7-2019

Exploiting mobility for predictive urban analytics & operations

Kasthuri JAYARAJAH
Singapore Management University, kasthurij.2014@phdis.smu.edu.sg

EXPLOITING MOBILITY FOR PREDICTIVE URBAN ANALYTICS & OPERATIONS

KASTHURI JAYARAJAH

SINGAPORE MANAGEMENT UNIVERSITY
2019

Exploiting Mobility for Predictive Urban Analytics & Operations

by

Kasthuri Jayarajah

Submitted to School of Information Systems in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Information Systems

Dissertation Committee:

Archan MISRA (Supervisor / Chair)
Professor of Information Systems
Singapore Management University

Rajesh Krishna BALAN
Associate Professor of Information Systems
Singapore Management University

LIM Ee-Peng
Professor of Information Systems
Singapore Management University

Shili XIANG
Research Scientist
Agency for Science, Technology and Research

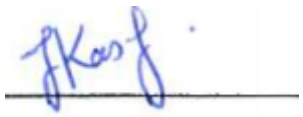
Ramayya KRISHNAN
Dean and W.W. Cooper and Ruth F. Cooper Professor of Management Science and Information Systems
Carnegie Mellon University

Anind DEY
Dean and Professor, iSchool,
University of Washington

Singapore Management University
2019

I hereby declare that this PhD dissertation is my original work
and it has been written by me in its entirety.
I have duly acknowledged all the sources of information
which have been used in this dissertation.

This PhD dissertation has also not been submitted for any degree
in any university previously.



Kasthuri Jayarajah
23 July 2019

Exploiting Mobility for Predictive Urban Analytics & Operations

Kasthuri Jayarajah

Abstract

As cities worldwide invest heavily in smart city infrastructure, it invites opportunities for a next wave of urban analytics. Unlike its predecessors, urban analytics applications and services can now be real-time and proactive – they can (a) leverage situational data from large deployments of connected sensors, (b) capture attributes of a variety of entities that make up the urban fabric (e.g., people and their social relationships, transport nodes, utilities, etc.), and (c) use predictive insights to both proactively optimize urban operations (e.g., HVAC systems in smart buildings, buses in the transportation network, crowd-workers, etc.) and promote smarter policy decisions (e.g., land use decisions pertaining to the positioning of retail establishments, incentives and rebates for businesses).

Individual and collective mobility has been long-touted as a key enabler of urban planning studies. With everyday artefacts that a city's population interacts with being increasingly embedded with hardware (e.g., contact-less smart fare cards that people tap-in and out of buses and metro), and due to the sheer uptake of location-based social media platforms in recent years, a wealth of mobility information is made available for both online and offline processing. This thesis makes two principal contributions – it explores how such abundantly available mobility information can be (a) integrated with other urban data to provide aggregated insights into demand for urban resources, and (b) used to understand relationships among people and predict their movement behavior (including deviations from normal patterns). Additionally, this thesis introduces opportunities and offers preliminary evidence of how mobility information can be used to support a more efficient urban sensing infrastructure.

First, the thesis explores how mobility can be combined with other urban data for better policy decisions and resource utilization prediction. It first investigates how aggregate mobility data from heterogeneous sources such as public transportation and social media, can aid in quantifying urban constructs (e.g., customer visitation patterns, mobility dynamics of neighborhoods) and then demonstrate their use, as an example, in predicting the survival chances of individual retailers, a key performance measure of land use decisions of a city.

In the past, studies have relied on the predictability of mobility to generate various urban insights. In a complementary effort, by demonstrating the ability to predict instances of *unpredictability*, sufficiently in advance, this thesis explores opportunities to proactively optimize urban operations by harnessing such unpredictability. First it looks at individual mobility at campus-scale, to discover and quantify social ties. It then describes a framework to detect episodes of future anomalous mobility using social tie-aware mobility information, and then use such early warnings to demonstrate its use in an exemplar smart campus application; task assignments of workers of a mobility-aware crowd-sourcing platform.

In a final exposition of emerging possibilities of using mobility for real-time, operational optimization, I introduce a paradigm for collaboration between co-located sensors in dense deployments that exploits human mobility, at short spatio-temporal scales. As preliminary work, this thesis investigates how associations between densely co-located cameras with partially overlapping views can reinforce inferences for better accuracy, and offers evidence of the feasibility to run adaptive, light-weight operations of deep learning networks that drastically cut down on processing latencies.

This thesis provides additional examples of real–time, in-situ, mobility-driven urban applications, and concludes with key future directions.

Table of Contents

List of Figures

List of Tables

Acknowledgments

A Ph.D. is a long and arduous journey, and I would have never completed this without the help and generosity of many.

First and foremost, I want to thank my advisor, Prof. Archan Misra, for his guidance, kindness and support all through these years. I consider myself extremely fortunate to have "stumbled" upon your profile while job hunting, 7 years ago. I cannot thank you enough for all the good times we've had both at work and outside. Your passion for work, attention to detail, the constant will to improve, and your compassion towards everyone, has certainly been inspirational. I still remember my first, albeit negative, performance review with you on-the-job – it was upsetting at first, but it was the perfect wake-up call. I constantly remind myself of that day to never fall into the trap of complacency. I'm grateful to you for accommodating all my requests – to continued to have faith in me even when I went mute without any updates for weeks. Whomever I've become and will become professionally, I owe it to you. I'm truly thankful to be able to fondly refer to you as my "Academic Papa".

I wish to express my sincere gratitude to my stellar dissertation committee. My thanks to Rajesh and Prof. Lim for their guidance and support since my early days at SMU as an engineer. Rajesh, I've learned a great deal about systems research from you all along[1]. I thank Shili Xiang, Anind Dey, and Ramayya Krishnan for it has been both a constructive as well as a humbling experience for me to have gone through multiple iterations of the dissertation with you. I'd like to extend my thanks to Youngki Lee, Cecilia Mascolo, Tarek Abdelzaher, Jennifer Mankoff, Reetika

[1] and, I don't believe your giggle-deciphering accuracy is that high!

Gupta, Shonali Krishnaswamy, Liu Guimei, Geeth de Mel and Anastasios Noulas for their guidance at various phases of my Ph.D. journey. I'm thankful to Azeem and Riju for their advice and critical feedback during my initial days in research.

Wanting to pursue the Ph.D. programme in Singapore was a long and hard decision. I'm every bit grateful to this great nation to have offered the opportunities to fulfil my academic goals as well as to raise my family here. I'm indebted to the generous support extended by A*STAR, LARC and LiveLabs. I'm thankful to Baihua, Pei Huan, Caroline, Sipei, Kazae, Phoebe and Lydia for all their support.

I'm happy to have shared my journey at SMU with amazing friends from the lab. Meera, Sougata, Jeena, Swetha, Amit, Kartik, Vignesh, Madhumita, Thivya, Vu, Sinh, Loc, Huy, Camellia, Ratan, Dulanga, Andrew, Lakmal, and Zaman, I'll forever cherish our memories of caffeine addiction, over-nighters and gaming evenings[2]. Meera, special thanks to you for always being that one person I could talk to, and for constantly keeping track of my plant and how well it "covered" me.

Finally, I sincerely thank my family for their love and support. I'm grateful to Amma and Appa for their unconditional love. You both have shown me and Anna what it means to be truly resilient people – starting with nothing, battling through tough situations, time and time again, and remaining ever so industrious to make ends meet (especially with the "oomai pappadam"). I'm extremely thankful to my in-laws for having shown immense kindness and patience towards me. I'm thankful to Dubi, Waffles Long-Ears and Bobby for their enduring support and for cheering me on, especially during "paper nights". Anna, I'm so grateful for your courage to have left home at just sixteen and giving our entire family a chance to "step-up", and for standing by me when I needed my family the most. I'm so grateful for the two adults we have grown to become. And, Sugirthan, my husband, my best friend. I'm ever so grateful for your love, patience with all my shortcomings and sensitivities and for never saying NO. This journey would not have been possible without your constant encouragement and sacrifices.

[2]P.S. I AM the Catan champion

To

Sugirthan, for true partnership and patience

And

Appa, Amma, Anna, for your unconditional love and support

Chapter 1

Introduction

In recent years, we have seen a rapid growth in the use of digital data and technologies for better urban planning and improved use of urban infrastructure. However, current uses of urban analytics principally rely on (a) offline assimilation of data, albeit at increasingly finer spatio-temporal resolution, larger scale and over longer time periods and (b) post-hoc studies and analyses to generate insights and action plans in a reactive manner. In this thesis, I describe a set of research innovations for next-generation, *predictive* urban analytics services made possible by (1) the large scale deployments of IoT devices by a combination of government initiatives, commercial businesses and individual citizens, and (2) technologies that enable easy generation, storage, analysis and distribution/communication of such data.

I envision the next wave of predictive urban analytics to embody the following attributes at the three stages of an analytics system (see Figure 1.1):

- **Low-latency Sensing:** High coverage, real-time, situational sensing capabilities from thousands of, often resource-constrained, devices.
- **Modeling and Analysis:** Ability to consume information from multi-modal (e.g.,video streams, unstructured text from social media posts) , multi-scale (i.e., incoming data could be coming at different spatial and temporal resolutions) data sources to model and generate actionable insights.

- **Proactive Optimization and Policymaking:** Ability to proactively optimize urban operations (e.g., disseminating surplus buses to relieve congestion during a highly disruptive event) and enable smarter policy decisions (e.g., uncovering and subsequently exploiting, for land use planning, the factors that affect the differential success rates of key retail trades).

In this thesis, I focus on a subset of smart urban analytics – specific to *urban resources* [9, 95, 89] that include but not limited to: information resources (e.g., sensors), transportation resources (buses, trains, roads, rail lines, traffic management infrastructure, etc.), land resources (e.g., retail spaces, car parks, residential zones, etc.) and human capital (e.g., crowd-workers). Given this focus, *smart urban resource analytics* pertains to both:

- predictive planning and policy support for urban resources, and the appropriate allocation of such resources, and,

- operational optimization of such resources leading to either energy savings, or improved accuracy (due to reduced operation time, or faster execution, or intelligent, preemptive ON/OFF scheduling), and/or improved utilization for the same incurred cost.

1.1 Motivating Scenarios

In this section, I take the example of three scenarios to motivate possibilities for proactive urban operations and smarter policy support for urban planning.

Scenario 1: Land Use Policy Planning Esther is a planner at the local urban planning authority of an Asia city. Esther and her team of planners have been assigned the task of revising the master plan of the city which dictates how the city's land will be used for the next 15 years. Esther, as a trained geographer, understands well the implications of land use choices and their long term consequences. The city

she works for, is a highly dense metropolitan where land is an extremely scarce resource. Locking down a parcel of land, for example, for use as an up-scale retail district whereas the catchment to that neighborhood are typically from low-to-medium income earners, would risk the choice of land use to be futile. Traditionally, Esther and her fellow planners exercise due diligence as to which neighborhoods would be in demand for which type of land use through limited number of surveys and observational studies. However, since the opening of the digitization wing of the authority, Esther believes that she can utilize publicly available transportation and social media data to predict which retail trades will be attractive to, or successful in, which neighborhoods, to guide her in allocating vacant land parcels for retail use and in renewing licenses to operate for existing parcels.

To realize this scenario, a retail planning assistance tool should be able to perform the following:

(i) identify data sources for quantifying urban constructs that can help in understanding the success of individual businesses,

(ii) identify static and dynamic attributes of individual businesses, the neighborhood they operate in, and the catchment of such neighborhoods, that can predict the survival likelihood, or chance of success of such businesses in the future, and,

(iii) provide predictive insights into the optimal mix of retail trades that should be allocated to greenfield or redeveloping land parcels within the city.

This thesis accomplishes items (i) and (ii) by studying the survival chances of individual venues in major cities around the world through the combined use of social media and urban mobility data, in Chapter 2. Aggregate insights on the suitability of retail trades to specific neighborhoods (e.g., item (iii)) is an example of what the findings of this thesis can aid with in future.

Scenario 2: Proactive HVAC Control for Smart Buildings – Carrier, Joule and Faraday are freshman undergraduate students of a smart campus and are part of

the same project group for the CS101: Programming Fundamentals course. Joule and Faraday have been friends since their childhood and have stuck together ever since. It is not unusual for Joule and Faraday to spend all their time outside classes together; they frequently have lunch together with a couple of their other friends, and often take part in co-curricular activities together, although they don't take the same courses. As required by the course, the project group meets every Wednesday afternoon for an hour of catch up on progress and planning to-dos for the subsequent week. Although the group has booked a meeting room for the entire term, the group some times decides to cancel the meeting when all three cannot make it at the scheduled time (for e.g., due to schedule disruptions), and re-group at a different time at a public space. The facility management department (FM) of the university observes that such booking defaults happen at a non-negligible level across all buildings on campus and decides that a more proactive control mechanism for the campus' cooling needs would be appropriate. After much discussions with the IT department of the university, the FM decides to use the university's live, indoor mobility monitoring system, which supports room-level localization of individuals who connect to the university's WiFi network, to improve the energy efficiency of its current HVAC system.

To realize this scenario, a smart HVAC control system should be able to perform the following:

(i) construct personalized models of mobility for residents of the campus based on observations of their historic mobility indoors,

(ii) unobtrusively learn residents' social ties based on mobility traces,

(iii) extract features of mobility that can predict deviant behavior from routine mobility, taking into account both a resident's and their close ties' combined mobility, with sufficient look-ahead time,

(iv) predict regular meetings that are likely to not take place at their usual time

and place, based on the prediction from the previous step, where it is inferred that all or most attendees will default,

(v) learn the relationship between energy savings and user comfort, accounting for environment factors as well as nominal setback temperature settings, and,

(vi) adjust temperature, proactively, based on the estimated likelihood of a scheduled meeting happening.

This thesis offers techniques and empirical insights related to items (i), (ii) and (iii) at both indoor and city-scale, in Chapters 3 and 4. Additionally, it provides early insights into how such techniques can improve operational efficiency in an alternate application scenario, i.e., in the task allocation of crowd-workers on a smart campus.

Scenario 3: Efficient Video Surveillance – Kwee Boon is a supervising engineer at the local authority in charge of the government housing blocks of the city. To ensure the safety of the residents of the blocks, the authority has installed multiple surveillance cameras in the public areas such as open corridors and lift landings. Since the installation, the authority has been successful in using the video feeds from the cameras to support the investigations of the local police – for example, an elderly gentleman was injured as a bystander attacked him to snatch his wallet as soon as he exited from the lift. The police were able to manually sift through multiple hours of raw video footage to identify and track the perpetrator as he moved from the lift landing where he attacked the elderly person to the nearby cafe that he frequents, several days later. Although it was a success, Kwee Boon is concerned that with the current mechanism which only supports investigations, post-hoc. He consults with his authority's technology department who reported back that it would be possible to provide a more dynamic system which will allow for near real-time, analytics of the video feeds from the multiple cameras, but also warns Kwee Boon of the potential pitfalls – that the energy and network communication costs involved with uploading raw footage from the hundreds of cameras to a powerful server that

can run the analytics could be too high which the authority may not be able to justify. As an alternative, the consultant suggested that if the analytics could be run on the camera hardware itself, it would save network costs, and if the cameras could communicate among themselves to avoid redundant processing, that it could save energy.

To realize this scenario, an intelligent camera network should be able to perform the following:

(i) run compressed models of state-of-the-art, often deep learning based, low-level video analytics (e.g., person detection) on each individual camera hardware,

(ii) communicate with neighbors in the network to identify possible collaborators, pass digests of their individual inferences intelligently, such that

(iii) the energy and communication costs are minimized, whilst

(iv) optimizing operation to maintain or improve the accuracy of their combined inferences.

This thesis offers preliminary insights into mechanisms for compressing DNNs, and improving the inference accuracy and processing latency for a group of networked cameras in Chapter 5. The process of discovering collaborators and the quantification of energy/communication cost savings are addressed as part of ongoing work.

1.2 Exploiting Mobility for Predictive Urban Analytics

This thesis draws on findings from previous studies in urban mobility and identifies the following properties of mobility in enabling the predictive urban analytics paradigm described.

1.2.1 Incorporating disparate sources of mobility

Urban digitization has led to an explosion in the volume and diversity of available data about how individuals daily interact with a city's infrastructure–i.e., a window into the "Pulse of a City". For instance, the GPS aboard the everyday smartphone and contactless smart fare cards together provide precise information on an individuals outdoor mobility. Similarly, appropriately instrumented WiFi systems in indoor buildings can locate its dwellers with room-level accuracy. Additionally, the population themselves voluntarily reveal information on their whereabouts when they post social media content that is geo-tagged via platforms such as Facebook and Foursquare. In addition to providing information on location, social media data, due to its richness in semantics, can provide additional data pertaining to its users such as the users' tastes, likes, and friends. Separately, as businesses and retailers take to such online platforms, aggregate mobility information such as patron visit patterns to individual businesses are also revealed.

In Sections 2, 3 and 4, I explore the use of various such mobility sources pertaining to (1) multiple spatio-temporal scales (e.g., check-in events at city-scale vs. continuous campus scale mobility monitoring;) and (2) different aggregation levels ("individualized" vs. "aggregated"). Mobility of each individual in a building ascertained from WiFi-based indoor positioning systems is an example of "individual mobility". On the contrary, crowd flows between neighborhoods extracted from taxi trips is a form of aggregated mobility.

1.2.2 Exploiting regularity in mobility

As demonstrated in several past works, population mobility is *highly* regular (see Chapter 6) and has supported the study of several urban phenomena, especially at the city scale. In Section 4, I explore the extent of regularity indoors, and exploit the predictability of non-conformance to such regularity, for proactive optimization of building resources such as the productivity of crowd-workers who voluntarily

perform a variety of resource monitoring tasks.

1.2.3 Exploiting similarity in sensed signals across co-located sensors

As I explore in Section 5, co-located sensors that are part of the same deployed network, inadvertently, share much similarity in the environment they sense – for example, two cameras pointing at the same scene from slightly different perspectives and angles, will share partially overlapping views. Entities moving within such views would hold space-time relationships that could be modelled/captured, thereby opening opportunities to (1) reinforce inferences across a set of sensors for better accuracy, (2) save on computations and latency by selectively triggering inferencing within and across a set of sensors and (3) detect failure in sensors by detecting deviations from such established relationships. Note that this part of the work is preliminary in the sense that it does not directly utilize mobility, but develops precursor collaborative sense-making techniques that mobility-aware IoT systems of the future can leverage.

1.3 Key Challenges

Influence of Social Ties on Personal Mobility: In this thesis, I focus on ways to exploit mobility for enabling many urban application scenarios. As I show in Section 4, mobility of individuals, at both city-scale and indoors, is influenced by the mobility or routines of their social ties. Hence, it becomes important to account for social interactions (with peers and friends) in mobility-based applications. In settings where friendship information is not explicit (contrary to social media where this information is typically explicit through 2-way friendship or 1-way following relationships), unobtrusive inference of social ties is a challenge. In Section 3, I show and validate how ties can be inferred from mobility traces.

Developing Meaningful Validation Measures: Whilst mobility information from disparate sources allow for the quantification and study of several urban constructs, the validation of the exactness of such quantification remains a challenge. For example, in Section 2, I consider the increase or decline in the volume of check-ins to individual retail establishments to approximate the performance of the business. Check-ins are events that are captured as and when a user of a location-based social media platform uses the feature for marking that he/she is at an establishment, and thereby, allow for quantification of many performance indicators (e.g., week-over-week uptake, hourly popularity, etc.). However, the official records of business performance are often sparse (e.g., annual tax filings), or private (e.g., Point-of-Sales data that is available only within businesses), both of which hinder the validation of how accurately the check-in based measures represent the construct. Hence, alternative, often crude approximations need to be identified for *validation*.

Bandwidth and latency requirements: Dense, large-scale deployments of sensors place a high order on the bandwidth required to transfer sensed data (e.g., video feeds from cameras on highways) to a centralized server in the traditional architectures deterring real-time application of analytics systems.

A solution then is to equip sensors at the edge with on-board processing capabilities such that not all sensed media has to be communicated back to a central server. However, unlike their centralized counterparts, the sensors at the edge have limited processing capabilities and energy capacities. To this end, I explore the concept of collaboration at the edge, between such low-end sensing devices, exploiting spatial correlation (as a building block of mobility-leveraging techniques), to support complex inference pipelines and real-time latencies, without compromise on accuracy.

Interaction effects between urban resources and their consumption: While traditional urban studies have focused on capturing the impact of various environment variables on the consumption or use of single resources, I argue that the urban resources such as those described here showcase interdependencies. For instance,

in Section 2, I show evidence of how visitation patterns of consumers and mobility dynamics of an area can dictate the retail performance of businesses – however, another key dimension is to understand how such retail use of urban spaces in turn affects movement patterns in that area.

This thesis aims to address the first three challenges outlined here.

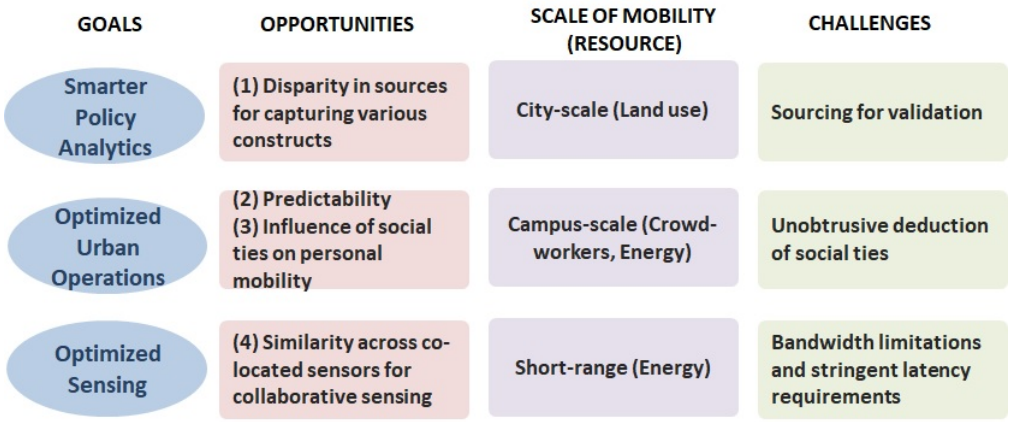


Figure 1.1: Summary of the Key Components of this Thesis.

1.4 Thesis Statement

Previous sections highlight the opportunities that arise from the availability of such mobility information and some of the key challenges involved in enabling the example scenarios put forward. In this dissertation:

I demonstrate novel applications of machine learning-based prediction techniques, to utilize individual & aggregated mobility behavior (as well as mobility-driven correlations in spatial occupancy), at different spatiotemporal scales, to (a) drive predictive urban policy insights; and (b) improve the operational efficiency of smart-city services at city, campus and building level.

This dissertation establishes the thesis through the following steps:

1. First, it identifies attributes of mobility, at different spatio-temporal scales,

city-scale, campus-scale (indoor) and build-scale (short range), that can be exploited to support proactive urban operations, and policy support. It identifies examples of proactive analytics-based applications that operate across 3 distinct spatial scales.

2. Then, it exploits the wide access to disparate mobility sources that represent both individualized and aggregate mobility; it quantifies various urban constructs using mobility datasets to demonstrate how predictive insights can be informative to urban authorities through the specific example of estimating the future survival likelihood of individual retail businesses.

3. Next, it reiterates the fact that human mobility, regardless of its scale, is influenced by the mobility of one's social ties, and provides a methodology to infer the social network of individuals, passively, using their mobility traces observed longitudinally. It then provides validation on how such knowledge of personal and collective mobility can aid in attaining situation awareness through the example of detecting transient events in both within campuses as well as at the city-scale.

4. It then exploits the nature of regularity in mobility and the influence of social ties on mobility to demonstrate the possibility to optimize urban operations, proactively. It first describes a framework that predicts, sufficiently in advance, instances where students in the campus deviate from their routine mobility behavior, and use such predictions in an exemplar campus-scale application: optimizing task assignments to crowd-workers in the university to proactively improve worker productivity.

5. Finally, it provides preliminary evidence that mobility-driven correlations between multiple sensors, deployed across a campus-scale space, can be used to optimize the sense making pipeline on individual IoT devices. In particular, it develops new methods of collaborative inferencing that enable resource-

intensive deep learning pipelines to be made more robust and resource-efficient for execution on pervasive devices.

Chapter 2

Incorporating Disparate Mobility Sources for Urban Analytics

The availability of digital data about a city's infrastructure and businesses (e.g., records of transportation transactions, purchase records at stores, etc.) has led to an explosion of work on urban analytics. In particular, such data often tells us about the economic vibrancy of neighborhoods, identifies civic issues of concerns (e.g., crime, rodent infestation) and reveals the lifestyle habits of its residents (e.g., commuting patterns). One such issue receiving significant attention, both from an economic and policy standpoint, relates to the causes of failure of neighborhood businesses, and the possibility of applying predictive analytics to anticipate such failures.

Broadly, an establishment's failure susceptibility can be ascribed to a variety of *controllable* and *uncontrollable* factors. Controllable factors could include the quality or price of the store's product offerings, its operating hours, and its customer satisfaction. Conversely, uncontrollable factors could include unemployment rates of the city, overall economic conditions, and urban policies. Establishing what constitutes *failure* is a challenge in itself and has had a critical role in limiting the number and extent of existing studies on business survival [99, 126]. Prior works have utilised financial records where they consider *bankruptcy* as failure. However,

this approach is limiting as it does not capture cases where a proprietor decides to shut down an establishment, nor is it capable of predicting cases of likely failure (where intervention mechanisms may have a salutary effect). Despite such efforts, the inherent low frequency of financial reporting lends itself to (1) studies that focus on static macro factors leading to failure and (2) a failure in recognising establishments that are at high risk of mortality in the near future.

The recent proliferation of urban datasets, especially related to urban mobility and social media activity, offers interesting opportunities for high-fidelity sampling of these controllable factors. For example, mobility data can reveal the *urban dynamics* of different locations (e.g., does a neighbourhood attract visitors from various other neighbourhoods?), whereas location-based social network (LBSN) data can elucidate *consumer interactions* at the individual venue level (e.g., how popular is the venue relative to others in its vicinity?). Content in such LBSN data (e.g., text reviews of establishments) can also provide insight into possible causes or determinants of failure. There has only been a modest amount of research on the use of social media for business analytics–e.g., Wang et al. [126] utilised LBSN data to predict the failure of food establishments using a set of over 600 restaurants in New York City (NYC) over a 6 month period, and Karamshuk et al. [72] provided empirical strategies for using LBSN-based features to find optimal locations for new stores.

In this work, we utilise two complementary, large-scale longitudinal datasets: (1) venue *check-ins* on Foursquare, observed in ten cities across the globe (75 million check-ins from Jan 2011 through Dec 2013), and (2) *taxi trip records*, observed across Singapore (38 million trips between Nov 2011 to Jan 2012) and New York City (143 million trips between Jan 2013 to Dec 2013), to develop a predictive model for retail business failures. We examine the role of a number of features on retail business survival, across both a broader swathe of retail categories, and specifically for food & beverage (F&B) establishments. As F&B is known to be a highly competitive and risky business in many cities, we examine this category

more closely for universal trends. We employ three classes of features: (a) *Static Locality Profiles*, capturing the properties of the locality in which an outlet operates; (b) *Visit Patterns*, reflected in the volume and spatiotemporal patterns of Foursquare check-ins; and (c) *Neighbourhood Mobility Dynamics*, reflected in visitation patterns across distinct neighbourhoods. Our specific prediction question is: *given observable features at a point in time, how likely is it that a retail establishment will close down within the next 6 months?*[1]

2.1 Key Questions and Overall Methodology

The motivation for our work is to build a predictive model for venue closure. To this end, we first identify a set of candidate features (section 2.3), pose the problem of predicting closure as a binary classification task, and report our findings in section 5.7. In this section, we formalise the key questions we answer in this work, introduce notations used throughout the work, and define what constitutes *closure*.

Using a combination of LBSN and transport data, we seek answers for the following logically-sequential questions:

1. Can a sustained and significant decline in the popularity of a venue serve as a proxy for its closure? (see Section 2.2.2)

2. Can metrics of the locality profile, visitation patterns, and mobility dynamics of a retail business be used as predictors its success or failure? (see Sections 2.4.1, 2.4.2 and 2.4.3)

3. How accurately can we predict business failure? Does the accuracy vary by the type of business? (see Sections 2.4.2 and 2.4.4)

4. Do factors that attribute to business failure vary by city, or geographies? (see Section 2.4.5)

[1]The 6 month duration can, of course, be varied: for now, we choose 6 month as it appears to be a natural time constant for retail businesses deciding whether to close down or not, and also because determining an establishment's operating state at finer timescales from Foursquare data is very noisy.

2.1.1 Notation

We consider the set $\mathbf{V}$ of venues in a city. A venue $v_i \in \mathbf{V}$ is represented with a tuple $< loc, date, gen, spec >$ where loc is the geographic location of the venue, $date$ is its creation date, gen is its general category, and $spec$ is its specific category (see section 2.2.1). Further, we define a venue's *neighbourhood*, $\mathbf{N_i}$, as the set of venues that are located within a given radius; we set this distance to 500m as prior work [28] has shown that a venue's operation is affected primarily by conditions within this radial distance. Formally, we define the neighbourhood as:

$$\mathbf{N_i} = \{v_j \in \mathbf{V} : dist(v_i, v_j) < D\} \tag{2.1}$$

where $dist(v_i, v_j)$ represents the distance between v_i and v_j and $D = 500$. We also define a venue's *competitive neighbourhood*, $\mathbf{CN_i} \subseteq \mathbf{N_i}$, as the subset of venues that belong to the same general category, gen, as the venue. Similarly, we define the specific competitive neighbourhood, $\mathbf{SN_i} \subseteq \mathbf{CN_i}$, as the subset of venues that share the same specific category, $spec$, within the radius D. Further, we define `established` venues as those that have existed for longer than a year and `new` venues as those that have existed for less than one year.

The administrative zone (e.g., Census tract, subzone, ward, etc.) a venue belongs to is referred to as the venue's *locality*, throughout this work. Acronyms used throughout the text are listed in Table 2.1.

2.1.2 Defining Closure

For this work, we classify venues as either opened or closed. Prior research on Foursquare data has shown that venues added after June 2011 were highly likely (probability above 0.8) to actually be new venues opening rather than existing venues being added to the system for the first time [28]. To uncover venues that are at risk of closure, we look at check-in metrics between June 2011 - December

2013. For a given month, we define $C_t(v_i)$ as the total number of check-ins to venue v_i in that month. Similar to prior work (Wang et al. [126]), we consider a significant decline in check-in volume as a sign of impending failure. We define a venue, v_i, as closed when $RemainsOpen(v_i) = 0$. The formal definition is as follows:

$$RemainsOpen(v_i) = \begin{cases} 0, & \text{if } \frac{\sum_{t=0}^{T} C_t(v_i)}{T} < K \times mean(v_i) \wedge \frac{\sum_{t=0}^{T} C_t(v_i)}{T} < N \\ 1, & \text{otherwise} \end{cases} \tag{2.2}$$

where $mean(v_i)$ represents the mean number of check-ins for venue v_i prior to June 2013 (starting from its first presence in Foursquare), $T = 5$ represents the 6 month window, and $N = 6$, denoting an average of less than one check-in per month for the venue v_i. In other words, we define a venue to be closed if it has less than an average of one check-in per month, for 6 months, and if this average represents a significant decline in demand for this venue. We examine the validity of this definition of closure in section 2.2.2 where we show it is consistent with ground truth. We experimentally determined the value for K by varying this scaling factor incrementally from 0.15 to 0.30 which resulted in minimal variations in the percentage of closed labels. For London, the percentage of new venues that closed was 6.1% when $K = 0.15$ and 6.7% when $K = 0.30$. Across all ten cities, our closed labels marginally by an average of $< 7\%$. As such, we experimentally converged to $K = 0.25$ as this was the optimal value for our analysis.

2.1.3 Operationalising the Venue Survival Problem

As previously described, we focus on predicting whether a venue is likely to survive the next six months. We define the "prediction date" (PD) as the fixed date of July 1, 2013 across all venues. We refrain from deciding on a prediction date per venue (based on factors such as age or actual date of failure) as seen in prior work [10] due to practical limitations in acquiring such information (see section 3.1).

Using the first Foursquare check-in as a proxy for activity, we define a starting date (SD) for each venue. For each venue, the period $(SD, PD]$, the time from the venue's opening until the prediction date, is considered as the past data and the period $(PD, PD + 6months]$ (July'13 to Dec'13) is considered the *virtual* future data. This 6 month period is used in Equation 2.2 to label whether a venue has closed. We depict this in Figure 2.1. The features used in this work (described later in section 2.3) are all based on data pertaining to an observation window which is uniform across all venues regardless of their SD. In this work, we consider an observation window of 6 months which immediately precedes the prediction date, i.e., $(PD - 6months, PD]$ (Jan'13 to June'13), and also investigate the sensitivity of our results to shorter observation periods. All venues considered in this work were operational by the commencement of the observation window.

Table 2.1: Acronyms used throughout Chapter 2.

Acronym	Detail	Acronym	Detail
F&B	Food and Beverage	SD	Starting Date
LBSN	Location based Social Network	PD	Prediction Date
AUC	Area under the (ROC) curve	ReLU	Rectified Linear Units
ROC	Receiver Operating Characteristic	SELU	Scaled Exponential Linear Units
FS	Foursquare	CBD	Central Business District

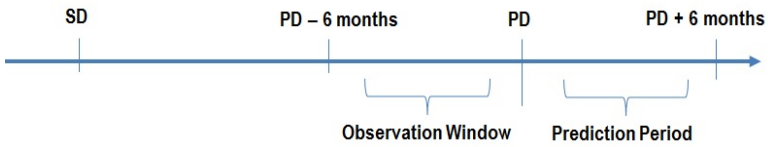


Figure 2.1: Definition of virtual past and future data used in this work. We use a fixed Prediction Date (PD) across all venues and answer the question, *which of these venues will close during the prediction period* $(PD, PD + 6]$*?* using features computed over $(PD - 6, PD]$.

2.2 Mobility Datasets

2.2.1 Dataset Description

We make use of two types of data for this work: one sourced from a location intelligence platform (Foursquare) on a multitude of cities across the world, and the other obtained from transportation authorities in two major cities.

Foursquare Data: Within the last decade, online Location-based Social Networks (LBSNs) have experienced a surge in popularity, attracting millions of users around the world. These LBSNs have created troves of data which describe, at a fine spatiotemporal granularity, the geographic position of users as they move in urban areas. Foursquare enables users to check in to different locations and share that information with their friend group. As of August 2015, Foursquare had more than 50 million active users and more than 10 billion check-ins [124].

In this work, we use a longitudinal dataset from multiple cities around the world, that spans three years (Jan 2011 through Dec 2013) and over 75 million check-ins. Table 2.2 includes the summary of statistics of the 10 cities we consider in this work. For each venue, we have the following information: geographic coordinates, specific and general category which fall within Foursquare's API of hierarchical categories [2], and the creation date. Additionally, the dataset also contains time-stamped check-ins captured in the form of *transitions*. A transition is defined as a pair of check-ins by an anonymous user to two different venues within the span of three hours and is identified by a start time, end time, source venue, and destination venue. In Figure 2.2 we provide a visualisation of the spatial distribution of venues in two cities, New York City and Singapore, that we label as closed and open according to Eq. 2.2 and the timeline described in section 2.1.3.

Transport Data: We rely on two transport datasets, one from New York City and the other from the city-state Singapore, that help us in understanding the movement dynamics and local catchment of *localities* within the city.

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

Table 2.2: Summary of city statistics. For each city, we report the total number of transitions, the number of established venues, the number of new venues, the percentage of established venues that closed, and the percentage of new venues that closed. Venues defined as *new* and *established* had been open for less or more than one year respectively (described in section 2.1.1). Venue closure was defined using Equation 2.2 (i.e. $RemainsOpen(v_i) = 0$).

City	Check-ins	Established Venues	New Venues	% Established, Closed	% New, Closed
Chicago	10,600,106	8,726	556	7.3	6.5
Helsinki	4,400,044	3,359	272	5.0	5.5
Jakarta	5,200,052	7,135	540	12.6	3.3
London	4,000,040	6,633	399	2.8	6.5
Los Angeles	3,300,033	5,652	263	6.1	2.7
New York	13,700,137	14,733	1048	8.5	7.4
Paris	3,600,036	4,653	189	5.1	6.3
San Francisco	4,100,041	5,407	336	5.4	6.0
Singapore	12,800,128	14,193	552	23.7	3.4
Tokyo	12,600,126	12,385	551	4.4	2.0

From New York City (limited to the Manhattan Borough), we obtain time-stamped records of dropoffs and pickups by yellow taxis for the period of January 2013 - December 2013 (which overlaps partially with the check-ins dataset), made available publicly by the New York City Taxi and Limousine Commission[3]. Each record contains the GPS coordinates of the pickup and dropoff points and the corresponding timestamps. We aggregate the pickup and dropoff points to Census tracts [4], where a tract typically houses at most 16,000 residents [5]. In the case of Singapore, we use data from a major taxicab company consisting of all trips occurring between November 2011 through January 2012 whose pickup and dropoff points we map to subzones [6] which are administrative boundaries. Table 2.3 summarises key statistics of these datasets.

[3]http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

[4]http://maps.nyc.gov/census/

[5]https://data.cityofnewyork.us/City-Government/2010-NYC-Population-by-Census-Tracts/si4q-zuzm

[6]https://data.gov.sg/dataset/master-plan-2014-subzone-boundary-web

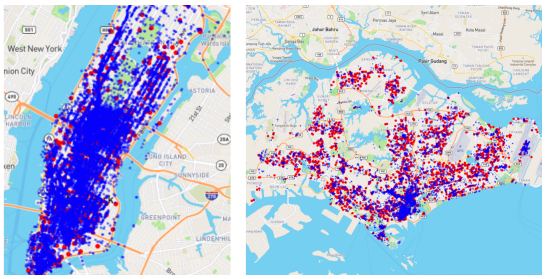


Figure 2.2: Spatial distribution of venues, labelled as "close" (in red) and "open" (in blue), according to Eq. 2.2 for the period of Jun'13 – Dec'13. The figure shows venues from two cities, New York City (left) and Singapore (right).

City	Spatial Aggregation	Total Trips	Observation Period
NYC	288 Census Tracts	143 million	Jan 2013 - Dec 2013
SG	323 subzones	38 million	Nov 2011 - Jan 2012

Table 2.3: Summary of taxi datasets used in the analysis.

2.2.2 Venue Closure

As mentioned previously, it is empirically hard to get the 'ground truth' of the closure of retail establishments across cities. In contrast to business openings, which are often advertised and announced on social media, venue closings often happen without fanfare. Moreover, F&B establishments sometimes exhibit "virtual closure"-a specific venue can simply re-brand itself (e.g., from a coffee shop to a restaurant/lounge), without actually changing owners.

