Transportation Research Part C 22 (2012) 88–102

Contents lists available at SciVerse ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Measuring the impact of opening the London shared bicycle scheme to casual users

Neal Lathia *, Saniul Ahmed, Licia Capra

Dept. of Computer Science, University College London, Gower Street, London WC1E 6BT, United Kingdom

ARTICLE INFO

Article history:
Received 4 July 2011
Received in revised form 12 December 2011
Accepted 14 December 2011

Keywords:
Shared bicycles
Data mining
Mobility
Policy

ABSTRACT

The increasing availability of sensor data in urban areas now offers the opportunity to perform continuous evaluations of transport systems and measure the effects of policy changes, in an empirical, large-scale, and non-invasive way. In this paper, we study one such example: the effect of changing the user-access policy in the London Barclays Cycle Hire scheme. When the scheme was launched in July 2010, users were required to apply for a key to access to the system. By December 2010, this policy was overridden in order to allow for "casual" usage, so that anyone in possession of a debit or credit card could gain access. While the transport authority measured the policy shift's success by the increased number of trips, we set out to investigate how the change affected the system's usage throughout the city. We present an extensive analysis of station data collected from the scheme's web site both pre- and post-policy change, showing how differences in both global and local behaviour can be measured, and how the policy change correlates with a variety of effects observed around the city. We find that, as expected, quicker access to the system correlates with greater week end usage; it also reinforces the week-day commuting trend. In both the pre- and post-change periods, the geographic distribution of activity at individual stations forms concentric circles around central London. However, upon policy change, a number of stations undergo a complete usage change, now exhibiting an opposite trend with respect to that which they had prior to the policy change.

1. Introduction

"Cities embody [the] political decisions made by their designers" (Zuckerman, 2011).

As the world population grows and an ever-increasing proportion of people live in cities, designing, maintaining, and promoting sustainable urban mobility modes is becoming of paramount importance. Shared bicycle schemes (Shaheen et al., 2010) are one such example: their proliferation throughout the world's metropolises clearly reflects the belief that providing easy access to healthy (and quick) modes of transport will lead cities away from the congestion and pollution problems they currently face. Shared bicycle systems operate in urban areas by providing public access to bicycles from a fixed number of stations that are distributed around the city. Travellers may pick up bicycles from any station they choose and return them to any of the station's free parking slots. No limitation is placed on the usage (in terms of origin/destination); instead, the bicycle usage is often limited by *time* (e.g., free for first *x* minutes; penalty fare imposed if the bicycle is not returned within *y* hours).

A key facet of building successful shared bicycle system, and, more broadly, any urban public transport system, is understanding how designed system characteristics, implemented as policies, affect usage. To see how design may clash with

* Corresponding author. Tel.: +44 20 7679 7214.
E-mail address: n.lathia@cs.ucl.ac.uk (N. Lathia).

0968-090X/$ - see front matter © 2011 Published by Elsevier Ltd.
doi:10.1016/j.trc.2011.12.004

policy (and, in doing so, affect usage), consider the case of Melbourne's (Australia) shared bicycles: the 600-bike system generates fewer than 70 trips per day.[1] This may be due to the laws legislating that cyclists wear helmets—while the system itself does not provide them; pedestrians would have to carry them in order to use the system. Urban planners not only need to understand the policies and domains where they deploy their systems, but, increasingly, also be able to quickly measure and respond to mobility demands as they are shaped by evolving policies.

The proliferation of ubiquitous and interconnected sensors and mobile devices in urban areas now means that transport operators have access to real-time data about the networks that they manage (Calabrese and Ratti, 2006; Ratti et al., 2006): cities of the future may thus be able to continuously monitor the flow of traffic throughout their areas, have a better understanding of individual travellers' behaviours (Bryan and Blythe, 2007), offer real-time travel information (Ferris et al., 2010) and personalised location-based services (e.g., mobile advertising as explored by Quercia et al. (2011)). Furthermore, access to fine-grained data about the status of the transport network will enable empirical measurement and analysis of the effects of both designed (particularly, policy-related) and unplanned (e.g., disruptions) events on the usage of the system. In this work, we focus on the former, and provide an extensive analysis of shared bicycle data from London, England, that covers a period where the access policy was changed. We first preamble our work by giving an overview of the system that we investigate (Section 1.1); we then outline the rest of this paper by enumerating our key insights and results (Section 1.2).

1.1. London's shared bicycle scheme

The "Barclay's Cycle Hire" was launched by London's public transport authority (Transport for London) on July 30th, 2010 with 5000 bikes and 315 stations distributed across a 44 km^2 area of central London.[2] Membership to the system was initially available to anyone who had applied for, received, and registered their membership key: over 12,000 people signed-up to join the system prior to its launch. Payment for the system is partitioned into two categories: a renewable *membership* fee (providing access to the system) and per-trip *usage* fee (charges for each trip taken by bicycle). Membership is priced by duration: £1 for 24 h, £5 for 7-days, £45 annual, along with a one-off cost of £3 to receive the membership key, that can be used to unlock bicycles. The first 30-min of usage is free, with prices then increasing for longer usage (from £1 for up to an hour to £50 for 24-h). Usage of over 24 h results in a late-return charge (£150), while non-return can incur the maximum penalty of £300. Note that the membership key's role in the system is only to facilitate access; cyclists will still be subject to the membership and usage fees.

By February 2011,[3] members had made over 2.5 million journeys using the system. The policy of required registration, however, lasted only the initial months of the scheme's operation (July–December). The scheme was opened[4] to the "casual users" on December 3rd 2010. These users can hire a bicycle by using a credit or debit card at the docking station terminal, as opposed to using an access key (note that the change does not affect usage pricing).

Transport for London claimed that the shift in policy toward casual use increased both the number of users and journeys taken each day. However, finer grained insight into the changes that the system underwent as these decisions were implemented is lacking: a simple increase in the number of trips does not uncover how the usage of the system changed after the policy shift. Since the system launched, an online map detailing bicycle and parking slot availability at each station throughout the day has been available. This means that the "pulse" of London, or spatio-temporal trends as measured by bike usage (Froehlich et al., 2009), is now open to analysis. The occupancy rates of all of the city's stations can be measured, and, when put alongside the shift in user-access policy, insights into changes in the spatio-temporal signature of usage of bicycles across the city can be uncovered.

1.2. Objectives and key results

In this paper, we examine the *measurable changes* in shared-bicycle station occupancy data across the period where the access policy was switched, and discuss how these changes impact all parties involved in using the system. Moreover, while we know that the shift in access policy led to an increase in the aggregate number of trips, we hypothesise that this increase masks an underlying shift in both the spatial and temporal usage patterns of the system (for example, of recreational usage of the system vs. utilitarian commutes). We set out to validate this hypothesis by means of sensor data mining and analysis.

