotrina, J. (2009). Activity recognition from accelerometer data on a mobile phone. In *International work-conference on artificial neural networks* (pp. 796–799). Springer.

Camilleri, G. M., Méjean, C., Bellisle, F., Andreeva, V. A., Sautron, V., Hercberg, S., & Péneau, S. (2015). Cross-cultural validity of the intuitive eating scale-2. psychometric evaluation in a sample of the general french population. *Appetite*, *84*, 34–42.

Can, G., Benkhedda, Y., & Gatica-Perez, D. (2018). Ambiance in social media venues: visual cue interpretation by machines and crowds. In *Proceedings of the ieee conference on computer vision and pattern recognition workshops* (pp. 2363–2372).

Carney, D. R., Jost, J. T., Gosling, S. D., & Potter, J. (2008). The secret lives of liberals and conservatives: personality profiles, interaction styles, and the things they leave behind. *Political Psychology*, *29*(6), 807–840.

Carter, A. C., Brandon, K. O., & Goldman, M. S. (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), 742–750.

Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S. J., & Bierut, L. J. (2015). "hey everyone, i'm drunk." an evaluation of drinking-related twitter chatter. *Journal of studies on alcohol and drugs*, *76*(4), 635–643.

Centers for Disease Control and Prevention. (2020). Accessed: 2020-01-13. Retrieved from https://www.cdc.gov/alcohol/faqs.htm

Chatterjee, S., Hovsepian, K., Sarker, H., Saleheen, N., al'Absi, M., Atluri, G., ... Nakajima, M., et al. (2016). Mcrave: continuous estimation of craving during smoking cessation. In *Proceedings of the 2016 acm international joint conference on pervasive and ubiquitous computing* (pp. 863–874). ACM.

Chatterton, P., & Hollands, R. (2002). Theorising urban playscapes: producing, regulating and consuming youthful nightlife city spaces. *Urban studies*, *39*(1), 95–116.

Chatterton, P., & Hollands, R. (2003). *Urban nightscapes: youth cultures, pleasure spaces and corporate power*. Psychology Press.

Chen, J., & Ngo, C.-W. (2016). Deep-based ingredient recognition for cooking recipe retrieval. In *Proceedings of the 2016 acm on multimedia conference* (pp. 32–41). ACM.

Chollet, F. (2017). Xception: deep learning with depthwise separable convolutions. *arXiv preprint*, 1610–02357.

Chon, J., & Cha, H. (2011). Lifemap: a smartphone-based context provider for location-based services. *IEEE Pervasive Computing*, *10*(2), 58–67.

Chon, Y., Lane, N. D., Kim, Y., Zhao, F., & Cha, H. (2013). Understanding the coverage and scalability of place-centric crowdsensing. In *Proceedings of the 2013 acm international joint conference on pervasive and ubiquitous computing* (pp. 3–12). ACM.

Chon, Y., Lane, N. D., Li, F., Cha, H., & Zhao, F. (2012). Automatically characterizing places with opportunistic crowdsensing using smartphones. In *Proceedings of the 2012 acm conference on ubiquitous computing* (pp. 481–490). ACM.

Christensen, M. A., Bettencourt, L., Kaye, L., Moturu, S. T., Nguyen, K. T., Olgin, J. E., ... Marcus, G. M. (2016). Direct measurements of smartphone screen-time: relationships with demographics and sleep. *PloS one*, *11*(11), e0165331.

Chung, C.-F., Agapie, E., Schroeder, J., Mishra, S., Fogarty, J., & Munson, S. A. [Sean A]. (2017a). When personal tracking becomes social: examining the use of instagram for healthy eating. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 1674–1687). ACM.

Chung, C.-F., Agapie, E., Schroeder, J., Mishra, S., Fogarty, J., & Munson, S. A. [Sean A.]. (2017b). When personal tracking becomes social: examining the use of instagram for healthy eating. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 1674–1687). CHI '17. doi:10.1145/3025453.3025747

Clapp, J. D., Madden, D. R., Mooney, D. D., & Dahlquist, K. E. (2017). Examining the social ecology of a bar-crawl: an exploratory pilot study. *PLoS one*, *12*(9), e0185238.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Computer Vision API on Microsoft Azure. (2018). Retrieved from https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/home

Cooper, M. L. (1994). Motivations for alcohol use among adolescents: development and validation of a four-factor model. *Psychological assessment*, *6*(2), 117.

Cooper, M. L., Russell, M., Skinner, J. B., & Windle, M. (1992). Development and validation of a three-dimensional measure of drinking motives. *Psychological assessment*, *4*(2), 123.

Cordeiro, F., Bales, E., Cherry, E., & Fogarty, J. (2015a). Rethinking the mobile food journal: exploring opportunities for lightweight photo-based capture. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 3207–3216). CHI '15. Seoul, Republic of Korea: ACM.

Cordeiro, F., Bales, E., Cherry, E., & Fogarty, J. (2015b). Rethinking the mobile food journal: exploring opportunities for lightweight photo-based capture. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 3207–3216). ACM.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273–297.

Countryman, C. C., & Jang, S. (2006). The effects of atmospheric elements on customer impression: the case of hotel lobbies. *International Journal of Contemporary Hospitality Management*, *18*(7), 534–545.

Courtney, K. E., & Polich, J. (2009). Binge drinking in young adults: data, definitions, and determinants. *Psychological bulletin*, *135*(1), 142.

Cox, W. M. (1990). *Why people drink: parameters of alcohol as a reinforcer*. Gardner Press.

Cox, W. M., & Klinger, E. (1988). A motivational model of alcohol use. *Journal of abnormal psychology*, *97*(2), 168.

Cramer, H., Rost, M., & Holmquist, L. E. (2011). Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services* (pp. 57–66). ACM.

Crowdflower Website. (2018). Retrieved from https://make.figure-eight.com/choose%5C_login

Culotta, A. (2013). Lightweight methods to estimate influenza rates and alcohol sales volume from twitter messages. *Language resources and evaluation*, *47*(1), 217–238.

Culotta, A. (2014). Estimating county health statistics with twitter. In *Proceedings of the 32nd annual acm conference on human factors in computing systems* (pp. 1335–1344). ACM.

Curtis, A., Droste, N., Coomber, K., Guadagno, B., Mayshak, R., Hyder, S., ... Miller, P. (2019). Off the rails—evaluating the nightlife impact of melbourne, australia's 24-h public transport trial. *International Journal of Drug Policy*, *63*, 39–46.

De Choudhury, M., Kumar, M., & Weber, I. (2017). Computational approaches toward integrating quantified self sensing and social media. In *Proceedings of the 2017 acm conference on computer supported cooperative work and social computing* (pp. 1334–1349). CSCW '17. doi:10.1145/2998181.2998219

Definition of home in English. (2019). Retrieved from https://dictionary.cambridge.org/dictionary/english/home#cald4-1-1-2

Definition of party in English. (2018). Retrieved from https://dictionary.cambridge.org/dictionary/english/party

Demant, J., & Landolt, S. (2014). Youth drinking in public places: the production of drinking spaces in and outside nightlife areas. *Urban Studies*, *51*(1), 170–184.