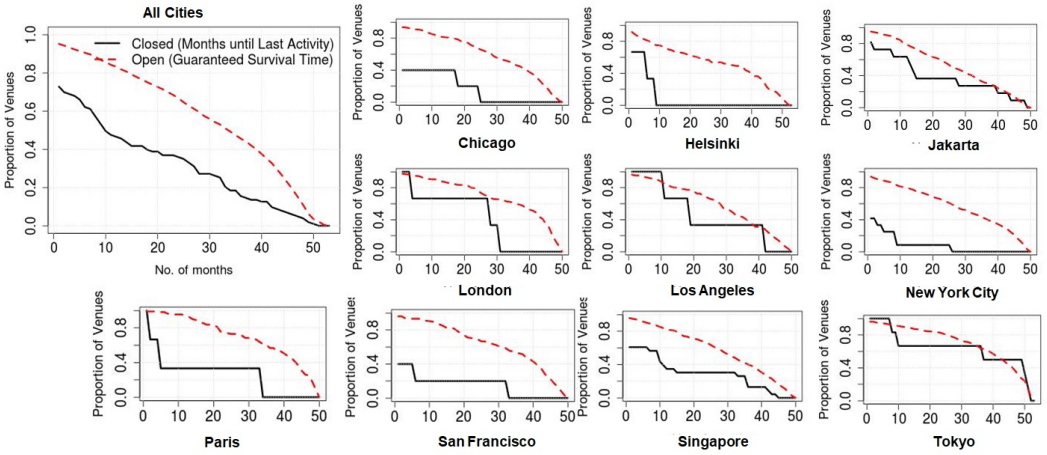


Figure 2.3: The survival curves (as KM plots) for all F&B venues considered in this work.

Given such phenomena and due to the fact that the check-ins dataset exists only until the end of 2013, we retrieve additional venue related data using the public Foursquare Venue API [1] to understand how credible the closure labels that we define are. This additional data retrieved consists of longitudinal observations of time-stamped, publicly shared activities at each of these F&B venues as at the query date (i.e., May 2018); these activities consisted of *tips*, public notes or short reviews users can share about a venue, and photo posts from visitors. This allows us to observe the activity at a venue beyond our prediction window and *proxy* the lifetime of the venues considered. We provide the survival plots of the 10 cities considered in this work in Figure 2.3 for F&B venues that started operation after June 2011. We use the Kaplan-Meier estimator [71] to estimate the survival function where the time to event is the time for which a venue remained active (based on tips and photos) beyond the prediction window (i.e, Dec'13). The $x-$axis represents the 53 month-timeline between Jan-14 till May'18 where the black line represents the "time till last activity or *closure*" for venues labelled as "closed" and the red dashed line represents the "guaranteed time alive or *open*" for venues labelled as "open". We see that less than $\approx$10% of restaurants (across NYC, Singapore and London), supposedly open during June-Dec'13, may be mislabelled, as they see no activity over the next 4 years. Conversely, approx. 50% (Singapore), 70% (NYC) and 65% (London) of restaurants, supposedly closed by Jan'14, cease all activity in the subsequent 2 years. On the other hand, approx. 60% of restaurants in Singapore that we labeled as "open" remained active beyond 2 years since the prediction period. This analysis lends credence to the reliability of our "failure" labelling process (i.e., the use of Equation 2.2), but also illustrates the challenge of perfect labelling. Later in section 2.4.5 we describe our efforts in curating a subset of data with higher quality labels and show that the performance of the models greatly improve with reduced noise.

Table 2.4: Summary of Features Investigated in this Work.

Feature Class	Feature	Definition	Source
Locality Profile	Competition	$\frac{CN_i}{N_i}$	Foursquare
	Specific Competition	$\frac{SN_i}{CN_i}$	Foursquare
	Place Entropy	$\sum_{i=1}^{k} p_i * \ln p_i / \ln k$	Foursquare
	Category Counts	$\lvert CN_i \rvert, \quad \lvert SN_i \rvert$	Foursquare
	Attractiveness to the Neighbourhood	$\lvert CN_i \rvert \times \ln\left(\frac{\lvert V \rvert}{\lvert V_C \rvert}\right)$	Foursquare
	Catchment of Locality	$\frac{\lvert D_l \rvert}{\lvert D \rvert}$	Transport
	Temporal Catchment of Locality	$\frac{\lvert D_{w,l} \rvert}{\lvert D_w \rvert}$	Transport
Customer Visit Patterns	Inflow & Outflow	$\frac{\sum_{j=0}^{\lvert V \rvert} t(v_j,vi)}{M}$, $\frac{\sum_{j=0}^{\lvert V \rvert} t(v_i,vj)}{M}$	Foursquare
	Distance Travelled to Reach Venue	$\frac{\sum_{j=0}^{N} dist(v_j,v_i)}{N}$	Foursquare
	Speed of Travel to Venue	$\frac{\sum_{j=0}^{N} dist(v_j,v_i) \times t_{i,j}^{-1}}{N}$	Foursquare
	Temporal Popularity Skew	$\sum_{i=1}^{24} h_i * \ln h_i / \ln 24$	Foursquare
	Visit Trend	$\frac{c_t(v_i)-b}{t}$	Foursquare
	Temporal Alignment with Competitors	$\sum_{j=1}^{24} (h_i(j) - H_i(j))^2$	Foursquare
Mobility Dynamics	Temporal Alignment with Locality	$\sum_{j=1}^{24} (h_i(j) - h_l(j))^2$	Both
	Reachability	$r_{(a,b)}$	Both
	Distance-weighted Reachability	$dr_{(a,b)}$	Both
Business Attributes	Cuisine Type	Categoric variables	Foursquare
	Price Tier		Foursquare

2.3 Feature Description

We next describe three classes of features which may play a role in the success of a business.

2.3.1 Profile of the Locality

Prior studies on restaurant failure in [99, 98, 96, 97, 126], indicate that an F&B venue's *locality* plays an important role in determining its success. We capture such intrinsic, largely *static*, properties of the locality (i.e, Census tracts in New York and Subzones in Singapore) using the following set of features.

We define the **Competition** of a venue as the proportion of competitors (in the case of F&B, this refers to all 'food' establishments enumerated in Foursquare) to the size of the neighbourhood, $\frac{|CN_i|}{|N_i|}$. Similarly, the **Specific Competition** is the proportion of neighbouring venues that serve the same cuisine in the competitive neighborhood–i.e., $\frac{|SN_i|}{|CN_i|}$. We also consider the counts **General Category Count**

($|CN_i|$) and **Specific Category Count** ($|SN_i|$) as separate features as they act as proxies for the overall *size* of the neighbourhood.

We next define **Place Entropy** of the area around a venue through the Shannon equitability index [114] from information theory. This metric is calculated as follows:

$$-\sum_{c=1}^{k} p_c * \ln p_c / \ln k \tag{2.3}$$

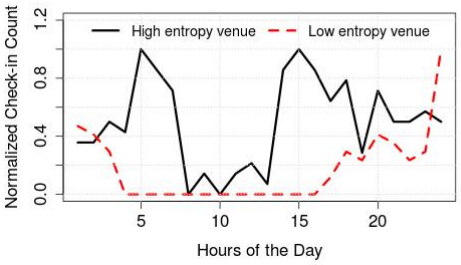
where p_c denotes the proportion of venues of category c and k is the total number of different categories in N_i.

Attractiveness of the neighbourhood: We measure this feature at both the general and specific category levels. We borrow the use of the $tf - idf$ weighting scheme from text mining literature, adopting the notion of neighbourhoods as *documents* and the venue categories as the *terms* that occur in them. The term frequency tf is simply the count feature defined above, and the Inverse Document Frequency, idf is computed as follows: $\ln\left(\frac{|V|}{|V_c|}\right)$ where $V_c \subset V$ is the set containing all venues belonging to that same category. Then, the attractiveness score of a venue to its neighborhood, N_i, is given by, $tf \times idf$.

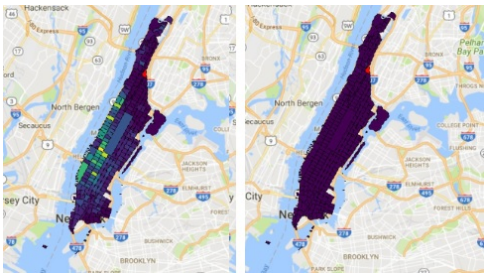
Catchment of the Locality: To capture the overall attractiveness of a locality, l, where a venue v_i is situated, we define its catchment using the taxi datasets as: ($\frac{|D_l|}{|D|}$), where D is the total number of taxi drop-offs across the city and D_l is the number of taxi trips that ended in a location within l. We also subdivide the previous feature and compute the **Temporal Catchment of the Locality** over four disjoint time partitions: morning (6 AM to 12 noon), afternoon (12 noon to 6 PM), evening (6 PM to 12 AM) and early morning (12 AM to 6 AM), separately across weekdays and weekends. During a window, w, the catchment for that window is then defined as, $\frac{|D_{w,l}|}{|D_w|}$.

2.3.2 Visit Patterns

The variation in trends of customer visits to such locations can reveal important insights into how businesses are faring. We define a number of visitation-driven features for capturing venue-level differences.



(a) The hourly popularity of a Pizza place in NYC which draws customers around the clock (high entropy) and a Frozen Yogurt place that draws customers mostly towards evening hours (resulting in low entropy).



(b) The reachability matrix (left) shows that the locality (marked in red) receives more visits from farther localities whilst its distance-weighted reachability matrix (right) takes the distance into account.

Figure 2.4: The difference in hourly popularity of two different venues (left) and the difference between the two reachability definitions (right) – darker regions represent higher numerical values of the respective features.

We define inflow as the number of Foursquare transitions that arrive at the venue of interest and outflow as the number of transitions that leave from the venue. Additionally, we examine transitions from/to all venues in the neighbourhood, N_i, to compute the **Surrounding Area Inflow** and **Surrounding Area Outflow**. Formally, where M is the lifespan of the venue in months and $t(v_j, vi)$ is the number of transitions from v_j to v_i, we define the average monthly inflow to venue v_i as $\frac{\sum_{j=0}^{|V|} t(v_j,vi)}{M}$ and outflow as: $\frac{\sum_{j=0}^{|V|} t(v_i,vj)}{M}$. We also calculate the ratio of inflow to outflow for both the venue and the surrounding area.

Prior research has shown the **Distance of Travel To** and the **Distance of Travel From** certain nodes in a network correlates with higher changes of connection between those nodes [109]. We thus measure the mean distance travelled to reach a venue, as well as to the surrounding area (i.e., N_i). Formally, if $dist(v_j, v_i)$ is the distance between v_i and v_j and N is the total number of transitions to the venue, we define this as: $\frac{\sum_{j=0}^{N} dist(v_j,v_i)}{N}$. As a possible measure of accessibility, we compute the

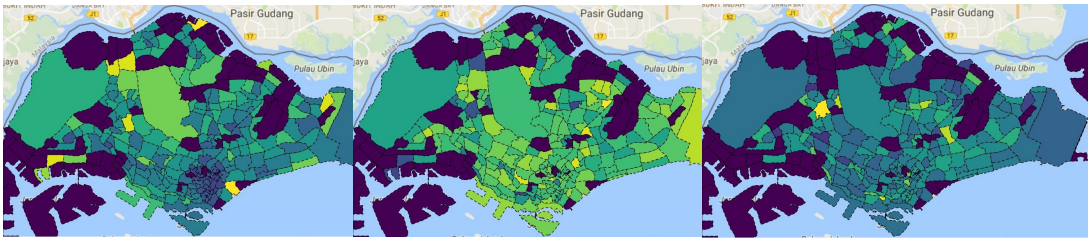
speed with which the venue can be reached from other locations. Formally, we compute the mean **Speed of Travel To** and the mean **Speed of Travel From** the venue of interest. We utilise a similar definition for the mean speed of the surrounding areas. We define speed as $\frac{\sum_{j=0}^{N} dist(v_j,v_i)\times t_{i,j}^{-1}}{N}$ where $dist(v_j, v_i)$ is the distance between v_i and v_j, $t_{i,j}$ is the time spent travelling between v_i and v_j, and N is the total number of transitions to v_i. While the travel distance $dist(v_j, v_i)$ can be computed given the location coordinates of the venue pair (v_i, v_j), the travel time is estimated by the difference in the transition start and end time. For these features, we also compute the standard deviations of these variables following standard definitions.

Temporal Popularity Skew: We define the hourly temporal profile of a venue, h_i, as a vector of 24 elements, with each element representing the proportion of check-ins the venue has received during that hour as compared to the total check-ins received over all hours. We measure the skew as the entropy of the venue's temporal profile (equivalent to Equation 2.3). As seen in Figure 2.4a, a venue that is popular across all hours would have a higher entropy compared to a venue with greater temporal skew (e.g., Frozen Yogurt place).

Temporal Alignment of Venue to its Competitors: Past work [92] has utilised the concept of *diurnal synchronisation* of a venue to demonstrate how a venue is able to observe a larger set of transitions when its operating hours are more closely aligned with its surrounding venues. Given a venue v_i with a temporal profile h_i and its competitive neighbourhood CN_i whose aggregate temporal profile is H_i, we define compute (mis)alignment as the Euclidean distance between the two vectors: $\sum_{j=1}^{24} (h_i(j) - H_i(j))^2$ over the 24 hours.

Visit Trend: We quantify a venue's temporal trend during the observation window as follows: given a time series of check-in values $C_t(v_i)$, we fit a linear regression model whose slope ($s(v_i) = \frac{C_t(v_i)-b}{t}$) represents the trend, with b being the intercept. Whilst our definition of closure (in Eq. 2.2) and the trend feature both look at the temporal profile of checkins, they are distinct in that while the trend captures drop/rise in checkins within the observation window whilst closure is decided

based on drop/rise before and after PD (based on average monthly volume of check-ins). To verify that there isn't potential leakage between these two constructs, we computed the correlation between "trend" and the quantity $\frac{\frac{\sum_{t=0}^{T} C_t(v_i)}{T}}{mean(v_i)}$ (see Eq. 2.2) and found that this was weak (varying between -0.018 and 0.06 across all cities, and 0.1 in the worst case for Los Angeles).



(a) Place Entropy (b) Temporal Popularity Skew (c) Alignment to Locality

Figure 2.5: Select features spatially aggregated over localities across Singapore. Darker regions represent higher numerical values of the respective features.

2.3.3 Mobility Dynamics

These features embody our intuition that the temporal patterns of movement of visitors to/from the venue, relative to the temporal pattern of visitors to the broader locality, helps capture the latent preferences of the urban population.

Temporal Alignment of Venue to its Locality: We hypothesise that a mismatch between the natural timings of draw of a locality and the venue of interest may impact a venue's chances of survival. We define the hourly temporal profile of a locality as a vector of proportional arrivals to the locality for each hour, h_l, during the observation period. Then, we define the (mis)alignment as $\sum_{j=1}^{24} (h_i(j) - h_l(j))^2$, i.e., the Euclidean distance between the hourly temporal profiles of the venue and its locality.

Reachability of Locality: We further hypothesise that a locality's accessibility plays a critical role in the survival of its venues. To quantify this, we first construct the transition matrix, R, whose elements $r_{a,b}$ represent the total number of trips that originated from locality a and ended at locality b during the observation period. A *reachable* locality is one that attracts trips from many localities. We measure the

reachability of a locality a as the entropy of the a^{th} column of R, R_a. Further, a *reachable* locality should attract visits from both distant and local regions. To account for this, we weigh the frequency of transitions inversely by the distance between the regions with $dr_{a,b} = r_{a,b}/d^2{}_{a,b}$ where $d_{a,b}$ is the Haversine distance between the localities a and b, and $dr_{a,b}$ are the elements of the modified transition matrix $\hat{R}$. In Figure 2.4b, we contrast the two features; the reachability vector shows that the locality receives much of its footfall from father localities (resulting in low entropy due to such skew) whereas the distance-weighted reachability vector is more uniformly distributed (resulting in high entropy).

Additionally, we also consider a number of control variables such as the **Specific Category** of the venue and the **Price Tier** of the venue. The tiers range from 1 (least pricey) to 4 (most pricey). In Figure 2.5, we visualize the spatial spread of three features aggregated over the different localities.

2.4 Evaluation

In this section, we report our findings on the predictive ability of individual factors that we consider in this work, and the performance of our methodology overall. We first discuss the influence of individual factors on predictability of survival likelihood in section 2.4.1, and summarise the overall performance of the combination of features in section 2.4.2 for the two cities New York and Singapore, for which we have both Foursquare as well as transport data, and further extend our analysis to *Retail* businesses at large.

Finally, in section 2.4.5, we study the robustness of our criteria for defining closure (see section 2.2.2).

Prediction task: We represent the venue closure prediction task as a binary classification task with the closure label (0 – closed and 1 – open) as the dependent variable and the features described in section 2.3 as independent variables and adopt a Logistic Regression model in all our analyses. Logistic regression also provides

the additional benefit of providing an understanding the relative influence of the features on the prediction outcome.

Experiment conditions: As our dataset consists of an unbalanced number of samples of positive (i.e., open venues) and negative (i.e., closed venues) classes with the negative class being much smaller (see Table 2.2), we first create a subset of all the negative samples and randomly sampled, equal sized positive samples, generating a balanced dataset. We then split the four groups into training and test sets with the training set consisting of 80% of the data on which we perform 10-fold cross-validation to pick the best performing model, and report the accuracy of prediction on the test sets. All features described were min-max normalised. The number of training samples in each of the four groups, (1) F&B venues in Singapore, (2) *Retail* venues in Singapore, (3) F&B venues in NYC and (4) *Retail* venues in NYC were 1450, 2794, 552, and 1062, respectively.

Performance metrics: In all our analyses, we report the accuracy based on precision, recall and AUC, following their standard definitions. Precision and recall represent the average over both the positive and negative classes.

Implementation: The computations related to logistic regression were performed using R (default package stats) and the ROCR [3] library for performance calculations.

2.4.1 Feature Selection and Pruning

In order to understand the ability of the features described in section 2.3 in predicting survival likelihood, we run logistic regression with each feature as the (only) independent variable and report the average AUC over 10-fold cross-validation in Table 2.5 of the top-5 influential features for the two cities, respectively. We also report the correlation between the variables in each case – here, we compute the correlation coefficient as the root of the coefficient of determination (R^2), with the sign (positive/negative) based on the estimated coefficient from logistic regression.

We avoid the use of the widely used Pearson's correlation coefficient [19] since the two-class dependent variable doesn't fit the linearity assumption that Pearson's requires. We see that the temporal popularity skew, and the temporal alignment with the competitors and the locality itself being top features consistently, each with a high $AUC \geq 0.75$. We apply the Boruta algorithm [76] for feature selection and consider the features that were consistent across F&B venues from both Singapore and New York City in subsequent analyses.

2.4.2 Predicting Venue Closure

In this section, we summarise our findings from running logistic regression [45] on the 20 confirmed features resulting from the Boruta search in Table 2.6. For brevity, we only show list features that were found to be statistically significant in the combined model. We run regressions separately for F&B venues, and extend to *Retail* venues in general. *Retail* venues consists of venues that belong to either F&B, Entertainment, Clothing Stores, Nightlife Spots, Food & Drink Shops, Gym/Fitness Centers and other Retail Shops.

We report the following key observations:

1. We see that a number of features consistently appear to have strong influence on the prediction outcome; namely, the (1) visit trend over the current period, (2) the skew in hourly temporal popularity, (3) temporal (mis)alignment of the venue to its locality, and (4) the (entropy) of the distribution of venue types in the vicinity of a venue, across both cities, and for both F&B and *Retail*, with very few exceptions.

Table 2.5: Features with the highest performance in predicting venue closure for Singapore (left) and New York City (right).

Feature	**AUC**	**Correlation**
Temporal Popularity Skew	0.788	0.528
Alignment with Neighborhood	0.786	-0.479
Alignment with Locality	0.758	-0.526
Inflow	0.66	0.270
Outflow	0.624	0.268

Feature	**AUC**	**Correlation**
Temporal Popularity Skew	0.794	0.589
Alignment with Neighborhood	0.782	-0.575
Alignment with Locality	0.732	-0.555
Trend	0.716	0.487
Inflow (Neighborhood)	0.626	-0.201

Table 2.6: Coefficients from Logistic Regression for two cities. *** represents $p < 0.001$, ** represents $p < 0.01$, and * represents $p < 0.05$. SG - Singapore, NYC - New York City.

Feature	**SG, *Retail***	**SG, F&B**	**NYC, *Retail***	**NYC, F&B**
Inflow	-58.76 .	-1.61	-21.01	-34.01 *
Outflow	**175.04 *** **	4.42	68.27 **	47.02 **
Speed Entering	-3.75 *	**-10.06 *** **	-3.51	-17.17 *
Hourly Temporal Skew	**9.80 *** **	**10.53 *** **	**9.13 *** **	**12.97 *** **
Visit Trend	**0.13 *** **	**0.13 *** **	**74.34 *** **	**58.67 *** **
Total Visits	0.82	0.55	5.41 ***	1.99
Place Entropy	-2.33 **	-2.31 *	4.55	-3.70
Distance Entering, Sur	-11.32 **	-1.19	-6.91	-3.31
Distance Leaving, Sur	7.22 *	1.64	-0.89 .	-0.48
Temporal Alignment to Locality	**5.67 *** **	**6.96 *** **	4.58 .	8.69 *
N	2794	1450	1062	552
R2CU	0.356	0.349	0.580	0.577

2. Based on the coefficients for the hourly temporal skew feature, it appears that venues that are popular *around the clock*, and not subjected to specific hours, may have a better chance at survival. This finding suggests that restaurants that only cater to specific customer segments (e.g., lunchtime office workers or dinnertime visitors) are more likely to experience failure. To further analyse this, we looked at the failure rate, in Singapore, for restaurants in two neighbourhoods with skewed visitor dynamics: the Central Business District (**CBD**) that has a dominantly lunchtime presence, and Clarke Quay (**CQ**) that is geared towards tourist and leisure traffic and is more active at night. We picked all restaurants from **CBD** and **CQ** (139 venues in total), and ranked them by their hourly popularity entropy. We compare the top 30 restaurants (highest entropy) and the bottom 30 restaurants (lowest entropy)–i.e., approximately, the top and bottom 20-percentile of such venues. We find a clear difference: whilst only 73% of the bottom-30 restaurants survived the next 6 months, 100% (all 30) of the venues in the top-30 survived. Notably, both **CBD** and **CQ** belong to the Downtown Core Planning Area[7] with no significant difference in the median rentals for retail space.

[7] `https://www.ura.gov.sg/Corporate/Guidelines/Urban-Design/Downtown-Core`

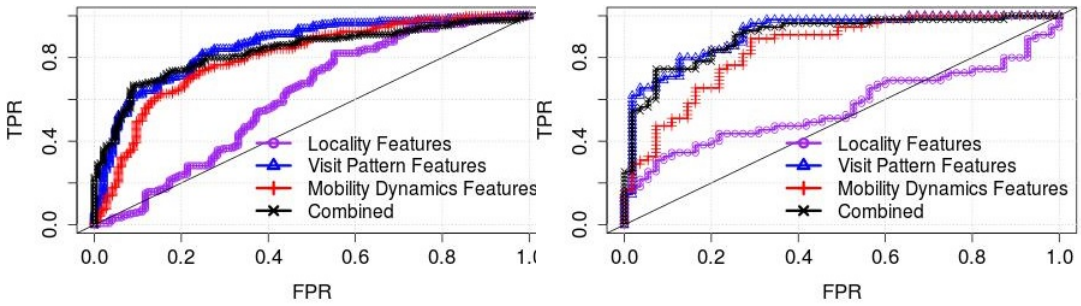
3. On the contrary, the estimated coefficient of *place entropy* (i.e., negative) suggests that a decrease in entropy improves the likelihood of survival. This seems to suggest that venues that are in the midst of more clustered neighbourhoods (such as ethnic enclaves) tend to survive longer.

4. Not surprisingly, the trend of customers check-in patterns during the current period is indicative of the venue's performance over the following 6-month period. And, as anticipated, we also note that the sign of the coefficient is positive, indicating that venues that experience an upward trend in check-ins have a much higher likelihood of survival.

5. Between the two cities, we observe that in the case of Singapore, several of the features tied to a venue's locality or neighborhood are found to be statistically significant – for instance, the (mis)alignment, measured by the Euclidean distance, suggests that venues that operate outside popular hours of the locality, have a distinctive advantage over their neighbours.

6. The goodness of fit is higher for NYC than Singapore, and suggests that $\approx$ 58% of the variance is explainable for both F&B and *Retail* venues.

Comparison with baseline: As we describe in section 6.1, the work of Wang et al. [126] is the closest to our work in that they rely purely on LBSN-based features to study decline in business performance for F&B venues in NYC, over a 3 month window. In Table 2.7, we compare our results against this baseline, reproducing the confusion matrix presented in [126]. The features we consider in this work achieve at least 15% better precision, and 7-10% increase in recall for venues in NYC. Our evaluation is on a balanced test set, whereas the baseline misclassification rates reported in [126] may be a bit misleading as they are based on a highly imbalanced dataset (with less than 20 samples in their "closed" class).



(a) Singapore, *Retail* (b) New York City, *Retail*

Figure 2.6: ROC Curves of *Retail* venues in Singapore and New York City. The Curves represent the performance for each class of features and for the combined model, respectively.

2.4.3 Accuracy across Feature Classes

To understand the influence of features classes on the prediction outcome, we run the logistic regression for each class separately, and in combination. In Table 2.8, we report the observed AUC scores for *Retail* venues from Singapore and New York, and note that in general, the visit pattern features alone reach AUCs ≥ 0.80, and the addition of mobility dynamics features lends to only a 4% and 3% increase in accuracy for Singapore and NYC, respectively. Figure 2.6 shows the ROC curves for the same. However, given that these two cities are amongst the world's top-rated in terms of public transit infrastructure [6], we expect that the mobility features across localities to be more uniform – for instance, the kurtosis of the reachability feature is negative for both cities (i.e., platykurtic, with -0.48 for SG and -1.07 for NYC). To understand the merits of this additional class of features, we construct a subset of venues (which we refer to as the *Contrast* set) which consists of those venues with the top-5% and bottom 5% value of the distance-weighted reachability feature, and re-run the analyses. As anticipated, we observe a significant improvement in the accuracy (i.e., 8% and 10% for SG and NYC, respectively) over using visit pattern

Table 2.7: Confusion Matrix Comparison against Previous Work [126].

	Wang et. al[126]		**NYC, Retail**		**NYC, F&B**	
Confusion Matrix	Labeled 0	Labeled 1	Labeled 0	Labeled 1	Labeled 0	Labeled 1
Observed - 0	160	5	87	19	46	13
Observed - 1	7	14	19	87	9	42
Precision (closed class)	66.67%		82.07%		82.35%	
Recall (closed class)	73.68%		82.07%		76.36%	

Table 2.8: AUC scores of the different feature classes with Logistic Regression against the Random Baseline. The *Contrast* set consists of venues with the top-5% and bottom-5% values of the reachability feature.

	SG, Retail	**SG, Retail (Contrast)**	**NYC, Retail**	**NYC, Retail (Contrast)**
N	2794	248	1062	102
Random Baseline		0.50		0.50
Locality		0.60		0.58
Visit Pattern		0.82		0.89
Mobility Dynamics	0.80	0.88	0.80	0.88
Combined	0.86	0.90	0.92	0.99

features alone.

2.4.4 Prediction Accuracy by Venue Category

In this subsection, we aim to understand if and whether the ability to predict closure, and the attributes that contribute towards the prediction, differ by the category, or cuisine, of the venue. In Figure 2.7, we plot the top-15 subcategories/types of cuisine appearing under the F&B category of venues in Singapore. We observe that the distribution is heavily skewed and that Coffee Shops appear to the predominant type of venue followed by Chinese Restaurants (as anticipated with Singapore's population consisting of 74% ethnic Chinese[8]).

In Table 2.9, we summarize the prediction results for the top-6 categories. We observe that whilst the accuracy is above our previous average for certain classes (e.g., Coffee Shops), the accuracy is only as good as the random baseline (e.g., AUC = 0.5 for Indian Restaurants) for certain other categories. However, as mentioned previously, as we maintain a balanced dataset of closed/open classes for the analyses, the total number of venues (N) drops as the percentage of venues being labelled as closed is low. Segregating the dataset by category means that the resulting sample size drops further. Hence, we caution the reader that the results may vary if tested on larger samples.

[8] `https://www.indexmundi.com/singapore/demographics_profile.html`

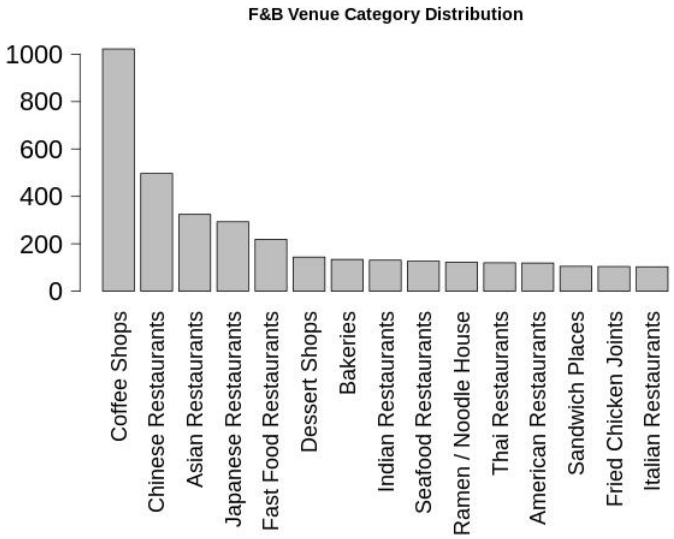


Figure 2.7: Distribution of Cuisines/Categories of F&B Venues in Singapore.

Category	N_{train}	N_{test}	Avg. Precision	Avg. Recall	Best AUC	R2CU	Mean AUC	Test AUC
Coffee Shops	184	24	0.93	0.92	0.99	0.36	0.82	0.74
Chinese Restaurants	112	14	0.85	0.79	0.88	0.43	0.69	0.67
Asian Restaurants	92	12	0.76	0.75	0.89	0.32	0.73	0.78
Japanese Restaurants	54	8	0.9	0.88	1	0.52	0.77	1
Fast Food	36	4	1	1	1	0.75	0.83	0.78
Dessert Shops	36	4	–	0.5	1	0.75	0.7	0.33
Bakeries	30	4	1	1	1	0.75	0.75	0.25
Indian Restaurants	32	4	1	1	1	0.75	0.6	0.5

Table 2.9: Venue Closure Prediction Accuracy by Category of F&B Venue.

2.4.5 Robustness Checks

In this section, we perform a series of checks to understand the robustness of our analyses under various conditions.

Dealing with Collinearity

In addition to the feature selection step carried out in section 2.4.1, here we test for collinearity across variables - to detect collinearity, we compute the Variable Inflation Factor (VIF) [93] of variables – a score greater than 12 suggests that there exists significant correlation across certain variables. We tabulate the standard errors and the VIF values, as an illustration, for the case of F&B venues in New York City. The table compares the *Combined Model* consisting of the select features

Table 2.10: Standard Error of Estimated Coefficients and Variable Inflation Factors of Selected Features for Retail Venues in New York City for the Combined Model (left) and Reduced Model (right). SE- Standard Error.

	Combined Model		**Reduced Model**	
Feature	**SE**	**VIF**	**SE**	**VIF**
Temporal Popularity Skew	3.85	11.83	0.84	1.06
Visit Trend	18.82	3.55	8.80	1.03
Place Entropy	2.77	3.52	1.86	1.58
Inflow	18.66	9.84	4.72	1.16
Outflow	18.17	10.00		
Distance Entering, Surrounding	2.61	3.60		
AUC	0.92		0.89	

from section 2.4.2, against a subset of features whose Pearson's correlation with any other feature is less than 0.5 which we refer to as the the *Reduced Model*, in Table 2.10.

We find that: (1) removing the uncorrelated features reduces the standard error in the estimated coefficients, and lowers the VIF significantly (all less than 2), and (2) in removing the correlated variables, the resulting AUC drops only marginally.

Labeling Methodology

As we saw in section 2.2.2, the labeling definition (as defined in Equation 2.2) is noisy. To provide a more plausible check on the validity of our 'failure' labels, we additionally retrieved the longitudinal data (from June'11-Dec'17). We then hypothesise that a truly-closed venue will not have any such photo posts or tips generated after its closure date; conversely, a venue that was open at the end of

Table 2.11: Results from Logistic Regression for each city on all F&B venues in the dataset versus only those we consider as *Confidently* labeled, on the *Reduced Model*. N is the total number of venues considered, balanced between the two classes open/close.

	Reduced Model on All Venues					**Reduced Model on *Confident* Venues**				
	N	**AUC**	**Precision**	**Recall**	R^2	**N**	**AUC**	**Precision**	**Recall**	R^2
Chicago	608	0.91	0.79	0.79	0.46	416	0.97	0.89	0.88	0.72
Helsinki	184	0.94	0.84	0.83	0.36	128	1.00	0.93	0.92	0.40
Jakarta	872	0.85	0.78	0.76	0.34	284	0.95	0.86	0.86	0.48
London	72	0.86	0.78	0.75	0.34	32	1.00	0.88	0.83	0.74
Los Angeles	158	0.87	0.78	0.78	0.37	64	0.97	0.83	0.75	0.60
New York City	660	0.87	0.79	0.78	0.40	458	0.90	0.80	0.80	0.45
Paris	184	0.87	0.76	0.76	0.37	52	1.00	0.88	0.83	0.68
San Francisco	140	0.91	0.84	0.84	0.31	66	0.98	0.94	0.93	0.45
Singapore	2980	0.82	0.75	0.74	0.30	1468	0.89	0.78	0.77	0.37
Tokyo	632	0.82	0.75	0.74	0.27	268	0.91	0.77	0.73	0.22
ALL	6490	0.86	0.78	0.78	0.42	3236	0.85	0.77	0.76	0.47

Dec'13 should have seen at least one instance of such activity over the entire period of Jan'14-Dec'17.

More specifically, we define the following labels, *Confident-Close* and *Confident-Open*, and extract the corresponding data subsets as follows:

- *Confident-Close:* is the subset of venues that are labelled 'failed' ($RemainsOpen(v_i) = 0$ according to definition of Equation 2.2) AND that have had zero activity (tips, photos) posted during the entire interval Jan'14-Dec'17.

- *Confident-Open:* is the subset of venues that are labelled 'not-failed' ($RemainsOpen(v_i) = 1$ according to definition of Equation 2.2) AND that have had least one activity (tips, photos) during the interval Mar'14-Dec'17. (We define this activity period from Mar'14, and not from Jan'14, to provide a buffer against possible delayed posting of an activity by a user who had visited prior to the venue's closure.)

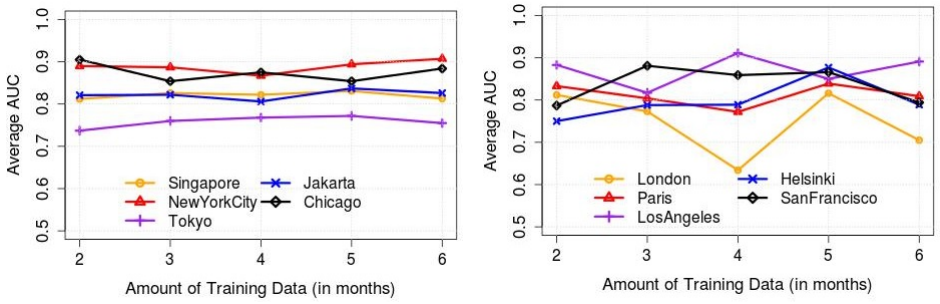
***Confident* Dataset:** We create this subset to only consist of the Confident-Close and Confident-Open F&B establishments to check if our results are largely invariant to such labelling errors. The astute reader will note that the Confident Dataset cannot be operationalized for prediction tasks (as the labels are determined based on a 5-year observation period that occurs *after* the prediction instant). However, an analysis using this dataset helps us investigate the robustness of our results.

In Table 2.11, we tabulate the the number of samples, AUC, precision, recall and the respective McFadden index observed with logistic regression, for the full dataset and the *Confident* subset, respectively, on the *Reduced Model* from the previous section. We do this to make sure that the number of *Confident* samples aren't too low for analysis. The results seem to suggest that better performance is seen with better quality labeling – for instance, for certain cities, we reach a perfect AUC score. However, we also note that despite inaccuracies in labeling, the model's performance is rather stable.

Impact of Amount of Past Data on Performance

In our analyses thus far, we consider an observation window of 6 months (prior to the prediction date) to predict the survival of a venue in the following 6 month period. A natural question then is: *how much data from the past is required to make a reasonable prediction?*

To answer this, we vary the length of the observation window between two (a minimum of two data points are needed for calculating the trend feature) and 6 months, immediately preceding the prediction date, and repeat our analysis. In Figure 2.8, we plot the amount of training data (in months) on the $x-$axis, and the mean AUC over 10-fold cross validation on the $y-$axis. For clarity, we plot the AUCs separately for (a) dense and (b) sparse cities (based on check-in volume) – in either case, we note that the AUCs are relatively stable. The worst case of a 20% drop occurs in the case of London besides which the variability is limited to within 2% for the denser cities. This analysis concludes that even modest amounts of training data (e.g., 2 months) is sufficient for achieving reasonable prediction performance.



(a) Dense Cities (b) Sparse Cities

Figure 2.8: Impact of the Length of Observation Window on Performance.

2.5 Discussion Points

Here, we discuss alternate solutions and problems that we considered in studying the performance of physical venues, our preliminary findings from analysing user reviews, and the limitations in the current undertaking.

2.5.1 Alternate Solutions

As deep learning models have been extensively used in a number of problem domains [10, 116, 73], we also conducted some preliminary studies using the PyTorch framework and the scikit-learn library [4]. Similar to Avati et al. [10], we slice the observation window (i.e., 6 months preceding the PD) into four, and extract features per slice as well as summary statistics for the entire window. We define the slices as: (1) **Slice 1**: $[PD - 2weeks, PD)$, (2) **Slice 2**: $[PD - 1month, PD - 2weeks)$, (3) **Slice 3**: $[PD - 3months, PD - 1month)$, and (4) **Slice 4**: $[PD - 6months, PD - 3months)$. For each slice, we consider the total number of check-ins each venue received, per day of the week, per hour of the day (leading to 168 features per slice) and summarize the total, mean, standard deviation, range, maximum and minimum over the observation window for each feature. The final feature set consists of 1680 features per venue with the outcome variable as defined previously (Eq. 2.2). We trained a deep neural network model with an input layer of 1680 dimensions and a two-dimensional output layer. The optimal number of hidden layers (each consisting of 512 dimensions, varied from 1 to 30) and the activation functions (varied between ReLU, SELU and Tanh) were fixed based on performance on the validation set. We take away from these early results, summarized in Table 2.12, that the configuration is not consistent across the cities, and the performance (AUC varying from 0.65 to 0.85) is generally poorer than that of considering hand-crafted features such as those in section 2.3.

At present, we don't investigate further the specific choice of a classification technique because: (i) our key focus is on establishing that the combination of social media & urban mobility data provides high predictive power, and (ii) the volume and diversity of data we use do not seem suitable for such data-hungry alternatives.