We begin with a description of the data collected and the cleansing process it underwent (Section 3), as well as a discussion of the limitations of urban-scale data-based enquiry into policy change (Section 4). Using well established data mining techniques, we then offer the first city-wide geographic and temporal analysis of the cycle hire usage in London, England, across a major access policy change (Section 5). The analysis allows us to observe both global and local changes. In particular, we find that:

- Simpler access to the system correlates with greater week end (i.e., recreational) usage; it also reinforces the week-day commuting trend, meaning that the shift actually encourages usage of the system for non-casual trips too.

[1] http://www.theage.com.au/victoria/helmet-law-makes-nonsense-of-bike-hire-scheme-20100722-10my2.html.
[2] http://www.tfl.gov.uk/corporate/media/newscentre/archive/16418.aspx.
[3] http://www.tfl.gov.uk/corporate/media/newscentre/archive/18060.aspx.
[4] http://www.tfl.gov.uk/corporate/media/newscentre/archive/17591.aspx.

- In both the pre- and post-change periods, the geographic distribution of activity at individual stations forms concentric circles around central London, with a further divide between north and south. However, upon policy change, some stations undergo a complete usage change, now exhibiting an opposite trend with respect to that which they had prior to the policy change.

These observations impact three different categories of users, as we shall discuss in Section 6: *bike users* (interested in finding available bikes/parking spots), *transport operators* (dealing with the problem of balancing load across stations), and *urban planners* (in deciding how to design social spaces). We then conclude in Section 7 with the future opportunities that the marriage of urban sensing and data mining presents: it enables the collection of rich data about urban mobility, and can offer detailed insights into the effect of policy change on transport systems.

2. Related work

In this section, we review the literature about shared bicycles and urban sensing, in order to place our work within the broader context of transportation, data mining, and sensor research. We classify related work into three main categories: research into bicycle usage and shared-bicycle systems (Section 2.1), transport policy evaluation (Section 2.2), and novel measuring techniques and algorithms (Section 2.3) that researchers have been using to gain insight into the dynamics of mobility in urban life. We summarise the key findings from our survey of related work in Section 2.4.

2.1. Cycling and shared bicycle systems

The use of cycling as a personal (non-motorised) mode of transport has been extensively researched within the transport domain. Historically, there has been a distinct focus on two aspects of cycling. First, understanding *why* people opt to travel by bicycle, by observing the environmental and socio-economic factors that correlate with bicycle usage: this is also a critical point that feeds directly into shared-bicycle system design. One of the key factors here, that differentiates cycling from other modalities, is the assumption that riders will be highly affected by the current weather conditions (both cold/rain and heat); a growing body of literature examines this relationship (Hanson and Hanson, 1977). For example, Thomas et al. (2009) found that weather affects recreational usage of bicycles more than utilitarian (e.g., commuting) usage; indeed, Nankervis (1999) showed that the influence of weather on students' cycling habits is not as strong as expected. Similarly, a study by Replogle (1992) claims that the adoption of cycling is not fully explained by climate, income levels or the proliferation of motorised vehicles: there are both physical, geographical, and social factors that are involved as well, such as the existence of unique sub-communities (e.g., colleges and universities: Baltes (1996)) and petrol prices (Smith and Kauermann, 2011). However, while these works seem to consistently undermine the effect that weather has in determining the uptake of cycling, other works seem to point in the opposite direction: for example, Miranda-Moreno et al. (2011) found that precipitation, temperature, and humidity (both separately and combined) have both immediate and lagged effects on daily ridership figures in Melbourne, Australia. Similarly, Saneinejad et al. (2012) found that an increase of temperature in Toronto, Canada, relates to an increase of cycling trips in groups of commuters who have the choice of alternating between different modes (albeit a 20% increase or decrease in precipitation resulted in a much smaller impact on travel mode choice). The role of weather in cycling continues to be an active area of research: in fact, the results may be city-dependent (Rose et al., 2011).

A second major factor that has been investigated is the relation between bicycle usage and infrastructure. Pucher and Buehler (2006) discuss the effect of cycling networks, bike parking, and urban density on bicycle usage. They find that denser cities, which are characterised by less urban sprawl, have residents who are more inclined to cycle, since short trips are more "bike-able." Dill and Carr (2003) make similar conclusions; in particular, they find a strong correlation between the number of bike lanes per square mile and bicycle usage.

Research on shared bicycle systems spans multiple fields. However, a number of common themes and goals emerge across the various domains. By uncovering facets of urban mobility, researchers aim to help either (a) *transport operators*, who may benefit from more accurate models of bicycle flows in order to appropriately load-balance the stations throughout the day, (b) *end-users*, who may benefit from both understanding and forecasting how the system will be used when planning their own trips, or (c) *urban planners*, who can leverage flow data when designing social spaces. In the following, we review and categorise related work according to the methodologies that have been adopted when analysing each city.

The first set of researchers have adopted a data-centric approach to shared bicycle systems, by applying a host of data mining techniques to uncover spatio-temporal trends in a city's data. Froehlich et al. (2009, 2008) were the first to do so; they performed an in-depth analysis of 13 weeks of Barcelona's *Bicing* system (Spain), clearly demonstrating the relationships between time of day, geography (particularly, clusters of stations within geographic areas of the city) and usage. Kaltenbrunner et al. (2010) performed a similar study of Bicing: in both studies, the authors focus on temporal properties of the bicycle station data in order to train and test classifiers that predict the state (availability of bicycles) of each station. Borgnat et al. (2009a,b, 2010) also infer statistical models from bicycle hire data, albeit from Lyon (France). Unlike the above, the focus of their prediction turns toward global features of the city's system—number of bikes hired per hour—and takes into account several factors relating to time (time of day, occurrence of holidays) and season (weather) when formulating a prediction. A similar analysis by Jensen et al. (2010) on the same city shows that shared bike usage in Lyon competes with car usage, in terms of trip speed: data that is sensed from station usage is thus likely to become imperative for urban planning and road development (i.e., designing for bicycle paths). Nair et al. (2012) have analysed data from Paris' (France) Vélib', relating usage to rail station proximity: they uncover the relation between bicycle usage and multi-modal trips (thus providing key insight into station placement policy).

A key problem of shared-bicycle schemes, observed both in the data analysis research (above) and user surveys is that stations may often be full (stopping people from returning bicycles they already have) or empty (preventing people from hiring a bicycle when they need one). In fact, these two problems, "availability of bikes" and "returning a bike," were the two categories that, proportionally, received the highest amount of negative votes in a recent survey about shared bicycles conducted by the London Assembly.[5] Solving the problem of *load balancing* a city's stations, which is often performed by vans moving excess bicycles between stations, is therefore of paramount importance. A number of researchers have proposed models that address the computational complexity of determining how to load-balance stations efficiently. For example, Raviv et al. (2010) consider the *static* repositioning problem: computing what the optimal route and number of bicycles each load-balancing van should have, given a number of constraints and a static snapshot of the state of the system. This body of research (including Benchimol et al., 2011; Vogel and Mattfeld, 2010; Nair et al., 2012) tends to include a pre-defined notion of *satisfaction* or *reliability* of the system from the users' perspective (defined as, for example, the probability of successfully hiring a bicycle), which can be expressed as an objective function that must then be optimised while taking into consideration other operational constraints.