Dictionary, C. (2018). Definition of ambiance in english. Retrieved from https://dictionary.cambridge.org/dictionary/english/ambience

Dietze, P. M., Livingston, M., Callinan, S., & Room, R. (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), 346–353.

Do, T. M. T., Blom, J., & Gatica-Perez, D. (2011). Smartphone usage in the wild: a large-scale analysis of applications and context. In *Proceedings of the 13th international conference on multimodal interfaces* (pp. 353–360). ACM.

Do, T. M. T., & Gatica-Perez, D. (2011). Groupus: smartphone proximity data and human interaction type mining. In *2011 15th annual international symposium on wearable computers* (pp. 21–28). IEEE.

Do, T. M. T., & Gatica-Perez, D. (2013). The places of our lives: visiting patterns and automatic labeling from longitudinal smartphone data. *IEEE Transactions on Mobile Computing*, *13*(3), 638–648.

Drinking culture. (2018). Retrieved from https://en.wikipedia.org/wiki/Drinking_culture

Duff, C. (2012). Accounting for context: exploring the role of objects and spaces in the consumption of alcohol and other drugs. *Social & Cultural Geography*, *13*(2), 145–159.

Duffey, K. J., & Popkin, B. M. (2011). Energy density, portion size, and eating occasions: contributions to increased energy intake in the united states, 1977–2006. *PLoS medicine*, *8*(6), e1001050.

Eagle, N., & Pentland, A. (2005). Social serendipity: mobilizing social software. *IEEE Pervasive Computing*, *4*(2), 28–34.

Ekholm, O. (2004). Influence of the recall period on self-reported alcohol intake. *European Journal of Clinical Nutrition*, *58*(1), 60.

Ellison, N., Heino, R., & Gibbs, J. (2006). Managing impressions online: self-presentation processes in the online dating environment. *Journal of computer-mediated communication*, *11*(2), 415–441.

Fang, H., Gupta, S., Iandola, F., Srivastava, R., Deng, L., Dollár, P., … Platt, J., et al. (2015). From captions to visual concepts and back.

Federal Law, S. (2012). *Loi fédérale sur le commerce des boissons alcooliques*.

Ffmpeg. (2010). Ffmpeg. Retrieved from http://www.ffmpeg.org

Food and Agriculture Pocket Statistic 2016. (2016). Accessed: 2017-08-21. Retrieved from https://www.bfs.admin.ch/bfs/en/home/statistics/agriculture-forestry/food.html

Foursquare Venue Categories. (2018). Retrieved from https://developer.foursquare.com/categorytree

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., & Bell, D. (2014). Analyzing the language of food on social media. In *Big data (big data), 2014 ieee international conference on* (pp. 778–783). IEEE.

Ganti, R. K., Ye, F., & Lei, H. (2011). Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine*, *49*(11), 32–39.

Gemmeke, J. F., Ellis, D. P., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., … Ritter, M. (2017). Audio set: an ontology and human-labeled dataset for audio events. In *Acoustics, speech and signal processing (icassp), 2017 ieee international conference on* (pp. 776–780). IEEE.

Global status report on alcohol and health from WHO. (2018). Accessed: 2018-12-03. Retrieved from http://apps.who.int/iris/bitstream/handle/10665/274603/9789241565639-eng.pdf

Gmel, G., Gaume, J., Faouzi, M., Kulling, J.-P., & Daeppen, J.-B. (2008). Who drinks most of the total alcohol in young men—risky single occasion drinking as normative behaviour. *Alcohol & Alcoholism*, *43*(6), 692–697.

Gmel, G., & Rehm, J. (2004). Measuring alcohol consumption. *Contemporary Drug Problems*, *31*(3), 467–540.

Gmel, G., Rehm, J., Kuntsche, E., et al. (2003). Binge drinking in europe: definitions, epidemiology, and consequences. *Sucht*, *49*(2), 105–116.

Goffman, E. (1959). The presentation of self in. *Butler, Bodies that Matter*.

Gosling, S. D. [Samuel D], Craik, K. H., Martin, N. R., & Pryor, M. R. (2005). The personal living space cue inventory: an analysis and evaluation. *Environment and Behavior*, *37*(5), 683–705.

Gosling, S. D. [Samuel D], Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: personality judgments based on offices and bedrooms. *Journal of personality and social psychology*, *82*(3), 379.

Gosling, S., Gifford, R., & McCunn, L. J. (2013). The selection, creation, and perception of interior spaces: an environmental psychology approach.

Graham, L. T., & Gosling, S. D. [Samuel D]. (2011). Can the ambiance of a place be determined by the user profiles of the people who visit it? In *Icwsm*.

Graham, L. T., Gosling, S. D., & Travis, C. K. (2015). The psychology of home environments: a call for research on residential space. *Perspectives on Psychological Science*, *10*(3), 346–356.

Guo, Y., Zhang, L., Hu, Y., He, X., & Gao, J. (2016). Ms-celeb-1m: challenge of recognizing one million celebrities in the real world. *Electronic Imaging*, *2016*(11), 1–6.

Hatuka, T., & Toch, E. (2017). Being visible in public space: the normalisation of asymmetrical visibility. *Urban Studies*, *54*(4), 984–998.

Hawks, D., Scott, K., McBride, N., Jones, P., Stockwell, T., Organization, W. H., et al. (2002). Prevention of psychoactive substance use: a selected review of what works in the area of prevention.

Hayden, D. (2004). *Building suburbia: green fields and urban growth, 1820-2000*. Vintage.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Hecht, B. J., & Stephens, M. (2014). A tale of cities: urban biases in volunteered geographic information. *ICWSM*, *14*, 197–205.

Hendriks, H., Van den Putte, B., Gebhardt, W. A., & Moreno, M. A. (2018). Social drinking on social media: content analysis of the social aspects of alcohol-related posts on facebook and instagram. *Journal of Medical Internet Research*, *20*(6), e226.

Holloway, S. L., Jayne, M., & Valentine, G. (2008). 'sainsbury's is my local': english alcohol policy, domestic drinking practices and the meaning of home. *Transactions of the Institute of British Geographers*, *33*(4), 532–547.

Holmberg, C., Chaplin, J. E., Hillman, T., & Berg, C. (2016). Adolescents' presentation of food in social media: an explorative study. *Appetite*, *99*, 121–129.

Holton, M. (2016). Living together in student accommodation: performances, boundaries and homemaking. *Area*, *48*(1), 57–63.