City	Mean (Std. dev) number of non-zero features	Sparsity	Num layers	Unit	AUC
Chicago	243.27 (234.66)	85.52%	27	SELU	0.76
Helsinki	268.63 (235.49)	84.01%	11	ReLU	0.77
Jakarta	101.87 (147.02)	93.94%	11	ReLU	0.68
London	190.02 (191.16)	88.69%	16	SELU	0.71
Los Angeles	177.72 (187.33)	89.42%	24	SELU	0.76
New York City	225.66 (247.02)	86.57%	16	SELU	0.77
Paris	150.48 (173.61)	91.04%	28	ReLU	**0.81**
San Francisco	242.70 (220.14)	85.55%	16	SELU	**0.85**
Singapore	141.24 (176.01)	91.59%	27	SELU	**0.81**
Tokyo	126.29 (184.43)	92.48%	11	Tanh	0.65
ALL			14	SELU	0.74

Table 2.12: Summary of Results from Deep Learning.

2.5.2 Predicting Upsurge in Popularity

Contrary to predicting venue survival, here we label venues as those that saw an upsurge in popularity (as defined in Equation 2.4), where the average monthly volume of check-ins during the test period is at least $K = 25\%$ more than the average seen over the current term.

$$Surges(v_i) = \begin{cases} 1, & \text{if } \frac{\sum_{t=0}^{T} C_t(v_i)}{T} > K \times mean(v_i) \\ 0, & \text{otherwise} \end{cases} \tag{2.4}$$

Table 2.13: Features with the highest performance in predicting upsurge in venue popularity for Food venues (left) and Retail venues, in general (right), for Singapore.

Feature	Mean AUC	Correlation
Trend	0.60	0.274
Popularity of Locality (Morning)	0.59	0.237
Temporal Popularity Skew (Ent.)	0.56	0.147
Speed Entering	0.55	-0.121
Alignment with Locality	0.55	-0.106

Feature	Mean AUC	Correlation
Speed Entering	0.62	-0.124
Distance Leaving	0.61	-0.158
Trend	0.59	0.266
Outflow	0.57	-0.167
Popularity of Locality (Morning)	0.56	0.242

In Table 2.13, we tabulate the top-5 features (that have the highest explanatory power over predicting upsurges), for F&B venues, and Retail venues, respectively. While we observe that the Visit Trend and Hourly Popularity Skew are features that were useful in predicting failures, previously, the values of the correlations/AUCs themselves (for upsurge prediction) are significantly less (e.g., highest AUC ob-

served is only 0.62 compared to 0.75, previously). We also note that different features such as the popularity of the locality in which the venues operates in, during the morning hours, are more useful in predicting upsurge.

Further, in Table 2.14, we summarize the estimated coefficients and the associated $p-$values from running logistic regression for the alternate upsurge-prediction problem, for F&B venues from Singapore. Note that the number of venues considered here for analysis is considerably less ($N = 348$), as only 7% of the total venues saw an upsurge of more than 25% during the prediction period (and were then subsequently labelled as belonging to the positive class), resulting in only 14% of the total venues in the dataset being considered for the analysis as part of a balanced data set. Consistent with previous findings, venues with positive visit trends and are operating out of band with the venues near them (i.e., positive coefficient for the misalignment feature) show chances of upsurge in popularity. We also note that Place Entropy which captures how heterogeneous a neighborhood is, and Category TF-IDF which captures how unique the venue's category is to its neighborhood, are both negatively correlated, but are useful in predicting popularity. Although, we observe a mean AUC of 71.9% (about 20% better the random baseline), the accuracy is less than what we observed for predicting survival with the same set of features. This brings us to the conclusion that investigations into other attributes will need to be required for predicting upsurges better.

2.5.3 Additional Insights from Mining Reviews/Tips

In our current approach, we are only aiming at predicting impending failures of venues using tangible attributes of the environment the venues operate in, and the venue's check-in patterns. As Parsa et al. [97] conceptualize, there are other intangible attributes of the business and how it is run that can dictate its fate – for example, the quality of service offered by the staff, the ambience of the venue, the relationship of the owner with its patrons, etc. In this section, we attempt to reveal,

Table 2.14: Coefficients from Logistic Regression for Predicting Upsurge in Popularity for F&B Venues in Singapore. *** represents $p < 0.001$, ** represents $p < 0.01$, and * represents $p < 0.05$.

Feature	**Estimate**
Visit Trend	10.62***
Alignment with Neighborhood	7.87**
Popularity of Locality during Morning Hours	2.39**
Distance Leaving	(-7.97)**
Place Entropy	(-7.18)**
Category TF-IDF	(-3.06)**
Temporal Popularity Skew (Entropy)	7.65*
N	348
Mean AUC	0.719
R2CU	0.3875

if any, such intangible factors by mining textual data (i.e., *tips*) that users of the platform have shared.

In particular, we aim to answer the following additional questions:

1. Out of those venues that are *similar* in terms of the observable factors, are there any causal factors (e.g., corrective actions such as promotions, deals and change in menus) that lead to some venues to thrive whilst others perish (and vice versa)?

2. Out of those venues that we label as at risk of failing in the six months to follow, are there any factors that lead to them to remain in business for longer?

To address the above, we first use $k-$means clustering for grouping venues (F&B venues from Singapore, in this case) by their similarity in the feature space. For ease of interpretation, we consider the subset of features that were part of the *Reduced Model* considered in Section 2.4.5. To decide on the optimum number of clusters, we use the Silhouette coefficient; in Figure 2.9, we plot the coefficient which is a measure of the quality of the resulting clusters (on the $y-$ axis) against the number of clusters on the $x-$ axis. We see that the $k = 6$ offers the greatest quality.

With $k = 6$, we then plot the normalized values for each considered feature of

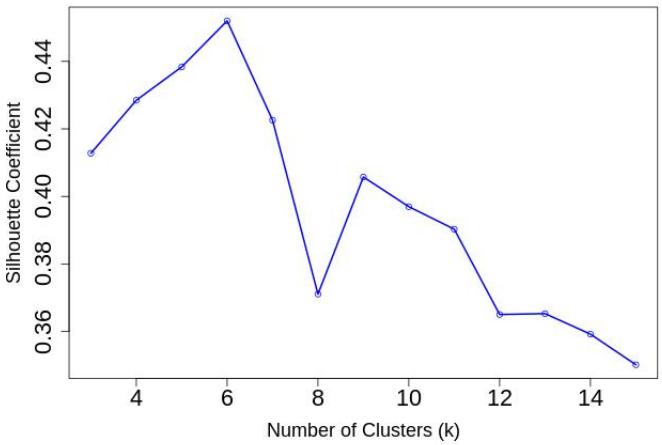


Figure 2.9: Impact of the choice of number of clusters (k) on clustering quality.

the centroid venue of each cluster in Figure 2.10. The number of venues in each cluster are 218, 282, 404, 375, 71 and 100, and the percentage of venues labelled as *RemainsOpen* in each cluster are 76.6%, 59.2%, 54.5%, 41.6%, 11.3% and 7%, respectively. As anticipated, we observe that the clusters have mixed representation of surviving and failing venues, with the clusters at the extreme representing the most surviving and failing venues, respectively. For further analysis, we pick the 24% venues that were labeled as failing in Cluster 1 and the 7% of the venues that were labeled as surviving in Cluster 6 – these two groups are essentially similar in terms of their observable variables to other venues in their respective clusters, but are failing (or surviving) unlike their peers.

To analyze the tips, we first pre-process them by removing standard stop words followed by tokenization. We compute the $TF - IDF$ of the words appearing in the tips – as we hope to understand the change in semantics related to the venues over time, here we consider each venue and year–month pair as a *document*, and the words appearing across all the tips during the specific months for each such venue as the *terms* appearing in that document.

As an aside, alternatively, we also tried to observe the trajectory of the tips over time using `word2vec` embeddings [82]. In addition to extracting the $TF - IDFs$

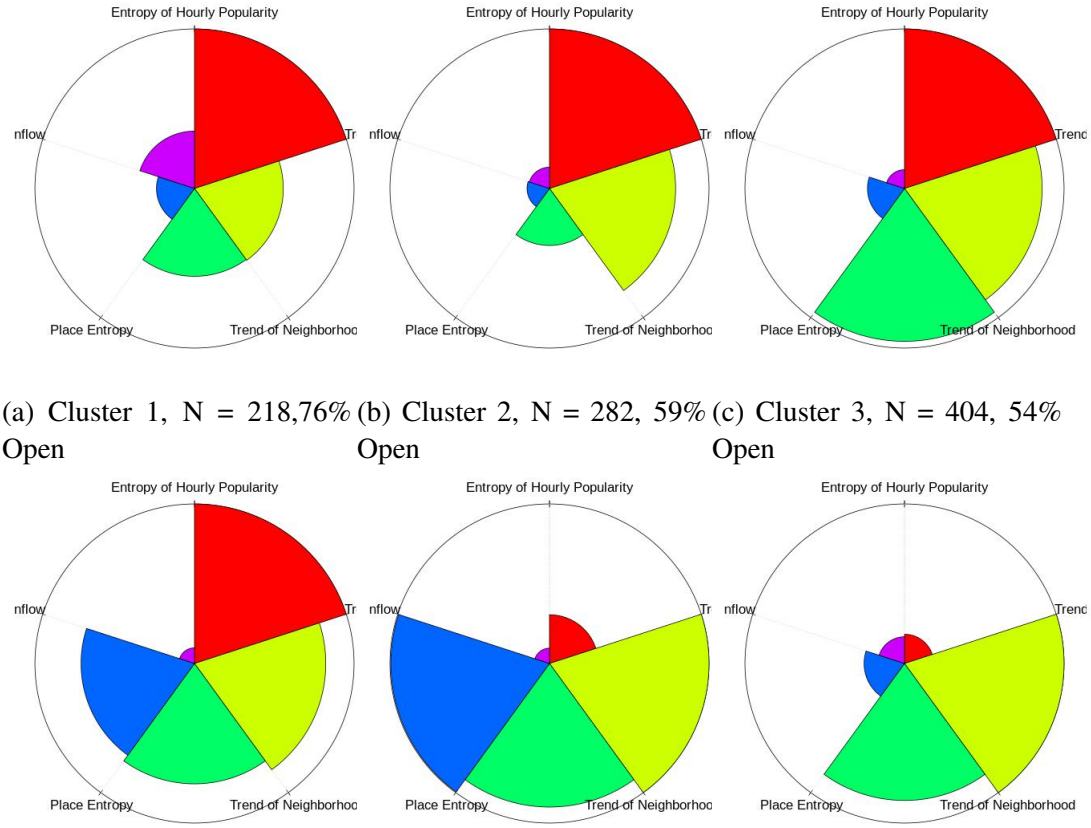
of words appearing in the text, we observed the $TF - IDF-$weighted position of the tips in a 300–dimensional word embedding space. For this purpose, we used a pre-trained embedding space trained on 300 million unique words appearing in Google News [9]. However, possibly due to the words used for training and the tips belonging to different domains, the positioning in the vector space did not lead to any better interpretation as hoped.

To answer our first question, in Figure 2.11, we provide top words appearing in Tips, at different time points, for three sample venues that were labelled as likely to fail whilst they were similar in terms of their attributes to a predominantly successful cluster of venues. The words suggest that these particular venues seem to have its patrons writing about their negative encounters with the owner and staff of the venues – for instance, multiple occurrences of the owners' (e.g., boss, aunty) attitude (e.g., as being rude) have been pointed out. The patrons also seem to show dissatisfaction towards "bad service" and "slow service". There are also occurrences of having to make reservations being pointed out in a negative tone. These anecdotal textual reviews suggest that despite a venue's higher chance of survival based on it's operating conditions such as its location, certain softer aspects of how the individual stores are run can indeed hurt their business.

On the contrary, in Figure 2.12, we provide sample top-words appearing over time for a chosen venue that was showing signs of survival despite its observable attributes being similar to predominantly failing venues, to study the latter question. It's observable from the Tips that the venue's patrons welcome the additional amenities provided by the venue such as "Free WiFi" and "wall sockets for USB charging". The patrons also have expressed that the venue provides "value for money" and that it "opens till late". Similar to before, these illustrative reviews suggest that despite a venue's low survival chances based on the environment and patron visits, intangible factors such as better customer quality can aid in keeping the business afloat.

[9] `https://github.com/mmihaltz/word2vec-GoogleNews-vectors`

These example shed light into some of the intangible factors (e.g., attitude of the staff) that matter when it comes to the survival of the business, as well as competitive positioning of a business that put them at an advantage (e.g., convenience factors). This qualitative analysis also highlights a key limitation of our current model that only takes into account observable, tangible factors.



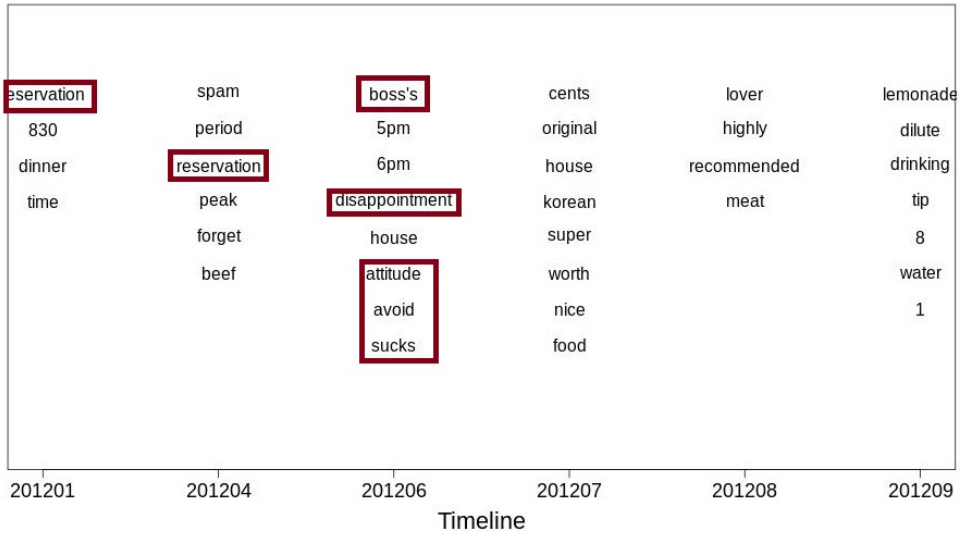
(a) Cluster 1, N = 218,76% Open (b) Cluster 2, N = 282, 59% Open (c) Cluster 3, N = 404, 54% Open

(d) Cluster 4, N = 375,41% Open (e) Cluster 5, N = 71, 11% Open (f) Cluster 6, N = 100, 7% Open

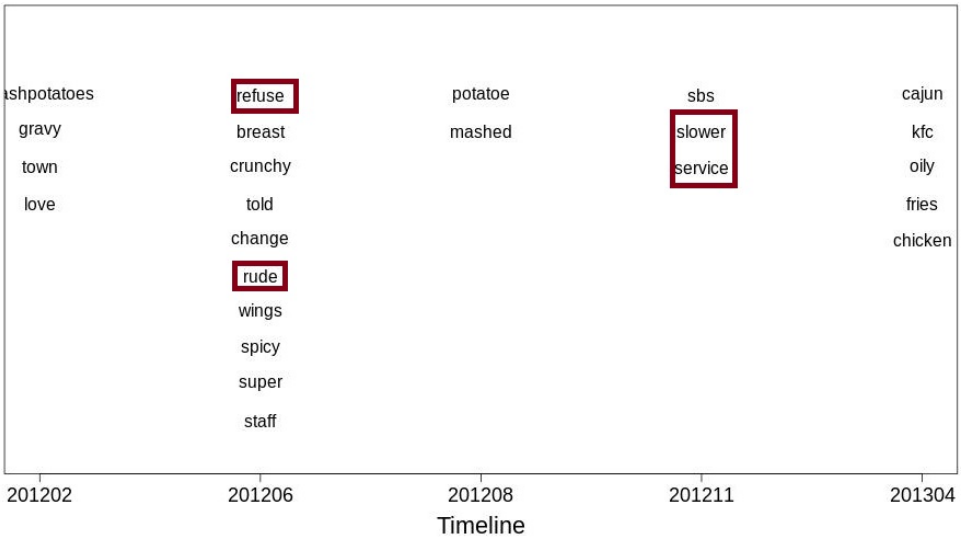
Figure 2.10: Clock Plot of the Centroid Feature Values of the Six Venue Clusters

2.5.4 Current Limitations

Data Availability: Although the geography based features from static venue information (accrued via the Venue API) is public knowledge, we have relied on private data for certain aspects of the analysis – for instance, the check-in dataset used in this work was shared by Foursquare, under an NDA agreement. However, to be able to build deployable systems that can help future business owners and govern-



(a) A Korean Grill BBQ Joint (Closed)



(b) A Fast Food Franchise (Closed)

Figure 2.11: Sample top words appearing in the *Tips* of venues that were failing but similar in characteristics to predominantly successful venues.

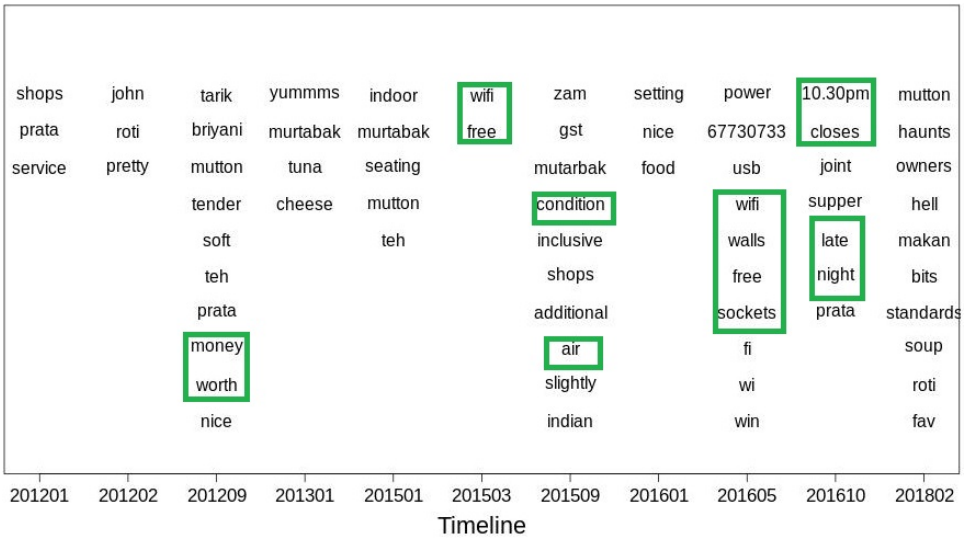


Figure 2.12: Sample top words appearing in the *Tips* of a thriving venue that was similar in characteristics to predominantly declining venues.

ment agencies, the alternative is to rely on current and publicly available check-ins. Hence, an analysis on how well the publicly sampled counterpart represents the population has to be conducted. Moreover, other sources of urban data are likely to provide additional improvement. For example, features such as the average rental price in different neighbourhoods, the water consumption of individual establishments (a great proxy for an F&B establishment's popularity) or the neighbourhood demographics and the economic profile of its residents (often available from census and tax data) are likely to be significant predictors of a retail business's longevity.

A Streaming Prediction Model: We have currently developed a non-continuous predictor: at present, we compute a variety of features using Foursquare and mobility data and then predict a venue's likelihood of survival over the *next* 6 months. Implicit in our approach is the belief that the majority of our features (such as the hourly temporal profile of localities) are stable, and do not vary significantly with time. As the next step, it would be useful to develop a streaming predictor–one which continually updates the survivability likelihood as time progresses, by appropriately incorporating up-to-date feature values. Another open question relates to the look-ahead horizon (currently 6 months)—we need to investigate how our pre-

diction accuracy degrades as we try to use current features to compute likelihood of survival further out in the future (e.g., 1 year).

2.6 Future Implications

In this section, I briefly describe key implications of our work to business owners and urban planners alike, and future work.

2.6.1 Implications to Stakeholders

While we see $AUCs \geq 0.82$ in general, which demonstrates the theoretical merit of this work, for a practical adoption by stakeholders, we further investigated the precision-recall trade-offs. In the case of impending business failure, a high recall would be warranted as retailers would be less sensitive to false-positives (i.e., the system predicting that the venue is likely to fail, but the venue survives in reality) than vice-versa. For a recall of 0.90, in the case of New York City, for example, a reasonably high precision of 0.83 and 0.73 can be achieved for F&B and over all retail venues, respectively. Whilst in Singapore, the achievable precision drops to 0.74 and 0.70 for the F&B and *Retail* classes. We share key takeaways from interviewing several stakeholders.

F&B Owners: The owners (referred to as Owner1 through Owner6) rated an average of 3.5 (on a scale of 1:not useful to 4: very useful) when asked whether understanding their businesses' survival likelihood in the next 6 months is useful to them. They all found the accuracy (of 80-90%) to be either sufficient, or good, for taking precautionary actions except Owner5 who said that the accuracy is low. "Too low to be useful" was also an option which none of the owners chose. Everyone found the prediction horizon (of 6 months) to be appropriate and mentioned that they use a combination of Point-of-Sales data, social media and third party services like Shopify [5] to monitor current health of their business. Only one respondent (Owner2) said that they perform trend analysis to forecast future performance. All

except one respondent said they are either "Likely" (1/6) or "Very Likely" (4/6) to take actions (such as revising the menu or run promotional campaigns) based on reports generated by a future survival prediction system – interestingly, the respondent (Owner2) who chose "Not Likely" runs a franchise of a fast food company and shared that such changes or decisions cannot be made by franchisees independently, but can only be made by the franchisor which is then implemented across the board by all franchisees.

Urban Authorities and Planners: We reached out to an experienced planner at the local authority in Singapore who responded that the agency is interested in knowing survival rates at both the individual (6 on a Likert scale from 1 to 7) and aggregate level (4 on a Likert scale from 1 to 7). The planner also found the prediction horizon of six months to be appropriate, although he felt that the level of accuracy (i.e., 80%) would be too low for the agency to make concrete interventions. He also shared that the agency sees potential in the overall methodology of combining data from LBSNs and urban transportation for informing planning decisions, and in studying people's behaviour, choices and patterns, in general.

Public Policy Expert: We spoke with a Professor of Public Policy in the UK whose response to our precision and recall results ($\approx 80\%$) were positive. He suggested the work could have implications on licensing agreements for new venues by local authorities who currently look for factors such as location and competition [8]. Analysing the likelihood of failure of that area could be considered as an additional factor in those agreements. He shared that the models could have commercial value for both technology intelligence companies and large retail businesses.

2.6.2 Shift towards e-Commerce

In this work, I specifically study the survival aspects of stores or venues with a physical presence by considering factors related to the physical location of the venues (e.g., mobility dynamics features). Given the active debate on whether eCom-

merce will completely take over physical retail[10,11,12], there is a natural question on whether such mobility and neighborhood features will continue to remain correlated with retail success or failure. I posit that the aspects of physical presence such as those studied in this chapter, will remain relevant due to a number of reasons. First, although online retail is a possibility for many types of products and services, there are still a sizeable chunk of the retail space that simply cannot go completely online – for e.g., businesses that still require the physical presence of the end customer such as gyms, barber shops, etc. Next, as more online businesses rely on shared economies for services such as delivery, especially in the F&B arena, the physical presence of the store still remains important as it directly impacts the possible availability of delivery agents and so on. Finally, I further posit that with careful design, the features we studied in terms of physical presence could be possibly translated to quantify the effects of "virtual presence". For instance, the reachability feature that captures the accessibility of a physical zone (where a store resides) from other zones, can now be improvised to capture the accessibility of an online platform or *zone* (for e.g., Zalora[13]) from other "virtual" zones (for e.g., popular search engines, social media site such as Facebook and so on).

2.6.3 Extending the Study of Business Survival

In this chapter, I demonstrated the feasibility of answering the specific question of "will a store survive the next 6 months?" using tangible attributes of the store itself and the characteristics of the neighborhood the store operates in. A natural extension of this question would then be to ask "when is this store likely to fail?", or "how long will this store continue to operate?" under the current operating conditions.

[10] https://www.cnbc.com/2019/04/02/online-shopping-officially-overtakes-brick-and-mortar-retail-for-the-first-time-ever.html

[11] https://www.forbes.com/sites/gregpetro/2019/03/29/consumers-are-spending-more-per-visit-in-store-than-online-what-does-this-man-for-retailers/#22f21b537543

[12] https://www.inc.com/associated-press/why-more-online-only-brands-are-embracing-the-brick-and-mortar-experience.html

[13] https://www.zalora.sg

Questions of this nature are traditionally studied using survival analysis [27, 38]. Whilst a limited number of studies have focused on the store survival problem, in general (e.g., survival rates [14, 129], survival analysis of storefronts on online platforms [127]), the methodology to estimate a *physical* store's "time to failure", to the best of our knowledge, has not been investigated thus far (possibly due to the difficulty in sourcing exact failure dates of physical storefronts, at scale and at fine temporal granularity). However, the demonstrated success of using survival analysis techniques in other domains (e.g., product survival in App stores [68], customer retention in the insurance industry [54] and survival outcome of patients in clinical trials [27]), shows promise that the same techniques could be translated for use in understanding survival rates and attributors of physical stores. As a first step in this direction, we further investigated the possibility to extract exact dates of failure for the venues in our current dataset. We randomly sampled 10% of the F&B venues in Singapore that we label as "Closed" (using the previous definition in Eq. 2.2), and manually searched through the web to find instances of mentions of venue closures. Out of the 25 venues for which we did find closure notices (advertised through disparate mediums such as Facebook[14], Burpple[15], and various blogs), we found that the date of closure was within ≈ 4.26 months of the last date of online activity (i.e., tips and photos) the store had seen on Foursquare. While it shows the future possibility to use longitudinal social media activity data as proxy to extract closure dates, it also reflects the difficulty in attaining such ground-truth labels, or notices of closure.

2.7 Acknowledgements

This work was done in collaboration with University of Cambridge and New York (NYU) University. I thank Dr. Cecilia Mascolo (Cambridge) and Dr. Anastasios

[14] https://www.facebook.com/
[15] https://www.burpple.com/

Noulas (NYU) for providing access to the Foursquare dataset and Dr. Rajesh Balan (SMU) for providing access to the Taxicab dataset used in this work. In Table 2.15, I summarize how the work was split between myself and Krittika d'Silva, a student collaborator, on this project.

	Kasthuri	**Krittika**
Problem scope formalization	50% Survival analysis in Section 2.1.3	50% Closure definition in Section 2.1.2
Feature engineering	50% (Temporal popularity skew, Mobility Dynamics features, Visit Trend, Catchment & attractiveness features)	50% (Place entropy, Inflow/Outflow/Speed/Distance network features, Competition features)
Analysis	100%	Analyses carried out by the author on studying the variability of prediction performance across different geographies and maturity of venues are not presented in this thesis, but are available in **[IMWUT18/UbiComp'18]**.

Table 2.15: Work and contributions split.

Chapter 3

Inferring Social Ties from Mobility Traces

As I mention in Chapter 1, a key challenge in exploiting mobility is in understanding the influence of social groups on a person's mobility. Past works have shown that an individual's mobility is dependent on, and affected by the movement of his/her social ties [125, 26], and that incorporating knowledge of one's friend's mobility can help improve prediction of a user's city-scale mobility behavior [31] (using GPS records). However, whether such influences hold at shorter mobility ranges (e.g., indoors) is unknown although there are compelling applications for indoor spaces (e.g., office buildings, smart campuses) that can benefit from accurate indoor mobility profiling. While friendships are often explicit in online social networking platforms (e.g., Facebook), such information is typically not available when considering an urban population whose mobility data is extracted indirectly from sources such as tap in/outs from public transportation rides. In this chapter, we study the following questions:

1. *Do social groups/friends influence a person's mobility at shorter scales?* To this end, in Section 3.2.2, we conduct statistical analyses on mobility constructs, such as dwell times in different indoor locations and the transitions between such locations, under different social contexts of students on a uni-

versity campus and find that social groups do in fact impact an individual's mobility.

2. *How can social relationships or ties of individuals be learned solely based on movement patterns of a group of individuals?*. To this end, in Section 3.3, we borrow and extend concepts from prior work [77] to quantify the social links between users sharing the physical space, and show that it is indeed possible to extract social ties from movement patterns.

Using indoor mobility data from close to 10,000 students on the SMU campus over multiple school terms, described in detail in Section 3.1, we first apply a state-of-the-art group detection system [112] to extract the social/group context of students (i.e., whether they are alone, or in groups of varying sizes). We study differences in the dwell time and transition behavior of students under the presence of such group contexts, at places with varying semantics within the campus (e.g., class related spaces such as seminar rooms, causal places such as the food court, or gym, or places that are designed for group work), in Section 3.2.2.

While the above group detector captures instantaneous groups, by observing such group formations over time, we define and extract the strength or intensity of the social relationship between pairs of students (referred to as the "tie strength"), in Section 3.3.1. We observe, in Section 3.3.2, that student pairs with *strong* tie strength share greater demographic similarity as opposed to random pairs of students which provides validation to our definition. Additionally, I also demonstrate that by accounting for properties extracted from such a *physical network of students*, analytics applications such as anomaly detection can benefit – in particular, I run a state-of-the-art anomaly detection technique, Local Outlier Factor [20] on longitudinal mobility data to detect anomalous locations and times. We see that while volume-based mobility features are useful in detecting only high-intensity anomalies (which cause a large increase or dip in people showing up at a specific place and time), social-network based features are capable of detecting both high and low

intensity anomalies (in Section 3.4).

3.1 Indoor Mobility Dataset

The dataset used in this work includes the location traces of individuals residing or visiting our university campus. The Singapore Management University (SMU) is located in downtown Singapore, and comprises approximately 10,000 students (undergraduate and graduate) and 1,500 staff. The university has no on-campus residential facilities; hence, all campus inhabitants commute to/from their residence. The university comprises 7 academic buildings, 1 administration building and an underground 'concourse' that acts as a publicly-accessible connector between the academic buildings.

The indoor location data is generated using a WiFi fingerprinting-based indoor location system [83], which has been operational on the campus for over 3 years, and which covers all the publicly-accessible parts of the 7 academic buildings and the underground concourse. The location system uses fingerprint measurements taken at *landmarks*: with modest exceptions (to accommodate irregular building layouts), landmarks are spaced 3 meters apart.The WiFi-based system utilizes the RSSI readings, from each WiFi-enabled device resident on campus, to compute the device's *medium-grained* indoor location, achieving a median accuracy of 6-8 meters (2-3 landmarks). Because of this medium-grained location tracking, we often express the location coordinates of each user at *section-level* granularity: a section typically corresponds to a collection of landmarks, and represents a logical partitioning of a building floor (e.g., a classroom, a group-study (meeting room), a food outlet, etc.).

New location estimates are generated once approx. 5-10 seconds (2-3 minutes prior to 2017). To focus primarily on the personal mobile devices of regular campus residents, we filter out (i) devices such as laptops and desktops (that exhibit only sporadic, intermittent mobility) and (ii) devices that belong to transient cam-

pus visitors (we require the device MAC address to be seen on campus at least for over 10 minutes over a day). Note that, due to the growing trend for devices to perform MAC address randomization when in a disconnected state, we effectively filter out those devices that do not connect to our WiFi network. Because we are using AP side measurements to compute locations, this approach captures all Wi-Fi connected client devices, incurs no additional energy overhead, requires no installation of additional software and thus eliminates resulting selection biases.

To filter out non-phone traces from our location dataset, we use the following heuristic: we consider a location to be *transient*, if a device had spent less than 5 minutes at any location. A device is likely to be a laptop if it *teleports*, i.e., it moves between places but shows no intermediate *transient* locations. For each unique device in the location trace, we computed its '*teleport ratio*, i.e., the the number of days the device was seen to 'teleport' divided by the total number of days it was observed on campus. Through empirical analysis, we found that, at a ratio of 0.3, the misclassification of known mobile devices (the 1468 devices that that registered testbed participants had explicitly provided) was only 0.8%. Accordingly, for the rest of this section, we considered only the subset of devices whose teleport ratio is lower than 0.3.

3.2 Influence of Social Groups on Indoor Mobility Behavior

We used the following three steps of analysis to determine the differences between individual and group behaviour: (1) application of a group labelling algorithm, (2) statistical hypothesis testing to identify differences, and (3) sensitivity analysis to determine the robustness of our results. We describe below each step in more detail.

3.2.1 Methodology

Step 1: Applying a Group Labelling Algorithm. With the location traces for the entire Fall 2014 semester, we next applied a group labelling algorithm to segregate the data into multiple mutually exclusive partitions – solo individuals and groups of various sizes (see below). This is not an easy task as our dataset does not have any explicit group or individual labels added by either the data collection process or by the users. As stated earlier, we used the *GruMon* system developed by Sen et. al [112] to do this partitioning.

GruMon extracts key features such as dwell time and place transitions from location streams, and declares a set of people as belonging to the same group if they have high feature correlations. It is pre-trained using a dataset of more than 250+ mobile users collected in other urban spaces, and is reported to have a $>$90% precision and $>$80% recall in detecting groups.

We provided *GruMon* with the entire Fall 2014 location trace and it partitioned the dataset into 4 partitions; (1) *Solo*: a partition containing just individuals moving by themselves, (2) *Small*: a partition containing groups with memberships of just two to three people, (3) *Medium* a partition containing groups with memberships of four to seven people, and (4) *Large* a partition containing groups with eight or more people. Each partition is mutually exclusive and contains the location traces for every person in that trace as long as the invariant for that partition holds. For example, *Solo* contains the trace for Person A as long as Person A is by themselves. The moment Person A becomes part of a larger group (say a group of size 2) and *GruMon* detects that, all of Person's A's traces (as long as they in the group of size 2) will appear in the *Small* partition only. **Note:** The group sizes were chosen for the following reasons: *Small* was chosen to represent the common case of students hanging out with their close friends, *Medium* was chosen to represent the common project group sizes we observed on the campus, while *Large* was chosen to represent larger groups that formed as part of class or extra curricular (CCA) activities.

Label	Membership Size	Likely Interaction Type
Solo	1	By Themselves
Small	2–3	With Close Friends
Medium	3–7	In Project Groups
Large	>7	Class or CCA Activities

Table 3.1: Output Partitions From *GruMon* Used In Section

Table 3.1 summarises the differences in the output of the group labelling process.

Step 2: Hypothesis Testing to Identify Differences. With the labelled outputs from *GruMon*,we explored the following the key hypothesis whether groups show significantly different mobility traces from individuals.

Note: when we say "Do groups do X", we actually refer to the aggregate behaviour of the individuals who make up that group. In particular, we show that individuals who are by themselves behave differently from individuals are part of a group – at an aggregate level. We do not perform any individual analysis comparing the behaviour of a specific individual when they are by themselves as compared to when they are in groups.

For the analyses in this section, we used the location trajectory data from 880 active users on the testbed (who were consistently observed on campus over the entire observation period). For every result we obtained, we conducted Kolmogorov-Smirnov tests (KS test), with alpha set to 0.05, to determine if the results were statistically significant. We used the KS test as it does not assume any underlying distribution pattern and is fairly robust to outliers.

Step 3: Determining the Robustness of the Results: An inherent problem with our approach is the errors caused by the labelling process using *GruMon*. The *GruMon* authors report a 91% precision and 82% recall. However, these values could result in 10% of the group data containing individuals and 20% of the solo data containing groups, thereby potentially invalidating the conclusions from our hypothesis testing.

To address this issue, we systematically subjected the *GruMon* output to various levels of random noise to determine robustness. Specifically, we flipped a random

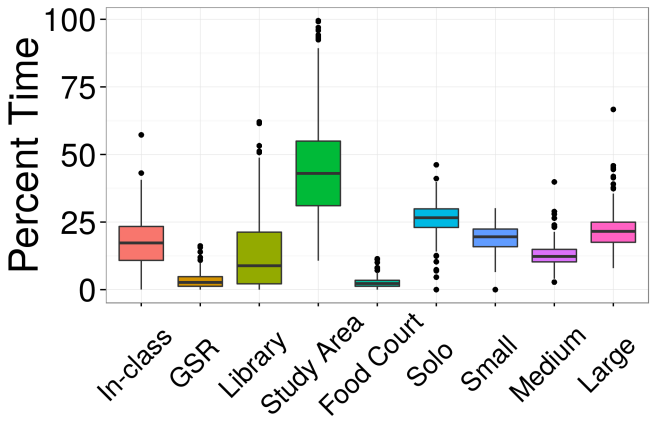
percentage, using values of 1%, 5%, 10%, 15%, and 20%, of the output of *GruMon*; i.e., if an element belonged to a group, it was now marked as Solo and groups (Small, Medium and Large) were now .

We then tested these new "noisy" distributions, using the KS test, to determine if these "noisy" distributions were significantly different from the original data partition. We also compared the hypothesis analysis of "noisy" versions of the Solo and group data with the original Solo and group data to see if the results had changed significantly. In Section 3.2.3, we only present results which were robust (i.e., there was no significant change observed) even under all five different noise levels.

3.2.2 Results

In this section, we present our observations on whether and how being in a group can affect the mobility pattern of individuals. In particular, we investigate two aspects of mobility; (1) dwell times and (2) semantic place transitions between the four distinct campus location earlier – *GSRs*, *Study Areas*, *Library*, and *Food Court*.

Time Spent in Groups



Each boxplot shows the avg, 25 and 75 percentile, and stdev. values.

Figure 3.1: Time spent in various locations and in groups

Before we present the results of our hypothesis, we first present an important

initial result. Namely, we show that users spend a significant proportion of their time in groups—and thus, reiterate the importance of studying behavioural changes in the group context. We observed the amount of time the users spent in various locations on campus, over the fall semester – differentiated by the group membership (Table 3.1).

Figure 3.1 shows the time spent by all the users in our location trace in four distinct campus locations – *GSRs* (Group Study Rooms that can accommodate up to 8 people), *Study Areas* (Public places on campus where student can congregate for project and study work), *Library* (study areas in the library), and *Food Court* (the main food court on campus). In addition, we show the percentage of time a user spent by themselves, and in *Small*, *Medium*, and *Large* groups, respectively. We make the following observations: (1) students spend a large portion of their time on campus, 9.8 hours on average over the Fall term, and out of 9.8 hours, they spent 84% of time outside classes (45.29% in *Study Areas*, 14.50% in *Library*, 2.56% in *Food Court*, 3.48% in *GSRs*, and 18.17% in multiple other areas around the campus), (2) excluding class times, students spent 64.62% of their time in groups (in any size of group), indicating the importance of understanding group behaviour, and (3) students are engaged in various social activities with different groups sizes. Out of the total time spent in groups, they spent 23.43% of their time in *Small* groups, 16.31% in *Medium* groups and 24.88% in *Large* groups.