2.2. Transport policy research

The proliferation, sustainability and modification of transport habits is governed by policy. To that end, there has also been a range of works related to measuring the effects of local policy. Issues here range from reconciling carbon-reduction targets with growing travel needs (Bows and Anderson, 2007), investigating the relationship between fares and quality of service on public transit demand (Paulley et al., 2006), and public acceptance of urban road pricing (Eliasson and Jonsson, 2011).

With regards to bicycles, Rietveld and Daniel (2004) inspect the extent that municipal policies hinder or encourage bicycle usage. Other than time- and geographical-related features of each city, they cover issues ranging from traffic safety to bicycle theft risk. By performing a cross-city comparison of policy-dependent factors (e.g., parking cost) and bicycle usage, they observe how the uptake of cycling correlates with increasing parking cost and reducing "hindrances" (locations where cyclists are required to stop). Pucher and Buehler (2006) examine policy – usage differences at a national level (Canada vs. United States): "in spite of their colder climate, Canadians cycle about three times more than Americans [...] factors result from differences between Canada and United States in their transport and land-use policies, and not intrinsic differences in history, culture, or resource availability."

Outside of the domain of bicycles, an important aspect of transport policy that has been researched is the relation between travel and pricing. Setting costs for travel is a key element of urban congestion control and incentivising public transport usage. Its effects have been studied both using spatial analysis Graham and Glaister (2006), and from a data-mining perspective, in order to build forecasting services for customers to manage their travel needs and spending more wisely, both for air fares (Etzioni et al., 2003) and public transport (Lathia and Capra, 2011b).

Lastly, Shaheen et al. (2011) performed a large scale study of bike sharing in Hangzhou, China, which hosts the largest system in the world (with over 60,000 bicycles). The authors surveyed 666 members and 140 non-members in order to understand the city residents' adoption trends. Most notably, requiring cyclists to register before using the system was viewed as detrimental to the system's uptake: respondents complained about "the hassle of the smart card application process (i.e., inconvenient office location, long lines)." This is particularly relevant to the work below since the policy change in London was implemented in order to facilitate adoption and usage of the bike sharing system.

2.3. Urban sensing

As the ubiquitous computing vision becomes closer to reality, thanks to advances in embedded sensor systems and the widespread adoption of smart-phones, new and rich datasets, recording detailed information about how people move through urban areas, are becoming available and amenable to investigation. In this section, we give a brief overview of this area, by describing a select number of examples that highlight how novel datasets are lending to transport research in particular.

There are two sources of implicit data that researchers have been investigating: fixed, embedded sensors (as is the case of shared bicycle systems), and mobile phones. An example of the former includes the automated fare collection systems that many public transport networks around the world have adopted to ease city residents' payment of travel. In doing so, they also hold a rich, large-scale dataset of people's travel habits. This kind of data is amenable to analysis; for example Lathia and Capra (2011a) studies the extent that travel incentives (e.g., peak-time fares) create a measurable change in people's travel habits. The latter, instead, takes advantage of the proliferation of mobile phones throughout the world: network operators

[5] April 2011. http://www.london.gov.uk/sites/default/files/FINAL%20REPORT.pdf.

can form a detailed view of their subscribers' whereabouts and movements. Calabrese and Ratti (2006), Ratti et al. (2006), Soto and Frias-Martinez (2011) are amongst a growing community of researchers who are investigating the extent that call detail records reveal details of urban usage and mobility.

One of the weaknesses of both of the data sources above is that, by being raw sensor readings, they lack qualitative descriptions of the context of people who are moving about urban spaces. There is thus a recent interest in using social media to fill this void. While social media has typically resided on the web (e.g., online social networks), there is a growing awareness about the fact that these datasets may offer contextually-rich data that is otherwise absent from sensor readings (Hayes and Stephenson, 2011). Recent research includes the use of "check-ins'," where users input their location to their mobile device (Bawa-Cavia, 2011), and geo-tagged photos (Rattenbury et al., 2007) to understand the relation between urban space, social events, and mobility. We will further elaborate on the limitations of sensor data-based analysis, with a particular focus on bike sharing data, in Section 4.

2.4. Summary

Bicycle usage (both from a shared-scheme and personal perspective) has motivated researchers to tackle a range of problems; an eminent characteristic of the analyses above is that there is a complex network of causal relations that determine whether people will use this modality at all. Unlike shared-bicycle systems, where data is readily available from station sensors, the body of research on non-shared bicycle usage is constrained in its ability to collect empirical usage data. Researchers must therefore opt to manually measure flows (e.g., counting parked bicycles as done by Nankervis (1999)), to examine surveys or census data like Shaheen et al. (2011), or install sensors (for example, pneumatic tube sensors in Thomas et al. (2009)) on cycle paths; invariably, cyclists who do not use these paths will elude analysis. Although the influence of weather seems to have been a favourite amongst the topics explored, recent research has shown that it does not determine bicycle usage as strongly as one would originally assume: structural, seasonal, social, and policy-based factors all play their part as well. In the following sections, we have kept as many of the above factors as possible constant, when sampling the bicycle data, with one exception, that is, the user-access policy, which is the main subject of investigation in this work.

3. London cycle hire: data collection and cleansing

We collected data from Transport for London's online interactive map[6] of the scheme's stations. Each marker on the map not only contains the location (latitude, longitude) of the station, but also carries the number of available bicycles and parking slots; we have been scraping data from this website every 2 min (except during moments of intermittent access to the map service and during down times of our own servers). This produces 720 observations of each station's state over any 24-h period. The key date in this dataset is December 3rd, 2010, when the scheme switched from requiring user registration to allowing for casual use. We therefore produce two datasets, as follows:

- The first dataset (which we call the *"Pre-Casual"* dataset) spans from October 17th, 2010 to December 3rd 2010. It covers 48 days, including 35 weekdays and 13 weekend days. Although the scheme was launched in July, we disregard all data until mid-October to (a) not bias our analysis with data throughout the weeks that the system was first available and (b) in order to alleviate the effect of varying seasons (i.e., the summer months). The dataset contains 11,440,440 observations collected over 33,325 two-minute time intervals of the scheme's 358 docking stations.
- The second dataset (which we denote as the *"Post-Casual"* dataset) spans from January 2nd, 2011 to February 22nd, 2011. As above, we eliminate the initial month of the new user policy to avoid the bias of both the transition phase and the holidays during December. The total number of valid days is the same as the pre-casual dataset, with matching quantities of working days and weekends (48 days, including 35 working days and 13 weekend days). The dataset contains 12,255,057 observations collected in 34,371 two-minute time intervals. The observations correspond to the scheme's 366 docking stations.

The changes between the two datasets are not only limited to the shift in access policy; in the time between the datasets, the hire scheme also grew by 15 stations. The stations themselves also increased in size, with the average station size increasing from 21.42 to 23.73 bicycles (the largest recorded station size has increased from 54 to 57); the estimated total number of docking spaces in the system has increased from 8200 to 8660. In order to minimise the limited (Thomas et al., 2009) effect of weather, both datasets are within similar seasons (beginning and end of winter).