Hossain, N., Hu, T., Feizi, R., White, A. M., Luo, J., & Kautz, H. (2016). Inferring fine-grained details on user activities and home location from social media: detecting drinking-while-tweeting patterns in communities. *arXiv preprint arXiv:1603.03181*.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... Adam, H. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Hubbard, P. (2008). Regulating the social impacts of studentification: a loughborough case study. *Environment and Planning A*, *40*(2), 323–341.

Hughes, K., Anderson, Z., Morleo, M., & Bellis, M. A. (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), 60–65.

Ikkala, T., & Lampinen, A. (2015). Monetizing network hospitality: hospitality and sociability in the context of airbnb. In *Proceedings of the 18th acm conference on computer supported cooperative work & social computing* (pp. 1033–1044). ACM.

Jang, J. Y., Han, K., Shih, P. C., & Lee, D. (2015). Generation like: comparative characteristics in instagram. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 4039–4042). ACM.

Järvinen, M., & Østergaard, J. (2009). Governing adolescent drinking. *Youth & society*, *40*(3), 377–402.

Jayne, M., Holloway, S. L., & Valentine, G. (2006). Drunk and disorderly: alcohol, urban life and public space. *Progress in human geography*, *30*(4), 451–468.

Jensen Schau, H., & Gilly, M. C. (2003). We are what we post? self-presentation in personal web space. *Journal of consumer research*, *30*(3), 385–404.

Johnston, L. (2010). *Monitoring the future: national results on adolescent drug use: overview of key findings*. Diane Publishing.

Kairouz, S., Gliksman, L., Demers, A., & Adlaf, E. M. (2002). For all these reasons, i do... drink: a multilevel analysis of contextual reasons for drinking among canadian undergraduates. *Journal of studies on alcohol*, *63*(5), 600–608.

Kao, H.-L. C., Ho, B.-J., Lin, A. C., & Chu, H.-H. (2012). Phone-based gait analysis to detect alcohol usage. In *Proceedings of the 2012 acm conference on ubiquitous computing* (pp. 661–662). ACM.

Karantonis, D. M., Narayanan, M. R., Mathie, M., Lovell, N. H., & Celler, B. G. (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), 156–167.

Kawano, Y., & Yanai, K. (2014). Food image recognition with deep convolutional features. In *Proceedings of the 2014 acm international joint conference on pervasive and ubiquitous computing: adjunct publication* (pp. 589–593). ACM.

Kazemi, D. M., Borsari, B., Levine, M. J., Li, S., Lamberson, K. A., & Matta, L. A. (2017). A systematic review of the mhealth interventions to prevent alcohol and substance abuse. *Journal of health communication*, *22*(5), 413–432.

Kelly, D., Smyth, B., & Caulfield, B. (2013). Uncovering measurements of social and demographic behavior from smartphone location data. *IEEE Transactions on Human-Machine Systems*, *43*(2), 188–198.

Kershaw, D., Rowe, M., & Stacey, P. (2014a). Towards tracking and analysing regional alcohol consumption patterns in the uk through the use of social media. In *Proceedings of the 2014 acm conference on web science* (pp. 220–228). WebSci '14. doi:10.1145/2615569.2615678

Kershaw, D., Rowe, M., & Stacey, P. (2014b). Towards tracking and analysing regional alcohol consumption patterns in the uk through the use of social media. In *Proceedings of the 2014 acm conference on web science* (pp. 220–228). ACM.

Keval, H., & Sasse, M. A. (2008). To catch a thief–you need at least 8 frames per second: the impact of frame rates on user performance in a cctv detection task. In *Proceedings of the 16th acm international conference on multimedia* (pp. 941–944). ACM.

Khajehzadeh, I., & Vale, B. (2015). How do people use large houses. In *Living and learning: research for a better built environment: proceeding of the 49th international conference of the architectural science association 2015* (pp. 153–162).

Kjærgaard, M. B., & Nurmi, P. (2012). Challenges for social sensing using wifi signals. In *Proceedings of the 1st acm workshop on mobile systems for computational social science* (pp. 17–21). ACM.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155–163.

Kotler, P. (1973). Atmospherics as a marketing tool. *Journal of retailing*, *49*(4), 48–64.

Kuntsche, E., & Gmel, G. (2013). Alcohol consumption in late adolescence and early adulthood–where is the problem. *Swiss Med Wkly*, *143*, w13826.

Kuntsche, E., Knibbe, R., Gmel, G., & Engels, R. (2005). Why do young people drink? a review of drinking motives. *Clinical psychology review*, *25*(7), 841–861.

Kuntsche, E., & Kuntsche, S. (2009). Development and validation of the drinking motive questionnaire revised short form (dmq–r sf). *Journal of Clinical Child & Adolescent Psychology*, *38*(6), 899–908.

Kuntsche, E., Kuntsche, S., Thrul, J., & Gmel, G. (2017). Binge drinking: health impact, prevalence, correlates and interventions. *Psychology & health*, *32*(8), 976–1017.

Kuntsche, E., & Labhart, F. (2012). Investigating the drinking patterns of young people over the course of the evening at weekends. *Drug and alcohol dependence*, *124*(3), 319–324.

Kuntsche, E., & Labhart, F. (2013). Icat: development of an internet-based data collection method for ecological momentary assessment using personal cell phones. *European Journal of Psychological Assessment*.

Kuntsche, E., Otten, R., & Labhart, F. (2015). Identifying risky drinking patterns over the course of saturday evenings: an event-level study. *Psychology of addictive behaviors*, *29*(3), 744.

Kusmierczyk, T., & Nørvåg, K. (2016). Online food recipe title semantics: combining nutrient facts and topics. In *Proceedings of the 25th acm international on conference on information and knowledge management* (pp. 2013–2016). ACM.

Labhart, F., Engels, R., & Kuntsche, E. (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), 644–648.

Labhart, F., Graham, K., Wells, S., & Kuntsche, E. (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), 284–291.

Labhart, F., Livingston, M., Engels, R., & Kuntsche, E. (2018). After how many drinks does someone experience acute consequences—determining thresholds for binge drinking based on two event-level studies. *Addiction*, *113*(12), 2235–2244.

Labhart, F., Santani, D., Truong, J., Tarsetti, F., Bornet, O., Landolt, S., … Kuntsche, E. (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), 721–736.

Labhart, F., Tarsetti, F., Bornet, O., Santani, D., Truong, J., Landolt, S., … Kuntsche, E. (2019a). Capturing drinking and nightlife behaviours and their social and physical context with a smartphone application - investigation of users' experience and reactivity. *Addiction Research and Theory*. Retrieved from http://infoscience.epfl.ch/record/263840

Labhart, F., Tarsetti, F., Bornet, O., Santani, D., Truong, J., Landolt, S., … Kuntsche, E. (2019b). Capturing drinking and nightlife behaviours and their social and physical context with a smartphone application–investigation of users' experience and reactivity. *Addiction Research & Theory*, 1–14.

Labhart, F., Wells, S., Graham, K., & Kuntsche, E. (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), 327–335.