Dwell times

We calculated the dwell times at all four places for every partition listed in Table 3.1. We first conjecture that the groups tend to stay for longer due to the increased interactions among the group members. For example, in the case of the *Food Court*, an individual's only objective is to consume the meal whereas when accompanied by a group of friends, conversations could last longer, prolonging the usual dwell time. Hence, we hypothesize the following:

HYPOTHESIS 1: *Groups Stay at Places Longer*

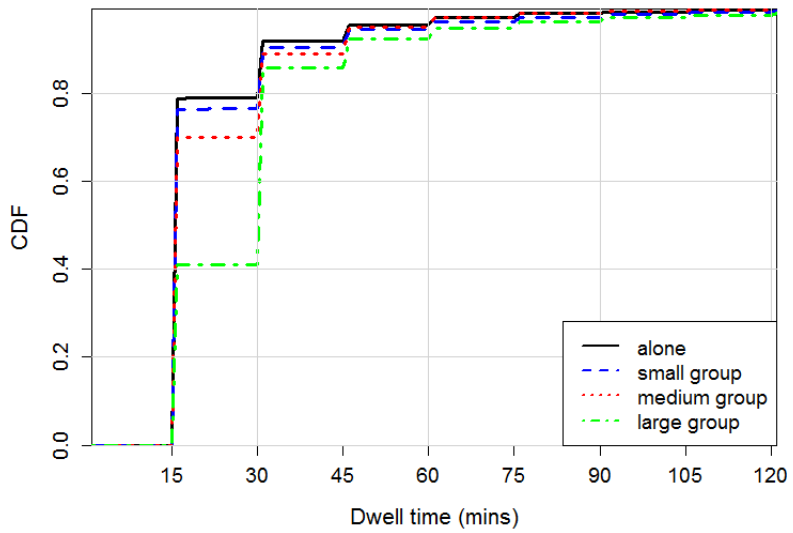


Figure 3.2: Comparisons of Dwell Times at a Food Court (Hypothesis 1)

Overall, we found that individuals and small groups behaved similarly with no significant differences (p-values of 0.9986, 0.9998 and 0.9992 at *GSRs*, *Food Court* and *Library*, respectively, from the discrete KS test), while medium-sized and larger groups tended to stay longer at most locations with the differences being statistically significant (p-value $<$ 2.2e-16). Figure 3.2 shows the dwell time CDF at a food court on campus for all four partitions. The difference is more drastic for *GSRs*; only 10% of *Large* groups spend less than 15 mins whilst almost 70% of *Solo* and *Small* spend less than 15 mins in *GSRs*. The *GSR* result also shows greater variability in the 15 to 120 min range after which all four configurations tend to merge exhibiting a long tail. In the case of *Study Areas*, 10% of *Solo* and 20% of *Large* stayed on for more than 2 hours, with the differences then becoming weaker. Note that the dwell times are shown in 15 minute increments as *GruMon* [112] only detects groups over 15 minute intervals.

The *Library* had the lowest difference between the partitions. This could be due to the intrinsic nature of the library; although groups may visit the library together, the activity of studying or reading in the library, by itself, is not a group activity and hence limits the conversations or interactions amongst the group. This results in the dwell time distributions of individuals and groups at the library to be similar.

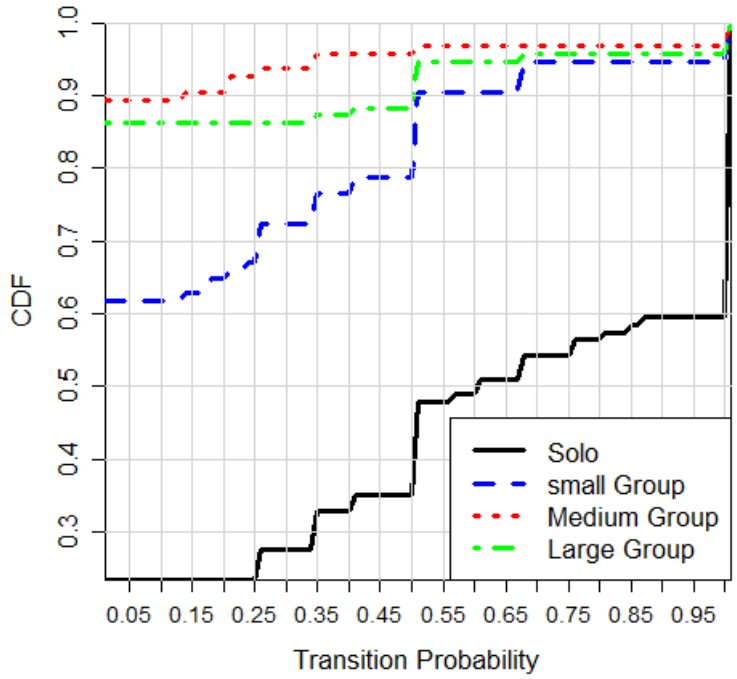


Figure 3.3: Transition Prob. from *Study Areas* to *Food Court* (Hypo. 2).

Semantic place transitions

Next, we consider the likelihood of groups versus individuals in making a transition from one place to another. For every pair of places (*p1* and *p2*), we calculated probabilities that a user moves from *p1* to *p2* based on past history calculated separately for when that user was in a group and when they were alone. Thus, for each place pair, we have two sets of probability values, one for group and one for individual, for all users. We used these two probability value sets serve as inputs to a KS test which showed that our results was significant. Note that, for simplicity of explanation, we only considered two cases above (groups and individuals). In reality, we computed probabilities for all four partitions. With these four probability sets, we hypothesise the following:

HYPOTHESIS 2: *Larger Groups Make Less Transitions*

As hypothesized, we observed significant differences in the transition probabilities between different partitions and pairs of semantic places (p-values $<$ 2.2e-16). Figure 3.3 shows the probability of transitions from *Study Area* to *Food Court*. We

observe, for example, that, at a transition probability of 65%, at least half of *Solo* had that probability of transitioning whereas only 18%, 5% and 2% of *Small*, *Medium*, and *Large* had a 65% probability of transitioning. Note that we model transition likelihood through an independent and identically distributed process and compute the transition probability over 15 minute windows, in isolation.

Overall, we consistently observed that *Medium* and *Large* moved significantly less than *Solo* and *Small* (p-value of the KS-test between *Medium* and *Small* was 0.001506). This means that although larger groups tend to spend more time together, they are less likely to move to different places. To invalidate an alternate explanation that the campus had fewer places that could accommodate larger groups (thus creating a negative for large groups to move), we ensured that all four places considered did in fact have the capacity to host at least *Medium* groups. Also, we note that the transition likelihood does not drop monotonically with increasing group size – for e.g., *Medium* is the least likely to transition from Study Areas to the Food Court, even less than the *Large* group. However, the difference between these two groups are not statistically significant ($p-$value = 0.885), whilst the difference between *Solo* and the other groups are consistently statistically significant.

3.2.3 Sensitivity Analysis

Finally, we test the robustness of our results to the underlying errors in the group detection algorithm.

Because the group detection algorithm only has a precision of $\approx 90\%$ and a recall of $\approx 80\%$, there will be errors in the group detection labelling process. This means that a fraction of group data might actually contain solo data, and vice versa. To determine the robustness of our results even with these errors, we injected additional random errors into the output labels and tested if the KS test still outputs the same conclusions between the *noise-added group data* and the *noise-added solo data* distributions. More specifically, we introduced errors by flipping a designated

percentage of errors	D-statistic	p-value
Original	0.1295	<2.2e-16
1%	0.1268	<2.2e-16
5%	0.1186	<2.2e-16
10%	0.1076	<2.2e-16
15%	0.1034	<2.2e-16
20%	0.0963	<2.2e-16

Table 3.2: KS results bet. the noise-added group and solo data

percentage of labels (chosen randomly using 1%, 5%, 10%, 15%, and 20% data flip levels, over a single run) from groups to solos or the other way. We chose this error range to match the various recall and precision levels of our group detector *GruMon*.

Table 3.2 shows the p-values and D-statistic values from the KS test between the original group dataset and solo dataset, as well as the noise-added group datasets and noise-added solo datasets with different error percentages, for Hypothesis 1. The table shows that the KS test still accepts the hypothesis even with 20% errors introduced; p-values for all the cases are much lower than 0.05, the well-known threshold of p-value for acceptance. Also, there is no meaningful difference across D-statistic values, which means the distances between the two distributions do not change noticeably with different levels of injected errors.

3.3 Constructing a Physical Social Network

In this section, we explore the second key question; can social ties between people be extracted from longitudinal observations of their mobility? We first describe the methodology to extract the strength of tie between pairs of students and later, in Section 3.3.2, validate the goodness of this measure. Herein, we contrast the demographic similarity (i.e., in terms of school and year of study) between pairs uncovered by the tie strength measure against such similarity of random pairs due to the unavailability of exact knowledge of the friendship network of the large number

D	Set of all spatio-temporal occurrences
ϵ	Set of all episodes
$e.A$	Episode e's actors
$e.s$	Episode e's location
$e.t^-$	Episode e's start time window
$e.t^+$	Episode e's end time window
$e.\delta$	Episode e's duration
$e.w$	Episode e's weight
$e.w_{p-s}$	Episode e's spatial precision
$e.w_{p-t}$	Episode e's temporal precision
$e.w_{u-s}$	Episode e's spatial uniqueness
$e.w_{u-t}$	Episode e's temporal uniqueness
h_s	Space s's spatial uniqueness
h_t	Time window t's temporal uniqueness

Table 3.3: List of notations used in defining tie strength.

of users in the dataset. Later, in Section 3.4, we provide further validation on how considering such tie strength and resulting network properties can be utilized in practical applications.

3.3.1 Definitions

Consider a database D of spatio-temporal events. Each $d \in D$ is a tuple of the form (a, s, t) where a is the identity of the actor (user), s is the location, and t is the discrete time window of a spatio-temporal occurrence. The spatio-temporal co-occurrence of two (or more) actors is defined as an *episode*. The set of all episodes is denoted by ϵ similar to Lauw et al. [77]. I extend and modify the original definitions for the physical space as follows.

Episode Definition

Our definition of *episode* follows from the definition of *events* in [77]. Note that the second condition is a discretized version of the original definition (where time was considered as continuous).

A spatio-temporal episode is a subset of tuples, $e \subseteq D$, meeting all of the following conditions.

1. tuples are of the same location:

   i.e., $\forall d_i, d_j \in e, d_i.s = d_j.s$

2. tuples are present together during the same time window:

   i.e., $\forall d_i, d_j \in e, d_i.t = d_j.t$

3. tuples involves at least two actors:

   i.e., $|\{d.a | d \in e\}| \geq 2$

4. each event is maximal:

   i.e., for any event $e' \subseteq D, (e' \subseteq e) \vee (e \subseteq e') \Rightarrow (e' = e)$

In effect, an episode denotes a contiguous time segment of collocation between two individuals.

Episode Weight

Similar to [77], I assign weights to episodes along two spatio-temporal properties, (1) precision and (2) uniqueness, where a higher weight represents both a higher likelihood of an interaction and also a higher strength between any two actors involved in the episode. Due to the repetitive patterns of student behavior in a campus setting, where students meet with other students multiple times across weeks, I introduce two additional properties: (3) frequency and (4) durability.

Precision: I modify the definitions of spatial and temporal precision in [77] to account for the differences in the two contexts considered (physical spaces vs. cyberspace).

Spatial Precision of an episode measures the closeness of the interaction between actors based on the granularity of the space. In [77], the granularity of the space reflects the size of the space – meaning that the larger the space, less precise is the underlying relationship between the actors. As the locations considered here are at equal logical granularity, the notion of closeness is expressed in terms of the number of actors involved in an episode. For example, the spatial precision of an

episode at the *Seminar Room* with 50 students taking a class should be lower than that of an episode at the *Meeting Room* involving three students alone. Hence, I define the spatial precision $e.w_{p-s} \in (0, 1]$ of an episode as the inverse of the number of actors in the episode, normalized by the maximum such value as in Eq. 3.1. Here, $e.A$ is the set of actors involved in the episode e.

$$e.w_{p-s} = \frac{\frac{1}{|e.A|}}{max_{e' \in \epsilon}\{\frac{1}{|e'.A|}\}} \tag{3.1}$$

Temporal Precision measures the confidence in considering two actors to have been part of a meaningful episode, and not co-occurred at the same time, merely by chance. Hence, longer the duration together, higher the chances are that two actors were meeting each other. I define the temporal precision $e.w_{p-t} \in (0, 1]$ of an episode e as its total duration (δ) normalized by the maximum such duration over all episodes as given in Eq. 3.2.

$$e.w_{p-t} = \frac{e.\delta}{max_{e' \in \epsilon}\{e'.\delta\}} \tag{3.2}$$

Uniqueness: I extend the definitions of uniqueness from Lauw et al. [77] by additionally factoring in the popularity (or rarity) of a place or time window, over all occurrences (and not co-occurrences alone).

Spatial Uniqueness measures the rarity of any episode occurring at location s. An episode occurring at a *unique* location, intuitively, is a stronger indication of a stronger relationship between the actors involved. In Eq. 3.3, I represent the spatial uniqueness of an episode, $e.w_{s-u} \in (0, 1]$, based on the number of other episodes that have occurred at the same location s and an additional spatial heat factor (h_s). For example, if a location is popular among all students, but students rarely meet as a group there, then the uniqueness of the episode is high. Conversely, if an episode occurring at a location is rare, but it is also rare for students to visit this location (individually, or as groups) in general, then the combined uniqueness of an episode

occurring at this location is weighed less. A nonzero minimum value is assured by counting episodes other than itself.

$$e.w_{p-u} = (1 - \frac{|e' \in \epsilon | (e' \neq e) \wedge (e'.s = e.s)|}{|\epsilon|}) \times h_s \tag{3.3}$$

I define the spatial heat factor (or the uniqueness of the space over all visits) $h_s \in (0, 1]$ as the number of total visits to the location s normalized by the maximum such visits to any location as in Eq. 3.4.

$$h_s = \frac{|d \in D | d.s = e.s|}{|D|} \tag{3.4}$$

Temporal Uniqueness is defined in a similar fashion to spatial uniqueness. The fact that an episode occurs at a *rare* time window indicates a stronger meaning of the underlying relationship among the participants. The temporal uniqueness $e.w_{t-u} \in (0, 1]$ is defined as the count of the number of episodes that overlap in time with the episode itself (Eq. 3.5). Additionally, I account for the temporal heat factor (h_t) which represents the rarity of the time window across all visits.

$$e.w_{t-u} = (1 - \frac{|e' \in \epsilon | (e' \neq e) \wedge (e'.[t^-, t^+] \cap e.[t^-, t^+] \neq \emptyset)|}{|\epsilon|}) \times h_t \tag{3.5}$$

I further define the temporal heat factor (or the uniqueness of the window over all visits) $h_t \in (0, 1]$ as the number of total visits during the time window t normalized by the maximum such visits during any time window as given in Eq. 3.6. Here t^- and t^+ are the start and end times of the window t.

$$h_t = \frac{|d \in D | d.t \in e.[t^-, t^+]|}{|D|} \tag{3.6}$$

Finally, the combined episode weight $e.w \in (0, 1]$ is expressed in the multiplicative form as in Eq. 3.7.

$$e.w = e.w_{s-p} \times e.w_{t-p} \times e.w_{s-u} \times e.w_{t-u} \tag{3.7}$$

Strength of Ties

It is established that a tie exists between two actors a_x and a_y if both a_x and a_y are actors of at least one common episode, and the cumulative weight over all episodes is at least $min - link - weight$. I define the weekly strength of tie between a_x and a_y ($< a_x, a_y > .w_j$) as the summation of the frequency-factored episode weights over the week j.

Frequency: Given that the two actors meet f_i times per day i during the week j, then we define the daily frequency $< a_x, a_y > .f_i \in (0, 1]$ as the total number of episodes during day i normalized by the maximum such frequency over all pairs of actors during the day. For simplicity, we assume a maximum frequency of 96 which is the maximum possible number of 15 minute episodes over a day.

Then, the cumulative weight between actors a_x and a_y for week j is given by Eq. 3.8.

$$< a_x, a_y > .w_j = \sum_{e \in \epsilon_{a_x, a_y, i}} (e.w_i \times \frac{f_i}{96}) \tag{3.8}$$

Durability: Further, given that the actors meet on c different days over the week j, then the durability $< a_x, a_y > .c_j \in (0, 1]$ is defined as the number of days in the week where the actors had at least one interaction episode, normalized by the maximum possible days (7) in a week that any pair of actors can meet, as in Eq. 3.9.

$$< a_x, a_y > .c_j = \frac{c}{7} \tag{3.9}$$

Finally, the link weight between two actors a_x and a_y over all weeks across the term (T, a time period of observation) is defined as in Eq. 3.10.

$$< a_x, a_y > .l = \sum_{j \in T} < a_x, a_y > .w_j \times < a_x, a_y > .c_j \tag{3.10}$$

3.3.2 Validation of the Tie Strength Measure

For the studies presented subsequently, location data pertaining to the academic period between August to November, 2014 is considered. The GruMon [112] algo-

rithm is used to detect the group interaction/movement episodes over this dataset, which thus provides the (start, end) times for a particular group episode, as well as the members (the $participantIDs$) belonging to that group episode. In addition to the movement data (of an average of 10,000+ distinct devices observed daily), the LiveLabs data also contains additional non-personally identifiable demographic information (specifically, the gender, age, school-affiliation and year of study) of a smaller set of approx. 1700 opt-in LiveLabs participants (all undergraduate students) which I use for validating the correctness of unobtrusively mined ties.

Determining $min - link - weight$

In an attempt to discard possible cases of *accidental strangers* being considered as genuine ties, I retain only the top 1/3 of the user pairs (links). From the indoor movement data, I choose $min - link - weight$ as the 67^{th} percentile of tie strength which is 3.352175e-05. This retained 143,118 ties between the 1,711 students. Further, I consider the top 10% of the retained ties as *strong ties* which I believe represent meaningful student pairs. We found 14,312 such pairs and construct a socio-physical network with the students as nodes and the tie strengths as undirected, weighted edges.

Further, to validate the goodness of the measure of ties described here, I compared the demographic similarity between pairs randomly chosen against the strong pairs retrieved by this technique. Table 3.4 lists the proportion of pairs of students who belonged to the same school, year of study, or both, respectively, averaged over ten runs. A key observation here is that the strong pairs consistently show greater similarity in comparison to the random pairs. A χ^2-test confirms that the distributions of the random pairs and strong pairs are statistically dissimilar (with p-value $< 2.2e - 16$ at 5% significance level).

Demographic Feature	Random Pairs	Strong Pairs
School	28.20%	39.83%
Year of Study	39.21%	56.35%
Combined	12.10%	25.53%

Table 3.4: Demographic similarity between student pairs.

3.4 Additional Validation: Anomaly Detection

In this section, as additional form of validation as to how considering social properties of mobility is useful, I consider the application of anomaly detection in indoor spaces. To this end, I introduce three classes of spatio-temporal features and summarize the overall approach for detecting anomalies (or outliers) as follows.

3.4.1 Features

Occupancy-based Features: The total count of users visiting specific locations over fixed time intervals as the *occupancy* feature (similar to the *magnitude* feature in Nayak et al. [88]).

Group Interactions based Features: Using GruMon[112] for detecting groups of students based on the time they spend together and their coordinated transitions between locations over discrete time intervals (e.g., 15 mins), I specify two features that capture the instantaneous group-membership properties: (1) the *group occupancy count* which is the total count of groups, and (2) the *average group size* which is the average size of the groups, visiting specific locations over specific time intervals.

Tie Strength based Network Features: In contrast to the group interaction-based features which considers groups of students appearing currently (with no information on the history of the group's previous visits), the tie strength based features aim to capture the historic relationships amongst those students present. Here, it is assumed that knowledge of tie strength is known a-priori (for example, from observations from past terms in the university setting). I consider four derived features: (1) the *proportion of strong ties*, and (2) *density*, (3) *diameter*, and (4) *mean*

degree of the sub-network of the pairs present, at the location and time of interest.

3.4.2 Detection Methodology

Given that the strength of tie between any two users is known, I take the following steps in detecting outliers, which I infer as anomalies or events.

Step 1: Feature extraction - For every 15 minute time window over the period of observation, the seven features described are extracted. The three network properties are extracted by considering the sub-network of user pairs present at the specified location during the specific time window.

Step 2: Compute outlier scores - To account for *day of week* effects, I consider each day of the week, separately (for a 14-week term, each day appears exactly 14 times). As such, there are 14 occurrences of any 15 minute window during a 24 hour day, for each day of the week (for example, there will be 14 such 7:15 PM to 7.30 PM windows across 14 Mondays). I compute the LOF score [20] for each combination of day of the week and time window, for each of the features, independently. I choose LOF as it is a *local*, density based algorithm for detecting outliers – this is due to the observation that in the campus setting, not all Mondays (say) are similar across the term; for example, there could be 4 weeks of Skating classes, 2 reading weeks and 8 other regular weeks. Hence, there is an inherent clustering of the weeks and the main objective here is to detect outliers that are one-time events and not repeating or regular events.

Step 3: Compute average scores - The features are extracted over 15 minute windows – however, events can last from as little as 30 minutes to as long as an entire day. As such, I compute the harmonic mean score of a particular day over specified time intervals. Other choices for the average were also considered (including mean and geometric mean) and the results were found to be fairly robust.

Step 4: Detecting outliers - In this final step, for each (window of observation, week of the day) pair, for each feature considered, I generate a relative ranking

DoW	Time	Event	Intensity	Occupancy	Group Occ.	AGS	PST	Diameter	Density	Mean Degree
Thursday	11 AM - 2 PM	Peace Day (18th Sep)	High	1 (6.98)	1 (2.17)	4 (1.13)	3 (3.52)	11 (0.99)	6 (1.08)	1 (10.07)
Thursday	11 AM - 2 PM	SIS Day (16th Oct)	High	2 (1.31)	2 (1.97)	5 (1.16)	9 (1.51)	7 (1.11)	11 (1.00)	4 (1.35)
Friday	11 AM - 5 PM	Vivace (22nd Aug)	High	1 (12.79)	1 (3.70)	6 (1.42)	3 (1.81)	6 (1.27)	5 (1.27)	1 (10.40)
Monday	7 PM - 8 PM	Fencing Clinic (1st Sep)	Low	8 (1.06)	6 (1.11)	10 (1.05)	9 (1.00)	2 (1.62)	9 (1.03)	8 (1.02)
Monday	7 PM - 9 PM	SMU Flare (25-Aug)	Low	9 (1.04)	7 (1.07)	5 (1.15)	11 (1.03)	9 (1.06)	12 (0.99)	4 (1.23)
Tuesday	7 PM - 8 PM	VPH Audition (2nd Sep)	Low	9 (0.97)	9 (0.98)	10 (0.96)	2 (1.62)	6 (1.00)	5 (1.01)	3 (1.17)
Tuesday	12 noon - 8 PM	Charity Piano (9th Sep)	Low	10 (0.97)	3 (1.19)	8 (0.97)	6 (1.02)	5 (1.01)	6 (0.99)	2 (1.45)
Thursday	3 PM - 6 PM	Martial Mayhem (4th Sep)	Low	4 (1.14)	4 (1.08)	2 (1.31)	4 (5.89)	6 (1.31)	5 (1.10)	1 (2.10)
Friday	4 PM - 8 PM	SMUX Skating (31st Oct)	Low	3 (1.35)	10 (1.11)	11 (1.02)	8 (1.25)	8 (1.34)	11 (1.01)	6 (1.14)

Table 3.5: List of known events at the *T-Junction* and relative ranking of outliers during term 1. DoW – Day of Week, AGS – Average Group Size, PST – Proportion of Strong Ties Present.

(and scores) for the respective 14 days. An $LOF \approx 1$ is considered normal, and $LOF \gg 1$ is considered an outlier. For the purpose of our evaluation, I declare that a feature has detected an *event* under the two conditions: if (1) the relative ranking of the (window of observation, week of the day) pair is 1 or 2 (in effect, we are extracting the top-2 days with highest outlier scores) and (2) the average outlier score is ≥ 1.2.

3.4.3 Results

Next, I evaluate the performance of the different features described in section 3.4 in detecting *known* events of differing intensity levels (both *high* and *low*) in the SMU campus. First, I retrieved a list of known events from an internal events portal in campus. The portal contains listings of various events (including academic talks/seminars, student club activities and campus-wide fairs) that took place on campus. These events were collected by a backend-infrastructure that interfaced with a variety of enterprise event-management systems/portals to retrieve such events. Each individual event is associated with a tuple $\langle eventID, \{locationID\}, startTime, endTime\rangle$–this corpus of events provides the *ground truth* needed to validate the event detection technique and the use of social network features. Then, I extracted the days of the week the events took place and their start and end times. These <day of the week, time intervals> serve as the windows of observation and are listed in Table 3.5. As such, I am interested in understanding which features are able to detect which types of events. For the analyses that follow, I focus on events that happened at the *T-Junction*, which is a

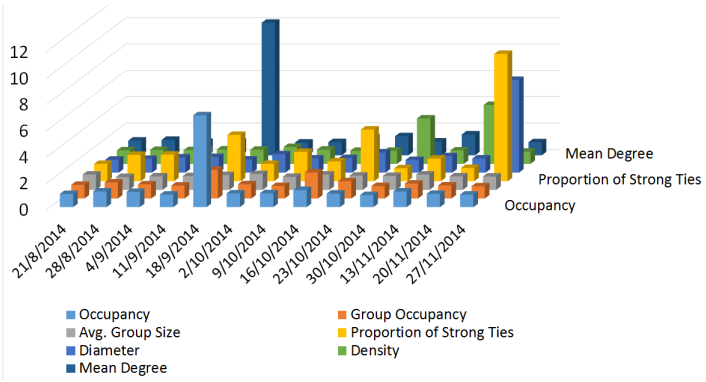


Figure 3.4: Outlier scores for time period between 11 AM and 2 PM on different Thursdays. Two events: Peace Day and SIS Day happened during this window.

popular location in the campus where both large campus-wide annual events as well as smaller events such as martial arts training sessions are organized. Note that the portal is not a comprehensive list of events, but only a source of partial ground truth.

Results: In Table 3.5, the relative rankings of outlier scores given by each feature independently, and the mean outlier scores (within brackets), for nine known events are tabulated. The shaded cells correspond to those events declared by each feature as an outlier (based on the combination of relative ranking and magnitude). The following key points are observed:

1. *High intensity events:* As observed in columns 5-6, as anticipated, the *occupancy* feature outperforms the tie-strength based network features in detecting *high* intensity events. It is clear that the *Group Occupancy* is also able to detect all three high intensity events.

2. *Low intensity events:* Whilst neither the *Occupancy* and *Group Occupancy* features were able to detect any of the low intensity events, the *Mean degree* feature is able to uncover both a mix of high and low intensity of events.

3. *Feature selection:* I also point out that the different network features are able to detect certain small events (such as CCA training sessions), exclusively. For example, the average group size, the proportion of strong pairs and the diameter, each detected a different small event.

Metric	Occupancy	Group Occ.	Mean Degree	Best Recall Combination
Precision	3/10 (30%)	3/10 (30%)	4/8 (50%)	7/38 (18.42%)
Recall	3/9 (33.33%)	3/9 (33.33%)	4/9 (44.44%)	7/9 (77.77%)

Precision is expressed as $\frac{a}{b}$ where a is the number of outliers that were true outliers, and b is the total number of detections. Recall is expressed as $\frac{c}{b}$ where c is the number of known events, which is equal to 9 during Term 1.

Table 3.6: Summary of event detection performance.

For illustration purposes, I refer to Figure 3.4. The plot shows the outlier scores of the 14 different Thursdays, as scored by seven features. The average outlier scores are on the y-axis and the different Thursdays are on the x-axis. It is seen that the three features, Occupancy, Group Occupancy and Mean Degree peak on the 18th Sep (Peace Day). Note that two features, proportion of strong pairs and mean degree, identify 27th Nov as an anomaly– this could have been an actual event that was not recorded in the database. In Table 3.6, the precision and recall values of the *Occupancy* feature is compared against the best performing *Group interaction feature* (i.e., Group Occupancy) and the best performing *network feature* (i.e., Mean Degree). The network feature provided an 11% increase in recall along with a 20% improvement in precision. The best recall is obtained for the combination, operationalized as a logical OR, of four features (Occupancy/Group Occupancy, Mean Degree, PST and Diameter) at 77.77% – this is a 44% improvement over using Occupancy feature alone. However, this also results in a 12% drop in precision. Note that unlike in the results presented in Table 3.5, the criteria for declaring as a detection was that the average LOF score be ≥ 1.2 alone, without any conditions set on the ranking. As such, between the four features that provided the best recall, there were 38 time instances that were detected as outliers.

3.4.4 City-scale Anomaly Detection

In this section, I extend the derivation of social ties and the validation of its use in the application of urban event/anomaly detection through the use of the secondary data source on outdoor mobility.

MRT/Bus data: This consists of MRT and Bus trip tap card[1] data from 4 million unique commuters during the three month period between November, 2011 and January, 2012. The trip data covers 127 unique MRT stations and 4873 bus stops (covering 353 unique bus routes). Each trip record is of the form $\langle ID, tap-in-time, tap-in-station, tap-out-time, tap-out-station \rangle$.

I adopt a simpler form of the tie strength measure (primarily using the *frequency* factor) and limit the focus on a subset of the data (weekends only). The derivation of the ties of strength is outlined below followed by a discussion of preliminary findings.

Generating User Pairs: I posit that two people who share an underlying relationship (friend or family) are likely to travel together. I consider those trips only made during the weekends to eliminate trips made by co-passengers who follow similar work commute patterns during weekdays.

A pair of users are considered as a *common trajectory pair* if both users *tapped in* and *tapped out* together within 20 seconds of each other, at the same entry and exit stations, during the same trip. The *strength of tie* between any two users is defined as the total number of common trajectories normalized by the maximum such count across all user pairs, over the 26 weekend days during the period of November 2011 and January 2012. A total of 24,154,562 such pairs were discovered out of which 1,571,167 pairs had traveled together at least twice. I consider this subset of pairs as *strong pairs*.

Results: I discuss the findings from detecting outliers for a specific station, *Esplanade*, for which we have curated ground-truth of organized events overlapping

[1] http://ezlink.com.sg/

with our period of observation. The station provides easy access to a theatre which hosts performances and concerts, as well as a waterfront area where the public congregates socially on holidays such as the New Year's day. We expect that the outlier scores would be high for the volume feature (number of people tapping-in at the station) only during events that attract a large number of people, whereas in the case of network features, the outlier scores would also be able to detect low intensity events. We first consider the time window between 6 PM and 8 PM on Saturdays. We pick this time window as this overlaps with the starting times of most of the events at the Esplanade theatre. We note that each Saturday (and Sunday) hosted events/concerts at the Esplanade and our objective here is two-fold: to discover which of the events are *more* social and which features help to uncover such events.

The *proportion of strong ties* feature detects the largest anomaly on 12th November, 2011 with a high LOF score equal to 44. This particular day hosted an array of Indian cultural events at the Esplanade as part of an annual, week-long stream of cultural events. We also note that two other features, degree and diameter, also found this day to be a top-1 outlier day. These events are primarily attended by the minority Indian community, and thus do not necessarily cause a large change in tap-out volumes at the station. However, the tie-strength based features were able to detect this event.

In Table 3.7, we report the precision and recall values for independent features and the combination of features that offers the best recall (which was found to be Tap-in Count, PST, Diameter and Mean Degree together), respectively, for a total of 11 known events (Saturday - 6, Sunday - 5). An event was declared to be detected if the relative ranking of the outlier was ≤ 5 and the mean outlier score was ≥ 1.2. We observe that the best combination is able to offer a recall of 72.3% although at a greater loss of precision compared to the case of the campus setting.

Metric	Tap-in	PST	Diameter	Best Recall Combination
Precision	4/4 (100%)	5/10 (50%)	4/6 (66.67%)	8/30 (26.67%)
Recall	4/11 (36.36%)	5/11 (45.45%)	4/11 (36.36%)	8/11 (72.72%)

Table 3.7: Summary of event detection performance with transport data.

3.5 Discussion Points

This chapter makes two key conclusions: (1) that social groups do influence how people move around, even within indoor settings, (2) and that it is possible to infer such contexts based on passive mobility data. Here, I outline some of the current limitations and future possibilities for this line of work.

Current Limitations: One key observation from the analyses is that the achieved precision of outlier detection is rather low (e.g., $\approx$ 20%). Many time instances are detected as outliers, however there are no corresponding ground-truth labels in the events data resulting in a large number of False Positives. However, we note here that the events data set does not include an exhaustive list of *ALL* events on campus, or in the city, as generating such a list is not practical. Hence, we caution that the precision values reported here are likely under-estimates. Similarly, it is also possible for the corresponding recall performance to be poorer than what we report here, for the same reasons.

Moreover, in the analyses presented in this chapter, we specically looked for outliers over specic durations (which were known in advance to contain events). In practical scenarios, we will need to develop appropriate expanding/ sliding window-based techniques to not only detect events, but also delineate their $\{$start, end$\}$ times.

Possible extensions: In this chapter, I described a composite tie strength measure that takes into account varying aspects of co-location, longitudinally, whose benefit is demonstrated through the application of event detection. One possible im-

provement over the current measure would be to consider the *semantics* of the tie, or *nature of the relationship* (e.g., close friends, gym buddy, etc.), additionally. It is possible that such semantic features might have different levels of use in discerning different types of events (e.g., academic, social, CCA based etc.). For example, traveling companions aboard buses (using the same smart-card dataset along with demographic identifiers in addition to the tap-in, tap-out transactions) of varying tie compositions could be detected and observed to generate lifestyle and activity indicators. For instance, two children accompanied by an adult (with the tie semantic being parent–child), visiting the Esplanade station, may be a high-confidence indicator that the group is attending a children-show at the theatre as opposed to a special concert at a popular night club.

Chapter 4

Exploiting Indoor Mobility for Smart Campuses

The ability to predict an individual's future location (or indirectly, her movement behavior) is a key enabler of many mobile computing applications and services. In the past decade, there has been an explosive growth in the availability of large-scale mobility datasets (e.g., [133, 91, 26, 18, 61, 134, 66, 64]), obtained via technologies such as GPS (e.g., [133, 91, 26]), cellular records (e.g., [18, 61]) and WiFi logs (e.g., [134, 66, 64]), that capture both campus and city-level movement. Researchers have used such datasets to empirically establish some of the scientific underpinnings of human mobility, including the predictability levels (or routine nature) of daily movement and the strong correlation between physical movement and social tie strengths. Two of the most typical prediction tasks investigated by researchers [15]) include NPP (*Next Place Prediction*–i.e., where will the user move next?) and RT (*Residency Time*–i.e., how long will the user stay at the current location?). Such work fundamentally looks to uncover a variety of latent mobility patterns, and can enable a variety of predictive applications, such as anticipatory temperature control of a home [110] or proactive delivery of relevant alerts (by digital assistants such as Amazon's Alexa™ or Google's Google Assistant).

In this work, I tackle a distinct question: *Likelihood of Future Non-Conformance*

(**LFNC**),–i.e., the odds that a user will not visit a routine place that she regularly visits. It is worth reinforcing, at the outset, the *distinctness* of our research question. Mobility prediction is principally about uncovering the underlying *routines* or patterns of a user, with the prediction accuracy being bounded by the inherent randomness (or predictability) of a user's movement behavior. Moreover, prediction algorithms focus on metrics such as minimizing average location error (i.e., the distance between the actual & predicted location coordinates). Instead, we embrace the fact that even the most predictable or routine user will, occasionally, diverge from such common mobility behavior, in an *apparently "random" or unpredictable* manner. For example, in a campus setting, a research group with a regularly scheduled meeting on Thursday afternoons will "suddenly" skip a meeting. We thus focus principally on the question: *How can we enhance the confidence of declaring, sufficiently in advance, that a user will not be visiting a particular location, based purely on historical traces of location data?* Unlike mobility prediction, *LFNC* is measured by a more binary outcome variable: will the user actually be at the most-likely predicted location or not?

Our research focuses principally on movement behavior in workplace environments (e.g., a university campus or office facility), and arises from the observation that user mobility in a workplace is largely a manifestation of underlying, often-routine activities. In particular, significant research [122, 121] has been conducted on inferring or understanding workplace behavior via routine, scheduled activities and calendar events (e.g., group meetings, research discussions and lunches). Our investigations are motivated by the realization that the ability to predict such (likely rare) cases of non-conformance may become increasingly important in an age of anticipatory services, where predicting the wrong context might lead to greater negative consequences (e.g., unwanted or misleading notifications by a virtual assistant) than simply declaring "I'm unable to predict".

We develop a methodology to demonstrate that such "random deviations" can, in fact, be predicted, with a surprisingly high degree of accuracy. More importantly,

we also explore the *lookahead capability* of such non-conformance prediction: i.e., we investigate the question "How far in advance can one reliably predict that a user (or users) will not be at a routine location?". Driven by past results that establish the strong social influence on mobility patterns, I shall investigate these questions by additionally incorportating the impact of peer movement behavior on the predictability of such non-conformance. In fact, our research is driven by the following two hypotheses:

- H1–*Temporal Correlation of Non-Conformance:* A user's propensity of deviating from a future routine location/activity pattern is correlated to the anomalousness of her current movement–if a user has been exhibiting anomalous movement patterns in the recent past, she is much more likely to deviate from her routine location/activity pattern in the future;

- H2–*Homophily of Anomalies:* In workplace environments, where users indulge in significant collaborative activities, anomalous movement behavior is often not isolated but *shared*: if a user's "friends" have been exhibiting non-routine movement as well, there is a significant increase in the likelihood that she will deviate from her future routine movement pattern.

4.1 Preliminaries

We describe the definitions and notations used in this section, and the indoor location data sets used in this work.

4.1.1 Approach at a Glance

Our primary goal in this work is in validating our two central hypotheses: (1) the deviations a user exhibits from his/her routine behavior in the past can be useful for predicting future non-conformance, and (2) that the concurrently, deviating behavior of his/her social network can improve that predictability. To this end,

- **Step 1:** We first study longitudinal indoor mobility data, from an urban campus, in order to shed light into two important prerequisites to operationalize our hypotheses (section 4.2). In particular, we investigate the (1) theoretical limits to predictability in indoor settings, and the (2) evolution and stability of physical social ties in the campus setting.

- **Step 2:** Next, we propose a Likelihood of Future Non-Conformance (LFNC) prediction pipeline that relies on a supervised learning classifier (section 4.3).

- **Step 3:** We then evaluate the proposed pipeline, extensively; we study the trade-off between the look-ahead distance (how far into the future), impact of social ties and the LFNC performance. We conduct several experiments to validate the robustness of the prediction pipeline (section 5.7).

- **Step 4:** We study the practical usefulness of LFNC predictions using data from a route-aware mobile crowd-tasking system operational in an urban university campus.

- **Step 5:** Finally, we study the extensibility of the proposed pipeline to the outdoor setting using city-scale, public transit trip data.

4.1.2 Definitions and Notations

We consider a trajectory $x_{(u,d)} := \{loc_1, loc_2, ..., loc_T\}$ whose elements are a sequence of staypoints loc_t, a user u visited during a day d, and $t \in [1, T]$. $x_{(u,d)}$ is a $T-$length vector each element representing equally sized, time bins over the day – for instance, for a trajectory considered at the hourly granularity, $|x_{(u,d)}| = 24$.

Next Place Prediction: Given a collection of $x_{(u,d)} \in X_{train}$ where X_{train} is the mobility training period (see Figure 4.1), and the trajectory of the same user u till time t, on a different day d_{test} outside the training period (i.e., $x_{(u,d_{test},1:t)} \in X_{test}$), the most likely next place at time $t+1$ denoted by $np_{u,d,t+1}$ can be predicted using

a next place prediction algorithm. Baumann et al. [15] provide a survey of 18 such prediction algorithms and report on their performance.

