

The Role of Urban Mobility in Retail Business Survival

KRITTIKA D'SILVA*, University of Cambridge

KASTHURI JAYARAJAH*[†], Singapore Management University

ANASTASIOS NOULAS, New York University

CECILIA MASCOLO, University of Cambridge and The Alan Turing Institute

ARCHAN MISRA, Singapore Management University

Economic and urban planning agencies have strong interest in tackling the hard problem of predicting the odds of survival of individual retail businesses. In this work, we tap urban mobility data available both from a location-based intelligence platform, Foursquare, and from public transportation agencies, and investigate whether *mobility-derived features can help foretell the failure of such retail businesses, over a 6 month horizon, across 10 distinct cities spanning the globe*. We hypothesise that the survival of such a retail outlet is correlated with not only venue-specific characteristics but also broader neighbourhood-level effects. Through careful statistical analysis of Foursquare and taxi mobility data, we uncover a set of discriminative features, belonging to the neighbourhood's static characteristics, the venue-specific customer visit dynamics, and the neighbourhood's mobility dynamics. We demonstrate that classifiers trained on such features can predict such survival with high accuracy, achieving approximately 80% precision and recall across the cities. We also show that the impact of such features varies across new and established venues and across different cities. Besides achieving a significant improvement over past work on business vitality prediction, our work demonstrates the vital role that mobility dynamics plays in the economic evolution of a city.

CCS Concepts: • **Information systems** → *Spatial-temporal systems; Data mining*; • **Human-centered computing** → *Empirical studies in ubiquitous and mobile computing*;

Additional Key Words and Phrases: Urban computing, spatio-temporal patterns, predictive modeling, location-based services

ACM Reference Format:

Krittika D'Silva, Kasthuri Jayarajah*, Anastasios Noulas, Cecilia Mascolo, and Archan Misra. 2018. The Role of Urban Mobility in Retail Business Survival. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 100 (September 2018), 22 pages. <https://doi.org/10.1145/3264910>

1 INTRODUCTION

There is a strong economic and policy interest in both uncovering the causes of failure of retail businesses and in predicting their likelihood. 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

*These two authors contributed equally & are listed alphabetically.

[†]Work was performed while a visitor at the University of Cambridge.

Authors' addresses: Krittika D'Silva, University of Cambridge, krittika.dsilva@cl.cam.ac.uk; Kasthuri Jayarajah*, Singapore Management University, kasthuri.2014@smu.edu.sg; Anastasios Noulas, New York University, noulas@nyu.edu; Cecilia Mascolo, University of Cambridge and The Alan Turing Institute, cecilia.mascolo@cl.cam.ac.uk; Archan Misra, Singapore Management University, archanm@smu.edu.sg.

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

© 2018 Association for Computing Machinery.

2474-9567/2018/9-ART100 \$15.00

<https://doi.org/10.1145/3264910>

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 [27, 33]. 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. 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?). Researchers have explored the use of social media for business analytics—e.g., Wang et al. [33] 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. [17] 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 and (2) *taxi trip records*, observed across Singapore and New York City, 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

Key Research Questions and Contributions: Our investigations require us to tackle and answer three key questions, enumerated in the sequence they are addressed here:

- *What are some of the key factors that explain business survival?* To address this question, we identify and analyse a range of features and perform a comprehensive study to demonstrate the predictability of survival or failure of a F&B business in the subsequent 6-month period, in two metropolitan cities, New York and Singapore. Overall, we achieve AUCs of 0.85 and 0.90, for Singapore and New York, respectively, with corresponding precision/recalls at $\approx 80\%$ —this represents an almost 10-15% improvement in accuracy, over similar past work [33]. We also found that the most important factor was the ability of an establishment to draw customers around the clock and not just during specific hours.
- *How generalizable is business survival predictability?* We answer this question in three parts, we: (1) extend our analysis to the broader category of *retail* venues, (2) repeat our analyses on multiple cities across the world, and (3) investigate differences between newly founded and established venues. Our results show consistent performance and also show that certain features have consistently high power (AUCs ranging from 0.82 to 0.84) for predicting survival likelihood, despite geographic differences among cities. Our results show that, across all ten cities considered and using the same classifier, established venues had an area

¹The 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.

under the curve (AUC) of 0.86 while newer venues had a somewhat lower AUC of 0.81. These results suggest that new venues may have more variability in their underlying causes of failure.

- *How robust are our results?* We perform a series of experiments to validate the robustness of our results. We study the collinearity across the set of features considered and demonstrate with a reduced model consisting of only a subset of features that the accuracy drops only by a few points (e.g., 3% for New York City, from 0.92 to 0.89). Additionally, we demonstrate that with more accurate labeling, our models perform much better (e.g., reaching perfect classification for certain cities). Our results also show that for cities with a high volume of check-in data, the prediction accuracy is relatively unaffected even under shorter observation periods.