Languages Used in Switzerland. (2020). Retrieved from https://www.eda.admin.ch/aboutswitzerland/en/home/gesellschaft/sprachen/die-sprachen---fakten-und-zahlen.html

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening?: privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 102.

Lee, D. K. (2016). Alternatives to p value: confidence interval and effect size. *Korean journal of anesthesiology*, *69*(6), 555–562.

Leech, R. M., Worsley, A., Timperio, A., & McNaughton, S. A. (2015). Understanding meal patterns: definitions, methodology and impact on nutrient intake and diet quality. *Nutrition research reviews*, *28*(1), 1–21.

Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, *40*(2), 61–77.

Liaw, A., Wiener, M. et al. (2002). Classification and regression by randomforest. *R news*, *2*(3), 18–22.

LiKamWa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013). Moodscope: building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on mobile systems, applications, and services* (pp. 389–402). ACM.

Lincoln, S. (2012). *Youth culture and private space*. Springer.

Lincoln, S. (2014). "i've stamped my personality all over it" the meaning of objects in teenage bedroom space. *Space and Culture*, *17*(3), 266–279.

Lincoln, S. (2015). 'my bedroom is me': young people, private space, consumption and the family home. In *Intimacies, critical consumption and diverse economies* (pp. 87–106). Springer.

Lincoln, S., & Robards, B. (2016). Being strategic and taking control: bedrooms, social network sites and the narratives of growing up. *new media & society*, *18*(6), 927–943.

Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., & Zimmerman, J. (2011a). I'm the mayor of my house: examining why people use foursquare - a social-driven location sharing application. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 2409–2418). CHI '11. Vancouver, BC, Canada: ACM.

Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., & Zimmerman, J. (2011b). I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 2409–2418). ACM.

Long, M., Cao, Y., Wang, J., & Jordan, M. I. (2015). Learning transferable features with deep adaptation networks. *arXiv preprint arXiv:1502.02791*.

Ma, J., Betts, N. M., & Hampl, J. S. (2000). Clustering of lifestyle behaviors: the relationship between cigarette smoking, alcohol consumption, and dietary intake. *American Journal of Health Promotion*, *15*(2), 107–117.

Mahajan, D., Girshick, R., Ramanathan, V., He, K., Paluri, M., Li, Y., ... van der Maaten, L. (2018). Exploring the limits of weakly supervised pretraining. In *Proceedings of the european conference on computer vision (eccv)* (pp. 181–196).

Malmi, E., Do, T. M. T., & Gatica-Perez, D. (2012). Checking in or checked in: comparing large-scale manual and automatic location disclosure patterns. In *Proceedings of the 11th international conference on mobile and ubiquitous multimedia* (p. 26). ACM.

Malmi, E., Do, T. M. T., & Gatica-Perez, D. (2013). From foursquare to my square: learning check-in behavior from multiple sources. In *Icwsm*.

Marczinski, C. A., Hertzenberg, H., Goddard, P., Maloney, S. F., Stamates, A. L., & O Connor, K. (2016). Alcohol-related facebook activity predicts alcohol use patterns in college students. *Addiction research & theory*, *24*(5), 398–405.

Marquard, D. (2014). Zurich, the party town. Accessed: July 15, 2016. Retrieved from http://bit.ly/28M8nbk

Marshall, D., & Bell, R. (2003). Meal construction: exploring the relationship between eating occasion and location. *Food quality and Preference*, *14*(1), 53–64.

Mathie, M. J. [Merryn J], Coster, A. C., Lovell, N. H., Celler, B. G., Lord, S. R., & Tiedemann, A. (2004). A pilot study of long-term monitoring of human movements in the home using accelerometry. *Journal of telemedicine and telecare*, *10*(3), 144–151.

Mathie, M., Celler, B. G., Lovell, N. H., & Coster, A. (2004). Classification of basic daily movements using a triaxial accelerometer. *Medical and Biological Engineering and Computing*, *42*(5), 679–687.

Matthews, H., Limb, M., & Percy-Smith, B. (1998). Changing worlds: the microgeographies of young teenagers. *Tijdschrift voor economische en sociale geografie*, *89*(2), 193–202.

Maurer, U., Smailagic, A., Siewiorek, D. P., & Deisher, M. (2006). *Activity recognition and monitoring using multiple sensors on different body positions.* CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE.

McCabe, S. E. (2002). Gender differences in collegiate risk factors for heavy episodic drinking. *Journal of Studies on Alcohol*, *63*(1), 49–56.

McCarty, D. (1985). Environmental factors in substance abuse. In *Determinants of substance abuse* (pp. 247–281). Springer.

Measham, F., & Brain, K. (2005). Binge drinking, british alcohol policy and the new culture of intoxication. *Crime, media, culture*, *1*(3), 262–283.

Mejova, Y., Haddadi, H., Noulas, A., & Weber, I. (2015). # foodporn: obesity patterns in culinary interactions. In *Proceedings of the 5th international conference on digital health 2015* (pp. 51–58). ACM.

menuCH: Resultats concernant la consommation alimentaire. Consommation des differents groupes d'aliments. (2017). Accessed: 2017-08-23. Retrieved from https://www.blv.admin.ch/blv/fr/home/lebensmittel-und-ernaehrung/ernaehrung/menuch/menu-ch-ergebnisse-ernaehrung.html

Meyers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., … Murphy, K. P. (2015). Im2calories: towards an automated mobile vision food diary. In *Proceedings of the ieee international conference on computer vision* (pp. 1233–1241).

Min, A., Lee, D., & Shih, P. C. (2018). Potentials of smart breathalyzer: interventions for excessive drinking among college students. In *International conference on information* (pp. 195–206). Springer.

Min, J.-K., Doryab, A., Wiese, J., Amini, S., Zimmerman, J., & Hong, J. I. (2014). Toss'n'turn: smartphone as sleep and sleep quality detector. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 477–486). ACM.

Mirtchouk, M., Merck, C., & Kleinberg, S. (2016). Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the*

2016 acm international joint conference on pervasive and ubiquitous computing (pp. 451–462). ACM.

Montoliu, R., Blom, J., & Gatica-Perez, D. (2013). Discovering places of interest in everyday life from smartphone data. *Multimedia tools and applications*, *62*(1), 179–207.

Moon, J. H., Lee, E., Lee, J.-A., Choi, T. R., & Sung, Y. (2016). The role of narcissism in self-promotion on instagram. *Personality and Individual Differences*, *101*, 22–25.

Nguyen, L. S., Ruiz-Correa, S., Mast, M. S., & Gatica-Perez, D. (2018). Check out this place: inferring ambiance from airbnb photos. *IEEE Transactions on Multimedia*, *20*(6), 1499–1511.

of Economic, D., & Affairs, S. (2018). 68% of the world population projected to live in urban areas by 2050. Accessed: 2020-01-01. Retrieved from https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html

of Lausanne, C. (2010). Drinking in public space in city of lausanne: a challenge for police. Accessed: July 15, 2016. Retrieved from http://bit.ly/28M8X8X

Office, F. S. (n.d.). Statistical data on switzerland. *Statistical Data on Switzerland 2018*. Retrieved from https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/publications.assetdetail.4522209.html

Ordonez, V., & Berg, T. L. (2014). Learning high-level judgments of urban perception. In *European conference on computer vision* (pp. 494–510). Springer.