Non-Conformance: We declare non-conformance at such time $t+1$, when $loc(u, d_{test}, t+1) \neq np_{u,d_{test},t+1}$.

Further, we define a look-ahead distance, K.

Future Non-Conformance: Then, non-conformance at a future time $t+K$ is defined by Equation 4.1.

$$Nonconformance(u, d_{test}, t, K) = \begin{cases} 1, & \text{if } loc_{(u,d_{test},t+K)} \neq np_{(u,d_{test},t+K)} \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

Trajectory Deviation: We define a current sequence, $x_{(u,d_{test},t-h:t)}$, of length h, at time t. A user's *deviation*, $d_{(u,d_{test},t,h)}$, from his/her routine trajectory, is then given by Eq. 4.2. Here, $dist(.)$ measures the time series distance between the two partial trajectories. Note that $\hat{x}_{(u,d_{test},t-h:t)} := np_{(u,d_{test},t-h)} \| ... \| np_{(u,d_{test},t)}$.

$$d_{(u,d_{test},t,h)} = dist(x_{u,d_{test},t-h:t}, \hat{x}_{u,d_{test},t-h:t}) \tag{4.2}$$

Social Ties: Separately, we construct a physical social network of users based on their movement trajectories in X_{Train}, as previously described in Chapter 3. For each user u, we consider his/her k ties with whom they share the highest cumulative tie strength as u's *ego network*, $U_k \subset U$. Further, we denote the deviation the user u's top-k ties concurrently by the set $D_{u,d_{test},t,h,k}$ whose elements are $d_{u_k,d_{test},t,h}$ where $u_k \in U_k$.

Likelihood of Future Non-Conformance: Then, we define LFNC as the probability of occurrence of a future non-conformance ($Likelihood(loc_{(u,d_{test},t+K)} \neq np_{(u,d_{test},t+K)})$), as at time t on d_{test}, given $d_{u,d_{test},t,h}$, i.e., the user's deviation from norm thus far (limited by h), and $D_{u,d_{test},t,h,k}$. As we describe later in section 4.3,

Table 4.1: Notations used in this section.

Notations	Meaning
U	Set of all users
K	Look-ahead distance
k	Number of social ties
U_k	User u's ego network of top-k users
h	Current sequence length
T	Trajectory length
$dow(d)$	Day of the week of d
$x_{(u,d)}$	Trajectory of user u on day d
$loc_{(u,d,t)}$	Location of u at time $t \in [1, T])$ on day d
$x_{(u,d,t-h:t)}$	Current sequence over which deviation is computed; partial trajectory of (u, d) during time $[t - h, t]$
$\hat{x}_{(u,dow,t-h:t)}$	Expected (or, routine) current sequence of user u for the same day of the week as d
$d_{u,d,t,h}$	Deviation during the current sequence, $dist(x_{(u,d,t-h:t)}, \hat{x}_{(u,dow,t-h:t)})$
$np_{(u,d,t+1)}$	Most likely *next place* of user u on day d given trajectory till t

we operationalize LFNC as a classification task whose outcome variable is a probability that is equal to $Likelihood(loc_{(u,d_{test},t+K)} \neq np_{(u,d_{test},t+K)})$.

Datasets used: We use the same indoor mobility dataset, previously described in Section 3.1, for the analyses presented in this work. In particular, we use data from two time periods: (1) *Dataset A* in which the set of users on campus whose periodic outdoor mobility information was also available (which we utilize in section 4.2 for comparison in predictability), and (2) *Dataset B* consisting of users who participated in a campus-wide crowd-tasking pilot (whose details we utilize in section 4.5). Table 4.2 summarizes key details. As illustrated in Figure 4.1, we split *Dataset B* into disjoint mobility model training (X_{Train}) and test (X_{Test}). As noted previously in section 4.1.2, Next Place Predictions are made over X_{Test} which then serves as the dataset for LFNC learning and classification.

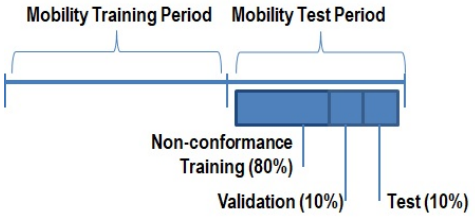


Figure 4.1: Illustration of the dataset segregation for different train/test purposes.

Table 4.2: Subsets of the Indoor Mobility Dataset used in this section.

	Observation Period	Users ($\|U\|$)	section
Dataset A	Aug–Dec, 2014	36	4.2
Dataset B	Feb–Mar, 2017	806	5.7, 4.5

4.2 Campus-Scale Mobility: Empirical Insights

Our broad goal is to predict deviations/anomalies from a user's regular movement pattern, using the collective movement pattern of an individual and her 'social ties' to improve the prediction accuracy. In addition, we would like to understand the lookahead time of such predictions, and how ongoing/past anomalous movement influences the prediction of future anomalies. However, for this approach to be successful, there are a few fundamental questions & challenges that we need to resolve first:

- What are the fundamental limits on predictability for indoor movement behavior in a workplace/campus setting? How does it differ from prior results on outdoor human mobility? In particular, for indoor environments, where the location trace itself has moderate error, the unpredictability is driven by both the random properties of human movement, and the noise introduced by the location traces.

- How do we reconstruct or infer social ties solely from the collective location traces of individuals? Recall that our goal is to utilize social ties built unobtrusively from the physical location traces, unlike past work that constructs such ties from other observations (e.g., online social network traces [26] or call records [94]). And, how do we verify that our inferred social ties are meaningful and stable?

We seek answers for the above questions in Sections 4.2.1 and 4.2.2 and summarize our answers at the end of this section.

4.2.1 Predictability Indoors

Fundamentally speaking, an individual's trajectory can be seen as a random sequence of symbols (with each landmark being a distinct symbol). The degree of randomness in such a movement pattern can be computed based on the notion of

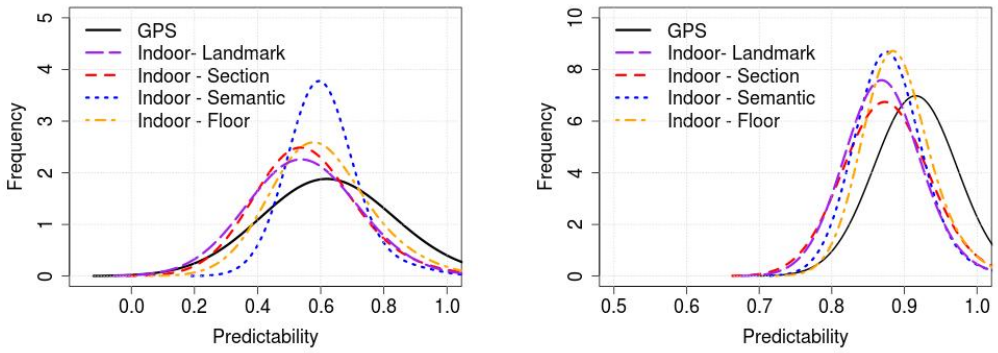
the entropy of this random sequence. We borrow concepts from Song et al.[118] in defining the theoretical upper bound on predictability *indoors*.

In Figure 4.2, I plot the distribution of predictability [118, 66] across users based on the (a) uncorrelated entropy where only the probability of a user turning up at a location is known (Figure 4.2a), and (b) the true entropy where the full history of a user's spatial and temporal patterns are known (Figure 4.2b), for users in *Dataset A*. We observe at varying spatial granularity of localization – in particular, at the (1) landmark level (most fine-grained, every 3 meters), (2) section level (6-8 meters), and (3) floor level, and compare against the predictability outdoors, observed via the continuous reporting of GPS coordinates (rounded to the third decimal which results in a granularity of $\approx$ 100 meters), for the same set of users.

In general, we observe that the maximum predictability is comparable to the outdoor case – for instance, the median predictability for GPS is 91% while the same is 87% at the section level. Noticeably, unlike what was reported in Jensen et al. [66] where predictability was looked at at the raw WLAN association levels indoors, we observe that at coarser spatial granularities (e.g., section level), the variability across users also reduces. For instance, while the previous work reported a non-negligible percentage of users showing maximum predictability in the range of 0.2 to 0.9, at the floor level, we observe that 99% of the users are bound to within a narrow range of 0.80 to 0.98 (i.e., within 3 standard deviations away from the mean).

4.2.2 Social Ties and Their Impact on Mobility

We first look at the formation and evolution of social ties in our campus setting in order to understand its practical usefulness for accounting friendship information as additional features that reveal insights into a user's mobility. We note that the 36 users in *Dataset A* are all freshman, and we observe their mobility since their first week on campus till the end of the term.



(a) Uncorrelated Trajectory Sequences (b) Infinite Length Sequences

Figure 4.2: Theoretical Maximum Predictability at Varying Spatial Granularity for (a) Temporally Uncorrelated, and (b) Correlated Trajectory Sequences

We use the technique described previously in Section 3.3.1 to infer the intensity or strength of tie between any two users. In Figure 4.3, we plot several metrics related to the evolving *friendship network* amongst the users, as the term progresses (represented by $x-$ axis). Between consecutive weeks, we consider the sub-network consisting of only the top-1 ties (i.e., the network consisting of each user and his/her closest tie, till current week), and observe that the Jaccard similarity [90] of the set of edges, steadily reaches its maximum at week 5 after which it plateau. Similarly, we see that the diameter (represented by the blue dotted line) of the complete network, undergoes a stark drop till week 3 after which it stabilizes. This shows that the network consisting of the closest ties stabilizes after the initial few weeks, and demonstrates that the passively captured social ties can in fact be reliably used as additional information in exploring the predictability of user mobility.

Additionally, we studied the *correlation* of mobility of a user and his/her top-1 tie. We divided *Dataset A* as X_{Train}: Aug-Oct, 2014, and X_{Test}: Nov, 2014. For each day d, and time t in X_{Test}, we computed the *probability of being at* $loc(u, d, t)$, for each user u, based on the visitation frequency distribution over all possible locations learned during X_{Train}. We then compute the Pearson's correlation coefficient of the time series of *probabilities* of user u and his/her close tie, as well as the correlation between the same user and any randomly chosen user from the dataset. In

Figure 4.4, we plot the CDF of the correlation values – we see that the similarity in being at likely locations (and not necessarily the same location) concurrently, is statistically significantly higher for top-1 pairs ($D = 0.395$, p-value = 0.005 on the Kolomogrov Smirnov test). We highlight here that this analysis suggests that there is similarity in the likelihood of the strong pairs being at their *respective routine locations*, simultaneously. We confirmed that this observation is not a mere artefact from our methodology to infer ties (which considers features extracted from co-location episodes of pairs of individuals). In fact, when considering the location trajectories of the user and his/her top-1 tie in the dataset, on average, nearly half the time (normalized Hamming distance of ≈ 0.53), the two were not found to be co-located, compared to 80% of the time for random pairs.

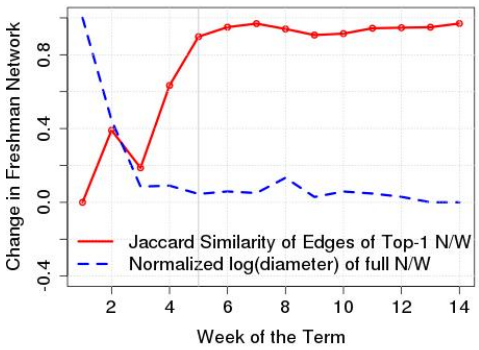


Figure 4.3: Evolution of the Physical Friendship Network of Students, as the Term Progresses.

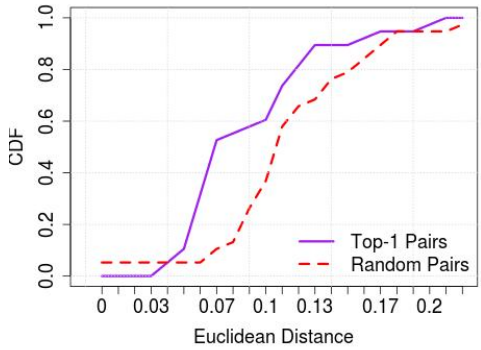


Figure 4.4: CDF of correlation between pairs of trajectories belonging to top-1 ties vs. random pairs.

Key Insights: Our empirical analyses establishes 3 key results:

1. User movement, in a mostly-indoor campus environment, has high predictability at the section-level and floor-level granularity, theoretically, and is reasonably consistent across users.

2. The set of top-K ties (derived from an initial observation period of 5 weeks) remains remarkably stable over the remaining 9+ weeks of the term. Accordingly, we've established that it is possible to derive the set of 'strong ties' of an individual, unobtrusively, using a modest period of observational data.

3. There exists significant correlation between the movement behavior/trajectory of close ties–i.e., "birds of a feather flock together with similar predictability". This finding corroborates similar insights previously presented for outdoor mobility [26, 31], and suggests that factoring in the movement behavior of close-ties should improve the location prediction accuracy for an individual.

4.3 LFNC Prediction Pipeline

In this section, we describe the overall working of the non-conformance monitoring pipeline.

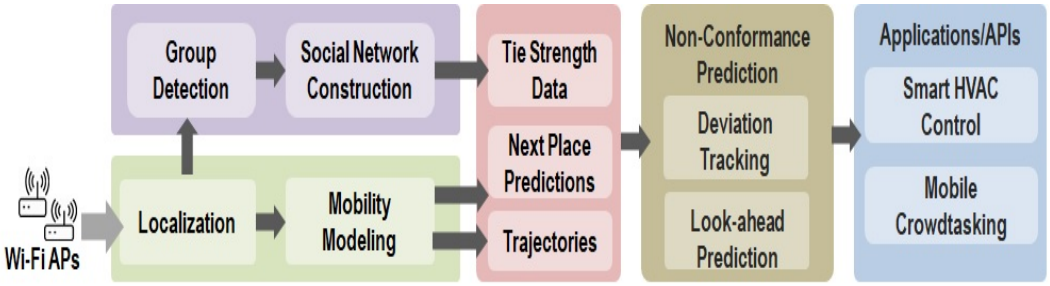


Figure 4.5: Proposed Pipeline for LFNC Prediction.

4.3.1 Trajectory Extraction

As described previously in Sections 3.1 and 4.1, the server-side indoor localization server receives reports on RSSI from individual APs as well as via the Real-Time Location Service (RTLS) running on the master controller allowing for frequent updates on the localization of individual mobile devices using an extension of the RADAR [11] algorithm. The *Localization* module then constructs trajectories per user, per day (i.e., $x_{(u,d)}$s) as defined in section 4.1.2.

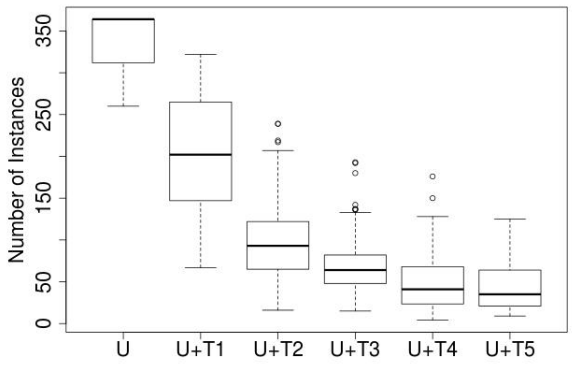


Figure 4.6: Number of instances in the dataset with user alone (U) vs. with his/her top ties ($U + Tk$, $k = 1, 2, 3, 4, 5$) simultaneously present on-campus, **per user**.

4.3.2 Mobility Modeling

This module implements three "next place" predictors: zeroR and Markov Chains of length 1 and 2 [46] (MC-1 and MC-2, respectively) taking into account the day of the week of day d (i.e., $dow(d)$) and time of the day t; in our implementation, we consider t at intervals of 15-minutes resulting in $T = 96$. Based on observations in X_{Train}, the constructed transition matrices for $MC - 2$ and $MC - 1$, and the visitation frequency matrix for $zeroR$ predictions allow for *next place predictions* in X_{Test}.

Similar to the implementation described in Kotz et al. [119], for each test sample, we roll back from $MC - 2$ to $MC - 1$ to $zeroR$, depending on whether the same context was *seen* during training. The matrices are stored as $key - value$ pairs, where the key is the concatenation of $< dow, t, loc_{t-1}, loc_t >$ for $MC - 2$, $< dow, t, loc_t >$ for $MC - 1$ and $< dow, t >$ for $zeroR$ predictions, respectively. During test time, for a given context at time $C_t :=< user, dow, t, loc_{t-1}, loc_t >$, if a MC-2 prediction is not possible (due to that case not seen during train time), the algorithms rolls back to $MC - 1$; and if such a prediction, too, is not possible, the algorithm rolls back to $zeroR$. If $zeroR$ is also not possible, a prediction is not made.

4.3.3 Tie Strength Extraction

Following from the definition of *group episodes* in Section 3.3.1, we use GruMon [112] to identify social groups from passive indoor location data. Based on longitudinal observations of such group episodes, as detailed in section 4.1, we measure the strength of tie between pairs of users in the dataset. In section 5.7, we restrict our evaluations to the impact of a user's ego network consisting of only the top-k ties with k set to $\{1, 2, 3, 4, 5\}$. Although higher values of k is possible, it reduces the size of the dataset on which evaluations are possible as the number of cases with top-k users are all present on-campus at the same time drops significantly, as k is increased. We show this in Figure 4.6.

4.3.4 Non-Conformance Prediction

Labelling Nonconformance: During the training phase, the *Look-ahead Prediction* module consumes pre-trained next place predictions for labeling whether the user will *conform* to, or be at the most likely place, at a time, $t + K$, i.e., K bins into the future.

As identified by the two central hypotheses of this work, we factor in: (1) the deviation, $d_{(u,d,t,h)}$ in trajectory a user has incurred during the *current sequence*, $x_{u,d,t-h:t}$, and (2) the deviation in trajectory the user's top-k friends have incurred, $D_{(u,d,t,h,k)}$.

Deviation Computation: We compute this deviation ($d_{(u,d,t,h)}$) at time t as the distance between the time series of the actual trajectory a user has traversed during time $[t - h, t]$ and the sequence of *most likely locations* for the same period. To measure this distance, among the alternatives considered, Dynamic Time Warping (DTW) [87], Hamming Distance [53] and Edit Distance [105], we found the former to be the most appropriate due to its flexibility in allowing for both (1) potential inaccuracies in the location data (e.g., an individual being located to the room next door), and (2) possibilities that an individual may be *slightly* late or early

to the routine places he/she visits. Algorithm 1 outlines the steps taken to label non-conformance and compute the corresponding user deviation.

Further, we express the deviation of the user's friend network consisting of his/her top-k friends, d, as the weighted (by the tie strength) linear sum of the individual deviations, for instances where all k friends are on-campus concurrently (and not necessarily co-located) and sum the individual deviations weighted by their strength of tie. In section 5.7, we vary k to observe its impact on the performance of non-conformance prediction. As a direct consequence, with increasing k, the number of instances where the user and the top-k friends are all concurrently present on campus drops drastically.

Predicting Nonconformance: We train and build a Gradient Boosting Classifier whose independent variables are $< dow(d), t, d_{u,d,t,h}, D_{u,d,t,h,k} >$ and the binary outcome variable represents conformance (or, nonconformance), for different look-ahead distances of K. Here, in addition to the deviation features which are derived from our key hypotheses, we consider the day of the week and time of the day as control variables as was common with past mobility prediction models.

Finally, the conformance predictions can be consumed by various applications including smart HVAC control and route-aware mobile crowdtasking (see Section 4.5).

4.4 Evaluation

In this section, we report our findings on the predictive ability of the individual factors based on the two central hypotheses that we consider in this work, in predicting future non-conformance of a user's mobility behavior. We first evaluate in section 4.4.1 the influence of trajectory deviations a user and his/her ties undergo over a day, in predicting future non-conformance. Then, in section 4.4.3, we explore the impact of time of the day and the types of places a user visits on such predictive performance.

noend 1 LFNC Labeling and Deviation Computation

Input: $MC2TransitionMatrix$, $MC1TransitionMatrix$, $ZeroRMatrix$, $x_{(u,d)}$, t, K and h
Output: $d_{(u,d,t,h)}, label_{u,d,t,K}$
$maxK \leftarrow K - t$ ▷ The number of look-ahead windows possible after current time t
$predictedtrajectory \leftarrow trajectory_{1:t}$ ▷ Input only known trajectory till time t
$d_{(u,d,t,h)} \leftarrow NULL$ ▷ Initiate deviation vector
$label_{u,d,t,K} \leftarrow NULL$ ▷ Initiate conformance vector
for $K = 1$ `to` $maxK$ **do**
    $nextlocation_{u,d,t+K} \leftarrow getNextLocation(x_{(u,d,1:t+K-1)})$ ▷ Predicted *next location* at time $t + K$
    if $nextlocation_{u,d,t+K} == loc_{(u,d,t+K)}$ **then**
        $label_{u,d,t,K} \leftarrow 1$ ▷ Label conformance
    else
        $label_{u,d,t,K} \leftarrow 0$
    end if
    if $K == 1$ **then**
        $\hat{x}_{(u,d,t+K)} \leftarrow nextlocation_{u,d,t+K}$ ▷ Append expected next place
        $d_{(u,d,t,h)} \leftarrow dist(x_{(u,d,t-h:t)}, \hat{x}_{(u,d,t-h:t)})$ ▷ Compute deviation
    end if
end for

Prediction task: We represent the non-conformance prediction task as a binary classification task with the conformance label (1 – non-conformance and 0 – conformance) as the dependent variable and the features described in section 4.3 as independent variables with a Gradient Boosted Model (GBM) for supervised classification. The choice of this classifier is motivated by the popularity of ensemble learning techniques and their frequent use in mobility prediction tasks in recent works [111, 39].

Experiment conditions: In all the experiments described in this section, we perform classification on a balanced set (with equal number of conformance, and non-conformance classes) derived from *Dataset B* (see section 3.1); to do this, we first create a subset of all the samples from the smaller class and randomly sampled, equal sized samples from the other class, generating the balanced dataset. Unless explicitly stated, we take X_{Train} (01-02-2017 – 28-02-2017) over which the next place prediction models are trained, and X_{Test} (01-03-2017 – 14-03-2017) over

which LFNC is trained and tested using a split of train (80%), validation (10%) and test (10%).

Parameter tuning and model selection: On the train set, we learn multiple GBMs assuming a Gaussian loss function and by varying the number of trees between 100 to 10,000 (in increments of 100). The shrinkage and interaction depth parameters are fixed to defaults (i.e., 0.01 and 4, respectively). We pick the best performing model as the one that has minimal error on the validation set. Finally, we evaluate the test set on this model to report findings.

Performance metrics: In all our analyses, we report the accuracy based on precision, recall and AUC, following their standard definitions. Precision and recall represent the average over both the positive and negative classes and we use 0.5 as the cut-off probability in declaring the binary outcome variable.

Implementation: The computations related to GBMs were performed using R's *gbm* package [2] and the ROCR [3] library for performance calculations.

4.4.1 Predicting Non-Conformance

We first evaluate the performance of predicting non-conformance under the three conditions: (1) $Model_{nodev}$, without any information of a user's deviation from his expected trajectory (i.e, with the day of the week, dow, and time bin, t, being the only input features), (2) $Model_{userdev}$, considering the deviation of the user in addition (i.e, $d_{(u,d,t,h)}$), and (3) $Model_{combidev}$, considering the deviation of both the user and his/her friends present on campus at that time (i.e., $d_{(u,d,t,h)}$ and $d_{(u,d,t,h,k)}$).

In Figure 4.7, we plot the accuracy (measured as AUC, on the $y-$axis) for LFNC predictions with $Model_{userdev}$, for varying look-ahead distances K, and current sequence lengths, h, or in other words, the deviations from how far back in the user's trajectory. We observe that observing trajectory deviations over short hs (e.g., 15 - 45 minutes), tend to result in better performance – for instance, for most values of K, the predictions corresponding to $h = 1$ and $h = 3$ (i.e., blue and red

lines) exhibit performance improvement of over 10% consistently.

In Figure 4.8, we plot the accuracy with the AUC on the $y-$axis for varying look ahead times, K, on the $x-$axis, for the three cases where both the user and his/her top-5 ties were present on campus. Note that each increment in K implies the addition of 15 minutes into the future from current time. We observe that among the three, considering both the user's and friends' deviation thus far (represented by the solid blue line) offers the greatest performance – even with a look-ahead time of 3 hours (i.e., $K = 12$), the combination provides an $AUC =\approx 0.9$. We further note that considering the user's deviation alone performs similarly well until around $K = 4$ after which the drop off rate increases resulting in at least a 15% drop in accuracy in comparison to one that additionally considers the friends' deviation. We also note that the performance of an approach that does not consider any of the deviation measures results in the poorest performance (relatively stable at $AUC =\approx 0.7$) – which means that the additional factors provide a performance improvement of $\approx$35%, 25% and 15% at look-ahead times $K = 1, 4, 12$, respectively. In Table 4.3, we tabulate the precision and recall values for the three cases.

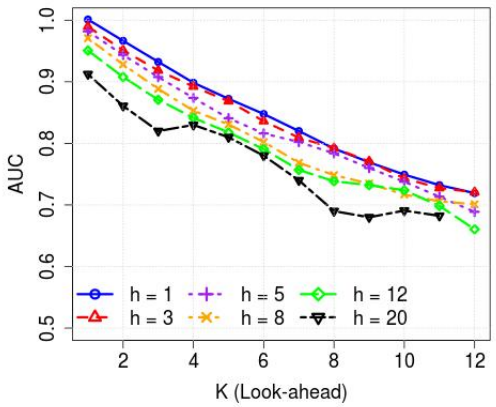


Figure 4.7: Comparison of LFNC Look-Ahead Capability for different h.

Table 4.3: Prediction Results with no deviation, using only user's deviation, and the combination of user+friends' deviation.

	$Model_{nodev}$	$Model_{userdev}$	$Model_{combidev}$
N @ K=1		1096	
Precision @ K=1	0.665	1	1
Recall @ K=1	0.654	1	1
N @ K=4		1072	
Precision @ K=4	0.642	0.862	0.897
Recall @ K=4	0.642	0.858	0.896
N @ K=12		958	
Precision @ K=12	0.69	0.631	0.792
Recall @ K=12	0.66	0.625	0.792

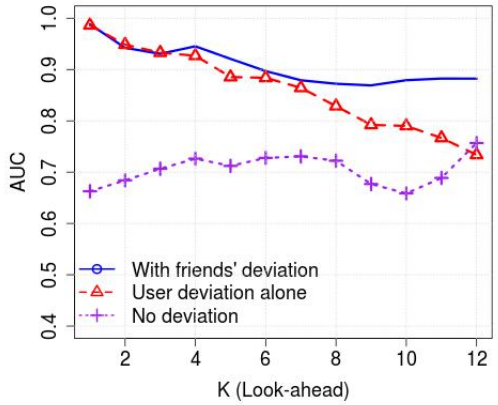


Figure 4.8: Comparison of LFNC Look-Ahead Capability of different Models.

4.4.2 Impact of Social Ties

Previously, we noted the utility in considering the combined deviation of a user's social ties from their respective "expected" routines in predicting non-conformance, ahead of time. Here, we explore the impact of the "size", or "extent" of the social ties considered on prediction performance. In Figure 4.9, we observe the performance of using user's deviation alone vs. user's top-k ties' combination of deviation, where we vary k from 1 to 5. Here, we note that the number of samples trained on and evaluated against for (i.e., the size of the dataset) is smaller for higher k's (as we previously saw in Figure 4.6). We make the following remarks:

1. Surprisingly, the consideration of a user's top-1 tie's deviation in addition to his/her own does *not* provide additional utility. However, we find that this is consistent with our prior observation (in section 4.2) that the user and the clos-

est ties' mobility behavior are too highly correlated that they do not provide additional information gain.

2. We further note that with increasing size of the ego network, the performance improves – for $K \geq 3$, we see that the AUC is generally ≥ 0.80 for as advanced as 10 hours of look-ahead (i.e., $K = 40$) whereas the the performance for the 'User-only' or 'User+Closest Tie' approaches both stabilize around 0.70.

Further, we investigated the performance in predicting non-conformance for times (i.e., $< dow, t >$ combination) during which the user has had differing levels of *regularity*, historically. The regularity is captured as the $zeroR$ probability as in [118]. We plot the performance for subsets of user instances thresholded by varying values of the regularity, S (ranging from 0.3 to 0.7), in Figure 4.10a and Figure 4.10b, for $Model_{userdev}$ and $Model_{combidev}$ with k=5, respectively. We highlight that the performance improves in general (about 5%) over "more regular" periods (i..e., higher values of S), in either case.

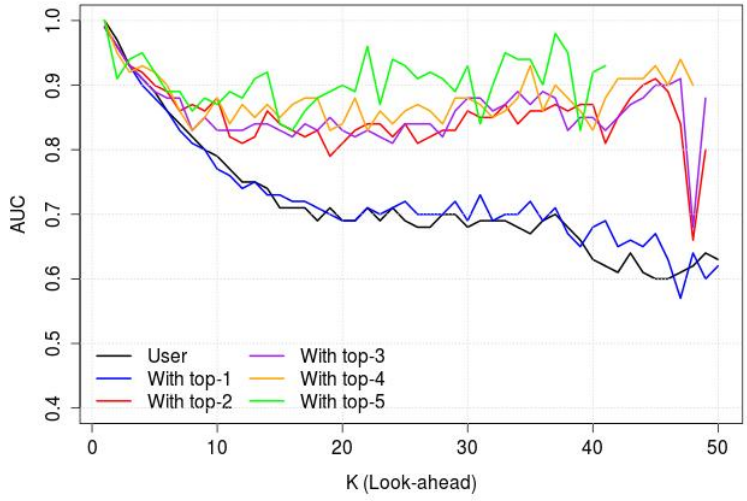
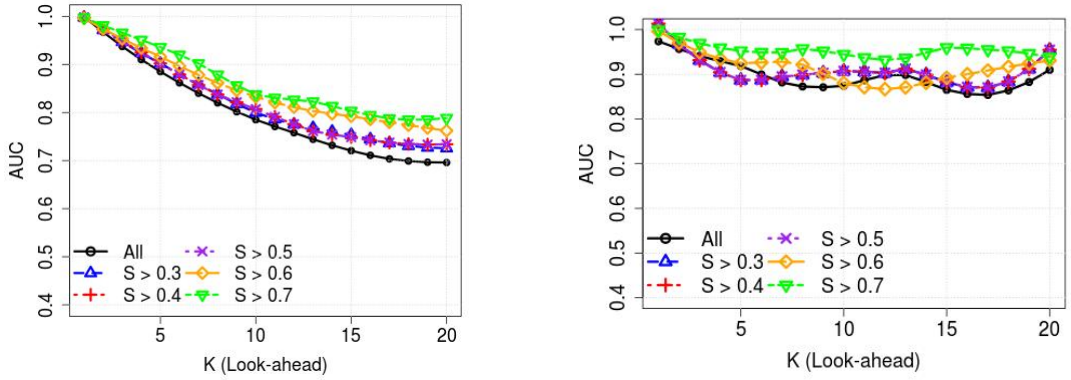


Figure 4.9: Performance by varying size of the ego network of a user considered (and hence, the resulting combination deviation feature).



(a) User's deviation alone $Model_{userdev}$

(b) User + Friends' deviation $Model_{combidev}$

Figure 4.10: Performance by differing levels of regularity of the predicted time instance, historically, for (a) $Model_{userdev}$ and (b) $Model_{combidev}$ with $k = 5$.

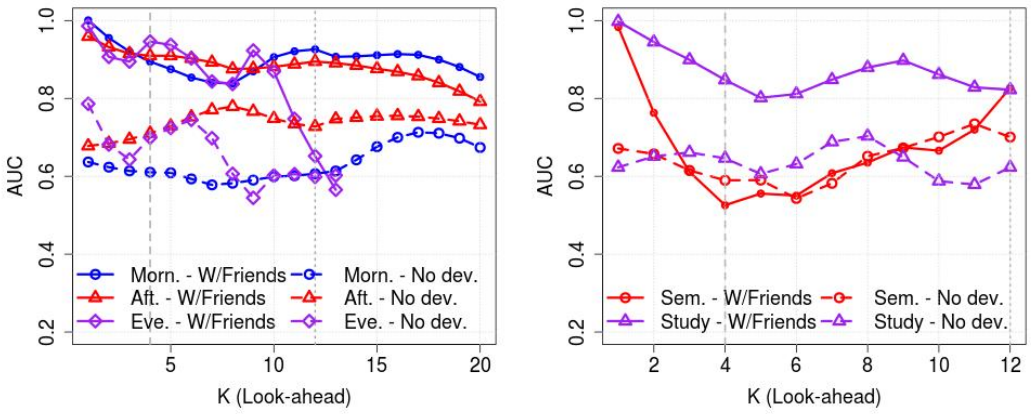
4.4.3 Performance by Time and Place

Thus far, we have reported the performance across all time windows and likely location types – here, we explore the role of time and place on performance, more closely.

In Figure 4.11a, we plot the performance of $Model_{nodev}$ and $Model_{combidev}$ stratified by the "time" – we consider Morning hours as between 8 AM to 12 Noon, Afternoon as 12 Noon to 5 PM and Evening as 5 PM to 9 PM. The "time" here represents the future time for which a prediction is made (i.e., current time t + look ahead time K). We observe the largest difference during morning hours where $Model_{combidev}$ performs at least 30% better in comparison, for up to look-ahead times of 3 hours ($K = 12$). We also note that the least difference in performance is observed for afternoon predictions – and that the gap keeps diminishing with increasing K. For both models, we observe the steepest drop in performance with K for evening predictions where the performance of both drops as low as 0.6 for $K = 12$. This observation could suggest that during evenings, when the users are more likely to engage in informal social activities, the deviations experienced thus far, seem to be less predictive of future anomalous movement.

Further, in Figure 4.11b, we plot the performance stratified by two of the most

common types of locations on campus – (a) seminar rooms (including any other scheduled teaching rooms such as class rooms), and (b) study areas (including group study rooms which are available to students as "booked" resource which is typically used for project discussions, and open study areas). Similar to the case above, the location here refers to the actual "future" location the user was at $t + K$. Between the two, the former represents a formal class of locations and the latter more a casual setting. As expected, we observe that the improvement in performance in considering the user's social ties' mobility behavior is evident for the more casual/social scenario – for e.g., a 40% improvement in prediction for $K = 1$ and which tapers down to roughly 25% with a look ahead time of 3 hours. For the most part, we observe that for the more formal setting, the performance of both models are comparable in that the impact of social ties' mobility has less impact on the predictability of class attendance.



(a) Accuracy by Time of Day

(b) Accuracy by Types of Places

Figure 4.11: Difference in Performance by (a) the time of day and (b) the type of places a user is at.

4.4.4 Robustness Checks

In this section, we report findings from a number of checks, in order to validate the robustness of non-conformance prediction under varying conditions.

Performance during Staypoint Transitions

Whilst previous work on outdoor mobility have demonstrated that the theoretical maximum predictability is achievable [118] in practice, there's a lack of evidence in the indoor setting. Kotz et al. [119] note that practically achievable accuracy of predicting next place is comparably less, and also that the prediction task is easy when the user is still (where the next place is the same as the current place) but suffers during *transitions*. To understand how this impacts future non-conformance predictions, we report on the performance for instances where *transitions* have occurred, in Figure 4.12. We distinguish between "actual" transitions where the user transitions in reality (at future time $t + K$), and "predicted" transitions where the user is "predicted" at time, $t + K$, to transition based on actual trajectory observations till $t + K - 1$. Consistent with our previous findings, we observe that $Model_{combidev}$ outperforms the baseline in both cases; we see approximately 35% improvement for $K = 1$ (i.e., in the next 15 minutes) and ≈20% with a 3 hour look-ahead time.

Non-overlapping Train/Test Time Series

As we deal with time series data in this work, a key concern during evaluation is the possibility of ground-truth leakage as a result of consecutive observation points from the time series becoming part of both train and test sets. This could potentially lead to an over-estimation of the performance observed.

In order to investigate this further, instead of randomly splitting the dataset to into 80-10-10% train-validation-test sets, we split the first half of the data (by date) into train and the remaining into equal parts of validation and test sets and re-ran the analysis. In Figure 4.13, we plot the resulting performance; we note that the performance remains relatively stable with only a $\approx$ 10% drop in performance for $K \in (1, 4)$. We also point out that this analysis was run on completely non-overlapping sets, although in practice, the performance should improve with online learning (i.e., a growing train set with each incoming test case and its correspond-

ing prediction with some notion of confidence whose discussion we defer to future work).

Dynamic Predictions

Thus far, we have discussed the performance over the dataset covering the entire observation period of the data with mobility training of 4 weeks between 01-02-2017 through 28-02-2017, mobility predictions over the 2 week period of 01-03-2017 through 14-03-2017 which was then split into train/validation/test for non-conformance predictions. The tie strength values used thus far are the cumulative strengths calculated as at the end of this period.

In order to understand how the performance would vary in practice where training data is acquired as the term progresses starting with zero data at the beginning of week 1 – i.e., the cold start problem, we study the online performance of week, w, using mobility training data from weeks, $[1, w-1]$, and tie strengths calculated as at the end week $w-1$. In Figure 4.14, we plot the performance of $Model_{combidev}$ and $Model_{nodev}$ for weeks 2 to 6. In both cases, we observe that the performance is relatively unstable for the first two weeks (with the least amount of training data) but that it stabilizes after week 4 with only marginal differences in performance beyond that period.

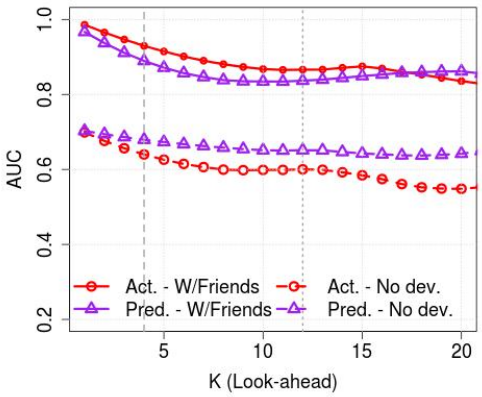


Figure 4.12: Performance for (a) actual and (b) predicted instances of locationing transition.

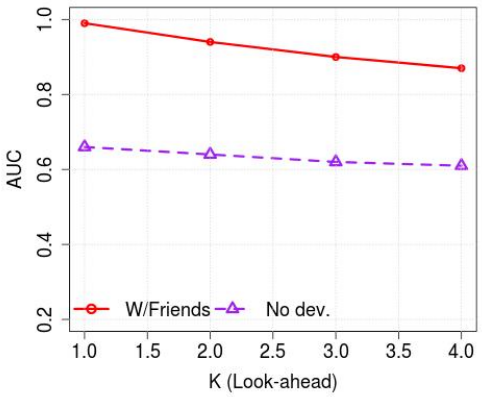


Figure 4.13: Impact on Performance using Non-Overlapping Train/Test Time Series.