Overall, we provide compelling new evidence of the power of combining venue-specific, location-related and mobility-based features in predicting the likely demise of retail/F&B establishments across different cities, despite lacking visibility over other factors (e.g., management quality, reviews, and broader economic trends) that plausibly influence such business outcomes.

2 OUR APPROACH AT A GLANCE

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 4), pose the problem of predicting closure as a binary classification task, and report our findings in Section 5. 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 questions:

- (1) Can metrics of the locality profile, visitation patterns, and mobility dynamics of a retail business be used as predictors its success or failure?
- (2) Do factors that attribute to business failure vary by city, or geographies?
- (3) Is failure similar for new and established businesses?

2.1 Notation

We consider the set V of venues in a city. A venue $v_i \in V$ is represented with a tuple $\langle loc, date, gen, spec \rangle$ 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 3.1). Further, we define a venue's *neighbourhood*, N_i , as the set of venues that are located within a given radius; we set this distance to 500m as prior work [10] has shown that a venue's operation is affected primarily by conditions within this radial distance. Formally, we define the neighbourhood as:

$$N_i = \{v_j \in V : dist(v_i, v_j) < D\} \quad (1)$$

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*, $CN_i \subseteq N_i$, as the subset of venues that belong to the same general category, *gen*, as the venue. Similarly, we define the specific competitive neighbourhood, $SN_i \subseteq CN_i$, 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. Later in Section 5.5, we use this distinction to examine the impact of a venue's age on prediction accuracy.

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 1.

2.2 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 [10]. 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. [33]), 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:

$$RemainsOpen(v_i) = \begin{cases} 0, & \text{if } \frac{\sum_{t=0}^T C_t(v_i)}{T} < K \times mean(v_i) \wedge \frac{\sum_{t=0}^T C_t(v_i)}{T} < N \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

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 3.2 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.

2.3 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 [8] due to practical limitations in acquiring such information (see Section 3). 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 to label whether a venue has closed. We depict this in Figure 1. The features used in this work (described later in Section 4) 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 1. Acronyms used throughout the text.

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

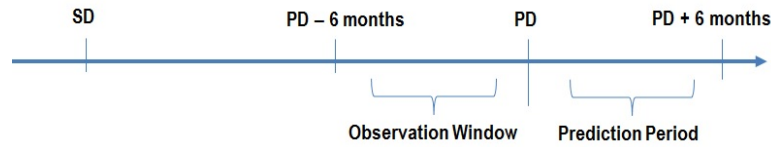


Fig. 1. Definition of virtual past and future data used in this work. We use a fixed Prediction Date (PD) across all venues and answer the question, *which of these venues will close during the prediction period (PD, PD + 6]?* using features computed over (PD - 6, PD].

3 MOBILITY DATASETS

3.1 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 [32].

In this work, we use a longitudinal dataset from multiple cities around the world, that spans three years and over 75 million check-ins. Table 2 includes the summary of statistics of the 10 cities we consider in this work—for the sensitivity analysis in Section 5.5, we also enumerate the count of new vs. established venues. For each venue, we have the following information: geographic coordinates, specific and general category which fall within Foursquare’s API of hierarchical categories², 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 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 and the timeline described in Section 2.3.

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.

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 Commission³. 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 tracts⁴, where a tract typically houses at most 16,000 residents⁵. 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 subzones⁶ which are administrative boundaries. Table 3 summarises key statistics of these datasets.

²<https://developer.foursquare.com/categorytree>

³http://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>

Table 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). Venue closure was defined using Equation 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

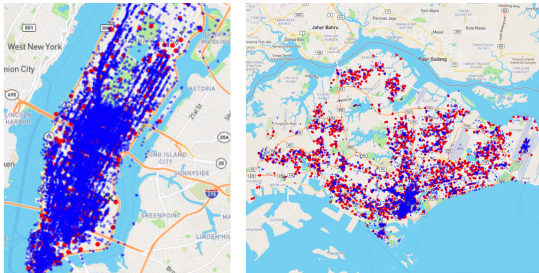


Fig. 2. Spatial distribution of venues in New York City (left) and Singapore (right). Blue represents “open” and red represents “closed”, as defined by Eq. 2.

3.2 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.

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 3 for F&B venues that started operation after June 2011. We use the Kaplan-Meier estimator [16] 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

Table 3. Summary of taxi datasets used in the analysis.

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

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), but also illustrates the challenge of perfect labelling. Later in Section 5.6 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.

