Published in: Transportation 2015, Volume 42, Issue 4, pp. 541–559.
(Special Issue: Emerging, Passively Generated Datasets for Travel Behavior and Policy Analysis)
DOI: 10.1007/s11116-015-9599-9

Comparing Cities' Cycling Patterns Using Online Shared Bicycle Maps

Advait Sarkar · Neal Lathia · Cecilia Mascolo

Received: 15 May 2014 / Accepted: 3 November 2014

Abstract Bicycle sharing systems are increasingly being deployed in urban areas around the world, alongside online maps that disclose the state (i.e., location, number of bicycles/number of free parking slots) of stations in each city. Recent work has demonstrated how regularly monitoring these online maps allows for a granular analysis of a city's cycling trends; further, the literature indicates that different cities have unique spatio-temporal patterns, reducing the generalisability of any insights or models derived from a single system. In this work, we analyse 4.5 months of online bike-sharing map data from 10 cities which, combined, have 996 stations. While an aggregate comparison supports the view of cities having unique usage patterns, results of applying unsupervised learning to the temporal data shows that, instead, only the larger systems display heterogeneous behaviour, indicating that many of these systems share intrinsic similarities. We further show how these similarities are reflected in the predictability of stations' occupancy data via a cross-city comparison of the error that a variety of approaches achieve when forecasting the number of bicycles that a station will have in the near future. We close by discussing the impact of uncovering these similarities on how future bicycle sharing systems can be designed, built, and managed.

Advait Sarkar
E-mail: advait.sarkar@cl.cam.ac.uk

Neal Lathia
E-mail: neal.lathia@cl.cam.ac.uk

Cecilia Mascolo
E-mail: cecilia.mascolo@cl.cam.ac.uk

Computer Laboratory, University of Cambridge
William Gates Building
15 JJ Thomson Avenue
Cambridge CB3 0FD, United Kingdom

1 Introduction

A growing body of research uses the digital records that we create while using web-based services to analyse and model behaviour in the physical world. These include, for example, studying online location-based social networks to understand user behaviours [24], rank nearby venues [22], build support for targeted advertising [20], facilitate mobile/local search [25], and using photo-sharing sites to uncover tourists' mobility throughout a city [11]; the common trend is that of using interactions with web services to learn about behaviour, and design future applications that leverage any predictability and insight derived from the data.

In this work, we examine how the online maps which were built to inform users of the state of urban bicycle sharing systems can be used to analyse, compare and predict mobility across the growing number of cities that are adopting these forms of transport. A bicycle sharing (or bike share) system is a service that makes bicycles available for use to urban commuters; users can pick up and return the bicycles to and from any one of the stations dispersed in the city. Notable examples include Barclay's Cycle Hire (London, England), Vélib' (Paris, France), Capital Bikeshare (Washington DC, USA), and Citi Bike (New York, USA) . While these systems provide a healthy, sustainable, and traffic-reducing means of navigating a city, they continue to suffer from a variety of shortcomings: most notably, the problem of balancing between system usage and demand, which leads to a lack of available bicycles or free parking spaces at stations at various times of the day [2].

These systems are often accompanied by online maps that give a snapshot of the current state of the system: for each station, they provide the number of available bicycles and free parking slots. In the following, we investigate the extent that these online maps can be used to uncover a variety of bike share systems' common patterns, and evaluate how the accuracy of a set of forecasting algorithms varies across cities—which differ in size, geography, and system usage. In particular, we make the following contributions:

1. We have collected a dataset from 10 cities' bicycle sharing systems; we describe how the data was collected and pre-processed prior to analysis in order to accommodate for errors that arise from sourcing data from online maps. The resulting data comprises 4.5 months and over 10^8 samples from 996 stations in 6 countries.
2. We apply unsupervised clustering techniques to address the extent that bike sharing stations (independently of the city they reside in) share similar patterns in their usage data. We find that stations across the 10 systems share substantial usage patterns; in fact, the smaller systems' stations cluster together, and heterogeneity is only apparent in those systems with more than 100 stations.
3. Finally, we use random forests and neural networks to compare the accuracy of forecasting how many bicycles will be at a given station and time to two baselines, and evaluate how prediction accuracy varies across cities. We find that, while prediction accuracy degrades for larger systems as models are queried for estimates that are further in the future, large systems' accuracy is not influenced by how recent the training data is (and the opposite results hold for smaller systems).

We close by discussing the analyses and applications that arise from this data, which include further support for cyclists (e.g., based on station occupancy prediction) and the transport authority (e.g., supporting redistribution and station placement decisions).

2 Related Work

Studies of shared bicycle systems have recently appeared in the data mining literature, and are often subjects for online data visualisations[1]. Froehlich, Neumann and Oliver [10] were the first to apply clustering techniques and forecasting models to identify patterns of behaviour in stations in Barcelona's "Bicing" system, explaining results according to stations' location and time of day. Similar clustering was used in a study that focused on London's system [13], in order to assess the effect of policy changes; notable differences in the way the system was used prior to and after the policy change were quantified. Guenther et al. [12] also focused on London, and built and validated a number of arrival forecasting models that predicted cumulative arrivals in small geographic clusters of stations (falling within 500m×500m squares) during peak hours; similar case studies have been published with data from Paris [17] and Singapore [23], and Lyon [7]. This family of recent work, which focuses on individual cities, demonstrates that repeated observations of these maps can be used to characterise a city's bike share spatio-temporal patterns: a recurring conclusion across analyses is that spatiotemporal system usage patterns are tied to, and reflect, city-specific characteristics. By focusing on single cities' systems, these works seem to indicate that each city has a unique pattern, and that forecasting algorithms applied to each one may not be generalisable across the world.

Beyond the data mining literature, two recent works address the multi-city scenario. O'Brien et al. [14] and Austwick et al. [4] characterise systems at the city-level, comparing them in terms of system size (both by station count and geographic area), daily usage, and compactness; they build a hierarchy of cities that share similar characteristics and apply community detection algorithms to analyse similarities within systems. Parkes et al. [15], instead, compare systems' policies, technologies, and reasons for bike sharing adoption (via both published data and qualitative interviews). Using diffusion theory, they comment on the importance of private sector operators as well as the influence of certain successful bike sharing systems such as the ones in Paris, Lyon, Montreal and Washington DC.