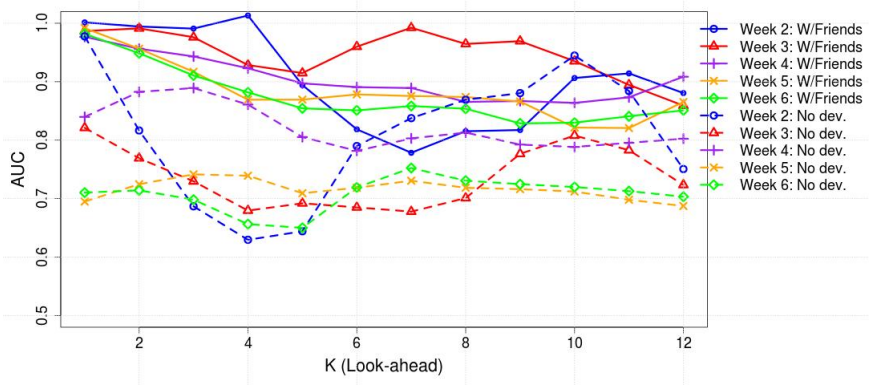


Figure 4.14: Performance as the Term Progresses.

4.5 Case Study: Location-Aware Mobile Crowd-Sourcing

There has been a significant body of research on the use of personal mobile devices to support various forms of participatory mobile sensing or crowdtasking in urban environments. A notable example of this paradigm is the use of campus users to report on the status of various campus resources/facilities (such as restrooms, cafeterias and office equipment). More specifically, the *Ta$ker* crowdtasking platform [69, 70, 24], operational on-campus since 2014, uses the predicted movement pattern of participating users to recommend tasks that are likely to minimize a worker's detour overhead; empirical results show that this paradigm of crowdtasking based on trajectory predictions increases worker productivity by 60% [24]. In the crowdtasking platform available on-campus, (i) the user's trajectory is derived based on identifying *staypoints* (the most likely location where the user spends the largest fraction of time within each 30 min window), and (ii) the task recommendations are made over distinct 3-hour windows. Empirical data also shows that 40% of tasks that are accepted by the platform workers are not eventually completed, with "unexpected" changes in the worker's movement pattern being cited as the most common cause for such non-completion. In such a scenario, the ability to better predict that the user is unlikely to be at specific predicted stay-points can be very valuable in

improving the recommendation process.

In this section, we shall show that the use of our LFNC prediction can lead to observable gains in overall task completion rates and investigate its impact on worker productivity.

Crowd-tasking Data: A crowd-tasking pilot was carried out on-campus using the *Ta$ker* platform during a 2-week period of 14th March, 2017 through 31st March, 2017 (overlapping with *Dataset B*). A total of 325 student users were assigned tasks based on their predicted trajectories out of which 242 of them completed at least 1 task. Out of these 242, 106 (44%) of them were *recommended* tasks based on their historical movement behavior and predictions for the assigned task window – such an assignment is expected to minimize the student's detour overhead. In total, out of 60,000+ tasks assigned, 3822 were completed with an overall task assignment-to-completion conversion rate of $\approx 6.2\%$.

4.5.1 LFNC Predictions

We utilize trajectory data of the *Smart Campus* users and their respective ego networks from 01-02-2017 to 13-03-2017 (i.e., X_{Train}) for mobility prediction training, and predict next place locations for $K-$ look-ahead distances over the pilot period. Out of this, as before in section 5.7, we split the set into train/validation/test sets and make LFNC predictions over the validation and test tests. The predictions are then carried over as input to the task assignment module – we emphasize that, as this is a post-hoc analysis on an existing pilot, we are unable to assign tasks based on the LFNC predictions, but are only able to analyze differences in the two groups that were predicted to have been *conformant* and *nonconformant*. As the task assignments are made over 3-hour windows (3 times over the day at 9 AM, 12 Noon and 3 PM), with students allowed to perform the tasks any time during each window, we consider $K = 4$, i.e., predictions an hour ahead of the task time, and $1 \geq k \geq 5$, i.e., considering the student and his/her top-k ties' deviation where k

Table 4.4: Summary of LFNC Prediction Results on the Mobile Crowd-Tasking Dataset.

	$K = 1$	$K = 2$	$K = 3$	$K = 4$
N_{sample}	114,470	113,544	112,316	111,036
AUC	1	0.96	0.93	0.9
Precision (@0.5)	99.11%	91.23%	86.42%	83.21%
Recall (@0.5)	99.11%	91.04%	86.16%	82.99%

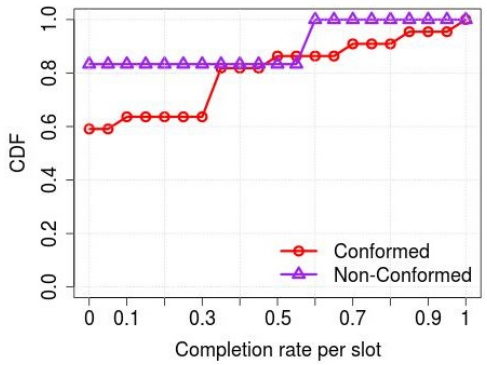
depends on the number of ties who are present on campus concurrently. We summarize the results in Table 4.4. Overall, we observe that with a look-ahead window of an hour, the predictions are correct $\geq 82\%$ of the time.

4.5.2 Key Take-Aways

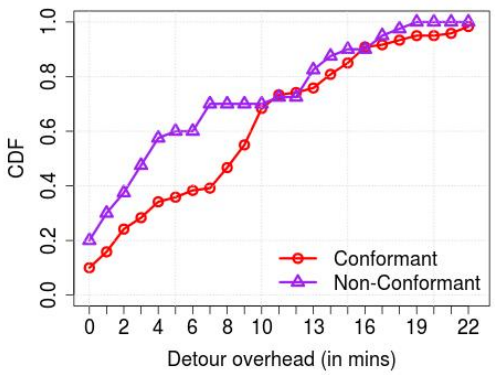
We first compare the detour incurred by students who were predicted to be *conformant* to their routine behaviour during the respective task window vs. those who weren't. In Figure 4.15b, we plot the detour, in minutes, the students from both categories incurred – the $y-$axis shows the CDF and $x-$axis represents the detour overhead. For a student actual trajectory was from $A \rightarrow B$, and the student completes a task assigned to a location C, then the detour overhead is computed as $distance(A \rightarrow C) + distance(C \rightarrow B) - distance(A \rightarrow B)$. Here, the distance function returns the temporal distance, or the time taken to reach one location from the other. Interestingly, we see that the *nonconformant* group incurred statistically significant less detour in comparison to the *conformant* group ($D = 0.30833$, p-value = 0.006) – for instance, nearly 70% of tasks completed incurred ≤ 7 minutes of detour for the former whereas only 30% of the completed tasks incurred detours ≥ 7 minutes for the latter. We computed the entropy over the distribution of task locations that the two groups of workers chose and found that the *conformant* group showed a 12.5% increase compared to the *nonconformant* ($Entropy_{conformant} = 3.39$, $Entropy_{nonconformant} = 3.00$). In effect, the *non-comformant* group were unable to assist with tasks distributed throughout the

campus, confining their task acceptance and execution to locations opportunistically close to their *actual* trajectories.

In Figure 4.15a, we plot the cumulative distribution function (CDF) on the $y-$axis and the average completion rate (over each unique user, time slot pair) on the $x-$axis (a student can be assigned multiple tasks during the same slot). We observe a statistically significant improvement in completion rates in the *conformant* group (represented by the red solid line) over *nonconformant* (blue dashed line) – a Kolmogorov Smirnov (KS) statistical test reveals a $p-$value of 0.0171 (i.e., ¡ 0.05) and a D statistic of 0.476. In particular, we note that there is at least 20% improvement in the percentage of students who had an average completion rate of zero – whilst 80% of the nonconformant students had a zero completion rate, only 60% of the conformant students incurred the same. Also, we note a marginal increase in the *overall task completion rate* of 9% in the *conformant* group which is 3% more than the overall population. These results demonstrate a possible use of such non-conformance prediction: a crowdsourcing platform aware of such users could choose to preferentially recommend tasks to other users, thereby increasing the overall task completion rate and the associated spatial diversity.



(a) Completion Rate



(b) Detour/Overhead

Figure 4.15: The difference in completion rates and detour incurred between the Tasker users whose behavior at task assignment time were predicted to be either "conformative" or "non-conformative", for $K = 4$ (i.e., with a look-ahead time of 1 hour) and $k = 2$.

4.6 Extending to the Outdoor Setting

In the previous sections, we have described and evaluated our central hypotheses for predicting future non-conformance, in a predominantly-indoor urban campus. Here, we extend our analyses to an outdoor, city-scale setting. Primarily, we hope to understand whether the inherent differences between indoor and outdoor mobility affect our capability to predict non-conformance.

Outdoor transit data: To study outdoor mobility, we exploit a public transit dataset from Singapore where each trip a commuter makes using the cashless payment card, on buses or trains, is captured along with the origin and destination station IDs and the corresponding timestamps. The dataset pertains to a period of 3 months from November, 2011 through January, 2012. In total, the dataset spans 300+ trips from over 5 million commuters, across 5000+ bus stops and train stations. For our analysis, we extract a set of 100-most frequent travelers (by total trip count on weekends over the period–see Figure 4.16), as well as their respective *co-travelers* (defined shortly, below). Whilst our indoor location data is updated periodically (every 2-3 minutes), the transit dataset is event-driven, containing location information only when a trip takes place. For the purpose of our analyses, we extrapolate the point-to-point trip data to construct trajectories (i.e., $x_{u,d}$, as defined in section 4.1.2)– for instance, if a user enters $station_A$ at t_1, exits $station_B$ at t_2, re-enters $station_B$ at t_3, exits through $station_C$ at t_4, then the users taken to have *stayed* at $station_A$ and $station_B$ during t_1 to t_2 and t_2 to t_4, respectively.

Strength of Ties: As the social network information among commuters is also unavailable in this dataset, we adopt an approach similar to [64]. Trips that originate and terminate at the same stations within 20 seconds of each other at both entry and exit, during weekends, are considered to be *co-trajectories*, and the respective commuters considered to be *co-travelers*. For the 100-most frequent travelers, we extract such ego networks where the pair shares at least 2 co-trajectories. The strength of tie is then computed simply as the number of co-trajectories shared between the

pair. We extract trajectories of 1024 travelers out of which 992 of them have taken a trip on at least 21 days. However, unlike in the case of the indoor dataset, we observe that a majority of travelers only have a single strong tie following this definition, and hence we limit our analysis to the *top-1* tie alone.

LFNC Prediction: As outdoor mobility is less frequent (longer stay duration) and outdoor location prediction is often at coarser granularity, we consider the locations of commuters at hourly intervals and at subzone level granularity[1]. More specifically, we map the geo-coordinates of the stations (the start and end points of a trajectory) to the corresponding subzone. To compute the *deviation* (i.e., the distance between the actual and expected trajectory), we sum up the Haversine distance between the corresponding locations in the two trajectories. Figure 4.17 plots the performance (AUC) for $Model_{nodev}$ and $Model_{combidev}$ with $k = 1$. Similar to our findings from the previous sections, we find that the deviations (from their normal routes), experienced by a commuter and the single strong tie, prove to be a reasonable early indicator of impending non-conformance – for instance, an $AUC \geq 0.85$ is observed for $K = 2$ hours. This represents a significant (30%) improvement in AUC over the deviation-unaware baseline, thereby demonstrating the power of our method.

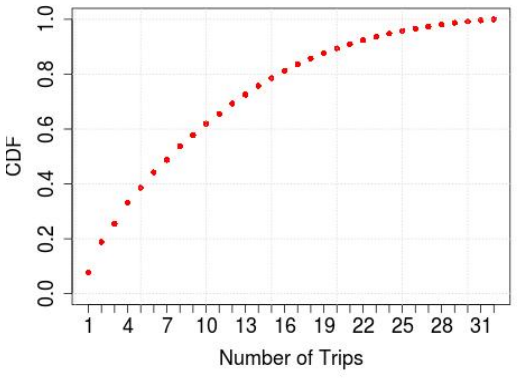


Figure 4.16: Distribution of number of *weekend* trips taken by users during the 2-month observation period.

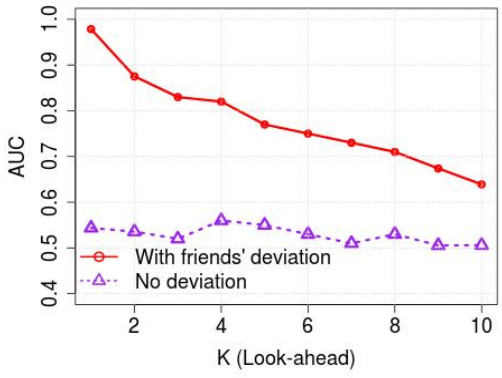


Figure 4.17: LFNC Performance with increasing K (in hours) on the Outdoor Mobility Dataset.

[1]https://data.gov.sg/dataset/master-plan-2014-subzone-boundary-web

4.7 Discussion Points

Possible extensions: At present, the non-conformance predictor takes into account routine defaults in the near past but is agnostic to the *semantics* of such routine location visits. One such semantic aspect is the nature of the location, or the activity that takes place at the location that an individual is expected to visit. For instance, previously, in Figure 4.11b, we noted that the performance of the predictor is dependent on the type of place – for e.g., between casual study areas and seminar rooms with scheduled classes, the performance of non-conformance prediction is weaker for the latter suggesting that despite routine deviations in the past, students are still likely to attend scheduled/class-related activities/locations. Further, another semantic aspect that could be considered is the *nature of the relationship* of individuals who are expected to be at a location together. For example, an individual who is less regular for large-meetings may be more regular in attending meetings where another individual is at a supervising capacity. Hence, a potential extension of the predictor would be to account for latent factors of both the spaces/locations and the individuals, additionally.

K^{th}-likely Next Place Prediction: We have presently focused only on identifying non-conformance–i.e., in making a binary declaration of whether a user will visit the highest predicted location or not. By itself, this does not directly answer the question: where is the user most likely to be instead? As a plausible alternative, we can consider an expanded range of *top-K* ($K \geq 2$) predicted locations, and identify an anomaly only if the user does not visit any of these K locations. It is likely that such anomalies represent *dramatic disruptions* to the user's regular mobility pattern–e.g., a special annual concert on campus. It is unclear whether our hypotheses of 'temporal correlation of non-conformance' and 'homophily of anomalies' are valid for such rarer anomalies.

On a related note, we attempted re-casting the non-conformance prediction problem as a NPP problem with $K = 1, 2, 3, and 4$. We do so by considering the tra-

Table 4.5: Comparison of Next Place Prediction performance under varying deviation conditions with $MC-2$.

	N_{all}	N_{low}	N_{high}	Acc_{all}	Acc_{low}	Acc_{high}
$K=1$	1350	677	673	0.41	**0.80**	0.00
$K=2$	1323	663	660	0.41	**0.75**	0.06
$K=3$	1296	649	647	0.42	**0.73**	0.10
$K=4$	1269	635	634	0.42	**0.71**	0.13

jectory deviation experienced by a user and his/her friends till the time of prediction as a feature in predicting the next place, in addition to the conventional features such as time of the day, day of the week and place (e.g., current and previous location 2nd order Markov prediction).

We first present the results for the NP predictor described in Section 4.3 (i.e., $MC-2$ using which we compute the trajectory deviation which is then used as input to the non-conformance predictor), in Table 4.5. We find that the accuracy, the percentage of cases where the next place was correctly predicted, to be $\approx 40\%$ overall. Further, if we separated the data into cases where the user experienced "low" or "high" deviation, based on (a) the user's deviation alone, or (b) the user and his/her ties' deviation, we see that the predictor is able to offer improved accuracy over the "low" deviation cases ($\approx 80\%$). This is anticipated as low deviation implies that the user has stuck to the *expected* or *normal* routine for that day.

Next, we train a Naive Bayes predictor with the time of the day, day of the week and last two places visited by the user as the independent features and the next place as the dependent feature (Baseline). Additionally, we consider the deviation (per day) by the user alone and the user and his/her friends, as additional features. We refer to this as the feature-enhanced NPP. We summarize the prediction results in Table 4.6. We see that the additional features provide only a marginal improvement over the baseline predictor. This brings us to the conclusion that even though it is possible to predict when a user might deviate from his/her routine mobility behavior, it may not be possible to predict where the user might deviate to, with the current set of features that we consider.

Table 4.6: Comparison of Next Place Prediction performance with/without deviation as additional features.

	Precision	**Recall**	**Correct Predictions**
Both User/Ties	0.652	0.652	65.19%
User only	0.66	0.66	66.04%
Baseline	0.659	0.659	65.88%

Table 4.7: Summary of LFNC Prediction Results using Different Classification Algorithms.

Classifier	$K=1$			$K=2$			$K=3$			$K=4$		
	AUC	Precision	Recall	AUC	Precision	Recall	AUC	Precision	Recall	AUC	Precision	Recall
Decision Tree	0.809	0.863	0.8	0.812	0.861	0.815	0.791	0.828	0.795	0.776	0.801	0.779
Naive Bayes	0.832	0.794	0.78	0.828	0.791	0.779	0.806	0.768	0.761	0.791	0.758	0.751
Logistic Regression	0.841	0.86	0.812	0.843	0.857	0.815	0.82	0.821	0.794	0.805	0.795	0.777
Random Forest	0.843	0.853	0.813	0.845	0.852	0.816	0.827	0.82	0.796	0.814	0.795	0.779

Other Applications and Alternate Anomaly Metrics: We believe that the ability to predict upcoming episodes of anomalous movement behavior can benefit many ubiquitous computing use cases, beyond the mobile crowd-sourcing application studied here. One such example is in dynamic calendaring applications, which can suggest schedule adjustments based on the attendance likelihood of participants. Another such example is *smart building energy management*, where predictions on the likely non-occurrence of regular meetings can help reduce energy consumption via proactive HVAC control techniques [30]. Both these cases, however, require more careful prediction of *collective LFNC* to determine the odds that multiple individuals will concurrently deviate from their normal mobility patterns.

Choice of Machine Learning Algorithm: All of our results presented to date use the GBM classifier. To study whether our insights on LFNC prediction are robust to the choice of classification technique, we conducted experiments (using the campus indoor mobility dataset) with additional shallow classifiers. Table 4.7 summarizes the key results for 4 different values of K (prediction window ranging from 15 mins—1 hour). We see that Logistic Regression and Random Forest classifiers seem to perform slightly better than the alternatives. However, the GBM classifier performs the best, achieving AUC of 0.9 for $K = 1$ (see Figure 4.8).

Chapter 5

Emerging Work: Exploiting Short-Range Mobility for Collaborative Sensing on the Edge

The accelerated growth in availability and deployment of consumer-scale Internet-of-Things (IoT) (e.g., cameras, microphone arrays & environmental sensors) presents many opportunities for predictive analytics for smart cities and homes.

Consider the scenario where community areas of housing developments of a futuristic 'smart" city are equipped with a large number of network-connected cameras. Such networked devices can potentially be used for sensing and analyzing various outdoor human activities. For example, as we motivate in the Introduction, they could be used, in the context of surveillance and public safety, to detect anomalous activities such as theft, or fall detection in the elderly. Such continuous streams of video data can also be used to further a city's understanding of how its space is utilized – for example, mobility behavior of the tenants of the housing estate can be observed over time to infer the specific neighborhood locations from which patrons visit–i.e., discover the local catchment profile. These applications require that the video streams are processed continuously. However, such heavy-duty processing often impose several resource bottlenecks.

Accurate inferencing of perceptual intelligence tasks increasingly involves the execution of computationally prohibitive machine learning (ML) pipelines – e.g., the state-of-the-art Deep Neural Networks (DNNs) for vision-based object detection require specialized, energy-hungry GPUs for low-latency operation. As a result, these devices (e.g., cameras) typically offload their computation to nearby, resource-rich computational entities such as cloudlets (over WLAN) or edge computing devices [106]. Whilst there are clear benefits to offloading computation, such augmentation suffers from several pitfalls: (1) in a multi-camera scenario, the sheer amount of raw video footage that needs to be offloaded can easily choke the available bandwidth for communication, (2) the relaying of footage to the cloud(let) and back can prohibit low-latency operation that is critical for certain classes of applications (e.g., tracking for surveillance), and (3) present edge computing models operate in an isolated fashion – i.e., each sensor node effectively utilizes its individual allocated computing slice on the edge device, and thus does not take advantage of the observations and intelligence that other nearby sensors possess.

However, a salient feature of these sensor-rich urban settings is that these sensors are often deployed with varying degrees of redundant coverage–e.g., cameras in buildings often have partially overlapping fields of view, implying that their sensed data are implicitly spatio-temporally correlated. This, we believe, opens up previously untapped opportunities for enabling scenarios such as those described here. We believe that through *Collaborative Sensing*, neighboring sensors can exploit short-range human mobility information to optimize their sensing and inferencing pipelines. For example, consider a scenario where two cameras located at consecutive intersections, a minute apart by foot. The first camera runs its heavy-duty inferencing pipeline to detect that there are persons walking within its view. In doing so, the first camera also captures characteristics or features of the *detected* persons – e.g., histograms of color intensity within the bounding box of the detected person that are less-computationally intensive to capture. Separately, due to the movement of people between the intersections, it is also possible for the two

cameras to *learn* over time, their spatio-temporal orientation – for e.g., based on the information that a person X leaves the view of the first camera from the top-left corner of the view at a particular time instant, it might be possible to *predict* the instant of arrival (e.g., 50 seconds later) and location of the same person X within the view of the second camera. Imagine a collaborative scenario where the pair of cameras *share* such intermediate *state* information such as locations of bounding boxes, timestamps and digests of the content (e.g., features, pixel values). Then, based on the information received from the upstream camera, the second camera can merely run the low-intensity processing of the image (e.g., color histogram) to do a simple match to detect that person. Only if the confidence of such a match is found to be low, then the camera can decide to run its full deep pipelines. Compared to the current edge computing paradigm where cameras offload raw footage, this mode of operation results in lower overhead to the communication infrastructure as only light-weight information is being shared among collaborating peers. From a high-level paradigm, this approach exploits inherent trade-offs between communication and computation in such a distributed sensing infrastructure. While this adaptive operation can potentially lead to savings in processing times, multiple peers can also share their inferences for improved joint-inferencing accuracy.

In this chapter, I describe preliminary work towards realizing this vision. We take the case of cameras with fully, or partially overlapping views, where the same persons appear within multiple views, concurrently. This is a special case of the collaborative scenario above where there's no longer a need for *predicting* likely locations and times of incoming person objects. Using person detection as an illustrative example, in the remainder of this chapter, I describe the following discoveries which are important building blocks of the *Collaborative Sensing Vision*.

1. **Improving accuracy:** We investigate alternative architectures for joint inferencing between collaborating peer cameras/views. We introduce two architectures: (1) $CSSD$ where inferences from a collaborator camera are passed

along as additional input to a modified SSD pipeline, and (2) $CNMS$ where inferences from a camera and its collaborators are post-processed by an additional Non-Maximum Suppression step to output the final inference. We show that by using a camera's own view and a simulated peer view on $CSSD$, accuracy improvements over 7-8% can be achieved for only a marginal increase in processing latency.

2. **Improving processing latency:** Next, I show that by identifying and exploiting redundancies in the SSD pipeline, a comparable inference accuracy can be achieved in running only as little as the first of the 23-layer SSD network (counting convolutional and fully connected layers alone) leading to a reduction of over 82% in processing time. I achieve this by identifying feature maps in the intermediate layers that correspond the highest with the final inferences of the pre-trained pipeline.

3. **Optimizing for accuracy and latency:** Further, I explore the possibility for joint improvements in accuracy and processing latency. I explore techniques for combining peer knowledge with the reduced version of the SSD pipeline (which I refer to as $SSD - 1$ due to the early discard of the pipeline at layer 1) for minimizing loss in the dimensions and localization of the inferences. The combined model ($ESSD$) achieves as much as 3-10% improvement in accuracy whilst saving 82% of the processing time.

As part of ongoing work, as a crucial next step, the learned correlations in human mobility, as observed by neighboring cameras, are being explored to extend the above contributions to the generalized scenario consisting of a mixed deployment of overlapping, as well as non-overlapping, but temporally correlated cameras.

5.1 Preliminaries

In this section, I describe some of the state-of-the-art object detection pipelines and the multi-view benchmark dataset that I use throughout this chapter.

5.1.1 Object Detection using Deep Pipelines

Earlier works in object detection typically employed sliding window approaches for classifying and localizing objects in images. They relied on hand-crafted features (e.g., Histogram of Gradients (HOG)[29], SIFT [80], SURF [17], integral channel features [35], etc.) that are calculated over such sliding windows. Then a supervised, albeit *shallow*, classification algorithm such as a Support Vector Machine, applied across the sliding window is used for detecting object locations. While the combination of HoG and SVM was primarily used for pedestrian detection, DPMs [42] further extended the detection task over more generic object categories. This was considered the state-of-the-art for many years, bench-marked on the PASCAL VOC dataset [40].

With the explosion of interest in deep neural networks (DNNs) in the last decade and their suitability for computer vision tasks, object detectors with superior performance have emerged. Earlier work on deep learning based object detectors exploited the idea of Selective Search [123]. In contrast to the sliding window approach which looks for objects across all locations in an image, selective search focuses only on a selected set of regions. In the R-CNN [50] network, these regions are extracted from region proposal networks (RPN). Improvements over R-CNN [49, 100] were shown to achieve higher accuracies (of up to 66% mean average precision, mAP, on PASCAL VOC 2012). Later, Faster R-CNN [104] combines the region proposal network and object localization into a single deep CNN network achieving state-of-the-art performance of 75.9 mAP on PASCAL VOC at the time.

The current state-of-the-art in object detectors are primarily "one shot detectors"

with Overfeat [113] being one of the first to emerge. Recent detectors such as YOLO [102, 103], SSD [79] are optimized for faster run times whilst preserving the accuracy of detectors such as RCNN which make them ideal for online, real-time needs. We take the SSD pipeline as our illustrative deep learner in this work with VGG16 [116] as the base network (which acts as a deep feature extractor). Together, the pipeline consists of a total of 23 convolutional and fully connected layers (with other layers such as max pooling, normalization, etc. acting as intermediaries). In the remainder of this chapter, we refer to these layers as Layer L from 1 to 23. We build on the implementation of the Single Shot Detector (SSD)[1], originally proposed by Liu et al.[79] and implemented using a Keras port[2] by Pier Luigi Ferrari.

5.1.2 Benchmark Data

We use video feeds from the the PETS 2009 dataset [43] that consists of the deployment of eight synchronous cameras. The cameras are placed outdoors and they capture pedestrian traffic under varying crowd density conditions. In most of our experiments, we limit our focus to two of the eight views – View007 is considered as the *reference* camera and View005 is the collaborator camera with respect to which we report all our findings in later sections. We pick these two cameras as their views share the highest overlap over all pairs of annotated views (see Section 5.4). The manually annotated ground truth was provided by Xu et al.[130] for 5 views (View001, View005, View006, View007, and View008). The videos were captured at an approximate frame rate of 7 FPS and a resolution of 720×576 resulting in a total of 795 frames per camera (3180 frames across the 5 cameras). I limit our experiments to the "person" detection task alone.

[1] Available from `https://github.com/pierluigiferrari/ssd_keras`
[2] `https://keras.io/`

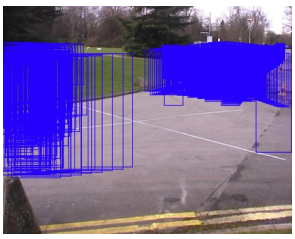
5.2 Opportunities for Collaborative Video Sensing on the Edge

In this section, I describe two key observations for enabling collaboration and processing at the edge between co-located cameras in dense deployments.

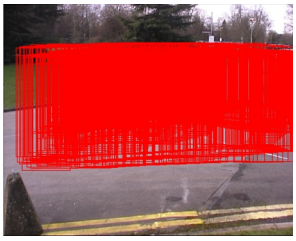
5.2.1 Correlations in Content between Co-located Cameras

As I described earlier, one key observation in camera deployments is that, typically, neighboring cameras share overlapping views. Hence, ideally, a group of cameras should be able to jointly process the images for better accuracy. For instance, occluded objects in a view may be visible from other views and combining both observations should lead to more accurate inferences. On the contrary, if a single camera view is processed to detect objects within its own view, then other cameras in the network shouldn't have to re-do the resource-intensive computations, at least for the parts of the view that there is significant overlap. In this subsection, I attempt to quantify the presence of such spatio-temporal overlap between views in the benchmark dataset.

I choose a pair of cameras (View007 and View005) from the PETS 2009 dataset for illustrating this. Figure 5.1 shows bounding boxes of persons in the view of "View007" that is (a) only visible to View007 alone, and (b) visible to both View007 and View005 (a neighboring camera), respectively. It is evident that there exists significant overlap between the views – for instance, at least 72% of the persons boxes appear in both views simultaneously, corresponding to a region of overlap of $\approx 12\%$ of View007's total area. In Table 5.1, I tabulate the overlap percentage of View007 with its neighboring cameras in the dataset. As I mention in the Introduction, while joint processing of overlapping views does not directly use human mobility, it helps establish the mechanics of collaborative inferencing. We anticipate that our future work will leverage on these established mechanisms for cameras that are not re-



(a) View 7 Only



(b) Both View 5 and 7

Figure 5.1: Bounding boxes of persons in the view of a representative camera (Camera #7 in PETS'09) (a) which were within the purview of only itself vs. (b) which were in the purview of both itself and one of its neighboring camera (Camera #5 in PETS'09).

	View005	View006	View008
Total person instances	1867	1689	1741
Co-occuring person instances	2562	2762	3919
ROI area	11.52%	5.28%	20.86%

Table 5.1: Overlap of views in the dataset with the reference view (View007) in terms of co-occurring person instances and overlap of area within the view.

stricted to overlaps spatially but have broader temporal correlation, and thus take into account human movement behavior.

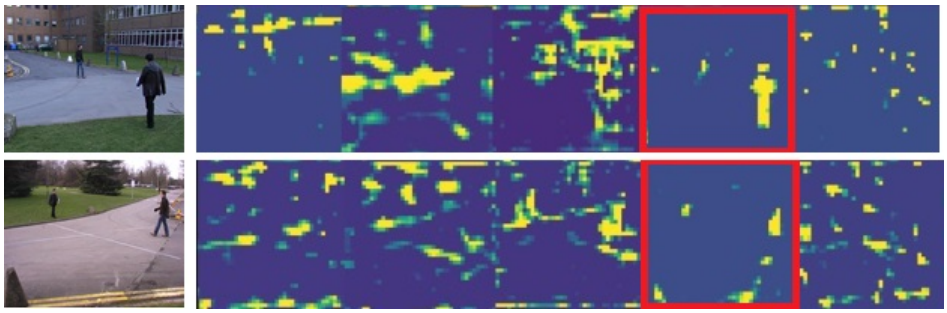
5.2.2 Interpretable Output from Intermediate Convolutional Layers of Deep Pipelines

The state-of-the-art learners for vision tasks such as object detection are currently all deep networks. During the training phase, input images go through a cascade of filters (e.g., 2D convolutions) over multiple intermediate layers, while the optimal weights of such filters are *learned*, iteratively, by minimizing the loss (e.g., cross-entopy) between annotated ground-truth bounding boxes and the predicted boxes at each iteration. One key aspect of such networks that we observe is that, despite the depth, trained networks hold interpretable, useful information in intermediate layers. For example, in Figure 5.2, the activation values of the final convolution layer of the VGG16 network (named as $conv4\_3$) are visusalized – the VGG16 network acts as the feature extractor, or base, network of the SSD architecture (amounting to a total depth of 23 convolutional and fully connected layers). The figure

shows two sample frames from View005 and View007, respectively, and activation values from five hand-picked feature maps output by $conv4\_3$. The network was pre-trained to detect the $20 + 1$ classes on the PASCAL VOC dataset. In the fourth feature map (bounded by the red rectangle border), we note that the high valued activations (marked in yellow) correspond well with the positions where the persons are present within the two sample frames. This observation leads us to believe that although execution through the "depth" of the learned network is vital to achieve high inference accuracy (compared to shallow learners), the intermediate layers, including the early layers of a deep network, may still contain useful information.

Further, in Figure 5.3, we plot the percentage completion of the deep SSD pipeline executed (in terms of time taken), on the $y-$axis as the pipeline progresses through its 23 convolutional layers (on the $x-$axis). I note that this particular network is non-linear, and that for example, by the time the VGG16 feature extractor completes execution, almost 70% pipeline would have finished running. This means that if we hope to save compute cycles by early discard (i.e., intercepting the network), the pipeline has to be interrupted by at least the 3rd convolutional layer to save at least 75% of the execution. In the case of SSD, we note that this nonlinearity exists due to the design of the architecture where it is composed of six prediction layers that detect objects at progressively increasing scales which results in progressively decreasing computations per layer.

In Figure 5.4, we plot similar execution timings of alternate, competing architectures for object detection – YoLo v2 and v3, respectively. Compared to SSD, we observe that the execution times are more linear with increasing depth for YoLo, especially for YoLo v3. This means that in order to achieve the same amount of processing time, the depth to which the network could be run can be more compared to SSD which may have implications for the accuracy with which we can infer at the intermediate layers. Hence, identifying deep networks that are compatible with the notion of "early discard" and profiling of depth vs. accuracy trade-offs of such networks is an important future line of inquiry.



The intermediate feature maps were extracted from the tenth convolutional layer of the VGG base network that SSD300 utilizes. The feature maps marked in red show higher correspondence to the "people" seen in the raw images.

Figure 5.2: The scene observed by two cameras, and representative feature maps extracted at an intermediate layer of the SSD300 pipeline for View005 (top) and View007 (bottom).

5.2.3 Key Challenges

In order to exploit these two key opportunities, the following challenges have to be addressed:

1. **Learn spatio-temporal correlations among devices/views:** To exploit such correlations, the individual camera nodes have to first discover such correlations by sharing appropriate features from their underlying sensor data streams with other camera nodes in the deployment.

2. **Enable adaptive operation of ML-based frameworks:** Based on context shared from peers, the individual cameras should learn to adapt their execution at run time, to improve accuracy and/or processing latency.

3. **Minimize bandwidth overheads:** As IoT devices such as the cameras start to share information amongst their peers, this could potentially pose network bandwidth bottlenecks. Hence, the optimal information to be shared needs to be identified which is useful but also minimizes the load on the network.

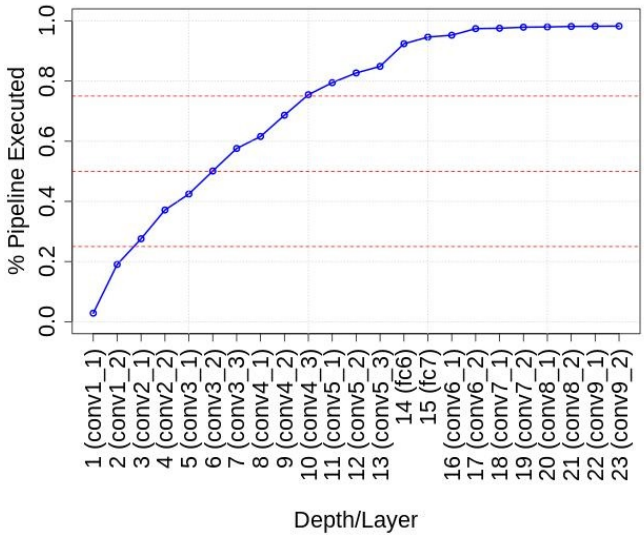
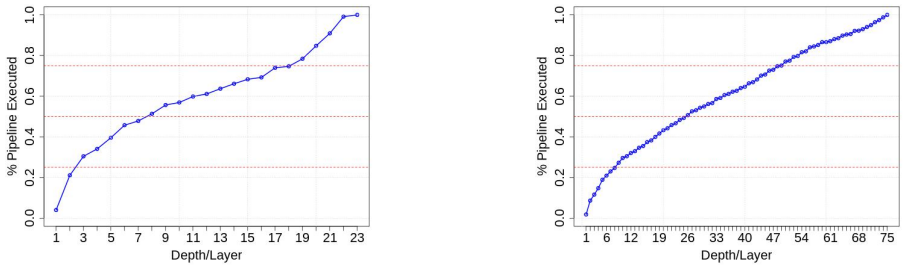


Figure 5.3: The execution progress timings (as percentage of the total run time of the full SSD pipeline) with increasing depth. The x-axis shows only the convolutional and fully connected layers for brevity.



(a) YoLo V2 (b) YoLo V3

Figure 5.4: Execution progress of the YoLo object detectors with increasing depth. The x-axis shows only the convolutional layers for brevity.

5.3 System Overview

In Figure 5.5, we illustrate our proposed system for collaborative sensing. The architecture involves a central control node and a network of peer cameras. Node X refers to a generic peer node which receives control commands from the Control Center and state information (e.g., inferred bounding boxes, statistical features, etc.) from other peer nodes in the network. I summarize the roles of the individual components below.

Central Controller: This module acts as the broker between peer camera nodes. The controller could be a camera node itself, or a more resource-rich device

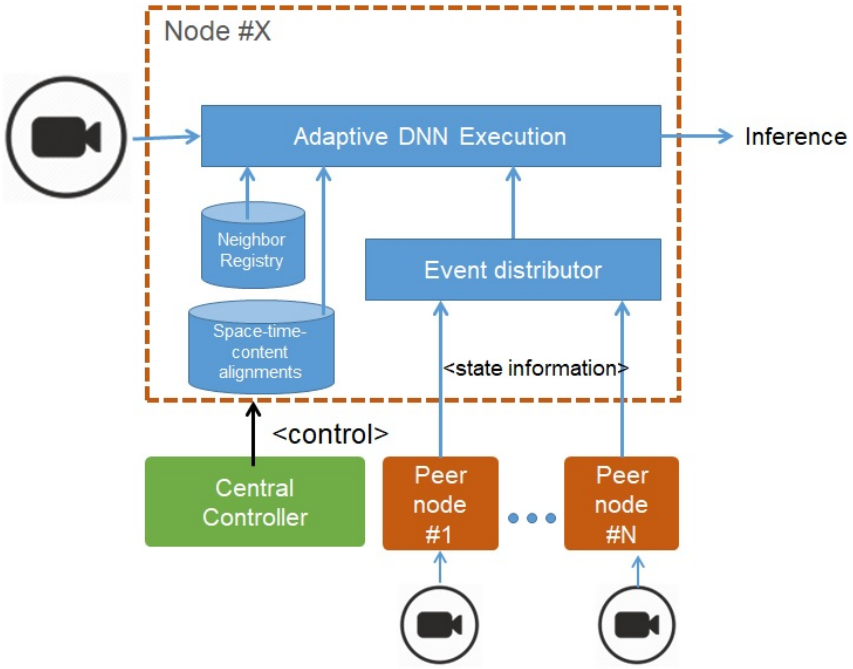


Figure 5.5: Proposed Architecture for Collaborative Sensing on the Edge

such as a cloudlet [106, 107] like we discussed, previously. The module frequently receives inference results and state information from the other peer nodes to perform several functions. It acts a match-maker by discovering *which* nodes should collaborate, based on the longitudinal observations of the individual camera's detections and the resulting correlations. In Section 5.4, we describe one such methodology to use such correlations to establish the orientation or mapping between views. It is also in charge of setting and updating the mode and various other configurations of the network – for instance, I discuss two alternate modes of collaborative operation in Section 5.5 that enables a camera to utilize additional information from its peers to improve its inferencing ability. In Section 5.6, I describe configuration parameters such as the depth to which cameras should run their DNN pipelines and the choices of feature maps that allow them to make the most accurate inferences that the Controller learns over time.