4 FEATURE DESCRIPTION

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

4.1 Profile of the Locality

Prior studies on restaurant failure in [27, 26, 24, 25, 33], indicate that a 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.

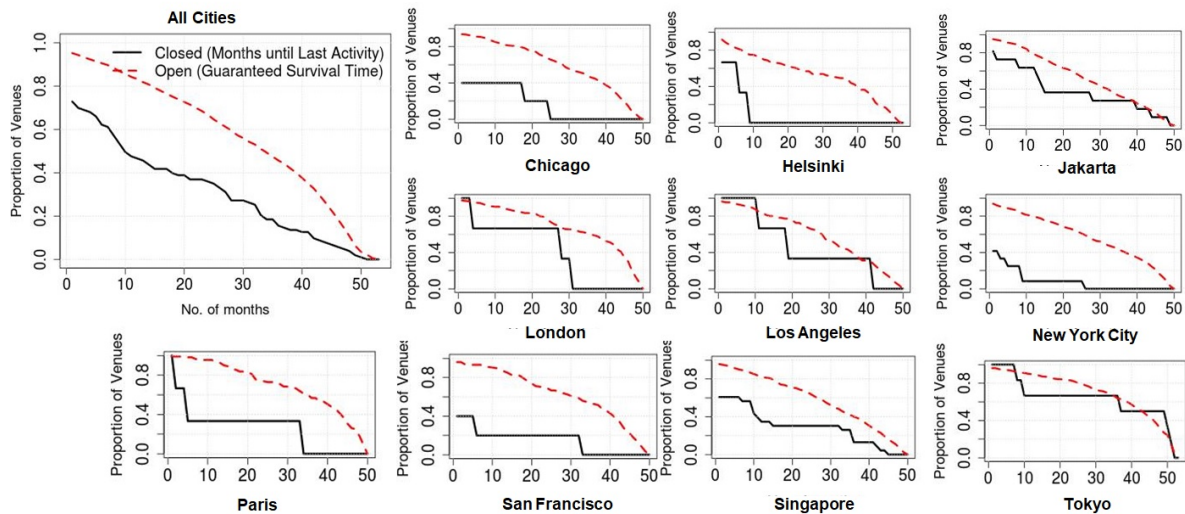


Fig. 3. The survival curves (as KM plots) for all F&B venues considered in this work.

Table 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	$ CN_i , SN_i $	Foursquare
	Attractiveness to the Neighbourhood	$ CN_i \times \ln \left(\frac{ V }{ V_c } \right)$	Foursquare
	Catchment of Locality	$\frac{ D_l }{ D }$	Transport
	Temporal Catchment of Locality	$\frac{ D_{w,l} }{ D_w }$	Transport
Customer Visit Patterns	Inflow & Outflow	$\frac{\sum_{j=0}^{ V } t(v_j, v_i)}{M}, \frac{\sum_{j=0}^{ V } t(v_i, v_j)}{M}$	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	$\sum_{j=1}^{24} (h_i(j) - h_j(j))^2$	Both
	Reachability	$r_{(a,b)}$	Both
	Distance-weighted Reachability	$dr_{(a,b)}$	Both
Business Attributes	Cuisine Type	Categoric variables	Foursquare
	Price Tier		Foursquare

We next define **Place Entropy** of the area around a venue through the Shannon equitability index [29] from information theory. This metric is calculated as follows:

$$- \sum_{c=1}^k p_c * \ln p_c / \ln k \quad (3)$$

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: $\left(\frac{|D_l|}{|D|} \right)$, 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, $\frac{|D_{w,l}|}{|D_w|}$.

4.2 Visit Patterns

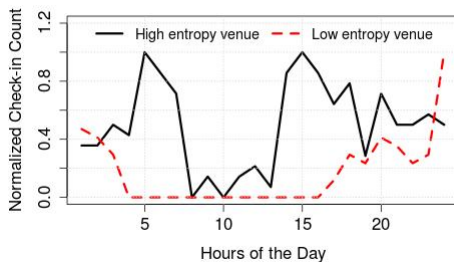
The variation in trends of customer visits to such locations can reveal important insights into how businesses are faring. We define a number of visitation-driven features for capturing venue-level differences.

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 $t(v_j, v_i)$ is the number of transitions from v_j to v_i , we define the average monthly inflow to venue v_i as $\frac{\sum_{j=0}^{|V|} t(v_j, v_i)}{M}$ and outflow as: $\frac{\sum_{j=0}^{|V|} t(v_i, v_j)}{M}$. We also calculate the ratio of inflow to outflow for both the venue and the surrounding area.