Pang, R., Baretto, A., Kautz, H., & Luo, J. (2015). Monitoring adolescent alcohol use via multimodal analysis in social multimedia. In *Big data (big data), 2015 ieee international conference on* (pp. 1509–1518). IEEE.

Papliatseyeu, A., & Mayora, O. (2009). Mobile habits: inferring and predicting user activities with a location-aware smartphone. In *3rd symposium of ubiquitous computing and ambient intelligence 2008* (pp. 343–352). Springer.

Parliament, E., & of the European Union, C. (2016). Regulation (eu) 2016/679 of the european parliament and of the council. Retrieved from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Petraitis, J., Flay, B. R., & Miller, T. Q. (1995). Reviewing theories of adolescent substance use: organizing pieces in the puzzle. *Psychological bulletin*, *117*(1), 67.

Petrill, S. A., Pike, A., Price, T., & Plomin, R. (2004). Chaos in the home and socioeconomic status are associated with cognitive development in early childhood: environmental mediators identified in a genetic design. *Intelligence*, *32*(5), 445–460.

Phan, T.-T., Florian, L., Muralidhar, S., & Gatica-Perez, D. (2020). Understanding heavy drinking nights through smartphone sensing and active human engagement. In *Proceedings of the 14th eai international conference on pervasive computing technologies for healthcare*. ACM.

Phan, T.-T., & Gatica-Perez, D. (2017). Healthy# fondue# dinner: analysis and inference of food and drink consumption patterns on instagram. In *Proceedings of the 16th international conference on mobile and ubiquitous multimedia* (pp. 327–338). ACM.

Phan, T.-T., Labhart, F., & Gatica-Perez, D. (2019). My own private nightlife: understanding youth personal spaces from crowdsourced video. *Proc. ACM Hum.-Comput. Interact*, *3*(189).

Phan, T.-T., Muralidhar, S., & Gatica-Perez, D. (2019a). # drink or# drunk: multimodal signals and drinking practices on instagram. In *Proceedings of the 13th eai international conference on pervasive computing technologies for healthcare* (pp. 71–80). ACM.

Phan, T.-T., Muralidhar, S., & Gatica-Perez, D. (2019b). Drinks & crowds: characterizing alcohol consumption through crowdsensing and social media. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *3*(2), 59.

Phuttharak, J., & Loke, S. W. (2018). A review of mobile crowdsourcing architectures and challenges: toward crowd-empowered internet-of-things. *IEEE Access*, *7*, 304–324.

Pliner, P., & Zec, D. (2007). Meal schemas during a preload decrease subsequent eating. *Appetite*, *48*(3), 278–288.

Porzi, L., Rota Bulò, S., Lepri, B., & Ricci, E. (2015). Predicting and understanding urban perception with convolutional neural networks. In *Proceedings of the 23rd acm international conference on multimedia* (pp. 139–148). ACM.

Quercia, D., O'Hare, N. K., & Cramer, H. (2014). Aesthetic capital: what makes london look beautiful, quiet, and happy? In *Proceedings of the 17th acm conference on computer supported cooperative work & social computing* (pp. 945–955). ACM.

Ragusa, F., Tomaselli, V., Furnari, A., Battiato, S., & Farinella, G. M. (2016). Food vs non-food classification. In *Proceedings of the 2nd international workshop on multimedia assisted dietary management* (pp. 77–81). ACM.

Rahman, T., Adams, A. T., Zhang, M., Cherry, E., Zhou, B., Peng, H., & Choudhury, T. (2014). Bodybeat: a mobile system for sensing non-speech body sounds. In *Mobisys* (Vol. 14, pp. 2–13).

Ravi, N., Dandekar, N., Mysore, P., & Littman, M. L. (2005). Activity recognition from accelerometer data. In *Aaai* (Vol. 5, *2005*, pp. 1541–1546).

Redi, M., Aiello, L. M., Schifanella, R., & Quercia, D. (2018). The spirit of the city: using social media to capture neighborhood ambiance. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 144.

Redi, M., Quercia, D., Graham, L. T., & Gosling, S. D. (2015). Like partying? your face says it all. predicting the ambiance of places with profile pictures. *arXiv preprint arXiv:1505.07522*.

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, *6*(1), 15.

Rehm, J., Room, R., Graham, K., Monteiro, M., Gmel, G., & Sempos, C. T. (2003). The relationship of average volume of alcohol consumption and patterns of drinking to burden of disease: an overview. *Addiction*, *98*(9), 1209–1228.

Results on Food Consumption Survey. (2017). Accessed: 2017-08-07. Retrieved from https://www.blv.admin.ch/blv/fr/home/lebensmittel-und-ernaehrung/ernaehrung/menuch.html

Rheingold, H. L., & Cook, K. V. (1975). The contents of boys' and girls' rooms as an index of parents' behavior. *Child development*, 459–463.

Rich, J., Haddadi, H., & Hospedales, T. M. (2016). Towards bottom-up analysis of social food. In *Proceedings of the 6th international conference on digital health conference* (pp. 111–120). DH '16. Montréal, Québec, Canada: ACM.

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

Sall, A., & Grinter, R. E. (2007). Let's get physical! in, out and around the gaming circle of physical gaming at home. *Computer Supported Cooperative Work (CSCW)*, *16*(1-2), 199–229.

Santani, D., Biel, J.-I., Labhart, F., Truong, J., Landolt, S., Kuntsche, E., & Gatica-Perez, D. (2016). The night is young: urban crowdsourcing of nightlife patterns. In *Proceedings of the 2016 acm international joint conference on pervasive and ubiquitous computing* (pp. 427–438). ACM.

Santani, D., Do, T.-M.-T., Labhart, F., Landolt, S., Kuntsche, E., & Gatica-Perez, D. (2017). Drinksense: characterizing youth drinking behavior using smartphones. In *Proceedings of the 2017 acm international joint conference on pervasive and ubiquitous computing*. ACM.

Santani, D., & Gatica-Perez, D. (2013a). Speaking swiss: languages and venues in foursquare. MM '13, 501–504. doi:10.1145/2502081.2502133

Santani, D., & Gatica-Perez, D. (2013b). Speaking swiss: languages and venues in foursquare. In *Proceedings of the 21st acm international conference on multimedia* (pp. 501–504). ACM.

Santani, D., & Gatica-Perez, D. (2015). Loud and trendy: crowdsourcing impressions of social ambiance in popular indoor urban places. In *Proceedings of the 23rd acm international conference on multimedia* (pp. 211–220). ACM.