Peer Nodes: We assume that each peer node has processing capabilities onboard (e.g., CPU or a mobile Vision Processin g Unit (VPU)) that allows them to execute full, or partial, DNN pipelines. The peer nodes directly exchange send/re-

ceive state information to/from chosen peers. Depending on the mode of operation, the nodes use the information from peers as input, in addition to the node's own view of the scene, to run collaborative DNN pipelines. The nodes consist of two data stores: (1) the Neighbor Registry which contains look-up information on local addresses of collaborating peers, additional information such as the peers' reputation scores [128], etc. and (2) the Space-Time Alignment Mapper which consists of information key to translating inferences of a peer's view to a node's own view. The Event Distributor in the node acts as a scheduling service – for instance, for a person predicted to arrive at a future time instant (using peer inference), the distributor marks the relevant information such that the node uses it only at the appropriate moment in future.

5.4 Spatial Mapping Between Views

In order for the cameras to collaborate, a prerequisite is in learning the mapping between the coordinate systems between the pairs of views. Existing techniques such as homography transformation [55] provide a way to map coordinates from one view to another under the assumptions that the world scene is planar, or that the two views are generated by a camera rotating around its center. However, as these assumptions do not hold true in our case, we propose here an alternative approach.

We rely on the ground-truth annotations provided to match the same individual across two views. Then, we generate a dataset which consists of only person instances that occur across both views *simultaneously* and within the overlapping Region-of-Interest. We train four regressors where the independent variables are: (1) bottom $y-$coordinate (representing the ground plane, $ymax$), (2) middle $x-$coordinate ($midx$), (3) width and (4) height of the bounding boxes appearing in the collaborator camera (for e.g., View005). The dependent variables are the same 4 attributes of the bounding boxes in the reference camera's view, one for each regressor. By trial and error, we find that while the $ymax$ and $xmin$ values can be

	View005	View006	View008
Adjusted R^2 (xmid)	0.977	0.981	0.993
Adjusted R^2 (ymax)	0.961	0.915	0.988
Adjusted R^2 (width)	0.946	0.982	0.991
Adjusted R^2 (height)	0.963	0.732	0.986
Overlap ratio	0.764	0.628	0.741

Table 5.2: Goodness of fit of the learned regressors for estimating bounding boxes from View005, View006 and View008 to View007.

estimated sufficiently accurately with single order regressors, the latter two require higher order polynomial representations (e.g., of order 2).

We randomly split the dataset into 80% training and 20% test, and use the training set to estimate the regression coefficients. For the remaining frames in the test set, we essentially *estimate* the bounding boxes in View007's view using inferences from View005 using the four regressors. To measure the correctness between the estimated and actual bounding boxes, we measure the overlap ratio (measured as Intersection-over-Union). In Table 5.2, we tabulate the adjusted R^2 values of the trained regressors and the average overlap ratio on the test set for the different collaborator views with the reference view, View007.

We observe that in all except one case the R^2 values are above 0.9 meaning that 90% of the variance in the reference camera's coordinates can be explained through the collaborator camera's bounding box coordinates. Moreover, the average overlap ratios are above 0.7 for View005 and View008 which is considered high – for instance, as we describe in Section 5.5, state-of-the-art object detectors use 0.7 as the threshold above which bounding boxes are assumed to be the same in the post-processing step of Non-Maximum Suppression.

We recognize that the mapping step requires high accuracy for the collaboration mechanism to be advantageous. Erroneous mappings can lead to worse performance than what an independent camera might be able to achieve by itself.

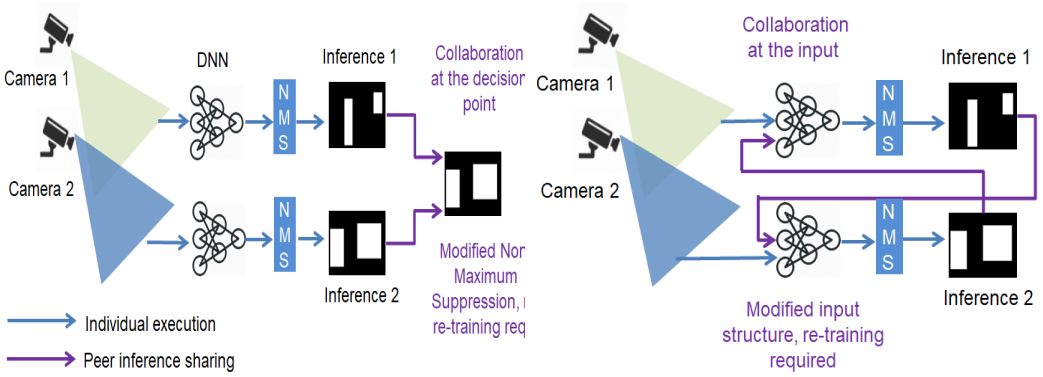
5.5 Collaborative Inference with Peer Nodes

In the **baseline** operation of the SSD pipeline, each camera first executes it's people detection algorithm–i.e., it feeds each input frame to the DNN to output a set S, representing a set of people objects, along with the rectangular bounding box coordinates for each of these objects. The SSD pipeline accepts raw image pixels as input and generates multiple feature maps at different scales, which are subsequently provided to a downstream regression and classification network for (a) detecting object classes, (b) computing bounding box coordinates and (c) deriving the associated confidence scores. Finally, the network combines these outputs from multiple feature maps to execute the Non-Maximum Suppression (NMS) step. NMS is used to avoid over-counting: it identifies bounding boxes that have significant spatial overlap and effectively unifies them into a single bounding box (corresponding to the one with the highest confidence value). In practice, such unification is performed by computing the Intersection over Union (IoU) metric, which is defined as: $Area\,of\,Overlap/Area\,of\,Union$, and which effectively describes the fraction of the area that is common to a pair of bounding boxes. Two boxes are declared to be equivalent if the IoU value exceeds a minimum threshold (e.g.,$\geq$0.7). Then, the inferred output is the remaining bounding boxes post-NMS.

For the collaborative operation, we explored two different architectures, as illustrated in Figure 5.6. In both designs, we first require that the inferences from a camera are transformed to the coordinate system of the collaborating camera (for e.g., as described in Section5.4).

5.5.1 Collaborative Non-Maximum Suppression (CNMS)

As mentioned before, Non-Maximum Suppression (NMS) is a post-processing step that suppresses (eliminates) bounding boxes that are deemed to be redundant. The CNMS approach performs collaborative fusion only by modifying this final NMS stage of the DL pipeline). As illustrated in Figure 5.6a, CNMS takes the output



(a) Collaborative NMS (b) Collaborative SSD

Figure 5.6: Architectures of (a) *Collaborative NMS* vs (b) *Collaborative SSD* models

bounding boxes of the SSD and *modifies* the output confidence value taking into account the overlap with bounding box coordinates obtained from each peer camera. Intuitively, CNMS will assign higher confidence scores to the bounding boxes (of the reference camera) that have high spatial overlap with the peer-generated bounding boxes. More formally, we first convert peer-generated bounding box coordinates to reference camera coordinates. Let us assume peer camera generates m bounding boxes of $P = \{p_1, p_2, ..., p_m\}$, while reference camera generates n bounding boxes $R = \{r_1, r_2, ..., r_n\}$ with n corresponding confidence scores represented by $C = \{c_1, c_2, ..., c_n\}$. We then compute $IoU(p_i, r_j)$ for all $i \in [1, m]$ and $j \in [1, n]$. If $IoU(p_i, r_j) > 0.8$, then we assign a new confidence score ${c_j}^{new} = (1 - \alpha) * c_j + \alpha * IoU(c_i, r_j)$, where $\alpha = 0.2$. Similarly, we execute the same confidence update procedure for multiple cameras and then we run the NMS step on this updated confidence scores.

5.5.2 Collaborative SSD (CSSD)

Unlike the CNMS approach, the CSSD technique modifies the structure of the SSD pipeline itself. Figure 5.6b illustrates the high-level CSSD approach: the input to the DNN is not just the individual video frames from the reference camera, but also an additional *input mask* with probable locations of objects (as indicated by the collab-

orating peer cameras). Assume that the peer camera generates m bounding boxes $P = \{p_1, p_2, .., p_i, .., p_m\}$ where $p_i = ({x_i}^{min}, {y_i}^{min}, {x_i}^{max}, {y_i}^{max})$ (coordinates of the top left and bottom right of the bounding box). We then initialize the input mask of size $(300, 300)$ with zeros. Assume $pixel_{i,j}$ denotes pixel value of input mask at location (i, j). Then $\forall\ p_k(k \in [1, m])$, assign $pixel_{i,j} = 1$, if ${x_k}^{min} < i < {x_k}^{max}$ and ${y_k}^{min} < j < {y_k}^{max}$. In effect, we complement the 3 channels of a regular RGB image with N additional channels (each a grayscale object mask from one of N collaborating cameras), creating an input image of $N + 3$ channels. Similarly, we do the same computation for multiple peer-cameras. Finally we pass the video frame from the reference camera concatenated with the computed input mask into the CSSD DNN network (see Figure 5.6b).

Contrary to the case of CNMS, the latter requires the SSD network to re-trained from scratch – as the input structure of the network is altered. To compare the performance of the collaborative models over the baseline in terms of accuracy, we conduct a **simulation study**. We considered all 3180 annotated frames in the dataset (from cameras 001, and 005 to 008) as image frames from the reference camera, and randomly perturbed versions of the corresponding ground-truth annotations (makred as bounding boxes) as the input from a "neighboring" camera. We tabulate inference accuracy under two the different designs in Table 5.3. These perturbations were random manipulations of the four coordinates of the ground-truth bounding boxes ($xmin$, $ymin$, $xmax$, and $ymax$) by 0 to 2 pixels (with equal probability) for half of the frames chosen at random, designed to *simulate* errors resulting from the coordinate mapping procedure. We see that the joint inference among the simulated peers lead to at least a 5 to 12% improvement in accuracy. However, we emphasize that this result is achieved only through a simulation. Later, in Section 5.7, we find that the accuracy improvement achievable, with inferences from actual peer cameras that are spatially mapped to the reference camera's view, was only marginal ($\approx$ 3% with $CSSD$) – this is in part due to the susceptibility of the network to the inaccuracies in the spatial mapping.

	Baseline	CNMS	CSSD
Accuracy/F Score	70%	75.5%	82.5%
Processing Latency	80ms	100ms	85ms

The processing latency values were recorded on Raspberry Pi 3 B+[a] running the DL execution on the Movidius Neural Compute Stick[b].

[a] https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/

[b] https://www.movidius.com/solutions/vision-processing-unit

Table 5.3: Comparison: CNMS vs CSSD

5.6 Early Discard of the Deep Pipeline

While we explored the opportunity to collaborate for improved accuracy in the previous section, here we focus on the possibility to improve on processing time by exploiting the redundancy within the deep pipeline.

As we speculated earlier, we believe that feature maps from intermediate layers can help in revealing inferences in the early layers. In order to discover feature maps at each layer that correspond the most to the final inference, we do the following:

1. For a selected frame(s), we compute the binary mask of activations for each feature map (fmap), at each convolutional layer of the network. We take the output of each feature map, re-target it such that the resulting resolution is the same as the original image, and mark the pixels with values greater than the 80^{th} percentile of the map with "1"s and the remainder with "0"s. Re-targeting will require interpolation as the resolution of the feature map is usually lower than that of the original image.

2. Then, for each binary fmap, we compute the correspondence with the final inference which is also a binary image with pixels within the detected bounding boxes marked with "1"s, and the remaining areas with "0"s. The correspondence is calculated as the Intersection Over Union (IoU) of the "1"s in the two binary masks.

3. Finally, for each layer, we pick the fmaps with the highest correspondence as the "best fmaps". These outliers are detected as those with IoU $\geq$ 1.5*inter-quartile range over the third quartile. In Figure 5.7, we provide a bar plot of the IoU values at each convolutional/fully-connected layer for the SSD network. We also note that the middle layers contain the fmaps that correspond the most to the final inference although the initial layers also contain a few fmaps that correspond with IoU $\geq$ 0.1.

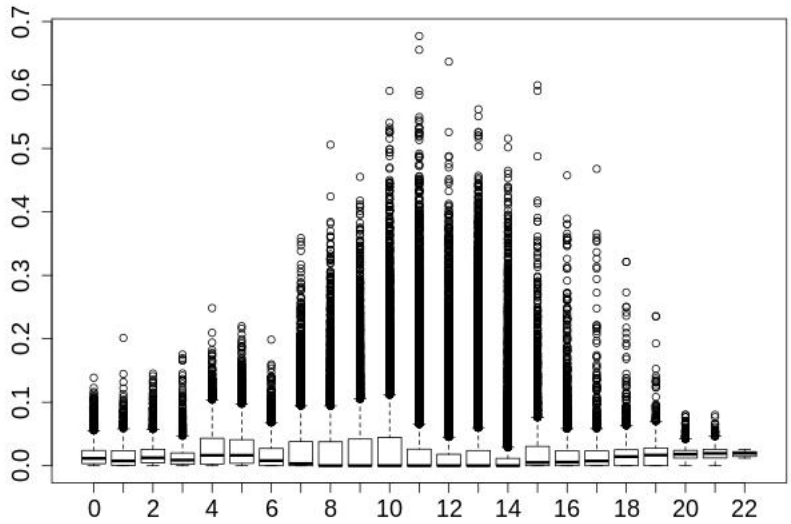


Figure 5.7: Distribution of IoU values of intermediate featuremaps at every convolutional/fully-connected layer of the SSD pipeline, against the final inference, for a sample image.

Now, in Figure 5.8, we plot the intermediate output of these chosen fmaps at each layer, averaged out. We observe that even at the early layers (e.g., $L = \{0, 1\}$), the fmaps are able to uncover parts of the image corresponding to person objects. The activations associated with the background scene (e.g., trees behind) are easily eliminated using an appropriate background subtraction method.

This observation leads us to believe that a simple blob detection routine applied on these intermediate outputs can match the output of the SSD pipeline. We refer to this improvised model as $SSD - l$ where l represents the l^{th} layer. The routine

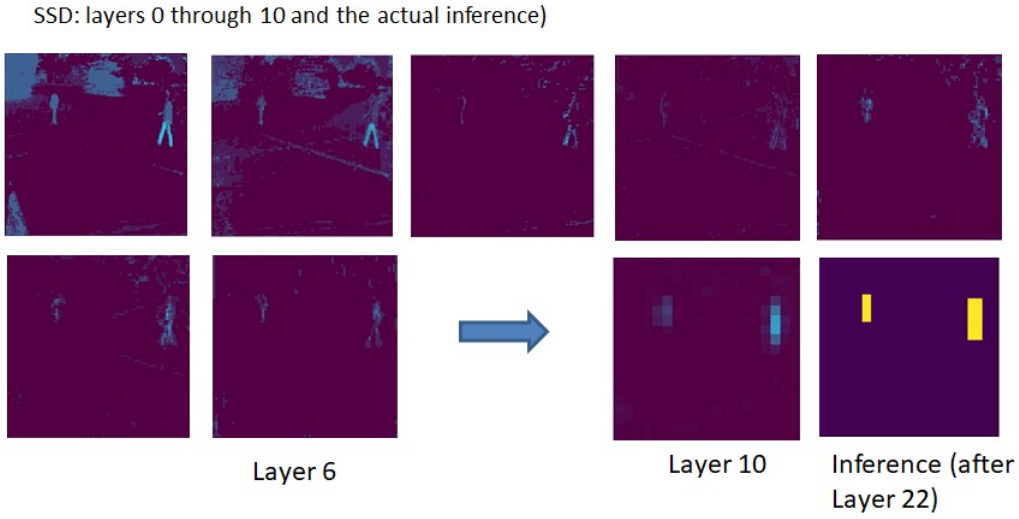


Figure 5.8: Intermediate activation values averaged over the chosen "best" fmaps for selected layers through 1 to 23.

is as follows:

1. Run the deep pipeline till Layer L, and compute the average mask of the chosen best fmaps from the previous stage.

2. Similarly, compute the average mask of the background image and perform background subtraction.

3. On the resulting image, perform blob detection and output as bounding boxes.

However, we note that an artefact of the blob detection process is that it would not be able to distinguish between individual persons appearing close together – i.e., their bounding boxes would be overlapping, and the detector would be declaring them as a single person as opposed to many.

5.6.1 Improving Accuracy through Collaboration

To overcome the loss of precision from the previous step, we propose $ESSD$. One of our main observations from the blob detection stage is that the output (i.e., from $SSD - l$)) is perfectly accurate in terms of localizing person objects. For the same

frame, if we relied on the *regressed* bounding boxes from a peer's inferences (see Section 5.4) within the common ROI, we see that although the localization suffers from errors, the dimensions of the detected person boxes are accurate. Hence, we devise a rudimentary dimension adjustment routine to combine both inputs as follows:

1. For each frame, we take bounding box coordinates from $SSD - 1$.

2. Then, we compute the average dimensions (height and width) of the bounding boxes from the peers, translated to the reference camera's view.

3. Height adjustment – bounding boxes that are at least $T_h = 50\%$ less than the average height of the peer inferences are considered as candidates for "merging". Such candidates that are vertically aligned (with an error margin of 10 pixels, for example), and are within half the average height apart are merged and the bounding box coordinates are revised. The process continues iteratively.

4. Width adjustment – bounding boxes that are at least $T_w = 50\%$ wider than the average width of the peer inferences, are considered as candidates for "splitting". Such candidates are split into two separate bounding boxes, and the process continues.

In Section 5.7, we share results in the improvement in accuracy of the $ESSD$ model over the baseline as well as the other models.

5.7 Evaluation

In this section, we present findings from improvements in accuracy and processing latency by using our approach. All experiments were run on a Intel Xeon server with 128 GB memory and up to 14 processors – but the experiments utilized a single core used for processing to mimic a typical edge node.

	Full view			Within ROI Only		
	Precision	Recall	F-score	Precision	Recall	F-score
Full SSD	95.72	56.98	71.44	93.76	52.38	67.21
$SSD-1$	82.86	63.28	71.76	85	59.27	69.94
$ESSD$	83.86	66.64	**74.26**	87.05	66.69	**75.53**

Table 5.4: Comparison of Person Detection Accuracy

5.7.1 Accuracy

In Table 5.4, we tabulate the person detection accuracy under various modified configurations of the SSD pipeline. Full SSD is the baseline where the full pipeline is run without any modifications. On View 007, the SSD achieves an $F-$score of 71.44% over the entire view, and 67.21% while only concentrating on the ROI which overlaps with the collaborator, View 005's view. The second row corresponds to $SSD-1$ where the SSD pipeline is executed only till the convolutional layer which achieves comparable accuracy. However, we note here that the precision of $SSD-1$ is at least 12–13% over the baseline whilst achieving a better recall. This is an artefact of the sensitivity of the blob detector as we describe previosuly.

Finally, the third row corresponds to the dimension–adjusted blob detection ($ESSD$) inferences enabled via collaboration. Compared to the baseline we see at least a 3% improvement in $F-$score over the entire view, and a more significant 8% increase within the ROI. In summary, we see that by including information from collaborators' inferences, that the accuracy can be improved.

We also note that $CSSD$ which is optimized for higher accuracy (but take longer processing time than the baseline SSD) achieves only a marginal improvement in accuracy over the baseline (F-score of 73.3%) under real conditions.

5.7.2 Processing Latency

While we see an improvement in the overall accuracy through collaboration, the ability to infer at an earlier stage of the SSD pipeline also significantly reduces the

Processing Step	Processing Latency
Baseline: Full SSD	427.96
Our Approach: ESSD	78.99
SSD till Layer 1 ($SSD-1$)	48.9
Fmap mask generation	15.83
Blob detection	0.45
Regression X 4	10.76
Dimension correction routine	3.05

Our approach provides greater accuracy at ≈ 82% reduction in processing time. The processing times of the individual steps of the pipeline are captured from row 4 onwards.

Table 5.5: Average processing latency (in milliseconds) of running person detection using the baseline SSD vs. our approach.

processing latency.

In Table 5.5, we tabulate the run time of the full SSD pipeline (row 1) against our approach (row 2). We see that the modified pipeline only takes 79 msecs, on average – this is a 82% reduction compared to the baseline (≈428 msecs). In the bottom rows, we tabulate the average run time of each of the individual steps of our approach. Although, theoretically we expect the run time of running $SSD-1$ (i.e., $conv1_1$) should only 1% of the full SSD pipeline (see Figure 5.3), in reality we only achieve a 89.55% savings (and not 99%). We believe this is in part due to the inner workings of the Keras wrapper that we use in our implementation, and a native implementation may allow for additional optimizations. Out of the remaining steps, the average mask generation and the regression take the next most highest run time. Since we run four separate regressions to detect the top-left corner coordinates, width and height, the time is multiplied by four. At Layer 1, the number of feature maps chosen as "best" is ≥ 5, currently. However the deeper in the pipeline we execute, it might be possible to choose fewer fmaps (see Figure 5.7 where the IoU increases towards the middle layers) for obtaining comparable inference accuracy.

5.7.3 Memory Requirements

By discarding the pipeline early, the deep model is now required to only execute the first few layers. Which means that only the weights corresponding to those early layers need to be loaded into memory for execution which is especially useful for devices (such as the Raspberry Pi) with limited computing resources. In Table 5.6, we record the size of the weights in memory and the number of parameters/weights involved for Layers 1 through 23 (i.e., corresponding to the 23 convolutional and fully connected layers). For instance, $L = 3$ refers to running the model *till* $L = 3$ which includes *ALL* layers till then (including max pooling, normalization, etc.). As such, $ESSD$ only requires 27.36 MB in memory which is 81.5% reduction compared to the baseline, allowing it to run 82% faster for comparable accuracy, as we saw previously.

5.8 Open Issues and Challenges to Be Tackled

In this chapter, I introduced the paradigm of collaborative sensing and how the redundancy in the sensed signals, across multiple sensors, can be exploited for improvements in the operation of the sensing architecture. In this preliminary work, I show early evidence of various possibilities to improve (a) the inference accuracy through collaboration, and (b) cut down on the processing latency of deep architectures using run-time optimizations, on videos from cameras with partially overlapping views. A key open challenge is thus extending these optimizations or *building blocks* to scenarios where such cameras do not necessarily possess overlap amongst them, but are merely correlated temporally due to close proximity. In such scenarios, we would thus require the explicit modeling of the *short-scale mobility* of moving subjects (e.g., humans for the person detection task). While the redundancy across cameras are assumed to be simultaneous in the current evaluations (due to the overlapping nature of the views where the same person appears across two cam-

Depth of Model	In-Memory Size of Model (MB)	Number of Parameters
conv1_1	27.36	1,792
conv1_2	50.40	38,720
conv2_1	67.68	112,576
conv2_2	79.20	260,160
conv3_1	87.84	555,328
conv3_2	93.60	1,145,408
conv3_3	99.36	1,735,488
conv4_1	103.80	2,915,648
conv4_2	106.75	5,275,456
conv4_3	109.71	7,635,264
conv5_1	111.19	9,995,072
conv5_2	111.93	12,354,880
conv5_3	112.67	14,714,688
fc6	114.89	19,434,304
fc7	116.36	20,483,904
conv6_1	116.73	20,746,304
conv6_2	117.39	21,926,464
conv7_1	117.44	21,992,128
conv7_2	117.54	22,287,296
conv8_1	117.55	22,320,192
conv8_2	117.56	22,615,360
conv9_1	117.57	22,648,256
conv9_2	117.57	22,943,424
Full	121.68	26,285,486

Table 5.6: Memory footprint of the deep learning model with increasing depth. *conv*–convolutional layer, *fc*–fully connected layer.

eras at the same time), in the extended scenario, the modeled temporal correlations would then be required to *predict* when a person appearing in one camera's view will appear in a nearby camera's view at a future time. Nevertheless, the designs presented in this current work form the basic building blocks for such future work. Below I outline some of the other open challenges in realizing this proposed work.

Scheduling, synchronization and unified inference: In a multi-camera network, the spatial and/or correlations of a camera with its neighboring cameras vary. For instance, the scene observed by a single camera will be a combination of (a) deterministic regions of interest shared by its neighbors with partially overlapping views, (b) probabilistic, future points of interest from its neighbors that are temporally correlated, and (c) regions that are only visible to the camera itself. An open challenge here then is in selectively triggering deep/shallow execution pipelines to piece together a final inference based on the combined observations in such a manner that the overhead in collaborating is kept minimal in comparison to the savings from avoiding redundant, complex computations.

Discovering collaborators: In this preliminary work, we assume that the cameras know a-priori which other cameras exist on the network and which cameras are candidates for collaboration (e.g., from the geographic positioning of the cameras, or ground-truth annotations of identified persons appearing across cameras). However, in practical deployments with a large number of cameras, collecting such manual annotations may not be scalable. A potential direction would then be for the cameras to run person re-identification networks [120, 56] to automatically detect the presence of the same person across multiple cameras (and perspectives) as substitutes for such annotations. Based on high confidence, longitudinal observations of such cross–camera person re-identification, spatial and temporal correlations can then be established to identify "collaborators". Such a *self-learning, self-calibrating* property would be highly appealing for large-scale, dense environments.

Generalizing to other Deep Models: In the current work, I present findings from experiments on the state-of-the-art SSD object detector. However, as I show

in Section 5.2, competing detectors have differing processing latency vs. network depth profiles. Investigations into the types of deep models that are *compatible* with operations such as early discard need to be explored.

Generalizing to other sensor modalities: In the current chapter, I discuss collaborative operation only in the context of video sensing, that requires deep neural network execution. However, opportunities for collaboration under different settings including various sensor modalities and execution complexity (e.g., shallow SVM-based prediction) warrant further investigations.

5.9 Acknowledgements

This work was done with the generous help of multiple research engineers from Singapore Management University. I thank Dr. Tarek Abdelzaher (University of Illinois, Urbana–Champagne) for valuable discussions on this work and future directions. In Table 5.7, I summarize how the work was split between myself and the research engineers, Dulanga Weerakoon and Randy Tandriansyah.

	Kasthuri	**Dulanga**	**Randy**
Joint inference (CSSD and CNMS)		100%	
Early discard of pipeline for latency improvement (SSD-1)	100%		
Accuracy improvement through dimension adjustment (ESSD)	100%		
Spatial mapping via regression	100%		
Analyses presented in the evaluation section	100%		
Implementation	45%	45%	10%
Research/system contributions not appearing this thesis			
Person re-identification across views for self-learning matching user pairs		100%	
Resilience mechanisms		100%	
Testbed at SMU		50%	50%

Table 5.7: Work and contributions split on the Collaborating Video Sensing project.

Chapter 6

Literature Review

In this chapter, I provide a summary of literature related to the key contributions of this thesis. I begin with an account of prior work that looks at business survival through traditional means (e.g., financial data) as well as through the use of social media, and a brief description of works that use a combination of data sources to study various urban problems (e.g., gentrification). Next, I provide a summary of works that have looked at modeling mobility, at various scales, for problems such as predicting the *next place*. Finally, I end the chapter with a brief summary of recent literature on edge computing, in particular, of those that provide early examples of collaboration, or cooperation, between edge nodes.

6.1 Use of Disparate Mobility Data in Land Use Studies

Retail Business Performance and Demand: Between 2004 and 2014, Parsa et al [99, 98, 96, 97], published a four-part series on *Why Restaurants Fail*. In [99], based on a quantitative study of 2400+ restaurants in Columbus, Ohio, followed by a qualitative study on restaurant owners who succeeded/failed, they present a framework for survival, composed of 4 main areas: environmental factors, family life cycle, internal factors and growth stage of the restaurant. In our work, we focus on the

environmental factors aspects, extending the study to explore a variety of aspects, in addition to the 4 factors the authors focused on – i.e., location (at the ZIP level), restaurant density at the ZIP level, whether the restaurant is privately owned and is a franchise. They provide useful empirical evidence that the restaurant industry is especially vulnerable in the early years with the highest mortality observed in the first year of operation. We build on this work to explore a variety of features, at the time of entry, for new/planned venues, and further explore the utility in using visitor information in improving the understanding of survival.

The closest to our work, is the analysis presented by Lei Wang et al. in [126]. Similar to our work, the authors explore the use of LBSNs in predicting the survival/failure of food establishments using a set 600+ restaurants in NYC over a 6 month period. The authors study the improvement customer visit information can provide accrued via check-ins publicly crawled. In particular, they investigate the influence of 5 business characteristics (number of competitors within a 3 mile radius, competitors with specials, price range and rating) and average daily check-ins of the restaurant and its competitors, and the trend of the average daily check-in growth, respectively. They report a number of significant findings including that considering the check-ins information of a restaurant and its neighbors, reduces the misclassification rate from 30% to around 10%. They also conduct a series of robustness checks where they vary the size of the hold out sample, change the threshold on what defines business failure and the size of the neighborhood. They report that beyond a radius of 1 km the influence of neighbors saturates. To detect the presence of nonlinear relationships across the covariates, the authors train a neural network with a single hidden layer and a varying number of units in the hidden layer and find that the minimum AIC/BIC occurs with zero units in the hidden layer, translating that the covariates do not exhibit any complex relationships. However, our work differs from this work in a number of different ways: (1) in what constitutes failure – in this work, failure is defined as a decline in performance; any restaurant that received less than 1 check-in per day (on average) in the last

3 months of the dataset is considered as a *failed* venture whereas in our work, we consider a restaurant to have failed if it has in fact *closed* for business, (2) we consider two scenarios explicitly – first, for venues that are yet to begin operation, i.e., those venues that don't have any check-in information as yet, we investigate the influence of a variety of geographic and mobility features in impacting the likelihood of survival within the first few years of operation. In other words, we try to answer the question, *how much does the environment and timing a venue starts in affects the chances of survival in its early years*, given that studies have shown that the longer a venue lasts, the better its chances of survival are and that the early years are the most vulnerable citeparse01 and second, for venues in operation, how much does the knowledge of checkins help understanding survival rates. Between the two works, we make a unique distinction in the fact that we consider those venues that have in fact started operation and are within their first 5-6 years of operation. (3) We extend our analyses to multiple cities around the world to understand the influences and extent of the different factors. The authors report a misclassification rate of 10% – however, due to the high imbalance in the number of restaurants that are declared as alive and failed (i.e., 90% vs 10%), even at a low misclassification rate, the precision and recall of the *failed* class becomes a low 72% and 66%, respectively, which is comparable to our results that use only geographic information (without the use of any checkins/visits).

Demand of Retail Businesses: A number of recent works have emerged since the proliferation of location-based social networking platforms in studying how modern economies and businesses prevail. In a 2015 study [92], the authors present a first-of-its-kind study which studied 100 cities around the world in terms of their place networks, temporal evolution (i.e., new venues being added and the change in transitions between venues), and their similarity/dissimilarity to other well-studied networks such as social networks and the WWW. They go on to build a link prediction model, an extension to well-studied gravity models, that essentially predicts at a future time snapshot, whether visitors to a particular venue, are also likely to visit

another venue of interest. Further, in [72], the authors provide empirical evidence to how LBSNs can help find optimal store placements based on a multitude of geographic (e.g., neighborhood, competitiveness, attractiveness of a neighborhood to a particular store category) and mobility features (e.g., inter- and intra-neighborhood transitions). They assume that the optimal placement for a venue is the location which can draw the most amount of checkins or footfall. However, the study is limited to three well-established retail chains and considers the number of checkins from the entire period of observation. Our focus is different in that, we (1) consider all types of venues (not just established venues with multiple outlets), (2) attempt to understand the longevity of the business, and (3) consider both cases where the visit patterns or, footfall, to the venue is known or unknown. More recently, in [28], the authors study a different aspect of business; whether certain categories of businesses *cooperate* or *compete* with other close-by businesses in the same category. Once again, the study is situated on the assumption that the business ecosystem has not changed within the 4-year observation period, and does not account for businesses that closed down during the same period. Further, in [36], the authors present a study on predicting the stable demand of a new venue from past visitation patterns of existing venues of the same category that are situated in the same ward, and/or wards that are found to be *temporally similar*. Our problem is different to this line of work as the focus is on estimating the longevity of new and existing venues, and not on the hourly/weekly demand patterns.

Use of Heterogeneous Mobility Sources: Recent works have also exploited the availability of spatio-temporally dense venue-level data, from social media, in understanding a myriad of pressing urban issues. In [57], the authors carry out a study on gentrification in the wards of London, and more recently in [135], the authors study the impact of cultural investment (e.g., new stadiums/museums) on the businesses and venues in proximity to those new investments. Further still, in works such as [47], a study on the long-term and short-term impact of large-scale, albeit transient, events. The authors present the case study of the London Olympics

which was held in 2014 where they provide empirical evidence of how local retailers benefited by the increase in footfall to the event-related areas.

Fusion of Social Media and Sensor Data in Urban Settings: In this work, we exploit the combination of both physical (e.g., related to the transport infrastructure of the cities) and social (e.g., related to the social relationships of users to venues and the voluntary, location checkins of users in social media platforms). Works in the recent past have studied the fusion of the two types of sensors in different domains. In [85], a vision for *socio-physical analytics* is presented which outlines challenges and opportunities in fusing the two. Further in [117, 41], the authors offer a rudimentary framework for fusing multiple types of web-based sensory data using the concept of an *Emage*, bringing heterogeneous data to a single scale using space and time boxes along with examples of extending the framework for popular operations such as addition, etc., and demonstrated its efficacy on a number of use cases in event detection in urban settings. Moreover, in [48], traffic anomalies detected using physical traffic sensors are matched spatio-temporally with anomalies detected on social media (i.e., unusual volumes of keywords) to *clarify* the cause or source of the anomaly. More recently, in [65], comparisons, across the two types of sensors, in the efficacy of detection and spatio-temporal localization of urban events are presented.

6.2 Personal and Collective Mobility

We first provide brief surveys of the body of knowledge on predictability of mobility and next place prediction from the literature, and then describe briefly, recent work that describe the interdependence of social ties and mobility, as well pointers towards context-aware crowd-tasking systems, which we use as a case study in this work.

Predictability of Human Mobility: While the ability to predict where a user will be next has many potential applications, works such as those of Song et al.

[118], Lu et al. [81] and Jensen et al. [66] have focused on quantifying the theoretical bounds of the predictability of human mobility. If mobility is not intrinsically *regular* or *predictable*, then the performance of next place prediction algorithms, however complex, will be limited. A natural choice for measuring randomness is entropy – Song et al. [118] use hourly mobility records of over 50,000 users to quantify (1) S_{rand} which is the entropy computed considering only number of different locations a user associates with, (2) S_{unc} which computes the entropy of a distribution over the different locations a user has been to and their associated frequency of visits, and finally, (3) S_{real} which considers both the frequency of visits and the sequence in which the locations were visited. S_{real} is computed as the Lempel-Ziv compression with the length tends to infinity. Using Fano's inequality, the authors derive an upper limit Π_{max} where Π is the probability of guessing the next location correctly with any algorithm. Further still, they discuss a lower bound R which captures the probability of the most likely location of a user during an hour of the day over a week. The authors make the key observation that mobility, taking into account past observations of frequency and sequence of visits, has very high median predictability (i.e., 93%) with significantly less variability across users, compared to using temporally uncorrelated location history. The authors investigate further the influence the distance traveled by users (through the radius of gyration) and demographics such as age and gender on predictability. Interestingly, even for users with high travel distances, the predictability remains high. While Song et al. [118] investigated outdoor mobility, Jensen et al. [66] study both outdoor (i.e., GSM and GPS) and indoor (i.e., via WLAN associations) by instrumenting the smartphones carried by 14 participants in Denmark. Whilst the GSM and WLAN records were sampled at faster rates (e.g., minutely), GPS, due to its large energy drain was sampled only 2-3 times every hour. Following from Song et al.'s work [118], the authors quantify the maximum predictability of mobility both indoors and outdoors; they find that even though the peak maximum predictability indoors is comparable to that of outdoor mobility, the variability across users is high. A key observation however

is that the authors look at indoor mobility at the WLAN association level, and not at a *location* around it, whereas for many meaningful applications, localization to up to room level (e.g., 6- 8 meters) is sufficient. Hence, in our work we explore the modeling of mobility at the room level and floor level. The authors also explore different time scales and find that 3-4 minutes gives the best performance although the reason for this is not well-justified. More recently, Lu et al. [81] explore the use of Markov chains of increasing orders to investigate whether the theoretical maximum predictability is achievable practically. They rely on another large scale, outdoor mobility data set to show that Markov Chains of order 2 reach comparable accuracies to the theoretical maximum, and further note that the accuracies surpass the theoretical maximum for non-stationary trajectories (identified using the Gewek diagnostic). However, in this work, the authors consider a loose definition of a user's trajectory where they consider only the last recorded location of a user as the user's location for the day and the trajectory being composed of daily locations – we believe that the achievable predictability with such a definition whilst high, would have reduced practical benefits.

Modeling Mobility: Many works in wireless systems have investigated the practicality of predicting the next cell or location a user or mobile device is likely to be next. One of the earliest works was Reality Mining [37] with about 50 users where high predictability was reported using an order 2 Markov Chain to predict the next location over a limited semantic set (i.e., Home, Work or Other). Later, Kotz et al. [119] conduct a large-scale study of over 6000 students on the Dartmouth campus with observations from over 2 years. In this work, the authors compare the next cell prediction accuracy of two families of predictors: Markov Chains and Lempel-Ziv compressor based, and note down several key observations. Overall, they find that the added complexity of the LZ-family of predictors does not necessarily afford higher accuracy, and simple enhancements such as falling back to less complex models when past history does not contain the current context and accounting for recency can improve performance marginally. They also note that

the accuracy is high only for users with long enough trajectories which might affect the practicality of such predictive algorithms. However, in this work, the authors only consider location changes as part of the user's trajectory and do not account for timing information – which again is a key attribute for practical applications.