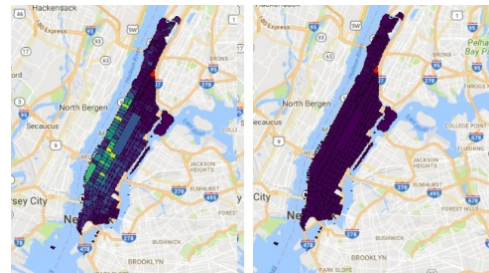
Prior research has shown the **Distance of Travel To** and the **Distance of Travel From** certain nodes in a network correlates with higher changes of connection between those nodes [28]. We thus measure the mean distance travelled to reach a venue, as well as to the surrounding area (i.e., N_i). Formally, 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 with 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 3). As seen in Figure 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 [22] 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



(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.

Fig. 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.

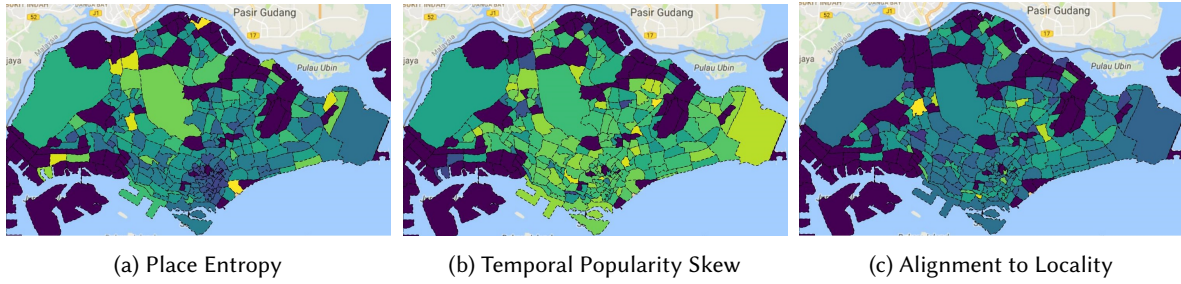


Fig. 5. Select features spatially aggregated over localities across Singapore. Darker regions represent higher numerical values of the respective features.

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) 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 $\frac{\sum_{t=0}^T C_t(v_i)}{\text{mean}(v_i)}$ (see Eq. 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).

4.3 Mobility Dynamics

These features embody our intuition that the temporal patterna 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 $dr_{a,b} = r_{a,b}/d_{a,b}^2$ 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 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 5, we visualize the spatial spread of three features aggregated over the different localities.

5 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 5.1, and summarise the overall performance of the combination of features in Section 5.2 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. Then, in Sections 5.4 and 5.5, we scale our analysis to ten different cities around the world (focusing on *Visit Patterns*, which can be derived from Foursquare data alone), to answer two additional key questions:

- (1) do factors that affect business survival vary from city to city?
- (2) is our prediction framework able to detect failure in both new and established businesses?

Finally, in Section 5.6, we study the robustness of our criteria for defining closure (see Section 3.2).

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 4 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), 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 [2]) and the ROCR [3] library for performance calculations. The comparison across multiple machine learning models (see Section 5.5) were built using Python and the scikit-learn library [4].

5.1 Feature Selection and Pruning

In order to understand the ability of the features described in Section 4 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 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 [9] 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* ≥ 0.75 . We apply the Boruta algorithm [19] for feature selection and consider the features that were consistent across F&B venues from both Singapore and New York City in subsequent analyses.

5.2 Predicting Venue Closure

In this section, we summarise our findings from running logistic regression [13] on the 20 confirmed features resulting from the Boruta search in Table 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.
- (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.
- (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 7, the work of Wang et al. [33] 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

Table 5. Features with the highest performance in predicting venue closure for Singapore (left) and New York City (right).

Feature	AUC	Correlation	Feature	AUC	Correlation
Temporal Popularity Skew	0.788	0.528	Temporal Popularity Skew	0.794	0.589
Alignment with Neighborhood	0.786	-0.479	Alignment with Neighborhood	0.782	-0.575
Alignment with Locality	0.758	-0.526	Alignment with Locality	0.732	-0.555
Inflow	0.66	0.270	Trend	0.716	0.487
Outflow	0.624	0.268	Inflow (Neighborhood)	0.626	-0.201

Table 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, <i>Retail</i>	SG, F&B	NYC, <i>Retail</i>	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