3 Data Pre-Processing

A typical bike sharing online map discloses the latitude, longitude, number of available bicycles and vacant parking slots of each station in the city. We gathered this station occupancy data for 10 bike sharing systems (Table 1) by scraping their web services every two minutes for a number of months. We define the full set of times

[1] E.g., `http://bikes.oobrien.com/global.php`, as of April 22, 2015

		Pre-Cleaning		Post-Cleaning		
City	**System**	**Observations**	**Stations**	**Observations**	**Stations**	**Data Retained** (%)
Barcelona, Spain	Bicing	38,674,016	415	37,087,351	409	95.90
Denver, USA	Denver B-Cycle	4,907,697	52	4,787,219	50	97.55
Girona, Spain	GiroCleta	956,510	10	945,154	10	98.81
João Pessoa, Brazil	SAMBA	389,260	4	381,760	4	98.07
London, England	Barclays Cycle Hire	37,954,996	410	35,414,070	390	93.31
Miami, USA	DecoBike	8,158,542	99	4,589,939	53	56.26
Rio de Janeiro, Brazil	BikeRio	2,132,812	22	2,091,210	22	98.05
Rome, Italy	Roma'n'Bike	2,941,530	30	2,924,098	30	99.41
Siracusa, Italy	GoBike	1,750,500	18	1,740,400	18	99.42
Taipei, Taiwan	YouBike	1,081,630	11	975,540	10	90.19

Table 1 Bike sharing system data collected from their online maps pre/post cleaning.

that we queried for data as T, and the set of stations as S. A sample of station $s_i \in S$ at time $t \in T$ is a tuple of the form:

$$s_i(t) : b_i(t), v_i(t) \tag{1}$$

Where b_i is the number of available bicycles, and v_i is the number of vacancies (free parking bays). The full dataset is the largest continuous sample of data that we obtained and spans the 4.5 months between March 23rd, 2011 and August 6th, 2011.

3.1 Station Size and Location Inference

It is important to note we have no further data; in particular, we could not obtain static meta-data about stations from the web service, which includes stations' actual size (in terms of number of parking bays) and precise geographic location. We thus had to infer these traits from the data samples we obtained. In doing so, we also pruned inconsistent and inaccurate samples and stations from the dataset.

Station Capacity. We refer to a station's *capacity* as the maximum number of bicycles that it is possible to park at a station. For each sample $s_i(t)$ we thus define an observed capacity, $b_i(t) + v_i(t)$. We note that for any given station, its observed capacity time series was not strictly constant or monotonically increasing: this may be explained by malfunctioning docks (i.e., erroneous data), or stations whose capacity is changed by new docks being added or removed (i.e., actual changes in station sizes). To allow for station sizes to grow, yet account for errors in the data, we remove those stations whose sizes appear to fluctuate at a higher than daily rate. Formally, we have 720 samples per station per day: if, across the data, we observe a station size that appears for fewer than 720 times, we remove the station from our analysis. Across the 10 systems, there are 996 stations in total which range from small (e.g., size < 10) to large (size > 50).

Sample Pruning. We found that the fluctuating sizes described above were temporally close to one another: a high number of invalid observations in a single day signals anomalous station behaviour and, therefore, potential problems with the remaining samples of that day. We thus removed all days with fewer than 504 (70% of

720) samples. Finally, we removed all stations with fewer than 62 remaining days, i.e. 45% of our total 4.5 month period.

Station Locations. We assume that a station's usage will be highly dependent on the characteristics and geography of the urban area that surrounds it; our analysis resides on these remaining constant throughout the period of observation. However, stations may be moved and, moreover, the station latitude/longitude data that was collected was often erroneous: occasionally, coordinates were reported as zero or outside of the geographic space where the system resides. We thus separated stations that have been moved (and are thus not amenable to analysis) from those that may have few erroneously reported locations. To do so, we computed the pairwise ground distances between all locations recorded for a single station using the Haversine formula [19]. If any of these distances was larger than 10m, the station was removed from our analysis. If all distances were less than 10 meters (the approximate accuracy of civilian GPS systems [1]), the most recently reported location was assumed the most accurate.

The data loss as a consequence of our preprocessing is presented in Table 1. With 720 daily observations for 996 stations over 150 days, we have over 10^8 data points. Fluctuations in Miami's reported locations led to the removal of nearly 50% of its observation data, while the remaining stations retained over 90% of theirs.

4 Cross-City Analysis

We now analyse the extent to which cities are comparable using the data from their online bike share maps. We first examined the data at an aggregate level showing how daily usage varies across cities.

4.1 System-Level Occupancy Time Series

Station Occupancy. In order to compare stations with different capacities, we refer to the current *occupancy* of a station as the proportion of a station's parking slots that are occupied (i.e., 0 for no available bicycles; 1 for no available parking slots); formally,

$$o_i = \frac{b_i}{b_i + v_i}. \tag{2}$$

When taken in aggregate across an entire city, occupancy, or "fullness of stations", is to be interpreted as the *inverse* of usage: it is low when few bikes remain in the stations, indicating that the bikes are being highly used. Conversely, occupancy is high when many bikes are idle in stations, indicating that the bikes are not being used. A negative slope in the occupancy series corresponds to increasing usage; a positive slope indicates decreasing usage.

To analyse the extent to which cities share similar temporal trends, we divide each 24-hour day into 240 6-minute bins and averaged the observations within each bin on a per-station basis. This occupancy series was then averaged across all stations in a

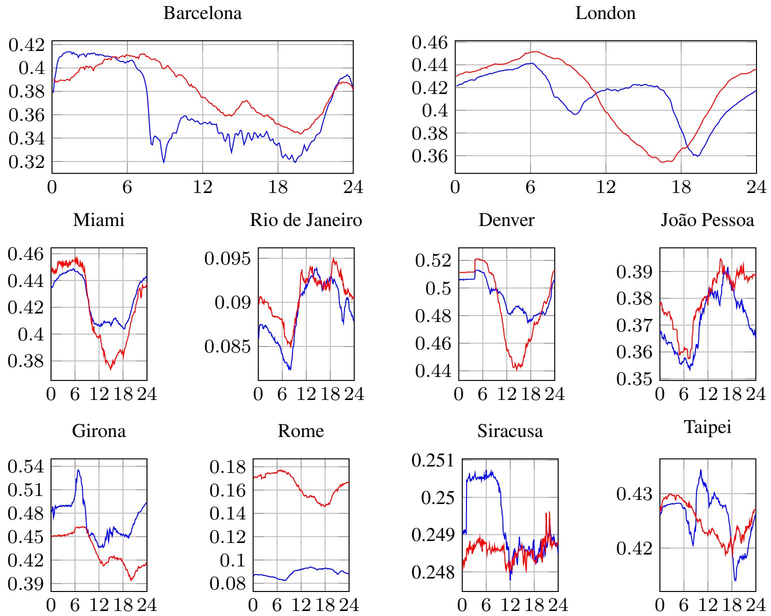


Fig. 1 Aggregate occupancy time series. On the x-axis is the hour of day, and on the y-axis is the average occupancy of the city's stations. The red line depicts the series for weekends, and the blue line depicts weekdays. Note that each plot differs in y-axis scale.