It is important to note that this data does not contain trips. In other words, using this data alone, we cannot say that a cyclist has taken a bicycle from station 1 to station 2; we simply observe that, at a given time period, a station has b bicycles and s available parking slots. To that end, we cannot use this data to count journeys; instead, the data gives us a snapshot of what is happening in a given location at a given time.

The data that was collected from the map is not a fully reliable representation of the state of each station: issues ranging from inconsistent observations to internet connectivity problems diminished the quality of the data we collected. We

[6] https://web.barclayscyclehire.tfl.gov.uk/maps.

therefore pre-process the data using a multi-step cleansing process (as done by Froehlich et al. (2009)). The cleansing is performed in three steps, and takes into consideration:

1. *Station Size Consistency*: the sum of the available bicycles and free parking slots should remain constant. We estimate the size of each station as the 95th percentile of the sums of all the bikes/free slot pairs that we observe. If an observation reports a station size that is greater than the inferred size for the given station it is considered invalid and removed.
2. *Day Data Threshold*: if a particular station has a high proportion of invalid data points during a single day (i.e., the 720 samples contain fewer than 70% valid data points), we remove the entire day's data for that station. This accounts for abnormal or anomalous station behaviour.
3. *Station Data Threshold*: based on the previous two points, if a station's data contains less than 45% of the possible weekdays' data, the entire station is pruned from the dataset.

The pre-processing resulted in a loss of less than 4% of the data (i.e., the pruned *Pre-Casual* dataset holds 11,060,481 data points, and the *Post-Casual* dataset 11,804,333). In the following sections, we will analyse and compare the differences between these two datasets, although we first preamble our analysis by discussing the limitations that this form of observational data imposes.

4. Limitations of urban-scale data-based enquiry

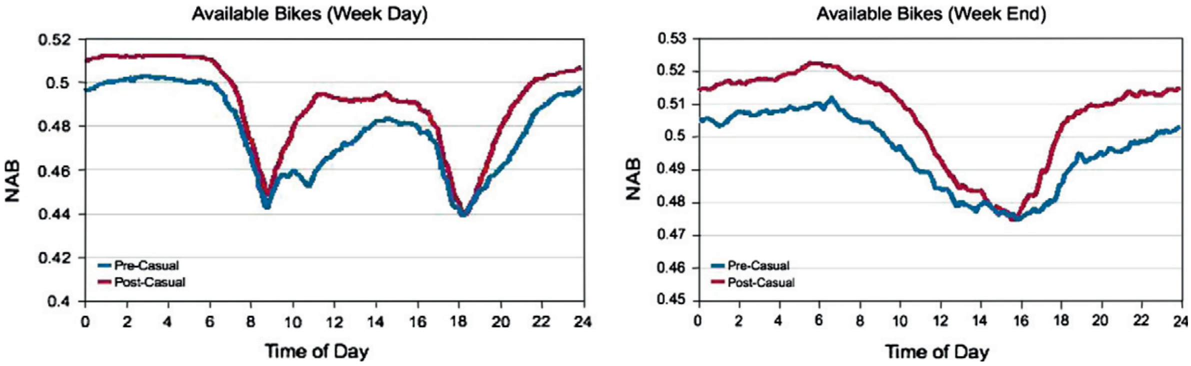
With the data described above, we set out to measure the spatio-temporal changes incurred throughout the city during the time that the user access policy was changed. Since our work focuses on sensor data that was collected from the entire city's system, this form of analysis greatly differs from a controlled laboratory experiment. In a traditional setting, researchers would control for (or randomise) any external factors that may influence the outcome they seek to measure. However, at this scale of deployment, we are inherently unable to control for all external factors; one may be inclined to conclude that it is difficult to speak of anything more than observed correlations rather than attribute causality to a particular intervention (in this case, a policy change).

Indeed, the results and system changes that we present below could be attributed to the growth of the system, increased media coverage and advertisements, changes in land usage (e.g., openings of new buildings, office closures), weather, or even simply aggregate changes in individual habitual behaviour. We have taken a number of steps to limit the effect of these factors. First, we limit the period of observation in order to minimise the variation in weather patterns (even though the literature consistently undermines the effect of weather on cycling – Section 2.1), and to avoid measuring the growth of the system. We remove the holiday period from our dataset, in order to ensure that cyclists will be following their regular mobility patterns; we will also only present aggregated results, in order to minimise day-to-day fluctuations in the system's usage. Finally, to the best of our knowledge, no further incentives (in the form of policies) had been put in place during the time of observation: the most significant change that occurred between our two spans of observation is the user-access policy change.

As outlined above, the approach that we adopt is to compare two datasets that were created by temporally splitting observations into a *pre*- and *post*-intervention period, surrounding the policy change. These so-called "before-and-after" studies, or "quasi-experiments," are not uncommon in the domain of transport research (Institute of Transportation Engineers, 2009); for example, evaluating the efficacy of safety policies on the reduction of car crashes will also fall under this banner. In particular, when interventions occur at the system-level, there is no way that a control group can be created (e.g., an area of the city where the new access policy does not apply); we are thus unable to see what would have happened had no intervention taken place. However, in this case, the transport authority already claims that the change in policy caused higher usage; our goal here is thus not to discover *that* a causal relation between the policy change and increased usage exists, but to use sensor readings to observe *how* this increase in usage was distributed across time and the city, while keeping in mind that other (constrained) factors may also be at play. In effect, what we present here amounts to an exercise in *measurement*, by using sensor data to uncover patterns at the city-scale, rather than evaluation, which would aim to validate that the policy change produced greater system usage. Note that we do not advocate for this methodology as a replacement for the traditional forms of investigative work such as surveys or expert interviews (boyd and Crawford, 2011); instead, we will show how this data, much like mobile phone data (Calabrese and Ratti, 2006; Ratti et al., 2006; Soto and Frias-Martinez, 2011), reveals trends that would otherwise remain invisible. We discuss improvements and extensions to this form of research in Section 6.

5. Data analysis: pre- vs. post-casual usage

We divide the following analysis into three parts: first, we consider *average* system-wide temporal trends in the data (Section 5.1); second, we look at the *spatio-temporal* differences that arise between the two samples, while still looking at the whole system (Section 5.2); finally, we zoom in and report differences between the datasets at the *local* (station) level (Section 5.3).



(a) Week Day Bicycles (b) Week End Bicycles

Fig. 1. Comparing the average Normalised Available Bicycles (NAB) across the whole system in both datasets: the change in user access policy re-enforces both week-day commuting and week-end recreational usage of the system.