Santani, D., Hu, R., & Gatica-Perez, D. (2016). Innerview: learning place ambiance from social media images. In *Proceedings of the 2016 acm on multimedia conference* (pp. 451–455). ACM.

Sapiezynski, P., Stopczynski, A., Gatej, R., & Lehmann, S. (2015). Tracking human mobility using wifi signals. *PloS one*, *10*(7), e0130824.

Sapiezynski, P., Stopczynski, A., Wind, D. K., Leskovec, J., & Lehmann, S. (2017). Inferring person-to-person proximity using wifi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(2), 24.

Scherer, K. R., Zentner, M. R. et al. (2001). Emotional effects of music: production rules. *Music and emotion: Theory and research*, *361*(2001), 392.

Schiano, D. J., Elliott, A., & Bellotti, V. (2007). A look at tokyo youth at leisure: towards the design of new media to support leisure outings. *Computer Supported Cooperative Work (CSCW)*, *16*(1-2), 45–73.

Schlenker, B. R. (2012). Self-presentation.

Scholl, P. M., Kücükyildiz, N., & Laerhoven, K. V. (2013). When do you light a fire?: capturing tobacco use with situated, wearable sensors. In *Proceedings of the 2013 acm conference on pervasive and ubiquitous computing adjunct publication* (pp. 1295–1304). UbiComp '13 Adjunct. doi:10.1145/2494091.2497284

Sharma, S. S., & De Choudhury, M. (2015). Measuring and characterizing nutritional information of food and ingestion content in instagram. In *Proceedings of the 24th international conference on world wide web* (pp. 115–116). ACM.

Shaw, R. (2015). Controlling darkness: self, dark and the domestic night. *cultural geographies*, *22*(4), 585–600.

Shrikhande, A. J. (2000). Wine by-products with health benefits. *Food Research International*, *33*(6), 469–474.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.

Siahaan, E., Hanjalic, A., & Redi, J. (2016). A reliable methodology to collect ground truth data of image aesthetic appeal. *IEEE Transactions on Multimedia*, *18*(7), 1338–1350.

Silva, T. H., de Melo, P. O. V., Almeida, J. M., Musolesi, M., & Loureiro, A. A. (2014). You are what you eat (and drink): identifying cultural boundaries by analyzing food and drink habits in foursquare. In *Icwsm*.

Skelton, T., & Valentine, G. (2005). Cool places. In *Cool places* (pp. 11–42). Routledge.

Sklearn with GridSearchCV. (2018). Retrieved from http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

Souza, F., de Las Casas, D., Flores, V., Youn, S., Cha, M., Quercia, D., & Almeida, V. (2015). Dawn of the selfie era: the whos, wheres, and hows of selfies on instagram. In *Proceedings of the 2015 acm on conference on online social networks* (pp. 221–231). ACM.

Stanesby, O., Labhart, F., Dietze, P., Wright, C. J., & Kuntsche, E. (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).

Stevely, A. K., Holmes, J., & Meier, P. S. (2019). Contextual characteristics of adults' drinking occasions and their association with levels of alcohol consumption and acute alcohol-related harm: a mapping review. *Addiction*.

Stroebele, N., & De Castro, J. M. (2004). Effect of ambience on food intake and food choice. *Nutrition*, *20*(9), 821–838.

Su, X., Tong, H., & Ji, P. (2014). Activity recognition with smartphone sensors. *Tsinghua science and technology*, *19*(3), 235–249.

Swiss Food Composition Database. (2017). Accessed: 2017-08-01. Retrieved from http://naehrwertdaten.ch/request?query=TopCategoryList&xml=MessageData&xml=MetaData&xsl=ListCategories&lan=de&range=0-19

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Aaai* (Vol. 4, p. 12).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2818–2826).

Taylor, B., Irving, H., Kanteres, F., Room, R., Borges, G., Cherpitel, C., ... Rehm, J. (2010). The more you drink, the harder you fall: a systematic review and meta-analysis of how acute alcohol consumption and injury or collision risk increase together. *Drug & Alcohol Dependence*, *110*(1), 108–116.

Thebault-Spieker, J., Terveen, L. G., & Hecht, B. (2015). Avoiding the south side and the suburbs: the geography of mobile crowdsourcing markets. In *Proceedings of the 18th acm conference on computer supported cooperative work & social computing* (pp. 265–275). ACM.

Thomaz, E., Essa, I., & Abowd, G. D. (2015). A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 acm international joint conference on pervasive and ubiquitous computing* (pp. 1029–1040). ACM.

Thrul, J., & Kuntsche, E. (2015). The impact of friends on young adults' drinking over the course of the evening—an event-level analysis. *Addiction*, *110*(4), 619–626.

Toch, E., & Levi, I. (2013). Locality and privacy in people-nearby applications. In *Proceedings of the 2013 acm international joint conference on pervasive and ubiquitous computing* (pp. 539–548). ACM.

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

Truong, J. (2018a). Attending to others: how digital technologies direct young people's nightlife. *Geographica Helvetica*, *73*(2), 193–201.

Truong, J. (2018b). Collapsing contexts: social networking technologies in young people's nightlife. *Children's Geographies*, *16*(3), 266–278.

Tuan, Y.-F. (1971). Geography, phenomenology, and the study of human nature. *Canadian Geographer/Le Géographe canadien*, *15*(3), 181–192.

Tucker, I. (2010). Everyday spaces of mental distress: the spatial habituation of home. *Environment and Planning D: Society and Space*, *28*(3), 526–538.

Valentine, G., Holloway, S., Knell, C., & Jayne, M. (2008). Drinking places: young people and cultures of alcohol consumption in rural environments. *Journal of Rural Studies*, *24*(1), 28–40.

Van Liempt, I., Van Aalst, I., & Schwanen, T. (2015). Introduction: geographies of the urban night. Sage Publications Sage UK: London, England.

Veltink, P. H., Bussmann, H. J., De Vries, W., Martens, W. J., Van Lummel, R. C., et al. (1996). Detection of static and dynamic activities using uniaxial accelerometers. *IEEE Transactions on Rehabilitation Engineering*, *4*(4), 375–385.

Vich, G., Marquet, O., & Miralles-Guasch, C. (2017). Suburban commuting and activity spaces: using smartphone tracking data to understand the spatial extent of travel behaviour. *The Geographical Journal*, *183*(4), 426–439.

Wang, H., Lymberopoulos, D., & Liu, J. (2014). Local business ambience characterization through mobile audio sensing. In *Proceedings of the 23rd international conference on world wide web* (pp. 293–304). ACM.

Wansink, B., Payne, C. R., & Shimizu, M. (2010). "is this a meal or snack?" situational cues that drive perceptions. *Appetite*, *54*(1), 214–216.

Watts, C. E. (2004). Predicting free time activity involvement of adolescents: the influence of adolescent motivation, adolescent initiative, and perceptions of parenting.