system to create two series, one with weekdays only, and the other with weekends only (Figure 1). From these, a number of common trends emerge:

First, a number of cities' week day data indicates usage for commuting purposes (Barcelona, London, Taipei, Rio de Janeiro), with drops in occupancy that correspond to typical commuting times. However, these dips in occupancy differ, uncovering how the usage of these systems is tied to their geography and local habits related to areas of work: for example, the morning commute dip in occupancy in London is later than that of Rio de Janeiro. This difference may be explained on two fronts: London's system is deployed in the centre of the city, while Rio's is more geographically dispersed, and each city may have its own work-day cultural habits (e.g., Barcelona's data indicates a third usage peak during the afternoon [10]). These patterns also show city-wide preferences for *when* to cycle: London and Taipei have a greater drop in occupancy in the later hours of the working day, while the data from cities in Brazil indicate greater morning usage. Other cities' system data seems dominated by their week-end patterns. These include Girona, where usage is at its highest during week-end evenings, as well as Miami and Denver (USA): these systems' use is likely to be driven by leisure rather than work-related reasons. Finally, comparing cities uncovers variances in bicycle culture; some aggregate patterns are dominated by what seems like redistribution activity (e.g., the jump in occupancy at 1:00AM in Siracusa, Italy),

City	Pearson Correlation
Miami	0.9625471
Rio de Janeiro	0.9094963
Denver	0.8494322
João Pessoa	0.8428481
Barcelona	0.723479
Girona	0.6216594
London	0.4265658
Siracusa	0.3715248
Taipei	0.2241003
Rome	-0.904212

Table 2 Pearson correlation between city's week day and week end aggregated data.

as well as the strongly variant week day/week end patterns in Rome. Further, the two Brazilian cities' week day and week end patterns are more similar than other cities.

To compare the extent to which cities' behaviours changes from week days to week ends, we computed the Pearson correlation between the aggregated week day and week end vectors (Table 2). A wide range of values emerges, ranging from cities with highly correlated behaviours (e.g., Rio de Janeiro, Denver), lower positive correlations (London, Barcelona), and negative correlations (Rome). This indicates that, across cities, variances between week day-week end seem to emerge from the amount, time, or way in which each system is used.

4.2 Cross-City Occupancy Clustering

We next investigated the extent to which individual stations share similar behavioural traits across different cities. We used hierarchical clustering [9], using an agglomerative strategy. In this bottom-up approach, a vector representation of each bicycle station is initialised as a singleton cluster. In each iteration of the algorithm, the distance between every pair of clusters is computed, and the two clusters which have the highest similarity between them are merged into a single cluster containing the stations from both. We discovered naturally-occurring behavioural classes for stations across all systems: certain behavioural classes are system-independent, and that there is, in fact, significant transferable knowledge between stations and between systems.

Recall that occupancy is defined as the fraction of a station's total slots currently occupied by bicycles. As with the preliminary analysis, we created a 240-point occupancy vector for each station by dividing each 24-hour day into 240 6-minute bins and averaging observations within each bin. We normalised the vectors by subtracting its mean from each element:

$$o'_i = o_i - \frac{1}{240}\sum_{i=1}^{240} o_i \tag{3}$$

and used this set of station vectors as initial input.

We next selected a metric to measure the similarity between station vectors. Typical similarity metrics (e.g., the sum of absolute pairwise differences) would not allow for stations that share similar patterns that are temporally displaced to be computed as similar; slight temporal distortions in the series are unduly penalised. To account for this, we used a distance metric based on the dynamic time warping (DTW) algorithm [6]. This is a well-known technique for finding the optimal alignment of two temporal sequences. It works by strategically inserting gaps in either of the two sequences to maximise their alignment. Our DTW implementation uses a 1-hour Sakoe-Chiba band [21], effectively limiting the extent to which such gaps are placed, allowing segments of the series to fall out of synchronisation by up to one hour before incurring a heavy distance penalty. Thus, DTW produces a distance measure from the sequences which have been optimally aligned subject to our constraints.

An important parameter with hierarchical clustering algorithms is setting when to stop clustering, and consequently the number of resultant clusters k. The choice of k is often guided by intuition, based on what level of clustering yields the most valuable analytical insight, or using heuristics [26]. We use a simple heuristic based on incremental merge distances. When initialised with n objects, the algorithm begins with n separate singleton clusters and ends with a single cluster containing n items, with k decreasing by one at each iteration. Thus there are $n-1$ iterations, and at the end of the i^{th} iteration, $k = n - i$. In the i^{th} iteration, the distance d_i between the two clusters which are selected to be merged is the *minimum* pairwise distance between any two cluster centroids in that iteration. We record d_i for $1 \leq i \leq n-1$. This series is differenced to yield a series $\Delta_i = d_i - d_{i-1}$ for $2 \leq i \leq n-1$. The series Δ_i is the incremental merge distance: how much *further* apart the two closest clusters in the i^{th} iteration are than the two closest clusters in the previous iteration.

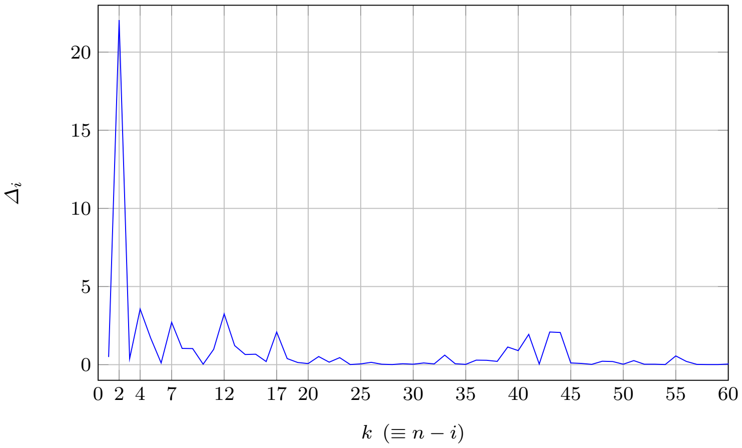


Fig. 2 Incremental merge distances for final 60 iterations of an example clustering study.