Given the set of data points that we have, there are two complementary views of the state of the system that are possible: available bikes and available parking slots. To be able to compare the usage of stations of (potentially) different size we normalise each station's data. Recall that a station i's size S_i is calculated as the 95th percentile of the sums of all observed pairs of bikes $B_{i,t}$ *and* free parking spaces $P_{i,t}$. We take one of the two views, and use a metric denoted the Normalised Available Bicycles NAB (Froehlich et al., 2009). It is calculated by dividing the current number of available bicycles B by the station size:

$$NAB_{i,t} = \frac{B_{i,t}}{S_{i,t}} \quad (1)$$

NAB may be interpreted as the percentage of the station's occupied docking spaces (0 is empty; 1 is full): a temporal trend with a positive slope denotes arrivals, while a negative slope means that people are leaving (hiring bicycles). As above, we note again that station occupancy data cannot be used to count journeys; we opt for this normalised metric in order to examine scale-free temporal and spatial trends over the city's landscape.

5.1. System average temporal analysis

We begin by considering the aggregate usage of the system. Fig. 1 shows the average NAB values across all stations in both datasets (pre-casual in blue, post-casual in red[7]), split into week day and week end. These plots depict an average of averages: first, we compute the average NAB values *per station*, then average them to produce the final visualisation. While we know that both the number of trips and size of the system has grown, the average system-wide data displays some surprising results. First, the change in week day pattern is not as expected: the system now conforms much closer to the pattern reflected by London's public transport infrastructure (Lathia et al., 2010). The morning and evening commuting spikes in usage are clearly visible, and much further away from one another; it would seem that day-time usage (between peaks) has decreased, as there are more available bicycles and parking slots in that time. Intuitively, it would actually make more sense the other way around: we would expect that, allowing *all* to access the system, would increase the recreational usage of bicycles in the week days as well as week ends. Instead, we find that the post-casual access dataset reinforces the two-spiked trend expected of utilitarian (commuting) trips; quick access to the system thus acts as an effective incentive for commuters to use the system as well. In comparison to week day usage, London week-end usage of the system has changed much less instead, with a distinct peak during the mid-afternoon hours (about 4:00 PM).

At an aggregate level, the system appears highly stable: at least 44% of the inferred bicycles are available during the commuting spikes; when commuting ends, the proportion of available bicycles increases to approximately 50%. There are two challenges that arise from viewing the data this way: first, since it is based on normalised values, it does not show performance related to the actual number of bicycles that are in use. In other words, given the data we have, it is difficult to view the extent that the system reaches its full capacity: we do not know the exact number of bicycles held by the system. Second, although the system appears to be stable at the aggregate level, we do not know the extent that this stability is uniformly distributed across the city. To that end, we present the results of a fine-grained spatio-temporal analysis on the data in the next section.

5.2. Spatio-temporal analysis

In this section, we analyse how the individual, per-station week day patterns are geographically distributed across the city, and which stations the change between the two datasets affects the most. To do so, we use clustering algorithms. This

[7] For interpretation of colour in Figs. 1–7, the reader is referred to the web version of this article.

class of unsupervised learning algorithms automatically group the stations based on their usage patterns; we can then put the results onto a map to see how usage relates to the city's geography.

We use a hierarchical clustering algorithm that has previously been used in traffic mining (Froehlich and Krumm, 2008). It works as follows: each station is represented as a time-series vector of normalised available bicycle (NAB) values; note that there is no geographic or positional information encoded into this data. We first smooth the data by applying a 2-sided moving average. Each station is then placed in its own cluster; the similarity between each pair of clusters (initially, stations) can then be measured, and the two most similar ones are merged to form a new cluster. This process is repeated iteratively until a pre-defined number of clusters is reached. We can then map the stations according to the cluster that they belong to.

More formally, this procedure requires three functions. First, a 2-sided moving average to smooth the data: each element p_i is smoothed by averaging it with its neighbours:

$$p_i = \frac{1}{3}(p_{i-1} + p_i + p_{i+1}) \tag{2}$$

Second, a similarity measure to compare clusters. We measure the similarity of a pair of time series p and q with the Euclidean distance:

$$sim(p,q) = \sqrt{\sum_i (p_i - q_i)^2} \tag{3}$$

Note that, in this case, greater similarity between clusters translates to lower scores (or similarity values). We use an agglomerative procedure to build a hierarchy of stations: we started off with a single cluster for each station and then successively merge the two most similar clusters, until the desired number of clusters have been produced. We merge two clusters p (with x stations) and q (with y stations) using a weighted average, where weights are the number of stations belonging to that cluster, to produce a new cluster n:

$$n_i = \frac{(x \times p_i) + (y \times q_i)}{x + y} \tag{4}$$

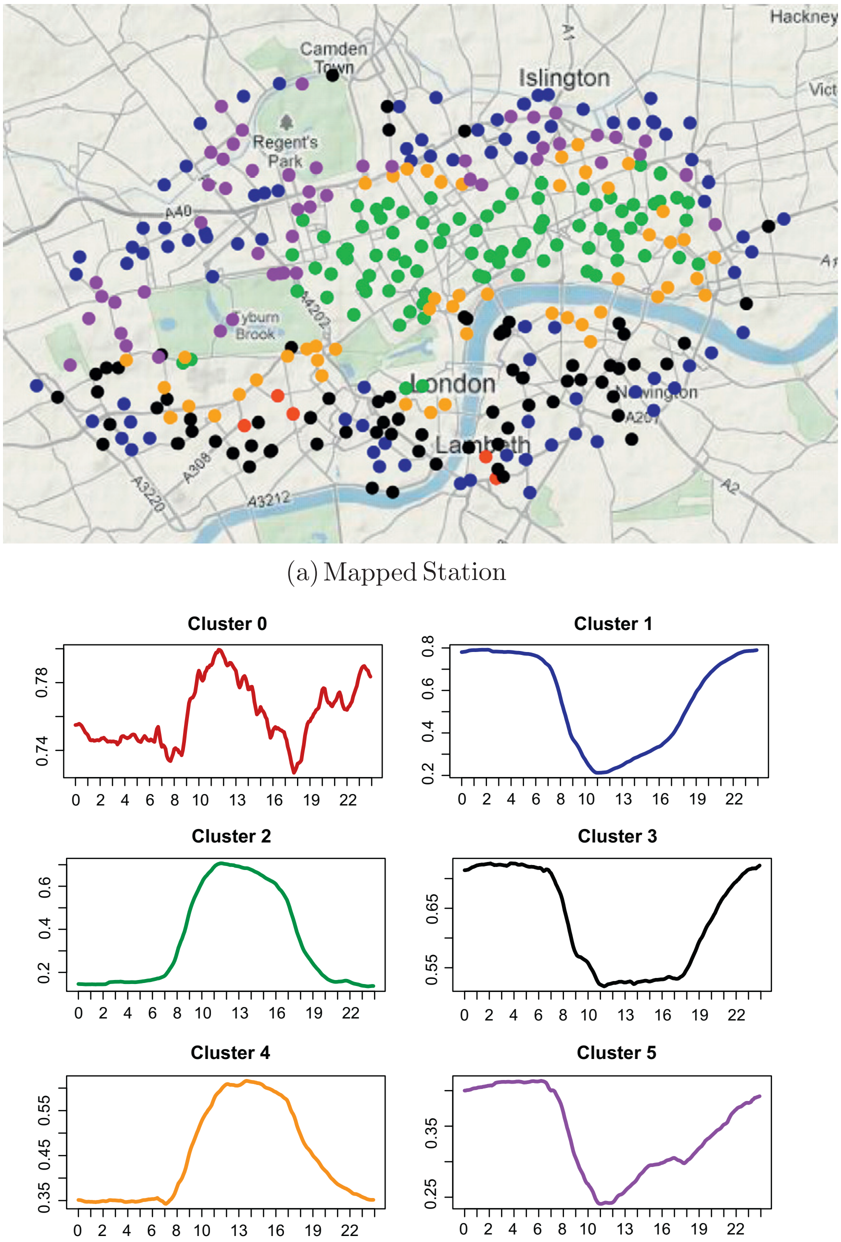
The time series n that is produced by merging clusters is referred to as the *centroid* of that cluster: it shows the average behaviour of all stations belonging to that group, for the period covered by the data. Fig. 2 shows the results of clustering the *Pre-Casual* dataset, showing both the set of centroids that we obtain and how the groups are geographically distributed across the city. Each circle represents a docking station, coloured in accordance to the cluster it belongs to. The six clusters represent the typical behaviour of various stations:

- *Day-Time Origins.* Clusters 1 (blue), 3 (black) and 5 (purple), on the right-hand side of Fig. 2(b), are locations that users take bicycles from in the morning and flow to in the evenings. Cluster 3 (black) differentiates itself by having a delayed resurgence of bicycles, which do not start arriving again until after 6 PM. Cluster 1 (blue) and 5 (purple) differ from each other in terms of scale: the former fluctuates between very full (80%) and very empty (20%), while the latter is, on average, consistently holding fewer bikes (25–40% full).
- *Day-Time Destinations.* Cluster 2 (green) and 4 (orange) are morning destinations, when bicycles arrive, and points of departure in the evenings. They differ in terms of magnitude: cluster 4's stations are, on average, more full than cluster 2's stations during the night time (approximately 35% vs. less than 20% full), and less full during the day (peaking at approximately 60% vs. over 65%).
- *Combined Origins/Destinations.* Cluster 0 (red) highlights stations that display a combination of behaviours: both morning and evening arrivals, and consistently full (over 70% full). There are only five stations in this cluster, all of which are within walking distance of underground stations; moreover, four of them are within a short distance from major national rail stations (London Victoria, Paddington, and Vauxhall stations).

Overall, the clusters of stations seemingly form rings of activity patterns surrounding central London; neighbouring stations are likely to share similar usage behaviours. We also note that (again, without leveraging any geographical knowledge), the clustering algorithm has separated outer-ring stations in north London, which mostly fall in clusters 5 and 1 (purple and blue), from those in the south, which tend to be in clusters 3 and 1 (black and blue).

In the above, we reported on results obtained when setting our algorithm to produce six clusters. This number has been chosen empirically, so to obtain high intra-cluster similarity, and low inter-cluster one. We note that the visual results may, at first glance, seem to indicate that a three cluster solution may be more appropriate (one with a morning "leaving" pattern, another with a morning "arrival" pattern; the last with combined results). We did test this setting while performing our experiments; presenting results for all the cluster results is beyond the scope of this work. However, we note that these results would mask the *scale* and *variability* with which different stations are used. For example, as described above, while cluster 1 (blue) and 5 (purple) have a very similar pattern, the former's bicycle availability varies between 80% and 20%, while the latter is more consistent (25–40% full).

We now repeat the same clustering algorithm on the *Post-Casual* dataset. The resulting centroids are plotted and mapped in Fig. 3, with colours in this set corresponding to the most similar cluster in the pre-casual cluster results. In practice, since hierarchical clustering is an algorithm that discovers similarity between stations by automatically organising the data, com-



(a) Mapped Station

(b) Cluster Centroids

Fig. 2. Pre-casual dataset cluster results. The *x*-axis is time (hour of the day), while the *y*-axis is the NAB value for the cluster centroid. The results clearly show the flow of bicycles is inward during mornings and outward in the evenings.

paring the cluster results from two independent datasets is challenging. We examine the relationship between each set of centroids by producing a one-to-one mapping between the pre- and post-dataset clusters, matching each of the six results in the former with the most similar neighbour in the latter (with similarity measured as per Eq. (3)). This may not produce a unique mapping; for example, two clusters in the former dataset may match a single cluster in the latter dataset. We thus ensure that a unique one-to-one mapping is produced, by resolving conflicts with manual inspection of pairwise cluster similarities. Unfortunately, this meant that, in our case, cluster 0 (red in Fig. 2b) was ultimately matched to another that is not very similar to (red in Fig. 3b). This occurred since the second set of clustering results did not include any group of stations that exhibit highly similar results to the former group for this one cluster.

As before, clusters exhibit one of two main behaviours: (a) day-time origin stations (mostly empty during the day, full at night), for clusters 0, 1, 3 and 5, with differences in terms of magnitude; or (b) day-time destination stations (mostly full