Weafer, J., & Fillmore, M. T. (2013). Acute alcohol effects on attentional bias in heavy and moderate drinkers. *Psychology of addictive behaviors*, *27*(1), 32.

Wechsler, H., & Nelson, T. F. (2001). Binge drinking and the american college students: what's five drinks? *Psychology of Addictive Behaviors*, *15*(4), 287.

Wilkinson, S. (2017). Drinking in the dark: shedding light on young people's alcohol consumption experiences. *Social & Cultural Geography*, *18*(6), 739–757.

Wilkinson, S., & Wilkinson, C. (2018). Night-life and young people's atmospheric mobilities. *Mobile Culture Studies. The Journal*, *3*(2017), 77–96.

Wong, R. Y., Mulligan, D. K., Van Wyk, E., Pierce, J., & Chuang, J. (2017). Eliciting values reflections by engaging privacy futures using design workbooks. *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 111.

Wright, C., Dietze, P. M., Agius, P. A., Kuntsche, E., Livingston, M., Black, O. C., . . . Lim, M. S. (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), e149.

Xu, Q., Erman, J., Gerber, A., Mao, Z., Pang, J., & Venkataraman, S. (2011). Identifying diverse usage behaviors of smartphone apps. In *Proceedings of the 2011 acm sigcomm conference on internet measurement conference* (pp. 329–344). ACM.

Yan, T., Marzilli, M., Holmes, R., Ganesan, D., & Corner, M. (2009). Mcrowd: a platform for mobile crowdsourcing. In *Proceedings of the 7th acm conference on embedded networked sensor systems* (pp. 347–348). ACM.

Yan, Z., Yang, J., & Tapia, E. M. (2013). Smartphone bluetooth based social sensing. In *Proceedings of the 2013 acm conference on pervasive and ubiquitous computing adjunct publication* (pp. 95–98). ACM.

Yarosh, S. (2013). Shifting dynamics or breaking sacred traditions?: the role of technology in twelve-step fellowships. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 3413–3422). ACM.

Yatani, K. (2016a). Effect sizes and power analysis in hci. In *Modern statistical methods for hci* (pp. 87–110). Springer.

Yatani, K. (2016b). Effect sizes and power analysis in hci. In *Modern statistical methods for hci* (pp. 87–110). Springer.

You, C.-w., Wang, K.-C., Huang, M.-C., Chen, Y.-C., Lin, C.-L., Ho, P.-S., . . . Chu, H.-H. (2015). Soberdiary: a phone-based support system for assisting recovery from alcohol dependence. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 3839–3848). CHI '15. Seoul, Republic of Korea: ACM.

Zepeda, L., & Deal, D. (2008). Think before you eat: photographic food diaries as intervention tools to change dietary decision making and attitudes. *International Journal of Consumer Studies*, *32*(6), 692–698.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: a 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2018). Places: a 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, *40*(6), 1452–1464.

Zong, J., & Matias, J. N. (2018). Automated debriefing: interface for large-scale research ethics. In *Companion of the 2018 acm conference on computer supported cooperative work and social computing* (pp. 21–24). ACM.

Thanh-Trung Phan

Ph.D. Candidate (expected in May 2020)

Avenue de la gare 64
Martigny, VA 1920, Switzerland
Working Permit B in Switzerland
+41(079) 529 6380
phanthanhtrung86@gmail.com
https://www.idiap.ch/~tphan/
phanthanhtrung.86
Trung T. Phan

Good at using machine learning in computer science to solve real-world problems in computational social science; interested in gaining insights of human behavior with social patterns using multi-dimensional data from social media and crowdsourcing ubicomp data including mobile sensors, videos/images, or textual contents; very good at coding ability (Python, R), algorithms (machine learning, deep learning), and applied statistics; bi-lingual; international working experience; good communications skill; a proven team-work with excellent problem-solving skills; committed to continuous learning and passion for new technology; enjoy challenges.

EDUCATION

2016–2020 **Ph.D.**, *Computer Science*, École polytechnique fédérale de Lausanne (EPFL), Lausanne–Switzerland.

- Advisor: Prof. Daniel Gatica-Perez (Idiap Research Institute and EPFL, Switzerland).
- Thesis: "Understanding Eating And Drinking In Context From Crowdsourced Data". Understanding food and drink consumption behavior of young people and its related consequences based on social media and mobile crowdsensing data. My dissertation is a pilot concrete step to understand how social media and mobile crowdsensing data could be used to study social issues and human behavior.

2011–2013 **Master of Engineer**, *Multimedia and Communications*, Telecom Paristech and Eurecom, Biot–France.

- Advisor: Prof. Ylä-Jääski Antti (Aalto University, Espoo, Finland) and Prof. Raphaël Troncy (Eurecom Institute, Biot, France).
- Thesis: "Diffusion rate on Twitter: Based on semantic relatedness and homophily of users".

2004–2009 **Bachelor**, *Software Engineer*, University of Science, Vietnam National University (VNU-HCMC), Ho Chi Minh City–Vietnam.

- Advisor: Ph.D. Nguyen Tran Minh Thu (University of Science, VNU-HCMC, Vietnam).
- Thesis: "Learning about Android OS and developing illustration application on Android OS - eSaleShopping".

PUBLICATIONS

JOURNALS

2019 IMWUT **Thanh-Trung Phan**, Skanda Muralidhar and Daniel Gatica-Perez. Drinks & Crowds: Characterizing Alcohol Consumption through Crowdsensing and Social Media. In Journal and Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT '19), Vol. 3, No. 2, Article 59 June 2019. PDF

2019 CSCW **Thanh-Trung Phan**, Florian Labhart and Daniel Gatica-Perez. My Own Private Nightlife: Understanding Youth Personal Spaces from Crowdsourced Video. In Journal and Proceedings of the ACM on Human-Computer Interaction (CSCW '2019), November 2019. *(Acceptance rate: 31%)*. PDF

CONFERENCES & WORKSHOPS

2020 PervasiveHealth — **Thanh-Trung Phan**, Florian Labhart, Skanda Muralidhar, and Daniel Gatica-Perez. Understanding Heavy Drinking at Night through Smartphone Sensing and Active Human Engagement. In Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare(Atlanta, USA, May 20-23, 2020). ACM, 2020.

2019 PervasiveHealth — **Thanh-Trung Phan**, Skanda Muralidhar, and Daniel Gatica-Perez. #Drink or #Drunk: Multimodal signals and drinking practices on Instagram. In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, pages 71–80. ACM, 2019. Honorable Mention Paper Award. PDF.

2019 FAT/MM — Daniel Gatica-Perez, Darshan Santani, Joan.-I. Biel, and **Thanh-Trung Phan** Social Multimedia, Diversity, and Global South Cities: A Double Blind Side in Proc. ACM Workshop on Fairness, Accountability, and Transparency in Multimedia (FAT/MM), Nice, Oct. 2019. PDF.