Consider Figure 2, which depicts Δ_i as a function of k for the final 60 iterations of an example run of hierarchical clustering. The spikes at k=2, 4, 7, 12 and 17 indicate unusual jumps in merge distances, suggesting that clusters which perhaps

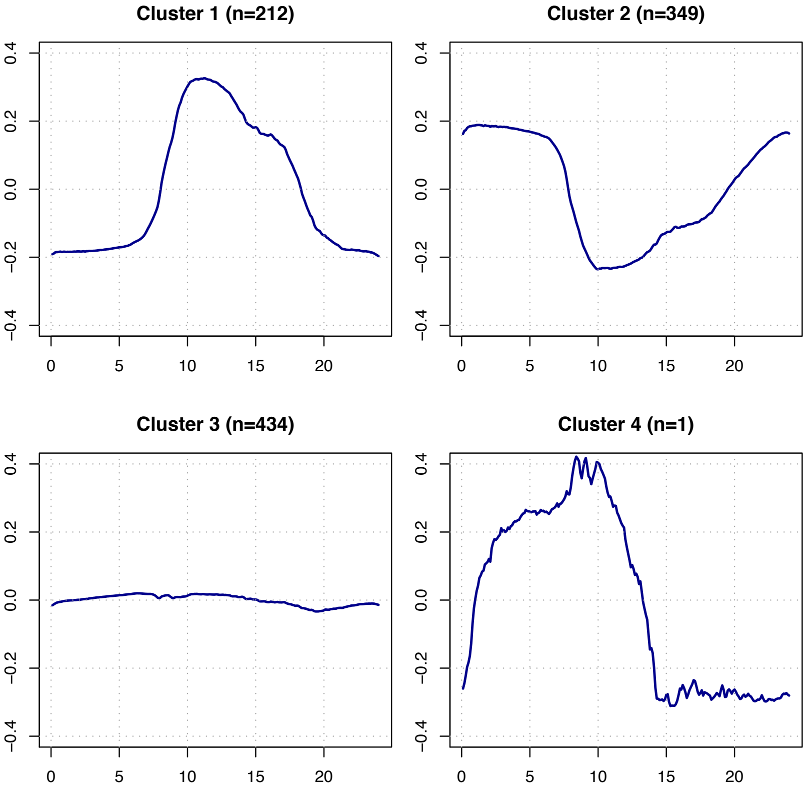


Fig. 3 Common trends across cities after clustering all the stations' occupancy vectors, four trends emerge (morning arrival (1), morning departure (2), flat (3), and the final, anomalous (4), result): these are the cluster centroids.

ought to stay separate are being lost. We were able to decide on an optimal k for all our clustering cases using a combination of the Δ_i heuristic and manual tuning.

Using this approach, we determined that 4 occupancy clusters would produce the most informative result. The centroids of the resulting clusters are plotted in Figure 3: there are 3 major clusters and one minor cluster. Of the 3 major clusters, the first, consisting of 212 stations, exhibits a sharp rise in mean-normalised occupancy starting at approximately 8:00AM. Occupancy peaks at around 11:00AM, then starts to decline until around 3:00PM. These stations can be considered "morning sink, daytime source" stations, as they act as bicycle sinks in the morning and as bicycle reservoirs during the day. The second major cluster consists of 349 stations, and is an almost perfect inverse of the first cluster. Stations in this cluster start the day relatively full, then are rapidly emptied in the morning and slowly refilled over the course of the day. These stations can be considered "morning source, daytime sink" stations, as they act as bicycle reservoirs in the morning, and are a sink for bicycles over the course of the day. The final major cluster consists of 434 stations. Stations in this cluster do not vary significantly in their levels of occupancy over the course of the day. These stations act neither as reservoirs nor sinks, but rather act equally as both. It is important to note that from a flat mean-normalised occupancy series, one cannot conclude that these stations are not very active, although that is one possible explanation. At most, one can conclude that the stations have roughly equal rates of bicycle inflow and outflow. We label the fourth cluster “minor” as it only contains a single station. The station in question lies in the heart of Barcelona and has a highly distinctive occupancy series. Judging by the fact that its occupancy starts climbing after midnight and peaks at 9:00AM, after which its occupancy rapidly drops, it is either close to a number of night-time attractions or it is being used as a depot for the redistribution scheme.

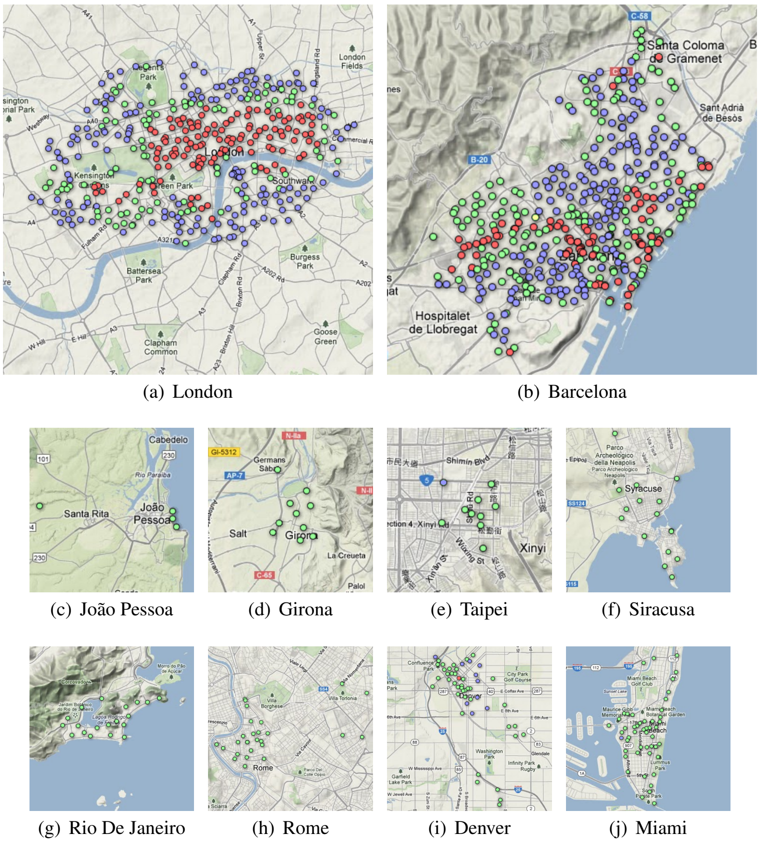
We map the stations, coloured by cluster results, in Figure 4. This uncovers the extent to which cities are homogeneous: several systems are composed entirely of stations belonging to the same cluster, namely the third cluster with the “flat” occupancy series. This implies that all stations in these small systems behave similarly to each other. It is likely that because the supply of stations is so constrained in these systems, demand for bicycles and vacancies is distributed more evenly, leading to the flat occupancy line. A different view emerges for the larger systems, namely London (Figure 4(a)) and Barcelona (Figure 4(b)), where the heterogeneity of station behaviour is clearly visible. London’s clusters look like concentric circles, with cluster one (morning sink) stations at the centre, surrounded by successive layers of cluster three and cluster two (morning source) stations. This reflects a morning surge of bicycles from outside the centre moving inwards, and a slow outwards flow over the course of the day. This suggests that bicycle redistribution vehicles should move outwards in the morning to counteract the morning surge, preventing depletion of the stations in the outermost cluster and the saturation of stations in the centre. Similarly, Barcelona’s morning sink stations run through the city centre and spread out along the coast, and the morning source stations are spread out over the rest of the city. This map corresponds well with Barcelona’s elevation; it is in a hilly region and the placement of the “sink” stations corresponds to lower elevations, while the “source” stations are at higher elevations, a consequence of the natural tendency of users to prefer riding downhill rather than uphill.