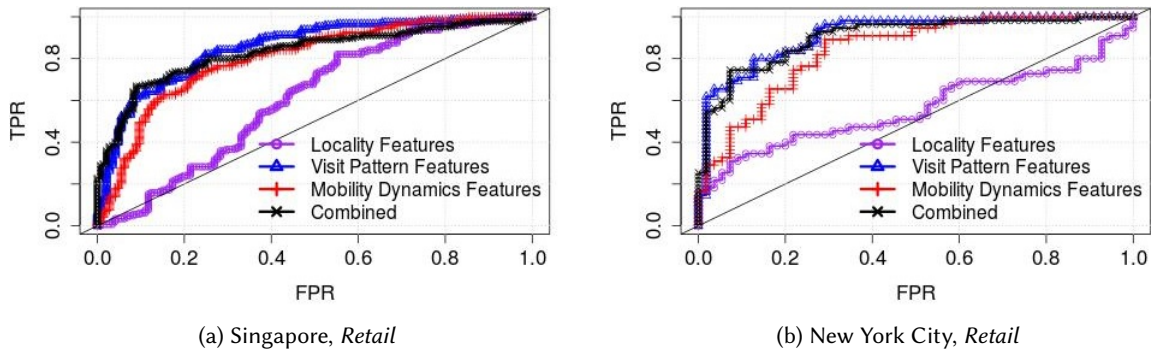


Fig. 6. ROC Curves of *Retail* venues in Singapore and New York City. The Curves represent the performance for each class of features and for the combined model, respectively.

in NYC, over a 3 month window. In Table 7, we compare our results against this baseline, reproducing the confusion matrix presented in [33]. 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 [33] may be a bit misleading as they are based on a highly imbalanced dataset (with less than 20 samples in their “closed” class).

5.3 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 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

Table 7. Confusion Matrix Comparison against Previous Work [33].

Confusion Matrix	Wang et. al[33]		NYC, <i>Retail</i>		NYC, F&B	
	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 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

addition of mobility dynamics features lends to only a 4% and 3% increase in accuracy for Singapore and NYC, respectively. Figure 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 features alone.

5.4 Individual Cities Versus All Cities

We next examine how similar the impact of features are for the ten different cities listed in Table 2. We analyse venues whose specific category are *Retail*, as defined above and consider features that are extracted using Foursquare alone (see Table 4 –i.e., the various *Customer Visit Pattern* features and the bulk of the *Locality Profile* features) due to the unavailability of transport features across all cities. We first examine the per city AUC scores using our Logistic Regression model. We summarise our results of coefficients and their significance level in Table 9. The results show a range in the coefficients for all features, suggesting that certain factors differently affect different cities. However, our results show that certain features are consistently significant across many or all our cities. We see that *Visit Trend* acts as a significant feature for all ten cities. Additionally, *Temporal Popularity Skew*, *Temporal alignment to Competitors*, and *Distance of Travel From* are significant features in many of the cities. These results suggest that the dynamics of urban environments have a strong influence on venues. Additionally, we note that the sign of the coefficient is consistent across all cities, suggesting the role these features play in the success of businesses is consistent. This analysis suggests that large-scale multi-city generalisations can be challenging as cities are often unique in their attributes. This suggests that although venue closure is predictable as a task, features vary by city in their contribution to business failure in individual cities.

5.5 The Impact of Venue Age on Prediction Accuracy

We examine next the predictability of two classes of venues: those that are established, *est*, and those that are new, *new*. We define established venues as those that have existed for longer than a year. Conversely, we define new venues as those have existed for less than one year. This analysis is performed across all the 10 cities, primarily because the number of new venues in any single city is too small for meaningful analysis. We analyse venues whose specific category are *Retail*, as defined above.

As described in more detail in Section 3.2, we examine venue closure between the time period of July to December 2013. For our analysis, we select new venues as those that opened on or after June 1 2012. As prior work has shown that venues added (in Foursquare) after June 2011 were highly likely to actually be new venues [10],

Table 9. Per city AUC score and the logistic regression coefficients for multiple cities for the top five most significant features. *** represents $p < 0.001$, ** represents $p < 0.01$, and * represents $p < 0.05$.

	Visit Trend	Temporal Skew	Alignment to Neighborhood	Distance of Travel From	Speed of Entering	AUC
Chicago	0.16***	2.05***	-3.66**	3.90	-0.019	0.8605
Helsinki	0.06*	1.72*	-6.47*	3.16**	-0.37	0.7725
Jakarta	0.13***	1.46***	-0.80	2.21*	-0.093	0.8326
London	0.19**	1.66	-7.68*	1.65*	-0.35	0.7689
Los Angeles	0.32***	2.23***	-3.45*	0.82	-0.033	0.8001
New York	0.16***	1.96***	-2.57*	0.76	-0.41**	0.8633
Paris	0.15***	2.07***	-1.86	0.72*	-0.24	0.8203
San Francisco	0.27***	1.72**	-3.44	1.14	-0.37	0.7775
Singapore	0.10***	3.26***	-0.34	0.32	-0.041	0.8355
Tokyo	0.13***	1.39***	-2.04*	0.48*	-0.26**	0.8155
All cities	0.12***	2.40***	-0.74	0.12*	-0.12*	0.8803

Table 10. Closure predictions for new and established venues.