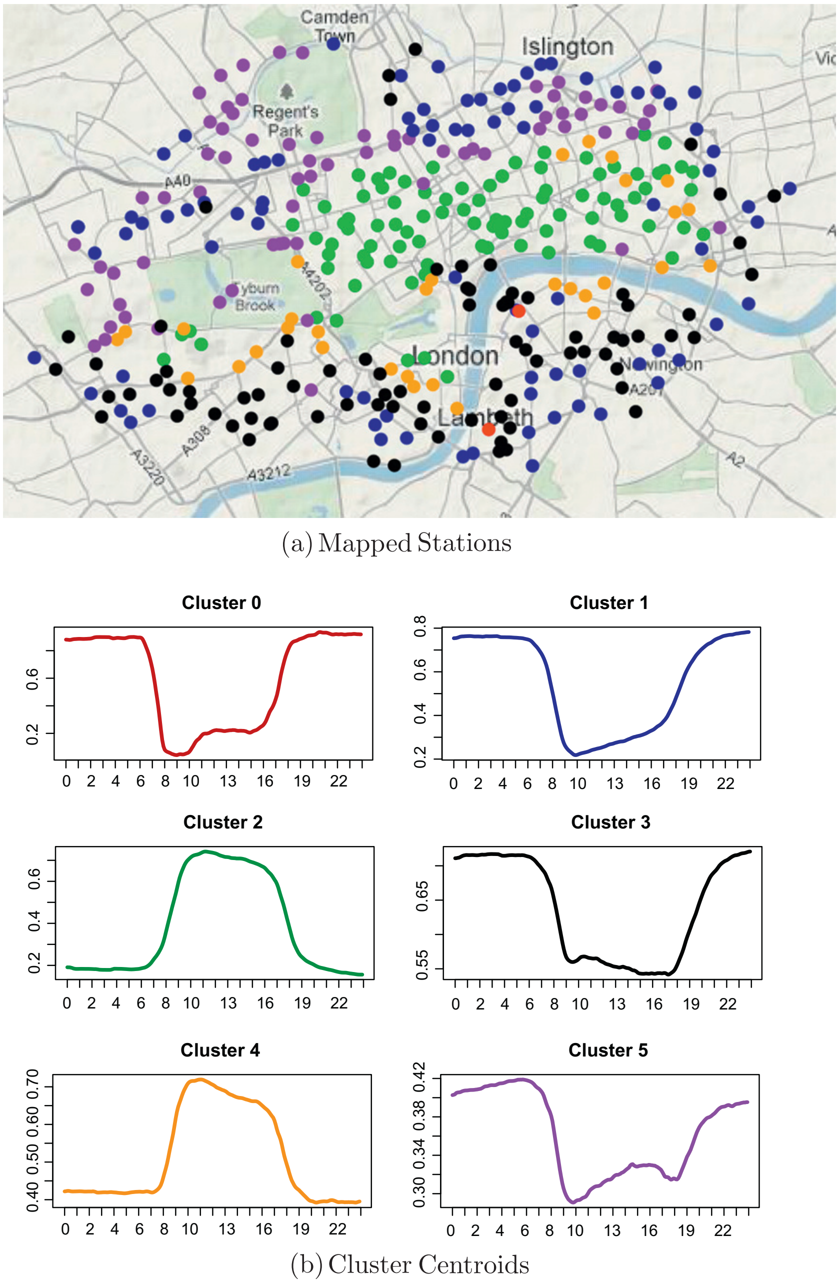
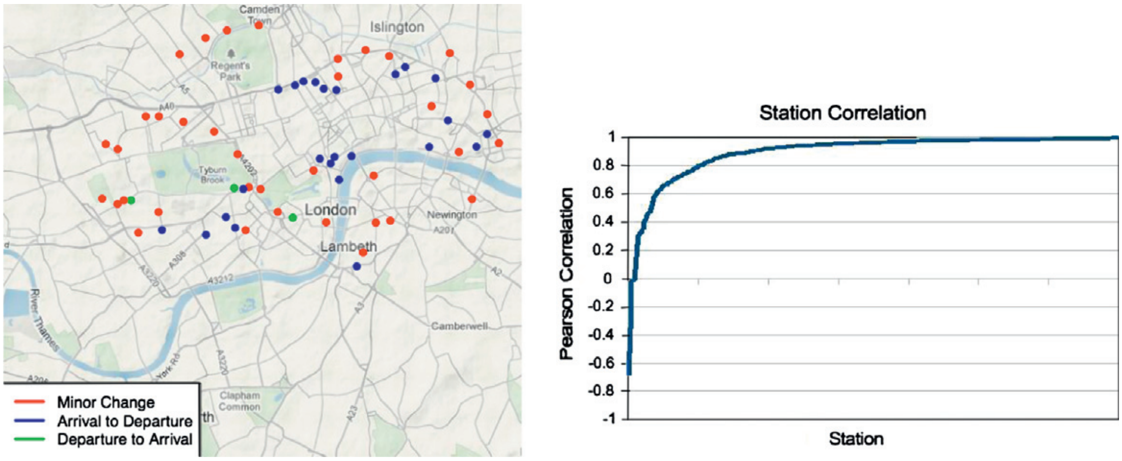


Fig. 3. *Post-Casual* dataset cluster results. The *x*-axis is time (hour of the day), while the *y*-axis is the NAB value for the cluster centroid. Note how the cluster centroids have changed, and that a number of stations have also changed behaviour.

during the day, empty at night), for clusters 2 and 4, again with differences in terms of magnitude. The third behaviour previously observed, of stations with combined day origins/destinations, has now disappeared. As with the pre-casual dataset, clusters form rings of activity patterns surrounding central London, with a further north/south divide. However, closer inspections reveals that a total of 64 stations changed their *cluster membership* between the datasets. More precisely:

1. *Minor Changes*. The station changes cluster (e.g., from 1 to 5, or from 2 to 4), but its overall behaviour remains the same (from day origin to day origin,or from day destination to day destination). 58% of the group-membership changes belong to this category.
2. *Departure Station Becoming an Arrival Station*. The station flips from being a day-time origin to being a day-time destination. Three stations (5%) belong to this category.



(a) Mapped Stations (b) Correlation Distribution

Fig. 4. Changed clusters: (a) a map of stations that have changed clusters: the changes largely appear on the perimeter of the system, as well as near parks, (b) the distribution of correlation values between the pre- and post-casual station NAB values: many stations remain highly correlated to their previous values, while a few become highly negatively correlated.

3. *Arrival Station Becoming a Departure Station*. The station changes from being a day-time destination to being a day-time origin: bicycles are now taken from there in the morning, and returned to there at night. The remaining 37% of the changes belong to this category.

To visualise how these changes are distributed across the city, we map the stations that have changed group membership in Fig. 4. The map clearly shows that changes mostly affected the outer stations of the system and those in close proximity of the city's parks (Hyde Park and Regent's Park). The stations that have changed from being destinations to origins (blue) are, in general, closer to central London although still on the outer perimeter of the system. The three outliers that change in the opposite direction, becoming day time destinations rather than origins (green), are all within a short distance from major London rail and underground stations (Victoria, High Street Kensington, Knightsbridge), with the latter also being adjacent to Hyde Park.

We also empirically quantify the changes that appear in each station by measuring the extent that a station remains self-correlated (i.e., the pair of station profiles, from the pre- and post-datasets, is correlated). Correlation measures the extent that the pair of profiles is linearly dependent. Given two vectors a and b, representing the pre- and post-change NAB values, correlation is computed with *Pearson's r*, as follows:

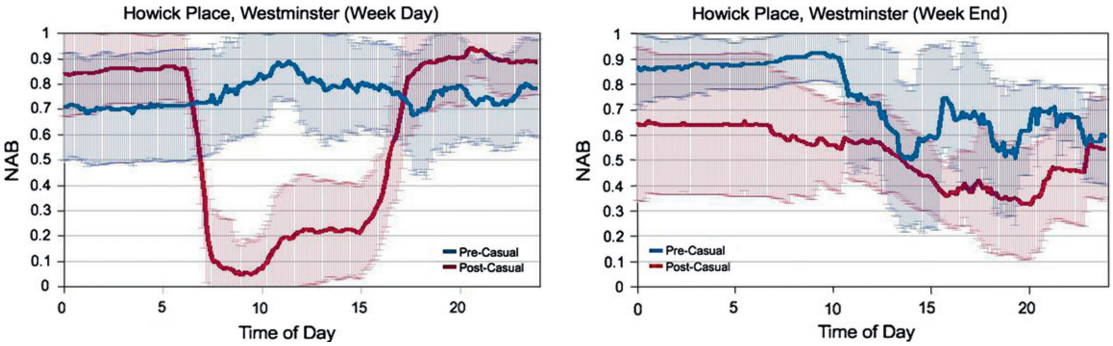
$$r_{a,b} = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2 \sum_i (b_i - \bar{b})^2}} \quad (5)$$

The correlation value r measures the extent that a linear dependence exists in a station's pair of profiles, pre-casual and post-casual. A value of 1 means that a strong linear relation exists: both profiles increase and decrease at the same time. A value of -1, instead, implies the inverse relation: one profile increases while the other decreases. A value of 0 indicates no linear relation. We note that, although we are comparing the results from the same station, we are comparing *two* profiles to each other, and thus are measuring self-correlation, not *autocorrelation*, which is typically used to measure a time series' relation to itself.