4.3 Cross-City Activity Clustering

To further investigate whether stations share trends across systems, we defined a station’s *activity* level (δ) as a single number generated by summing the absolute differences between every consecutive pair of points in its occupancy series o:

$$\delta(o) = \sum_i |o_i - o_{i-1}| \tag{4}$$

This value represents a notion of average “churn”, or “turnover;” we clustered stations on this value. For example, if a station’s δ is 3, then over the course of the day it has



(a) London

(b) Barcelona

(c) João Pessoa

(d) Girona

(e) Taipei

(f) Siracusa

(g) Rio De Janeiro

(h) Rome

(i) Denver

(j) Miami

Fig. 4 Mapping stations by cluster: heterogeneity only appears in large systems. We use a small red circular marker to denote a station belonging to cluster one (morning arrival), a blue marker for cluster two (morning departure), a green marker for cluster three (flat) and a yellow marker for cluster four (anomalous).

seen *approximately and no less than* 3 times as many borrow/return events as its capacity. It is accurate only to the granularity of our sampling: we only account for activity that occurs between one window and the next, losing any activity that has occurred within window.

We again cluster stations, using the same technique as above, on these values. To calculate the centroid δ_c of a cluster of stations S, we simply averaged the deltas of all stations in the cluster:

$$\delta_c = \frac{\sum_{s \in S} \delta(s)}{|S|} \tag{5}$$

This resulted in 6 clusters (Table 3). The first and largest cluster contains the majority of stations, 824. The value of δ_c for the centroid of this cluster is 0.224,

Cluster	n	$delta_c$
1	824	0.224
2	134	0.512
3	31	0.884
4	5	1.470
5	1	3.323
6	1	3.047

Table 3 Activity clustering results. The $delta_c$ is the average level of activity of stations in each cluster; clusters with values greater than 1 have, on average, more borrow/return events than their size.

or approximately 20%. Stations in this cluster see approximately 20% of their total capacity in bike borrowing and returning events over the course of a typical day. So a bike station in this cluster with a capacity of 10 experiences approximately 2 daily borrow/return events on average. A station with a capacity of 40 experiences approximately 8 borrow/return events, etc. The reasons for membership to this large cluster may be unique to each station: there may be a demand for the station to be a source, but it cannot act as one because it runs out of bikes too quickly; there may be a demand for it to be a sink, but it cannot be because it runs out of vacancies easily; or it may simply be too far away from anything of interest to be actually useful. Clusters then progressively get smaller (in membership) and more active. The second cluster contains 31 stations and has a centroid δ_c of 0.884, or $\sim$90%. By our measure, this is the most "active" of the three major clusters. A station in this cluster with a capacity of 10 sees around 9 daily borrow/return events. These stations are well-utilised with respect to their capacity. The values of δ_c for the smallest clusters are 1.470, 3.323, and 3.047, or approximately 150%, 330% and 300%. A station in these clusters with a capacity of 10 sees over 10 borrow/return events daily. Stations with extreme levels of activity occur where demand is high throughout the day, and supply and demand are consistently well matched for *both* bikes as well as vacant slots. These stations may not necessarily cope well with sudden changes in activity pattern, suggesting that auxiliary stations should be built nearby to balance the load.

5 Predicting Station Occupancy

The analysis above demonstrates how temporal patterns differ between aggregated cities' systems, as well as similarities that emerge from clustering stations based solely on their behaviour. We now investigate the extent to which stations' occupancy across different cities can be accurately learned and predicted. Forecasting the number of bicycles at any given time would seem dependent on a number of factors: local geography and urban surroundings (e.g., dwellings or offices), as well as fluctuations in demand (e.g., from events or weather). In this section, we do not attempt to use any external datasets to forecast station occupancy; we rely solely on the historical patterns observed from the online maps.

5.1 Methodology and Metrics

Given a set of timestamps T, and a station $s \in S$, we have a univariate time series of the number of bicycles the station currently holds. For each station, we considered 2 separate series: the series of observations sampled every 2-minutes, and the same series of observations as averaged, 6-minute samples. Unlike in the clustering analysis, the series were not averaged across all days. Thus, for example, a 2 day series at 2-minute samples contains 1440 observations. The same 2 days as a 6-minute averaged series contains 480 observations.

For each test timestamp t, we computed forecasts at four horizons: 6, 12, 24 and 48 minutes ahead; given the history of a station's observations as 2-minute samples, intervals of 6, 12, 24 and 48 minutes correspond to 3, 6, 12 and 24 samples respectively. To test the predictive models (described below), we trained them with data from week 10 and sampled for test week day data from weeks 11-19 (9 weeks) of our dataset: we chose 120 equispaced points at the edges of these intervals to be the points in time from which to run forecasts. That is, we simulated the situation where each of these points was the "present" state of the station, and queried the models for values of (b_i, v_i) at each forecast horizon. We then compared the predictions against the actual values to evaluate them.