Classifier	New venues			Established venues		
	Precision	Recall	AUC	Precision	Recall	AUC
Logistic regression	0.701	0.758	0.788	0.767	0.804	0.861
Gradient boosting	0.717	0.737	0.812	0.794	0.806	0.882
Support vector machine	0.693	0.765	0.817	0.771	0.801	0.863
Random forest	0.742	0.7	0.821	0.802	0.732	0.855
Neural Networks	0.737	0.668	0.809	0.779	0.767	0.858

we have strong confidence in our classification. We compare the prediction performance with multiple classifiers and present our results in Table 10.

There is variability across different classifiers; the best performance in terms of AUC is given by a gradient boosting model, which resulted in an AUC of 0.882. Further examination of the results also shows that established venues have a notably higher AUC scores than new venues, suggesting, as intuitively expected, that new venues may be more susceptible to high variations in their causes of failure.

5.6 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 5.1, here we test for collinearity across variables - to detect collinearity, we compute the Variable Inflation Factor (VIF) [23] 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 from Section 5.2, against a subset of features whose Pearson's correlation with any other feature is less than 0.5 which we refer to as the *Reduced Model*, in Table 11.

Table 11. 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.

Feature	Combined Model		Reduced Model	
	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	

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 3.2, the labeling definition (as defined in Equation 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 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) 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) 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.)

Table 12. 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 <i>Confident Venues</i>				
	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

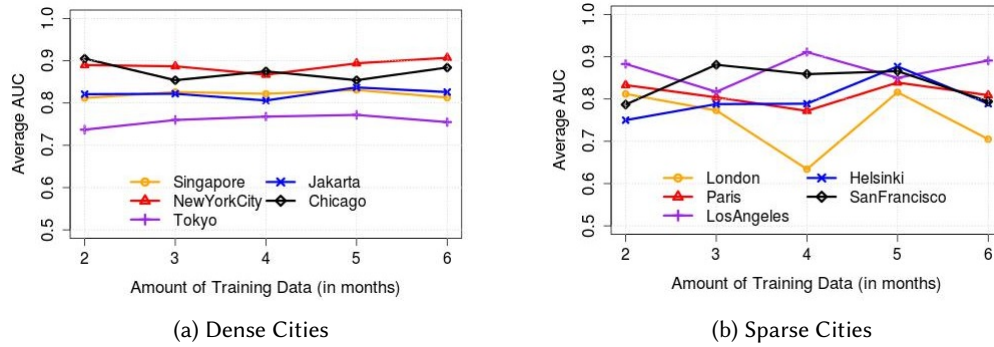


Fig. 7. Impact of the Length of Observation Window on Performance.

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 12, 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. Note that we consider all F&B venues in the dataset here, and not only those that were created after June 2011 as was the case in our previous analyses (See Section 5.5). 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 7, 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.

6 DISCUSSION AND FUTURE WORK

Here, we describe key implications of our work to business owners and urban planners alike, discuss the limitations in the current study and our plans for future work.

6.1 Implications

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 [7]. 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.

6.2 Limitations and Future Work

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 government 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.

Alternative Solution Approaches: As deep learning models have been extensively used in a number of problem domains [8, 30, 18], we also conducted some preliminary studies using the PyTorch framework and the scikit-learn library [4]. Similar to Avati et al. [8], 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). 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 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 4.

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.

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

7 RELATED WORK

We categorise the related work along several dimensions.

Studies on Business Survival and Mortality: Between 2004 and 2014, Parsa et al. [27, 26, 24, 25], published a four-part series on *Why Restaurants Fail*. In [27], based on a quantitative study of 2400+ restaurants in Columbus, they present a framework for survival, composed of four main areas: environmental factors, family lifecycle, internal factors and growth stage of the restaurant. In our work, we focus on the environmental factors (driven by the retail establishment's location) and additionally include venue-specific visitor dynamics to build a quantitative predictor of an establishment's 6-month failure likelihood.

LBSN-driven Business Analysis: The analysis presented by Lei Wang et al. in [33] is closest in spirit to our work. In [33], 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. Among other findings, their analyses show that: (i) the competitive analysis matters: considering the check-in information of a restaurant and its neighbours reduces the misclassification rate from 30% to around 10%, and (ii) such competitive effects dissipate beyond a radius of 1 km. In our work, we not only perform a more comprehensive study (10 cities, on multiple categories of venues), and also study the factors which govern business failure at a global level, incorporating several new key features, especially those related to a neighbourhood's mobility dynamics.