The Next Place Prediction problem has been studied extensively due to its multitude of applications, but mostly in the context of outdoor movement derived from GPS traces from smartphones, taxicabs and social media check-ins. Noulas et al. [91] study coarse-grained next place prediction using check-in data from over 1 million Foursquare (a popular Location-based Social Network platform) users where they consider transitions between categories of places, mobility flow between individual venues and share insights from spatio-temporal characteristics of check-in patterns. Gambs et al. [46] investigate the ability of Markov chains of order n to predict the next place of users, both indoors, using a phonetic dataset consisting of voice traces of 6 users, and outdoors, using the *GeoLife* dataset [133] consisting of GPS traces from Shanghai. Further, Baumann et al. [16] evaluate the problem extensively using 18 algorithms and their combination (using majority voting) on the Lausanne/Nokia MDC data set[1]. They report that although the accuracy is typically high, most errors are encountered during transitions from one place to the next. Spawning off from the Lausanne/Nokia MDC challenge, Gomes et al. [52], in addition to considering spatio-temporal history of traces, augment contextual information accrued through sensors such as accelerometer, bluetooth and call/sms logs for better prediction. Further, Do et al. [34] discuss a variation of the problem, the probability of being at a specific location at a time in future using a dataset consisting of 133 smartphone users where they use kernel density estimation accounting additionally for day of the week and weekday/weekend effects. Our work uses similar methodologies such as those discussed in previous work in identifying possible next places, but is different in that the goal is in predicting, with sufficient look-ahead time, the possibility of the default next places predictions be incorrect – in

[1] https://www.idiap.ch/dataset/mdc

other words, the problem reduces to providing a confidence measure of the predictions based on the current trajectory of a user. Separately, in Koehler et al. [75], the authors study the problem in two folds: (1) will the user stay at the current location for the next m time (i.e., temporally), and (2) if not, where will the user transition to next (i.e., spatially) using a number of machine learning techniques.

Influence of Social Interactions on Mobility: The expansive growth of Location-based Social Networks (LBSNs) such as Foursquare, and other popular mediums such as Twitter that allow for geo-tagging of posts, has led to many large-scale studies on urban mobility. The additional information declared by the users of the platform through features such as *follows* and explicit bidirectional friendships, makes it possible to infer the social relationships of the users, both offline and online. A number of works have focused on understanding the impact of such relationships on a user's mobility [125, 26]. Using physical trajectory data along with shared social relationship information, De Domenico et al.[31] report that incorporating knowledge of one's friend's mobility can help improve prediction of a user's mobility behavior. Recent works such as [21, 86] explored the use of body-worn social badges to infer and quantify face-to-face interactions of users in working environments. In these works, either friendship information is explicitly shared, or the participants are required to wear/carry additional sensors. In this work, we focus on inferring social ties *passively* using systems such as those described in past work [63, 64], and then utilize such social data-infused mobility information for predicting uncertainty in mobility behaviour.

Mobility-Aware Crowd-Sourcing Location-aware mobile crowdsourcing has recently been employed to support the execution variety of reporting-centric tasks across both indoor and outdoor environments. The Ta$ker mobile crowdsourcing platform [70] uses predicted location trajectories, of university students to proactively recommend tasks that minimize a worker's travel detour. The user trajectories are computed as a series of staypoints–i.e., places where the user resides for large time periods. Using this platform, Kandappu et al. [69] showed that unexpected

changes in trajectory caused around 6% of workers to "cheat"–i.e., report on tasks even though they did not visit the task location. Thus, LFNC prediction is useful not just in reducing unnecessary worker detour, but also in improving the reliability of responses.

6.3 Collaborative Sensing on the Edge

Analytics at the Edge: Edge computing technologies are at the forefront of enabling situation awareness via real-time analytics. Many early examples of traditional edge computing have been in the domain of video-based applications and services due to the high resource demands required in processing images using state-of-the-art deep learners [106, 107, 115]. In recent times, multi-device cooperation, at the edge, has received considerable attention. For example, recent literature in multi-camera systems [101, 78, 62], and unmanned aerial vehicle swarms [25, 23] have explored early ideas of cooperation among peer devices owing to their natural advantage in terms of improving accuracy as well as in reducing the bandwidth and latency overhead involved in communicating with a central server.

Collaborative Intelligence: Early efforts in describing the need for enabling collaborative intelligence among heterogeneous IoT devices, complementary to our vision, have been advocated recently [101, 78, 62, 7]. Qiu et al.[101] describe a scenario where cameras of differing capabilities co-exist in the same network: fixed surveillance cameras and resource-constrained mobile devices with cameras. The authors demonstrate that moving vehicles can be tracked seamlessly across this heterogeneous camera network through selective actuation of devices without overly draining the mobile devices. In essence, the resource-intensive video analytics pipeline is performed on the cloud and the mobile cameras are consulted intermittently, only to resolve ambiguities. Further, Lee et al. [78] demonstrate significant savings in bandwidth needs (of *dumb* cameras that offload raw footage to a central cloud) – they show that by establishing space-time relationships, apri-

ori, between co-existing cameras, that they can be selectively turned ON (and OFF) leading to as much as 238 times savings in bandwidth at a miss rate of only 15% for a vehicle detection task. Similarly, Jain et al. [62] also show that significant correlations exist between co-located cameras, and discuss configurations of video analytics pipelines that can be triggered by peer cameras leading to both cost efficiency and superior inference accuracy. Unlike such past work, we focus explicitly on using collaboration to modify or abort the inferencing pipeline itself, instead of selectively activating nodes or performing fusion of the outputs from multiple nodes. Most recently, we describe our vision [7] for providing machine intelligence as a service at the edge for resource-constrained devices – in addition to outlining core capabilities required for enabling such a service (e.g., scheduling, caching, resource profiling), we also describe opportunities for the convergence of the idea of collaboration between devices and deep intelligence as a service.

DNN Optimizations for Resource-Constrained Systems: One of our main goals in enabling collaboration is in reducing the processing latency of running DNNs on resource-constrained devices. While we advocate for information sharing between co-located devices to run altered pipelines, at run time, to save time, separately, there have been efforts in recent times to cut down on execution time for device operating in isolation – i.e., without collaboration. For example, Yao et al.[132, 131] demonstrate that compressing DNNs offline through selective pruning of the network can lead to significant savings in run time without compromise in accuracy. Exploiting the fact that scenes do not change so drastically between consecutive frames in streaming videos, Huynh at al. [60] show that intelligent caching (and, hence, avoiding redundant) computations at intermediate layers of the deep networks can help in saving time. While these frameworks are prescribed for individual devices, we believe that our proposed vision is complementary to this line of investigations, and that convergence would lead to greater savings.

Optimizations in Sensor Networks: Apart from optimizing individual deep pipelines, several recent works, especially in the domain of video analytics at the

edge [59, 58, 67], have put forward various performance optimizations *across a group of sensors* in a network. For instance, VideoEdge [59] supports efficient querying of video streams from multiple cameras by intelligently deciding, at run time, where (e.g., at the camera, on a nearby private cluster, or the cloud) to run which parts of the computation. FOCUS [58] supports efficient querying of high volume, pre-recorded data through the combination of cheap CNNs for data ingestion and indexing (optimized for speed), and complex CNNs for querying (optimized for accuracy). Recently, Jiang et al. [67] discuss an approach to profile individual camera nodes for application-dependent requirements (e.g., accuracy) and the corresponding resource needs (e.g., resolution of images, frame rate, etc.). These works, however, do not consider redundancy across cameras in the vicinity for collective performance optimization. As an exception, Jiang et al. [67] discuss how cross-camera correlations could be used to amortize the *cost of profiling* – for instance, they argue that for two nearby cameras (with similar views) running the same application, it is sufficient for one of the two to go through the costly profiling process.

Separately, past works [32, 33] in sensor networks have looked at optimizations, at the network level. For example, Deshpande et al. [32, 33] discuss model-based querying for data acquisition in a network of sensors. Similar to our goals, the authors note that there exist correlations between the observations that sensors in proximity observe – they use such correlations to *model* attributes of a nearby sensor based on another sensor's observations (such as temperature readings). In such networks, they also assume that querying a nearby sensor is less costly as compared to a farther way sensor (e.g., as more number of hops are required). Eventually, the authors propose algorithms to optimize query plans that select the best combination of sensors to acquire readings from, balancing the quality of the answers against communication and acquisition costs in the network. In contrast, in our work, we assume that the sensors are only partially correlated with nearby sensors (e.g., partial overlap of camera views) that they require to be *always-on* and con-

stantly queried. The optimizations that we propose in this work are designed for optimizations of the *always-on processing pipelines* that utilize partial information from neighboring sensors.

Chapter 7

Conclusions and Future Directions

In this chapter, I conclude this thesis by summarizing the main contributions, the impact of the research work carried out, and outline some of the key future directions.

7.1 Summary of Contributions

Applications and Problem Domains: In this thesis, I demonstrate the potential of large-scale mobility behavior in enabling urban applications, at different spatio-temporal scales, short-range, indoor and city-scale, for predictive policy insights and daily operations. I study the problem of survivability of retail businesses that is an important input for appropriating land resources for city planners by exploiting multi-modal mobility data. I propose the non-conformant mobility prediction problem, for both indoor and outdoor mobility, that has various opportunities for resource utilization improvements (e.g., task assignment of crowd-workers). Finally, I propose the concept of collaborative sensing among co-located IoT devices and provide early evidence of opportunities in terms of improving inference accuracy and significantly cutting down processing pipelines taking the example of video sensing.

Empirical Insights: The solutions for these problems required extensive study

of individual and aggregate mobility behavior – I highlight below some of the key insights from these studies.

Retail Business Survivability:

1. Several indicators attributes a store's current operation can influence its success in the future. For instance, a restaurant that is appealing around the clock – one that has patrons visiting throughout the day – have a better chance of surviving.

2. Beyond a store's daily operation, the neighborhood composition and mobility dynamics of that neighborhood can influence the fate of the store. For example, stores operating in regions which have better reachability (i.e., the patrons who visit the store come from all over the city) fare better than stores operating in regions that only attract locals.

Social Ties and Mobility:

1. Mobility, both indoor and outdoors is predictable, and mobility behavior of individuals are greatly affected by the presence of social groups – people tend to spend longer times at places (e.g., at food courts) when they are with friends, and there exists significant correlation between the movement behavior/trajectory of close ties–i.e., "birds of a feather flock together with similar predictability".

2. A person's propensity of deviating from a future routine location/activity pattern is correlated to the anomalousness of her current movement–if a person has been exhibiting anomalous movement patterns in the recent past, she is much more likely to deviate from her routine location/activity pattern in the future.

3. In workplace environments, where users indulge in significant collaborative activities, anomalous movement behavior is often not isolated but *shared*: if a user's "friends" have been exhibiting non-routine movement as well, there

is a significant increase in the likelihood that she will deviate from her future routine movement pattern.

System Artefacts: The LFNC framework described in Chapter **??** provides the design for enabling non-conformant mobility prediction with look-ahead times. We explore and report on the trade-offs in the accuracy of predicting non-conformity for different look-ahead times and other parameters such as the sequence length of the trajectory observed. For varying look-ahead times, we test the ability of the non-conformance predictor to distinguish between crowd-workers of a campus-scale crowd-tasking platform who may (or may not) complete a task assigned to them.

7.1.1 Publications

The research work described in this thesis have led to publications in peer–reviewed journals and conferences. Below is a list of selected publications.

[IMWUT18/ UbiComp19] **K. Jayarajah** and A. Misra. Predicting Episodes of Non-Conformant Mobility in Indoor Environments, PACM IMWUT, Volume 2, Issue 4, Dec 2018.

[IMWUT18/ UbiComp18] K. D'Silva, **K. Jayarajah (co-primary)**, A. Tassos, A. Misra, and C. Mascolo. The Role of Urban Mobility on Retail Business Survival, PACM IMWUT, Volume 2, Issue 3, 2018.

[UbiComp15] **Jayarajah, K.**, Y. Lee, A. Misra, R. K. Balan. Need Accurate User Behavior? Pay Attention to Groups!, ACM Ubicomp 2015.

[ASONAM15] **Jayarajah, K.**, A. Misra, X. W. Ruan, E. P. Lim. Event Detection: Exploiting Socio-Physical Interactions in Physical Spaces, IEEE/ACM ASONAM 2015.

Additionally, as I briefly describe in Section 7.2, the following publications pertain to other works that I pursued during my candidature, but do not contribute to the thesis statement.

[MobiSys19]	Meegahapola, L., T. Kandappu, **K. Jayarajah**, L. Akoglu, S. Xiang and A. Misra. BuSCOPE : Fusing Individual & Aggregated Mobility Behavior for Live Smart City Services.
[PURBA18/ UbiComp Adjunct]	**Jayarajah, K.**, A. Tan, and A. Misra. Exploiting the Interdependency of Land Use and Mobility for Urban Planning, ACM UbiComp/ISWC Adjunct Proceedings, 2018.
[MobiSys16]	**Jayarajah, K.**, R. K. Balan., M. Radhakrishnan, A. Misra and Y. Lee. LiveLabs: Enabling In-Situ Behavioural Experiments, ACM MobiSys 2016.
[ACR2017]	Gupta, R., S. Mukherjee, **K. Jayarajah**. Role of Group Cohesiveness in Consumers Responses to Mobile Promotions (Working Paper), Association For Consumer Research Conference North America Advances.

7.2 Additional Work and Technical Achievements

In this section, I describe research problems in the general theme of mobility based urban applications that I worked on during my PhD candidature, outside my main thesis contributions.

7.2.1 In-Situ Behavioral Experimentation

Previously, in Chapter 3, I describe the indoor location dataset that was made available through the LiveLabs Lifestyle Innovation Testbed [84, 12]. The LiveLabs ecosystem consisted of (a) a passive, server-side system for monitoring indoor mo-

bility of thousands of users who connect to the university's WiFi infrastructure, (b) a suite of mobile applications running on opt-in participants' phones that capture interactions of its participants with their respective phones, and (3) a behavioral intervention engine that allows experimenters to use additional real-time contexts such as location and group to send *interventions*, in the form of mobile notifications, to the participants' mobile phones. Together, the eco-system allows for *in-the-field, randomized behavioral experimentation*. This new class of experimentation overcomes several limitations of the traditional experimentation settings such as response bias [108] where the subjects feel the pressure to provide socially acceptable answers to a survey, and the manual effort required in shadowing subjects.

In **[MobiSys2016]**, I use the intervention engine to conduct various behavioral experiments in collaboration with faculty from the social science and marketing science disciplines where I use mobility-based attributes. The experiments ranged from (1) re-validating classic behavioral constructs such as the "foot-in-the-door" effect [44], (2) extending classic experiments such as studying the effects of priming mindsets [51] on people's future mobility captured passively and continuously (e.g. **[ACR2017]**), and further still, (3) innovating experiment designs, for e.g., capturing both the situation selection and perception simultaneously in a real world setting, in studies of personality traits. We determined that the system–induced experiments were comparable to traditional methods in terms of the response rates (21.85 - 34.27% during non-exam periods compared to 17 - 24% of alternate mediums such as paper-based, SMS/email based). More importantly, we observed that the system was void of non-response bias [108].

7.2.2 Enabling "Live" Smart City Services

I reiterate the fact that, thus far, while urban commuting data has been used extensively in providing useful insights into human mobility behavior, such analysis have been performed largely in offline fashion and to aid medium-to-long term urban

planning. One key contribution of this thesis has been in advocating near real-time (Chapter 5), and predictive (Chapter 4) mobility-driven urban applications. In **[MobiSys19]**, we demonstrate the power of applying predictive analytics on real-time mobility data, specifically the smart-card generated trip data of millions of public bus commuters in Singapore, to create novel and live smart city services. We see that vast majority of public bus trips are predictable, and driven by routine commuting patterns. Such predictability manifests in two aspects: (a) individual-level regularity, which allows us to predict an individuals point of disembarkation, as soon as she boards a bus, and (b) aggregate-level conformity, which allows us to use historical commuting flows (over all commuters) to identify a relatively small set of likely disembarkation points, even for commuters with no relevant prior travel history.

We propose and evaluate *BuScope*, a live mobility analytics platform, which enables making operational decisions or generating neighborhood-level insights on streaming mobility data, with $O(mins)$ responsiveness. The platform is flexible enough to recompute the analytical insights, at both individual and bus-level specificity, very frequently for peak city-scale workloadse.g., it incurs 17.33 msecs average latency to process each of $\approx$ 270,000 boarding and alighting transactions generated by 221,217 commuters on 3777 buses, during a typical weekday, 30 minute peak period. We illustrate the usefulness of such real-time processing through two exemplary applications: last-mile, on-demand, shared taxi provisioning and city-scale event detection.

7.2.3 The Interdependency between Land Use and Mobility

Earlier, in Chapter 2, I investigated a number of attributes related to customer visits and mobility of a neighborhood, on the survival chances of an individual business. A complementary problem is then understanding the impact of land use – i.e., the mix of retail venues of differing categories in a neighborhood – on the mobility of

that neighborhood. In preliminary work **[PURBA18]**, we study the question: *given the mix of land use in a neighborhood, can we estimate the utilization pattern of a planned carpark?*.

To this end, we collected longitudinal car park lot availability data from over 1500 car parks from across Singapore, and extracted the land use around each individual car park using social media, and predict the utilization pattern of a car park using machine learning algorithms. Although we limit the scope to car park utilization in the current work, the problem is generalizable to traffic and congestion levels in the neighborhood, in general, as it may be observable through taxi pickup/-dropoff patterns, public transport utilization and so on. We first explore the use of unsupervised clustering in extracting categories of car parks based on their weekly, temporal utilization patterns. We share preliminary insights from the resulting clusters of car parks and the land use around those. We then predict the utilization patterns of carparks, given the neighborhood land use mix, posing the problem as a multi-class classification task. We achieve an AUC $\approx$ 0.84 suggesting that our primary hypothesis that the activity or land use of an area can be indicative of the people flow in that area.

7.3 Future Directions

Capturing complex interdependencies across urban entities: As I describe in Chapter 1, one of the key challenges in the realization of the vision of a smart urban resource eco-system, is the modeling of the interaction effects between the different types of urban resources and mobility. One possibility to capture such complexities is in the representation of the different *kinds* of urban entities (e.g., carparks, retail businesses, consumers, etc.) as a multi-layer network [74] that captures both the inter- and intra-layer relationships by the edges between the entity nodes. Such a representation, with advances in network science, can help in studying effects such as interdependencies [13], cascading effects [22] (for example, in studying how the

failure of a transport nodes, for instance, the sudden breaking down of a train line, can cause repercussions in the other layers such as commuters' mobility.

Cross-modal Adaptation in collaborative environments: In this thesis, I describe and demonstrate how the existence of redundancy in the sensed environment between co-located sensors have can aid in significant improvements in terms of operational efficiency. However, this thesis only provides evidence under unimodal sensor settings (i.e.,, a network of cameras).

A natural extension of this idea is thus to consider a cross-modal sensor setting. I pose the question, *Given a deployment of sensors of a specific type (e.g., cameras) of a specific spatial configuration, can a new deployment of sensors of a different type (e.g., audio sensors) with the exact same positioning configuration,* adapt *to collaborate with minimal re-training?* In other words, can the learned correlations in one feature space (.e.g., 2D convolutions over pixel values) be mapped to a complete different feature space (e.g., Mel frequency cepstral coefficients)?

Collaborative actuation of IoT devices: This thesis focuses on collaboration at the sensing nodes for faster and accurate sensing. However, this leaves us with open questions on *whether and how collaboration at the actuation nodes* can exploit mobility for their optimized operation. In particular, I wish to investigate possibilities in the domain of office settings to study its impact on worker productivity. Take the example scenario of a typical open plan office space where multiple desk workers co-inhibit with various types of actuators. Typically, a common work space would require multiples of fluoroscent lights, AC vents, etc. and the specific ambiance attributes such as lighting, noise levels, temperature of the work space, depends largely uniform across the entire space with minimal personalizations. However, productivity research shows that individuals respond differently to variables in the environment and that it impacts each person's productivity to different extents. So I ask *can the co-inhabiting actuators collaborate to provide individualized, optimal working conditions to workers in an open plan office* such that the collective productivity can be optimized?

Bibliography

[1] Get details of a venue. https://developer.foursquare.com/docs/api/venues/details. [Online; Last accessed 09-May-2018].

[2] The r stats package. http://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html. [Online; Last accessed 13-May-2018].

[3] Rocr: Visualizing the performance of scoring classifiers. https://cran.r-project.org/web/packages/ROCR/index.html. [Online; Last accessed 13-May-2018].

[4] scikit-learn machine learning in python. http://scikit-learn.org/stable/. [Online; Last accessed 13-May-2018].

[5] Shopify. https://www.shopify.com.sg/. [Online; Last accessed 14-May-2018].

[6] These ten cities have the best public transit in the world. https://jalopnik.com/these-ten-cities-have-the-best-public-transit-in-the-wo-1610824583. [Online; Last accessed 13-May-2018].

[7] T. Abdelzaher, S. Yao, Y. Hao, Y. Zhao, A. Piao, H. Shao, D. Liu, S. Liu, S. Hu, D. Weerakoon, K. Jayarajah, and A. Misra. Eugene: Towards deep intelligence as a service. In *The 39th IEEE International Conference on Distributed Computing Systems (ICDCS 2019), Dallas, Texas USA. July 710, 2019*.

[8] U. S. B. Adminstration. Apply for licenses and permits, 2018. https://www.sba.gov/business-guide/launch-your-business/apply-licenses-and-permits.

[9] C. M. Agudelo-Vera, W. R. Leduc, A. R. Mels, and H. H. Rijnaarts. Harvesting urban resources towards more resilient cities. *Resources, conservation and recycling*, 64:3–12, 2012.

[10] A. Avati, K. Jung, S. Harman, L. Downing, A. Ng, and N. H. Shah. Improving palliative care with deep learning. *arXiv preprint arXiv:1711.06402*, 2017.

[11] P. Bahl and V. N. Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784. Ieee, 2000.

[12] R. K. Balan, A. Misra, and Y. Lee. Livelabs: Building an in-situ real-time mobile experimentation testbed. In *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (HotMobile '14)*, 2014.

[13] J. Banerjee, A. Das, and A. Sen. A survey of interdependency models for critical infrastructure networks. *arXiv preprint arXiv:1702.05407*, 2017.

[14] T. Bates. Analysis of survival rates among franchise and independent small business startups. *Journal of Small Business Management*, 33(2):26, 1995.

[15] P. Baumann. Adaptive sensor cooperation for predicting human mobility. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, 2014.

[16] P. Baumann, W. Kleiminger, and S. Santini. The influence of temporal and spatial features on the performance of next-place prediction algorithms. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 449–458. ACM, 2013.

[17] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

[18] R. A. Becker, R. Caceres, K. Hanson, J. M. Loh, S. Urbanek, A. Varshavsky, and C. Volinsky. A tale of one city: Using cellular network data for urban planning. *IEEE Pervasive Computing*, 10(4):18–26, 2011.

[19] J. Benesty, J. Chen, Y. Huang, and I. Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

[20] M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. Lof: Identifying density-based local outliers. In *In Proc. Of SIGMOD'00*.

[21] C. Brown, C. Efstratiou, I. Leontiadis, D. Quercia, C. Mascolo, J. Scott, and P. Key. The architecture of innovation: Tracking face-to-face interactions with ubicomp technologies. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 811–822. ACM, 2014.

[22] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025, 2010.

[23] A. Bürkle. Collaborating miniature drones for surveillance and reconnaissance. In *Unmanned/Unattended Sensors and Sensor Networks VI*, volume 7480, page 74800H. International Society for Optics and Photonics, 2009.

[24] C. Chen, S.-F. Cheng, A. Gunawan, A. Misra, K. Dasgupta, and D. Chander. Traccs: a framework for trajectory-aware coordinated urban crowd-sourcing. In *Second AAAI Conference on Human Computation and Crowdsourcing*, 2014.

[25] X. Chen, A. Purohit, C. R. Dominguez, S. Carpin, and P. Zhang. Drunkwalk: Collaborative and adaptive planning for navigation of micro-aerial sensor swarms. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, SenSys '15, 2015.

[26] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, 2011.

[27] D. R. Cox. *Analysis of survival data*. Routledge, 2018.

[28] M. L. Daggitt, A. Noulas, B. Shaw, and C. Mascolo. Tracking urban activity growth globally with big location data. *Open Science*, 3(4), 2016.

[29] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *international Conference on computer vision & Pattern Recognition (CVPR'05)*, volume 1, pages 886–893. IEEE Computer Society, 2005.

[30] P. Davidsson and M. Boman. Distributed monitoring and control of office buildings by embedded agents. *Inf. Sci.*, 171(4):293–307, 2005.

[31] M. De Domenico, A. Lima, and M. Musolesi. Interdependence and predictability of human mobility and social interactions. *Pervasive Mob. Comput.*, 9(6), Dec. 2013.

[32] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong. Model-driven data acquisition in sensor networks. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pages 588–599. VLDB Endowment, 2004.

[33] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong. Model-based approximate querying in sensor networks. *The VLDB journal*, 14(4):417–443, 2005.

[34] T. M. T. Do and D. Gatica-Perez. Where and what: Using smartphones to predict next locations and applications in daily life. *Pervasive and Mobile Computing*, 12:79–91, 2014.

[35] P. Dollár, Z. Tu, P. Perona, and S. Belongie. Integral channel features. 2009.

[36] K. D'Silva, A. Noulas, M. Musolesi, C. Mascolo, and M. Sklar. If i build it, will they come? predicting new venue visitation patterns through mobility data. In *Proceedings of the ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, 2017.

[37] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

[38] B. Efron. Logistic regression, survival analysis, and the kaplan-meier curve. *Journal of the American statistical Association*, 83(402):414–425, 1988.

[39] V. Etter, M. Kafsi, E. Kazemi, M. Grossglauser, and P. Thiran. Where to go from here? mobility prediction from instantaneous information. *Pervasive and Mobile Computing*, 9(6):784–797, 2013.

[40] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[41] A. Farseev and T. Chua. Tweetfit: Fusing multiple social media and sensor data for wellness profile learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 95–101, 2017.

[42] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645, 2009.

[43] J. Ferryman and A. Shahrokni. Pets2009: Dataset and challenge. In *2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*. IEEE, 2009.

[44] J. L. Freedman and S. C. Fraser. Compliance without pressure: the foot-in-the-door technique. *Journal of personality and social psychology*, 4(2):195, 1966.

[45] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

[46] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, page 3. ACM, 2012.

[47] P. Georgiev, A. Noulas, and C. Mascolo. Where businesses thrive: Predicting the impact of the olympic games on local retailers through location-based services data. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014.*, 2014.

[48] P. Giridhar, M. T. A. Amin, T. F. Abdelzaher, D. Wang, L. M. Kaplan, J. George, and R. K. Ganti. Clarisense+: An enhanced traffic anomaly explanation service using social network feeds. *Pervasive and Mobile Computing*, 41:381–396, 2017.

[49] R. Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[50] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[51] P. M. Gollwitzer, H. Heckhausen, and B. Steller. Deliberative and implemental mindsets: Cognitive tuning toward congruous thoughts and information. *Journal of Personality and Social Psychology*, 59(6):1119, 1990.

[52] J. B. Gomes, C. Phua, and S. Krishnaswamy. Where will you go? mobile data mining for next place prediction. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 146–158. Springer, 2013.

[53] R. W. Hamming. Error detecting and error correcting codes. *Bell System technical journal*, 29(2):147–160, 1950.

[54] T. Harrison and J. Ansell. Customer retention in the insurance industry: using survival analysis to predict cross-selling opportunities. *Journal of Financial Services Marketing*, 6(3):229–239, 2002.

[55] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[56] A. Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person reidentification. *arXiv preprint arXiv:1703.07737*, 2017.

[57] D. Hristova, M. J. Williams, M. Musolesi, P. Panzarasa, and C. Mascolo. Measuring urban social diversity using interconnected geo-social networks. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, 2016.

[58] K. Hsieh, G. Ananthanarayanan, P. Bodik, S. Venkataraman, P. Bahl, M. Philipose, P. B. Gibbons, and O. Mutlu. Focus: Querying large video datasets with low latency and low cost. In *13th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 18)*, pages 269–286, 2018.

[59] C.-C. Hung, G. Ananthanarayanan, P. Bodik, L. Golubchik, M. Yu, P. Bahl, and M. Philipose. Videoedge: Processing camera streams using hierarchical clusters. In *2018 IEEE/ACM Symposium on Edge Computing (SEC)*, pages 115–131. IEEE, 2018.

[60] L. N. Huynh, Y. Lee, and R. K. Balan. Deepmon: Mobile gpu-based deep learning framework for continuous vision applications. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '17, 2017.

[61] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky. Identifying important places in peoples lives from cellular network data. In *International Conference on Pervasive Computing*, pages 133–151. Springer, 2011.

[62] S. Jain, G. Ananthanarayanan, J. Jiang, Y. Shu, and J. Gonzalez. Scaling video analytics systems to large camera deployments. In *In Proc. of HotMobile*. ACM, 2019.

[63] K. Jayarajah, Y. Lee, A. Misra, and R. K. Balan. Need accurate user behaviour?: Pay attention to groups! In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, 2015.

[64] K. Jayarajah, A. Misra, X. W. Ruan, and E. P. Lim. Event detection: Exploiting socio-physical interactions in physical spaces. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2015.

[65] K. Jayarajah, V. Subbaraju, D. Weerakoon, A. Misra, L. T. Tam, and N. Athaide. Discovering anomalous events from urban informatics data, 2017.

[66] B. S. Jensen, J. Larsen, L. K. Hansen, J. E. Larsen, and K. Jensen. Predictability of mobile phone associations. In *Inter. Workshop on Mining Ubiquitous and Social Environments*, 2010.

[67] J. Jiang, G. Ananthanarayanan, P. Bodik, S. Sen, and I. Stoica. Chameleon: scalable adaptation of video analytics. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, pages 253–266. ACM, 2018.

[68] E.-Y. Jung, C. Baek, and J.-D. Lee. Product survival analysis for the app store. *Marketing Letters*, 23(4):929–941, 2012.

[69] T. Kandappu, N. Jaiman, R. Tandriansyah, A. Misra, S.-F. Cheng, C. Chen, H. C. Lau, D. Chander, and K. Dasgupta. Tasker: Behavioral insights via campus-based experimental mobile crowd-sourcing. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, 2016.

[70] T. Kandappu, A. Misra, S.-F. Cheng, N. Jaiman, R. Tandriansyah, C. Chen, H. C. Lau, D. Chander, and K. Dasgupta. Campus-scale mobile crowd-tasking: Deployment & behavioral insights. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, 2016.

[71] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[72] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo. Geo-spotting: Mining online location-based services for optimal retail store placement. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, 2013.

[73] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[74] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[75] C. Koehler. Indoor location prediction through modeling of human spatiotemporal behavior. *PhD Thesis*, 2015.

[76] M. B. Kursa, W. R. Rudnicki, et al. Feature selection with the boruta package. *J Stat Softw*, 36(11):1–13, 2010.

[77] H. W. Lauw, E.-P. Lim, H. Pang, and T.-T. Tan. Stevent: Spatio-temporal event model for social network discovery. *ACM Trans. Inf. Syst.*, 28, 2010.

[78] J. Lee, T. Abdelzaher, H. Qiu, R. Govindan, K. Marcus, R. Hobbs, N. Suri, and W. Dron. On tracking realistic targets in a megacity with contested air and spectrum access. *MILCOM*, 2018.

[79] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. SSD: Single shot multibox detector. In *ECCV*, 2016.

[80] D. G. Lowe et al. Object recognition from local scale-invariant features. In *iccv*, volume 99, pages 1150–1157, 1999.

[81] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson. Approaching the Limit of Predictability in Human Mobility. *Scientific Reports*, 3, Oct. 2013.

[82] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[83] A. Misra and R. K. Balan. Livelabs: Initial reflections on building a large-scale mobile behavioral experimentation testbed. *SIGMOBILE Mob. Comput. Commun. Rev.*, 17(4):47–59, 2013.

[84] A. Misra and R. K. Balan. Livelabs: Initial reflections on building a large-scale mobile behavioral experimentation testbed. *SIGMOBILE Mobile Computing Communications Review*, 17(4):47–59, Dec. 2013.

[85] A. Misra, K. Jayarajah, S. Nayak, P. K. Prasetyo, and E.-p. Lim. Socio-physical analytics: Challenges & opportunities. In *Proceedings of the 2014 Workshop on Physical Analytics*, WPA '14, 2014.

[86] A. Montanari, Z. Tian, E. Francu, B. Lucas, B. Jones, X. Zhou, and C. Mascolo. Measuring interaction proxemics with wearable light tags. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1), Mar. 2018.

[87] M. Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[88] S. Nayak, A. Misra, K. Jayarajah, P. K. Prasetyo, and E.-P. Lim. Exploring discriminative features for anomaly detection in public spaces. In *In Proc. of SPIE DSS'15*, DSS'15, 2015.

[89] P. Newton and D. Meyer. The determinants of urban resource consumption. *Environment and behavior*, 44(1):107–135, 2012.

[90] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu. Using of jaccard coefficient for keywords similarity. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1, 2013.

[91] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. Mining user mobility features for next place prediction in location-based services. In *Data mining (ICDM), 2012 IEEE 12th international conference on*, pages 1038–1043. IEEE, 2012.

[92] A. Noulas, B. Shaw, R. Lambiotte, and C. Mascolo. Topological properties and temporal dynamics of place networks in urban environments. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, 2015.

[93] R. M. O'brien. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5):673–690, 2007.

[94] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the national academy of sciences*, 104(18):7332–7336, 2007.

[95] M. A. Pagano and A. O. Bowman. *Vacant land in cities: An urban resource*. Brookings Institution, Center on Urban and Metropolitan Policy Washington, DC, 2000.

[96] H. Parsa, A. Gregory, and M. Terry. Why do restaurants fail? part iii: An analysis of macro and micro factors. *Dick Pope Sr. Institute Publications*, 15, 2011.

[97] H. Parsa, J.-P. I. van der Rest, S. R. Smith, R. A. Parsa, and M. Bujisic. Why restaurants fail? part iv: The relationship between restaurant failures and demographic factors. *Cornell Hospitality Quarterly*, 56(1):80–90, 2015.

[98] H. G. Parsa, J. Self, S. Sydnor-Busso, and H. J. Yoon. Why restaurants fail? part ii - the impact of affiliation, location, and size on restaurant failures: Results from a survival analysis. *Journal of Foodservice Business Research*, 14(4):360–379, 2011.

[99] H. G. Parsa, J. T. Self, D. Njite, and T. King. Why restaurants fail. *Cornell Hotel and Restaurant Administration Quarterly*, 46(3):304–322, 2005.

[100] P. Purkait, C. Zhao, and C. Zach. Spp-net: Deep absolute pose regression with synthetic views. *arXiv preprint arXiv:1712.03452*, 2017.

[101] H. Qiu, X. Liu, S. Rallapalli, A. J. Bency, K. Chan, R. Urgaonkar, B. Manjunath, and R. Govindan. Kestrel: Video analytics for augmented multi-camera vehicle tracking. In *Internet-of-Things Design and Implementation (IoTDI), 2018 IEEE/ACM Third International Conference on*, pages 48–59. IEEE, 2018.

[102] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[103] J. Redmon and A. Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[104] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[105] E. S. Ristad and P. N. Yianilos. Learning string-edit distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(5):522–532, 1998.

[106] M. Satyanarayanan. Edge computing for situational awareness. In *Local and Metropolitan Area Networks (LANMAN), 2017 IEEE International Symposium on*, pages 1–6. IEEE, 2017.

[107] M. Satyanarayanan, Z. Chen, K. Ha, W. Hu, W. Richter, and P. Pillai. Cloudlets: at the leading edge of mobile-cloud convergence. In *2014 6th International Conference on Mobile Computing, Applications and Services (MobiCASE)*, pages 1–9. IEEE, 2014.

[108] L. J. Sax, S. K. Gilmartin, and A. N. Bryant. Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, 44(4):409–432, 2003.

[109] S. Scellato, A. Noulas, and C. Mascolo. Exploiting place features in link prediction on location-based social networks. KDD '11. ACM, 2011.

[110] J. Scott, A. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar. Preheat: Controlling home heating using occupancy prediction. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 281–290. ACM, 2011.

[111] I. Semanjski and S. Gautama. Smart city mobility applicationgradient boosting trees for mobility prediction and analysis based on crowdsourced data. *Sensors*, 15(7):15974–15987, 2015.

[112] R. Sen, Y. Lee, K. Jayarajah, A. Misra, and R. K. Balan. Grumon: Fast and accurate group monitoring for heterogeneous urban spaces. In *In Proc. of SenSys'14*.

[113] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[114] A. L. Sheldon. Equitability indices: Dependence on the species count. *Ecology*, 50(3):466–467, 1969.

[115] P. Simoens, Y. Xiao, P. Pillai, Z. Chen, K. Ha, and M. Satyanarayanan. Scalable crowd-sourcing of video from mobile devices. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pages 139–152. ACM, 2013.

[116] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[117] V. K. Singh, M. Gao, and R. Jain. Social pixels: Genesis and evaluation. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, 2010.

[118] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[119] L. Song, D. Kotz, R. Jain, and X. He. Evaluating location predictors with extensive wi-fi mobility data. *SIGMOBILE Mob. Comput. Commun. Rev.*, 7(4):64–65, Oct. 2003.

[120] Y. Sun, L. Zheng, W. Deng, and S. Wang. Svdnet for pedestrian retrieval. 2017.

[121] J. Tullio. Intelligent groupware to support communication and persona management. In *Proc. UIST 2003*, 2003.

[122] J. Tullio, J. Goecks, E. D. Mynatt, and D. H. Nguyen. Augmenting shared personal calendars. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 11–20. ACM, 2002.

[123] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. *International journal of computer vision*, 104(2):154–171, 2013.

[124] VentureBeat. Foursquare by the numbers, 2015. `https://goo.gl/Vi1UUf`.

[125] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, 2011.

[126] L. Wang, R. Gopal, R. Shankar, and J. Pancras. On the brink: Predicting business failure with mobile location-based checkins. *Decision Support Systems*, 76:3 – 13, 2015. Analyzing the Impacts of Advanced Information Technologies on Business Operations.

[127] Y. Wang, S. Wang, Y. Fang, and P. Y. Chau. Store survival in online marketplace: An empirical investigation. *Decision Support Systems*, 56:482–493, 2013.

[128] D. Weerakoon, K. Jayarajah, R. Tandriansyah, and A. Misra. Resilient collaborative intelligence for adversarialiot environments. In *Proceedings of the 22nd International Conference on Information Fusion*, 2019.

[129] A. W. Wicker and J. C. King. Employment, ownership, and survival in microbusiness: a study of new retail and service establishments. *Small Business Economics*, 1(2):137–152, 1989.

[130] Y. Xu, X. Liu, L. Qin, and S.-C. Zhu. Cross-view people tracking by scene-centered spatio-temporal parsing. In *AAAI*, pages 4299–4305, 2017.

[131] S. Yao, Y. Zhao, H. Shao, S. Liu, D. Liu, L. Su, and T. Abdelzaher. Fastdeepiot: Towards understanding and optimizing neural network execution time on mobile and embedded devices. In *The 16th ACM Conference on Embedded Networked Sensor Systems (ACM SenSys), 2018*.

[132] S. Yao, Y. Zhao, A. Zhang, L. Su, and T. Abdelzaher. Deepiot: Compressing deep neural network structures for sensing systems with a compressor-critic framework. In *The 15th ACM Conference on Embedded Networked Sensor Systems (ACM SenSys), 2017*.

[133] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800. ACM, 2009.

[134] M. Zhou, M. Ma, Y. Zhang, K. SuiA, D. Pei, and T. Moscibroda. Edum: classroom education measurements via large-scale wifi networks. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 316–327. ACM, 2016.

[135] X. Zhou, D. Hristova, A. Noulas, C. Mascolo, and M. Sklar. Cultural investment and urban socio-economic development: a geosocial network approach. *Open Science*, 4(9), 2017.