The stations across the 10 cities we are considering have a very diverse range of capacities. To account for this, we used error metrics based on occupancy. For a series of n predicted observations of the form $(\hat{b}_i, \hat{v}_i)$, and the corresponding series of actual data (b_i, v_i), where $1 \leq i \leq n$, we calculate the Mean Absolute Error (MAE) in predicted occupancy as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i}^{n} \left| \frac{\hat{b}_i}{\hat{b}_i + \hat{v}_i} - \frac{b_i}{b_i + v_i} \right| \tag{6}$$

Similarly, we compute the Root-Mean-Square Error (RMSE):

$$\text{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^{n} \left(\frac{\hat{b}_i}{\hat{b}_i + \hat{v}_i} - \frac{b_i}{b_i + v_i} \right)^2} \tag{7}$$

5.2 Predictive Models

We formulated the problem of predicting a station's occupancy at a given time as a univariate time series forecasting problem. In this section, we describe the set of predictors that we tested to examine the predictive similarity shared between different city's bike share systems. Our main objective is to compare the performance of predictive models across cities, and we therefore rely on methods that have historically been well suited to this kind of problem.

Random and Static Models. Our baseline was a random model. For all horizons, this simply predicts a random value drawn from the discrete uniform distribution ranging from 0 to the station capacity. The second predictive model was a static model. This assumes that the time series is horizontal, and predicts that the currently observed number of bikes will persist indefinitely into the future. That is, if there are currently b bikes at the station, the static model will predict b bikes at all points in the future. These serve as the benchmark against which other models are compared.

Multilayer perceptron with backpropagation. A multilayer perceptron, or feed-forward artificial neural network with multiple layers of interconnected perceptrons, has previously been shown to be effective for univariate time-series forecasting [3,8]. We trained 8 separate networks with a single output dimension: one for each forecast horizon, for the 2-minute sampled series and the 6-minute averaged series. We train the multilayer perceptrons to predict the *occupancy* (in the closed interval between 0 and 1) based on previous values of occupancy, as neural network regression is more effective when the range of the input and output dimensions correspond well to the sensitive range of the activation functions [18].

Our neural networks had 10 input dimensions (neurons) and a single output neuron. We experimented with the following parameters: activation function of input layer (linear or sigmoid), number of hidden layers (0 or 1), number of neurons in the hidden layer (0, 3, or 10), activation function of hidden layer (linear or sigmoid), and activation function of output layer (linear, sigmoid, or softmax). The combination of parameters which consistently yielded lowest training error for a subset of our testbed was as follows: (a) an input layer of 10 neurons, each with a linear (or identity) activation function: $\phi_I(x) = x$, (b) a single hidden layer of 3 neurons, each with a sigmoid (logistic) activation function: $\phi_S(x) = \frac{1}{1+e^{-x}}$, and (c) an output layer of 1 neuron with a linear activation function.

Decision Tree Ensemble. Finally, we also tested an ensemble of decision trees with random feature selection and bootstrap aggregation, (i.e., a random forest). As with the multilayer perceptron, we trained 8 separate models per station, corresponding to 4 horizons across 2 time series. We tested the performance of ensembles containing 1, 5, 10, 25, 50, and 100 decision trees. Increasing the number of trees beyond 10 resulted in an extremely small improvement in training error, at the cost of a heavy penalty in the time taken to train the forest, so we set the number of trees in each ensemble to 10. Each tree was built using a modified version of Quinlan's C4.5 decision tree learning algorithm [16]. The predicted regression output of an input vector is computed as the mean predicted regression outputs of all the trees in the ensemble.

	Mean Absolute Error			
City	**Random**	**Last Value**	**Perceptron**	**Random Forest**
Barcelona	0.2003	0.0271	0.0773	0.0521
Denver	0.1772	0.0108	0.0925	0.0312
Girona	0.1683	0.0083	0.0725	0.0423
João Pessoa	0.1949	0.0019	0.0439	0.0549
London	0.1951	0.0145	0.0727	0.0397
Miami	0.1735	0.0098	0.0737	0.0263
Rio de Janeiro	0.2637	0.0003	0.0312	0.0175
Rome	0.2387	0.0021	0.0337	0.0215
Siracusa	0.2123	0.0003	0.0396	0.0314
Taipei	0.1678	0.0035	0.0843	0.0263

Table 4 Mean absolute error results when predicting stations' occupancy 6 minutes in the future, using the 2-minute time series.

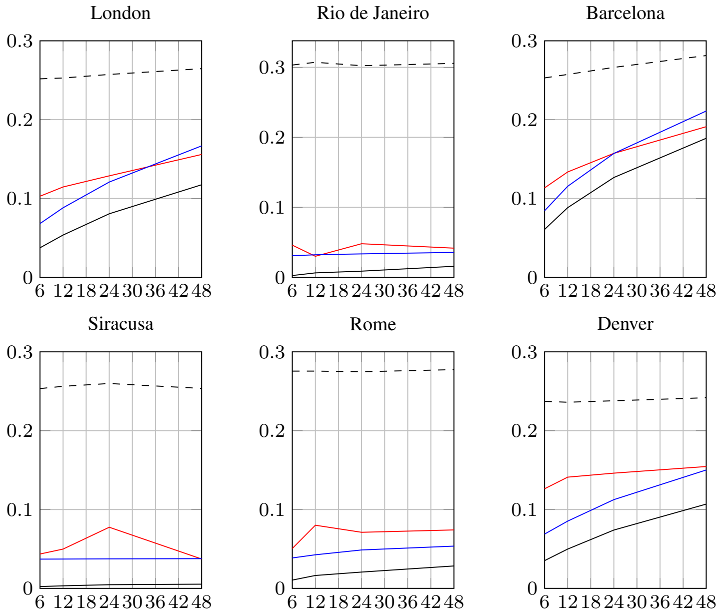


Fig. 5 Predictor performance comparison for 2-minute series. Averaged root-mean-square error in predicted occupancy on the y-axis. Forecast horizon, in minutes, on the x-axis. Random model in dashed black, static model in black, multilayer perceptron in red and decision tree ensemble in blue.

5.3 Prediction Results

The prediction error of all models, given a 2-minute sampled time series and a 6-minute horizon, is shown in Table 4. Across cities, the random benchmark has an error between 0.16 and 0.26. While the decision tree ensemble consistently outperforms the multilayer perceptron, the static model achieves the best average performance. This result can be explained by comparing to the clustering results, where we observed that majority of stations' temporal patterns remain flat. Moreover, even the bike stations with moderate levels of activity usually experience the bulk of their activity in certain concentrated times of day. Finally, the series themselves are only subject to incremental changes in value. That is, except for occasional large changes in occupancy due to redistribution vehicles, a single borrow/return event has only a very minor effect on the value of occupancy; even if the static predictor gets the number of bikes wrong, it is unlikely to be off by a large margin.

Effect of Prediction Horizon. We analysed the extent to which prediction accuracy degrades as we query each model for forecasts of time further in the future. In the smaller systems, with the exception of Girona, the prediction error generally remains stable as the forecast horizon is increased. That is, the models are approximately as good at predicting occupancy levels 6 minutes in the future as they are at predicting occupancy up to 48 minutes ahead.