A variety of works have explored facets of LBSN-based prediction of business demand. In [20], the authors develop *ZoneRec*, a framework for recommending zones for a new F&B venue. In this framework, a new venue's features are computed using TF-IDF scores on the venue label, and then matched to zones with similar features. More recently, in [10], the authors used changes in Foursquare check-in volumes, before and after opening of a new venue, to study whether certain categories of businesses *cooperate* or *compete* with other close-by businesses in the same category. Finally, in [11], the authors use Foursquare check-ins to investigate whether the stable temporal demand of a new venue (that presently has no historical trends of check-in data) can be predicted 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*.

Urban Dynamics using Social Media: In this work, we saw that the mobility dynamics of individual locations, at the neighbourhood level, turned out to be a key feature in our failure prediction model. Recent work has used social media data to investigate various aspects of such urban dynamics. In [15], the authors carry out a study on gentrification in the wards of London, and more recently in [34], 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, in works such as [14], 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.

Fusing Social Media and Physical Sensor Data for Urban Analytics: Our predictive model utilises features that are derived from a combination of social media and physical mobility (from taxi and bus traces) data. In [31], the authors tackled the problem of unifying multiple streams/modalities of Web-based sensory data by introducing the concept of ‘social pixels’, which aggregates user interest across multiple channels at a particular geo-location. In [21], the authors introduced a vision for *socio-physical analytics*, and outlined the challenges and opportunities in fusing these disparate data sources, from the social Web and physical sensor streams. In [12], the authors applied the concept of socio-physical analytics to wellness profiling, combining data from social media posts (Twitter tweets, Instagram pictures and Foursquare check-ins) and wearable sensor data (captured by Endomodo, an exercise tracking App) to infer the BMI (Body Mass Index) profile of users.

8 CONCLUSIONS

We have presented an approach to the prediction of venue closure. The approach uses a variety of features over spatio-temporal data crowd-sourced from location-based social networks and transport data. Our results show that it is important to go beyond just static features, of a retail establishment and its neighbourhood, and include mobility-derived features, related to both the visitor patterns to the venue and the aggregate movement patterns over the venue’s neighbourhood. Using a variety of such features we show, through an analysis of 10 cities that an appropriate classification model can generate higher prediction accuracy (*AUC* of ≈ 0.80) than previously reported. In ongoing work, we are attempting to better understand how (i) such features vary across different neighbourhoods, based on other factors such as resident demographics, (ii) additional features, from other information sources (e.g., electricity or water consumption) help enhance the accuracy, and (iii) how the classification accuracy will vary with changes to the prediction time horizon.

ACKNOWLEDGMENTS

This material is supported partially by the National Research Foundation, Prime Minister’s Office, Singapore under its International Research Centers in Singapore Funding Initiative. K. D’Silva’s work was supported through the Gates Cambridge Trust. K. Jayarajah’s work was supported by an A*STAR Graduate Scholarship. A. Noulas would like to thank the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation for their continued support at the Center for Data Science, New York University. A. Misra’s work was undertaken with the support of the Alan Turing Institute (EPSRC Grant Number EP/N510129/1). We thank Foursquare for supporting this research by providing the dataset used in the analysis. We also thank Mike Kenny for sharing insights from a public policy perspective, and Andrew Tan and Lakmal Meegahapola for conducting interviews with F&B owners.

REFERENCES

- [1] [n. d.]. Get Details of a Venue. <https://developer.foursquare.com/docs/api/venues/details>. [Online; Last accessed 09-May-2018].