2017 MUM — **Thanh-Trung Phan**, Daniel Gatica-Perez. #Healthy #fondue #dinner: Analysis and inference of food and drink consumption patterns on Instagram. In Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia, pages 327–338. ACM, 2017. PDF

EXPERIENCE

Jan 2016 – Apr 2020 — **Research Assistant**, *Data Science*, Idiap Research Institute and EPFL, Switzerland.
Uncovering alcohol drinking and eating in context by using mobile crowdsensing and social network data in the framework of the multiple projects (IMI, Dusk2Dawn, IFNC) funded by Swiss National Science Foundation (SNSF) and Swiss Government Excellence Scholarship, under the supervision of Prof. Daniel Gatica-Perez. (See Publications).

Nov 2014 – Dec 2015 — **Research Trainee**, *Data Visualization*, Idiap Research Institute, Switzerland.
Visualizing of communication and mobility patterns from big phone data (D4D Challenge) in the framework of the project "Social Media Africa Mirro" under the supervision of Prof. Daniel Gatica-Perez. (See Talk, Media News, and Visualization).

May 2014 – Sep 2014 — **Senior Developer**, *Team Lead*, FPT Software Ho Chi Minh, Ho Chi Minh City, Vietnam.
- Outsourcing with DirectTV Company, a big television company in US.
- Team lead in coordinating discussion of database evaluation and researching on SQL/non-SQL with US team.

Mar 2013 – Sep 2013 — **Master Thesis Intern**, *Data Science*, School of Computer Science, Aalto University, Espoo, Finland.
- Working on 6 months master thesis: "Diffusion rate on Twitter based on semantic relatedness and homophily of users".
- Using text mining to calculate semantic relatedness of textual contents of users and predict the information diffusion through users' topography and discover factors effecting on homophily and anti-homophily.

Oct 2012 – Feb 2013 — **Research Intern**, *Javascript Developer*, Eurecom, Biot, France.
- Working on 5 months semester project: "Grab your favourite TV show and share it" with Assistant Prof. Raphaël Troncy.
- Building up simulated application of synchronizing video playing between master and slaves by using user gestures on Kinect device (Kinect tool, Kinesis) combined with web browsers (Websocket, WebRTC, BinaryJS, NodeJS, JavaScript, and HTML5).

Oct 2012 – Feb 2013 **Developer Intern**, *Android Developer*, GridPocket, Sophia Antipolis, France.
- Working on 5 months intern: "Energy Diet Application".
- Member of master team used Android programming and database design on SQLite to develop Energy Diet application, resulting at saving energy expense for users.

Sep 2010 – Jul 2011 **R&D Employee**, *Data Science*, VinaGame Ho Chi Minh, Ho Chi Minh City, Vietnam.
Working on textual information retrieval of Vietnamese teenager language on ZingMe social network.

TEACHING ASSISTANT AND SUPERVISOR

2018–2020 **Assistant Supervisor**, EPFL, Lausanne–Switzerland.
- Master Student Saint-Supéry Santiago, "Learning food patterns from social media and crowdsourced data". (2019–2020)
- Master Student Ada Pozo Pérez, "Mining Behaviour in Everyday Life from Mobile and Social Data". (2019)
- Bachelor Student Antille Sarah Nicole. (2018)

2018–2019 **Teaching Assistant**, EPFL, Lausanne–Switzerland.
Computational Social Media (Spring 2018, 2019, 2020)

2010–2011 **Teaching Assistant**, University of Science, VNU-HCMC, Ho Chi Minh City–Vietnam.
Database (Spring 2010), Database Design (Fall 2010), Analysis and design of information system (Spring 2011).

LANGUAGES

Vietnamese	Native
English	Professional working
French	Conversational

KNOWLEDGE AREA

Keywords	Data Science, Quantitative Research, Human-centered Computing, Social Computing, Drinking Behaviour, Eating Behaviour, Alcohol consumption, Ambiance, Mobile Sensors, Motion Sensors, Video/Image/Audio Processing, Social Media, Mobile Crowdsensing, Mobile Crowdsourcing, Deep Learning, Natural Language, Causal Inference, Time Series, Statistical Modeling, Machine Learning
Machine Learning and Statistics	Deep Learning (Tensorflow, Pretrained CNN models), Predicting/classification models (Random Forest, SVM), Clustering, Statistics (e.g. T-test, chi-square, ANOVA, correlation, regression, etc.)
Programming	Python (NumPy, Matplotlib), R, Java, HTML, CSS, Javascript, C++, C#
Database	MySQL, NoSQL, Microsoft SQL Server, MySQL, Oracle, MongoDB
Frameworks	Hadoop, MapReduce, Mahout
IDE	PyCharm, Eclipse, NetBean, Visual Studio
OS & Office	Windows, MacOS, Linux, MS Office
Outros	SVN, Git, Latex, UML

FELLOWSHIPS, HONORS, AWARDS

May 2019 **Honorable Mention Paper Award**, PervasiveHealth2019, Trento–Italy.
One of two prizes, awarded at PervasiveHealth'19 (see Publications)

2015–2019 **Swiss Government Excellence Government Scholarship**, Switzerland.
The most honor Swiss-government full scholarship for postgraduate scholars or researchers at one of the public funded university or recognized institution, granted by Swiss Federal Commission.

2011–2013 **Eiffel Scholarship Program of Excellence**, France.
The most honor French-government full scholarship for outstanding international students to study master of engineering in France, granted by French Ministry of Foreign Affairs.

2010 **Top 1 graduate of Honors Degree**, VNU-HCMC, Vietnam.
Certificate of the Merit on being the best student of graduation 2010 (1/500 students).

2005–2007 **Scholarship**, VNU-HCMC, Ho Chi Minh City, Vietnam.
Scholarship for excellent students in each semester at VNU-HCMC, Spring–Fall 2008, Spring–Fall 2007, Fall 2006, Spring 2005.

TALK, MEDIA NEWS, AND VISUALIZATION

2017 **Media4Sec**, *Talk*, Barcelona, Spain, Cities through the lens of social media: basic insights.

2017 **Food and Drink**, *Media News*, Switzerland, Schweiz: gesundes Essen, ungesunde Postings.

2016 **Swiss Cities**, *Visualization (Youtube)*, Switzerland, Local/Non local.

2015 **D4D Project**, *Visualization (Youtube)*, Senegal, Big Mobile Phone Data in Senegal.

INTERESTS

Collector Collecting mugs in each city I have visited all over the world, a collection of 35+ mugs.

Volunteer Former vice dean of Thien Duyen Club, a non-profit English school for teaching poor children, coordinate the volunteer teachers and build curriculum for charity English class.

Travelling Visiting 3/5 continents with 20+ countries (America, Asia, Europe).

Cooking Cooking multiple Asian dishes because food can deeply talk thousands of words.

Photographer Taking nature photos with my heart feeling.

Updated: April 1, 2020