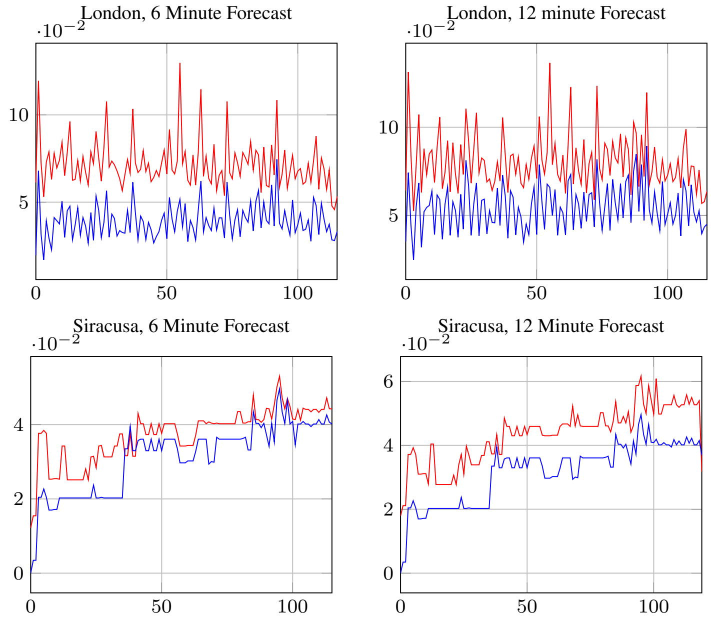


Fig. 6 Predictor performance over time for London's 2-minute series. Averaged absolute error in predicted occupancy on the y-axis. 120 testing points, equispaced within a 9-week testing interval, on the x-axis. Multilayer perceptron in red and decision tree ensemble in blue.

This is not true of the larger systems, where predictor performance deteriorates as the prediction window increases. This suggests that the borrow/return patterns for stations in the smaller system are more consistent, and that the larger the system gets, the more stochastic this pattern appears, at least from the perspective of a univariate time series.

For the larger systems, prediction errors for the 2-minute series are almost identical to those of the 6-minute series. However, for the smaller systems, a variety of differences is observed. In general, the performance of the static model is unaffected, and the performance of the decision tree ensemble and multilayer perceptrons are worse in the 6-minute series.

One explanation for this result is that patterns in a station's occupancy are straightforward in the 2-minute series, but have a periodicity that is lost when averaging into 6-minute bins. This explanation supports the idea that larger systems are more stochastic than smaller ones. The performance of the predictors was unaffected by averaging for the larger systems, but negatively impacted by averaging for the smaller systems, suggesting that the models had been exploiting regular patterns in the smaller systems which were then being thrown off by averaging.

Temporal Effect of Training Data. As above, we chose one week of training data (week 10 of the dataset) to train our multilayer perceptron and decision tree ensemble models. We evaluated the models on 10 weeks of test data (weeks 11-19 of the dataset). Consequently, as we progress through the test set, the testing point gets temporally further from the training data: we assume that this may need to be offset by temporally re-training models with the latest data, and explored the effect that the proximity to training data had on our predictions. Our initial hypothesis was that, as the training data becomes further outdated, it becomes less relevant and results in poorer predictions. Our results show that this is true for the small systems; however, for larger systems, such as that of London, there is no observable deterioration of performance, as shown in Figure 6. Each plot serially presents the error in our models' predictions for each of the 120 testing points. Each point is approximately 516 minutes later than the previous point in the series. The rightmost point of the plot is temporally furthest from the training data, and the leftmost point is temporally nearest.

5.4 Discussion & Applications

The primary purpose of the online maps that support bike sharing systems is to provide web users with a snapshot of the instantaneous state of the system and, in doing so, to provide information that supports cyclists' momentary decisions about whether and where to cycle to and where to collect and plan to park a rented bicycle. Monitoring, analysing, and forecasting the temporal trends in this data could support a variety of applications. Broadly speaking, these related to two domains: (a) managing and (b) planning new bicycle sharing systems.

Managing Bike Sharing Systems. There are several open problems suffered by existing systems, including bicycle redistribution [5] and meeting customers' demand for bicycles and parking slots [2]. Clustering methodologies such as ours could inform the planning of efficient redistribution schemes, by categorising areas of cities as "sinks" or "sources", or by identifying optimal times for triggering the dispatch of redistribution vehicles based on the aggregate occupancy behaviour. For instance, two major clusters in Figure 3 exhibit sharp changes in their occupancy levels between 9:00AM and 12:00PM. This suggests that a minimal redistribution scheme, requiring only a single daily redistribution occurring within this time window, would still be very effective. Maps of these clusters can be used to determine rough directions for redistribution vehicle routes, as we have demonstrated in Figure 4.

Forecasting algorithms may support customer information and gaming systems that incentivise cyclists to participate in bicycle redistribution, identify anomalous behaviour (e.g., due to nearby events), and given them predictions of station occupancy based on the time they are from their intended destination. Furthermore, accurate forecasts would enable preemptive action such as redistribution to address scarcity of bicycles/vacant parking bays due to an anomalous event in the city before the scarcity becomes disruptive.

Planning and Design. Our inter-city analysis uncovered that bicycle sharing systems across countries share similar patterns. Throughout our clustering and forecasting analyses, we found evidence to support the new idea that heterogeneity and variability in station behaviour are functions of the size of the system: the larger the system, the greater the spread of station types; and the larger the system, the greater the variability of an individual station's behaviour over time. These results could be used when planning on deploying a new bicycle sharing system in another urban area, in order to forecast its aggregate utilisation.

Our methodology for clustering on the borrow/return activity levels of stations (Table 3) is useful for deciding whether to add or remove stations. Physical groups of low-activity stations suggest the removal of one or more stations within those groups. Similarly, lone high-activity stations suggest the addition of one or more auxiliary stations to reduce the load and improve resistance to sudden changes in user activity.

5.5 Limitations

Future research into this domain may benefit from gathering data that is rid of the two main limitations we faced in this work. First, by being sourced from the web, rather than directly from the bicycle-station sensors, our data is noisy and subject to inaccuracies that had to be addressed (Section 3) by making inferences from the data. This includes metadata such as station location and its state, since stations can be closed for maintenance. In fact, more granular behavioural patterns could be uncovered with data that contains both user and bicycle identifiers. Further, we do not have access to any operational data (e.g., redistribution schemes or station maintenance schedules): including this data would allow for data-mining based analysis of vehicle redistribution schemes in different cities, which will be affected by the intersection of road traffic conditions and bicycle demand. The inclusion of origin-destination data would allow us to validate and augment studies such as the activity-level clustering.