- [2] [n. d.]. The R Stats Package. <http://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>. [Online; Last accessed 13-May-2018].
- [3] [n. d.]. ROCR: Visualizing the Performance of Scoring Classifiers. <https://cran.r-project.org/web/packages/ROCR/index.html>. [Online; Last accessed 13-May-2018].
- [4] [n. d.]. scikit-learn Machine Learning in Python. <http://scikit-learn.org/stable/>. [Online; Last accessed 13-May-2018].
- [5] [n. d.]. Shopify. <https://www.shopify.com.sg/>. [Online; Last accessed 14-May-2018].
- [6] [n. d.]. 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] U.S. Small Business Administration. 2018. Apply for licenses and permits. <https://www.sba.gov/business-guide/launch-your-business/apply-licenses-and-permits>.
- [8] Anand Avati, Kenneth Jung, Stephanie Harman, Lance Downing, Andrew Ng, and Nigam H Shah. 2017. Improving palliative care with deep learning. *arXiv preprint arXiv:1711.06402* (2017).
- [9] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 1–4.
- [10] Matthew L. Daggitt, Anastasios Noulas, Blake Shaw, and Cecilia Mascolo. 2016. Tracking urban activity growth globally with big location data. *Open Science* 3, 4 (2016). <https://doi.org/10.1098/rsos.150688>
- [11] Krittika D’Silva, Anastasios Noulas, Mirco Musolesi, Cecilia Mascolo, and Max Sklar. 2017. 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)*.
- [12] Aleksandr Farseev and Tat-Seng Chua. 2017. 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*. 95–101.
- [13] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. 2001. *The elements of statistical learning*. Vol. 1. Springer series in statistics New York.
- [14] Petko Georgiev, Anastasios Noulas, and Cecilia Mascolo. 2014. 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*.
- [15] Desislava Hristova, Matthew J. Williams, Mirco Musolesi, Pietro Panzarasa, and Cecilia Mascolo. 2016. Measuring Urban Social Diversity Using Interconnected Geo-Social Networks. In *Proceedings of the 25th International Conference on World Wide Web (WWW ’16)*.
- [16] Edward L Kaplan and Paul Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association* 53, 282 (1958), 457–481.
- [17] Dmytro Karamshuk, Anastasios Noulas, Salvatore Scellato, Vincenzo Nicosia, and Cecilia Mascolo. 2013. 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)*.
- [18] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 1725–1732.
- [19] Miron B Kursu, Witold R Rudnicki, et al. 2010. Feature selection with the Boruta package. *J Stat Softw* 36, 11 (2010), 1–13.
- [20] Jovian Lin, Richard J Oentaryo, Ee-Peng Lim, Casey Vu, Adrian Vu, Agus T Kwee, and Philips K Prasetyo. 2016. A business zone recommender system based on Facebook and urban planning data. In *European Conference on Information Retrieval*. Springer, 641–647.

- [21] Archan Misra, Kasthuri Jayarajah, Shriguru Nayak, Philips Kokoh Prasetyo, and Ee-peng Lim. 2014. Socio-physical Analytics: Challenges & Opportunities. In *Proceedings of the 2014 Workshop on Physical Analytics (WPA '14)*.
- [22] Anastasios Noulas, Blake Shaw, Renaud Lambiotte, and Cecilia Mascolo. 2015. 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)*.
- [23] Robert M O'brien. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity* 41, 5 (2007), 673–690.
- [24] HG Parsa, Amy Gregory, and Michael Terry. 2011. Why do restaurants fail? Part III: An analysis of macro and micro factors. *Dick Pope Sr. Institute Publications* 15 (2011).
- [25] HG Parsa, Jean-Pierre I van der Rest, Scott R Smith, Rahul A Parsa, and Milos Bujisic. 2015. Why restaurants fail? Part IV: The relationship between restaurant failures and demographic factors. *Cornell Hospitality Quarterly* 56, 1 (2015), 80–90.
- [26] H. G. Parsa, John Self, Sandra Sydnor-Busso, and Hae Jin Yoon. 2011. 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 (2011), 360–379.
- [27] H. G. Parsa, John T. Self, David Njite, and Tiffany King. 2005. Why Restaurants Fail. *Cornell Hotel and Restaurant Administration Quarterly* 46, 3 (2005), 304–322.
- [28] Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. 2011. Exploiting Place Features in Link Prediction on Location-based Social Networks (*KDD '11*). ACM, 9. <https://doi.org/10.1145/2020408.2020575>
- [29] Andrew L. Sheldon. 1969. Equitability Indices: Dependence on the Species Count. *Ecology* 50, 3 (1969), 466–467.
- [30] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [31] Vivek K. Singh, Mingyan Gao, and Ramesh Jain. 2010. Social Pixels: Genesis and Evaluation. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*.
- [32] VentureBeat. 2015. Foursquare by the numbers. <https://goo.gl/Vi1UUf>.
- [33] Lei Wang, Ram Gopal, Ramesh Shankar, and Joseph Pancras. 2015. On the brink: Predicting business failure with mobile location-based checkins. *Decision Support Systems* 76 (2015), 3 – 13. Analyzing the Impacts of Advanced Information Technologies on Business Operations.
- [34] Xiao Zhou, Desislava Hristova, Anastasios Noulas, Cecilia Mascolo, and Max Sklar. 2017. Cultural investment and urban socio-economic development: a geosocial network approach. *Open Science* 4, 9 (2017). <https://doi.org/10.1098/rsos.170413>

Received February 2018; revised May 2018; accepted September 2018.