6 Conclusion

This work examines the extent to which cities' bicycle sharing systems across the world share common traits, by sourcing data from monitoring online maps that contain the state of each system. Using 4.5 months of data from 10 cities, we conducted an analysis of the system-wide occupancy series, looking at the average daily behaviour of stations at each system. While aggregate results indicate variations in temporal trends across cities (alongside evidence that bike sharing system usage varies between weekdays and weekends), applying hierarchical clustering to all the stations in the system exposed that small systems are largely homogeneous—with heterogeneous behaviours appearing in large systems like London and Barcelona.

Next, we framed the problem of predicting the future behaviour of a bike station as a univariate time series forecasting problem and used the data to test four kinds of predictive models. We trained our models and evaluated their performance based on the forecast horizon as well as their proximity to the training data. While our more sophisticated models were not able to outperform the simple static model, the multilayer perceptron did nonetheless achieve less than 10% root-mean-square error in most cases.

Future work would benefit from considering additional datasets, in order to examine complementary aspects of urban mobility and bicycle sharing; many of these datasets are also available online. For example, weather data would reveal how adverse or favourable conditions impact system usage, information from other transport system APIs (e.g., subways or public buses) would reveal how transport systems interact and affect one another. Terrain-related data, topographical data and point-of-interest information (e.g., from Wikipedia or Foursquare) could explain unusual or unique station behaviour. Incorporating safety data could improve our understanding of the dynamics of safety in bike sharing systems.

References

1. In *Global Positioning System Standard Positioning Service Performance Standard (4th Edition)*, Department of Defense, United States of America (September 2008).
2. Pedal Power: the Cycle Hire Scheme and Cycle Superhighways. *Greater London Authority* (Nov. 2010). available at `http://www.london.gov.uk/sites/default/files/FINAL%20REPORT.pdf` (as of April 22, 2015).
3. Ahmed, N. K., Atiya, A. F., Gayar, N. E., and El-Shishiny, H. An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews 29*, 5-6 (2010), 594–621.
4. Austwick, M. Z., O'Brien, O., Strano, E., and Viana, M. The structure of spatial networks and communities in bicycle sharing systems. *PloS one 8*, 9 (2013), e74685.
5. Benchimol, M., Benchimol, P., Chappert, B., Taille, A. D. L., Laroche, F., Meunier, F., and Robinet, L. Balancing the Stations of a Self-Service "Bike Hire" System. *RAIRO Operations Research 45* (January 2011), 37–61.
6. Berndt, D., and Clifford, J. Using Dynamic Time Warping to Find Patterns in Time Series. In *KDD Workshop*, vol. 10, Seattle, WA (1994), 359–370.
7. Borgnat, P., Abry, P., Flandrin, P., and Rouquier, J.-B. Studying Lyon's Vélo'V: A Statistical Cyclic Model. In *European Conference on Complex Systems* (Warwick University, Coventry, UK, September 2009).
8. Crone, S. F., Hibon, M., and Nikolopoulos, K. Advances in Forecasting with Neural Networks? Empirical Evidence from the NN3 Competition on Time Series Prediction. *International Journal of Forecasting 27*, 3 (2011), 635–660.
9. Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification and Scene Analysis: Part 1, Pattern Classification*. Wiley, 2000.
10. Froehlich, J., Neumann, J., and Oliver, N. Sensing and Predicting the Pulse of the City through Shared Bicycling. In *Proceedings of the 21st International Joint Conference on Artifical Intelligence*, Morgan Kaufmann Publishers Inc. (2009), 1420–1426.
11. Girardin, F., Calabrese, F., Fiore, F. D., Ratti, C., and Blat, J. Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing* (2008).
12. Guenther, M. C., and Bradley, J. T. Journey Data Based Arrival Forecasting for Bicycle Hire Schemes – Preprint.
13. Lathia, N., Ahmed, S., and Capra, L. Measuring the Impact of Opening the London Shared Bicycle Scheme to Casual Users. *Elsevier Transportation Research – Part C (Emerging Technologies) 22* (2012), 88.
14. O'Brien, O., Cheshire, J., and Batty, M. Mining Bicycle Sharing Data for Generating Insights into Sustainable Transport Systems. *Journal of Transport Geography* (July 2014).
15. Parkes, S., Marsden, G., Shaheen, S., and Cohen, A. Understanding the Diffusion of Public Bikesharing Systems: Evidence from Europe and North America. *Journal of Transport Geography*.
16. Quinlan, J. R. *C4. 5: Programs for Machine Learning*, vol. 1. Morgan Kaufmann, 1993.
17. Randriamanamihaga, A., Côme, E., Oukhellou, L., and Govaert, G. Clustering the Vélib' Origin-Destinations Flows by Means of Poisson Mixture Models – Draft.
18. Ripley, B. D. *Pattern Recognition and Neural Networks*. Cambridge University Press, 2008.
19. Robusto, C. The Cosine-Haversine Formula. *The American Mathematical Monthly 64*, 1 (1957), 38–40.
20. Saez-Trumper, D., Quercia, D., and Crowcroft, J. Ads and the City: Considering Geographic Distance Goes a Long Way. In *ACM RecSys* (Dublin, Ireland, 2012).

21. Sakoe, H., and Chiba, S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on 26*, 1 (1978), 43–49.
22. Shaw, B., Shea, J., Sinha, S., and Hogue, A. Learning to Rank for Spatiotemporal Search. In *ACM WSDM* (Rome, Italy, 2013).
23. Shu, J., Chou, M. C., Liu, Q., Teo, C.-P., and Wang, I.-L. Models for Effective Deployment and Redistribution of Bicycles within Public Bicycle-Sharing Systems. *Submitted to Operations Research* (2011).
24. Vasconcelos, M., Ricci, S., Almeida, J., Benevenuto, F., and Almeida, V. Tips, Dones, and ToDos: Uncovering User Profiles in FourSquare. In *ACM WSDM* (Seattle, Washington, 2012).
25. Venetis, P., Gonzalez, H., Jensen, C., and Halevy, A. Hyper-Local, Directions-Based Ranking of Places. In *VLDB Endowment* (Seattle, USA, 2011).
26. Yan, M., and Ye, K. Determining the Number of Clusters Using the Weighted Gap Statistic. *Biometrics 63*, 4 (2007), 1031–1037.