We show the distribution of self-correlation values across all stations in Fig. 4(b). While the majority of stations remain highly self-correlated (with values above 0.8), there are a number of stations that are negatively self-correlated. This indicates that the usage of the station has, in fact, flipped from how it was previously used (for example, an arrival trend becomes a departure trend). In the following section, we investigate these local differences in more depth by focusing the analysis on three stations.

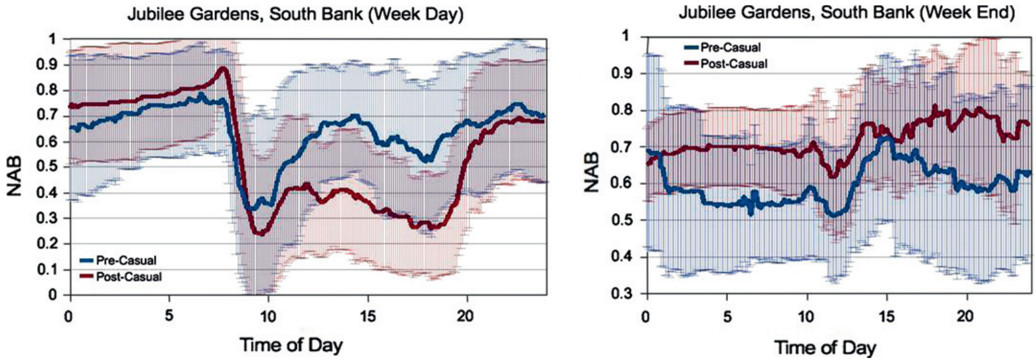
5.3. Local station activity change analysis: three case studies

While the aggregated data certainly shows changes that are happening on the city-wide scale, it says nothing about the effect of the policy change had on *local* temporal trends. We therefore continue the analysis with a comparison of the patterns exhibited by three stations in different areas of the city. We selected two stations randomly, as well as explicitly selecting the station with the highest measured change between the two datasets. The aim of the latter is to visualise the upper



(a) Week Day (b) Week End

Fig. 5. Focus on the Howick Place, Westminster Station: the station that displays the largest change between the two datasets.

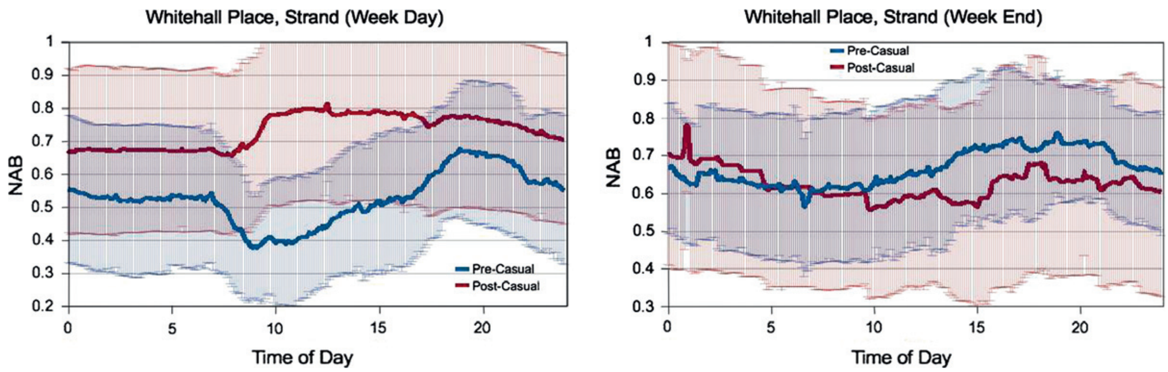


(a) Week Day (b) Week End

Fig. 6. Focus on the Jubilee Gardens, South Bank Station: the station differs in mid-morning/early afternoon behaviours; the pre-casual dataset has a positive slope, while post-casual has a negative trend.

bound to the changes we observe. In the following analysis, we compute and display the average NAB for each station, and plot it alongside its standard deviation (shaded regions). The pre-casual average is represented in blue, while the post-casual trend is red. The stations we discuss in this analysis are as follows:

Howick Place, Westminster (Fig. 5): this station is very near Westminster Cathedral, a typical tourist attraction, as well as Victoria Street, which is full of shops and restaurants. It is the station that displays the highest change between the two datasets (Pearson's $r = -0.69$). In fact, the week day behaviour (Fig. 5a) shows that what was once a relatively stable station, which remained, on average, between 70% and 90% full throughout the day, is now quickly emptied of bicycles in the morning and filled in the afternoon/early evening hours. The pattern has changed so sharply that even majority of the fluctuations, captured by the standard deviation (shaded regions), no longer overlap.



(a) Week Day (b) Week End

Fig. 7. Focus on the Whitehall Place, Strand Station: the week day departure (negative) trend changes to an arrival (positive) trend.

Jubilee Gardens, South Bank (Fig. 6): this station is one of the ten that is within less than 200 m from Waterloo station; the Jubilee Gardens themselves are a small park adjacent to the Thames river. Proximity to such a large transport hub correlates well with the sharp dip in bicycles held at the station in the morning hours; it may be explained by commuters who continue their journey to work from the station by bicycle. However, the lower correlation value (0.71) between the two profiles is explained by the fact that, in the pre-dataset, the station displayed, on average, a positive trend after the morning dip. The post-dataset, instead, has a negative trend: bicycles continue to leave the station until the early evening hours, when a sharp positive trend occurs. The week end trend shows less difference instead (recall, though, that the clustering was performed using the week day profiles, while week end data was excluded).

Whitehall Place, Strand (Fig. 7): this station is adjacent to the Embankment tube station. This station is characterised by the reverse change from what was observed above: the morning departure trend becomes a sharp arrival trend, and the station remains relatively constant throughout the day (while before, arrivals continued until early evening). In particular, the pre/post-values share nearly no correlation at all ($p = -0.02$). The week end pattern, instead, shows no significant change.

In the above analysis, we have visualised three individual stations' changes between the two datasets. While the first was significant – a station that was seemingly not used before becoming a morning origin/evening destination – the other two changes were less striking: the trends changed, while standard deviations continued to greatly overlap. This suggests that, while changes were measured in the means of a number of stations, there is no statistically significant difference between the pre and post-behaviour. This, again, serves to emphasise the value of station capacity data: the changes between the two datasets are *not* uniformly propagated across all stations.

6. Discussion: limitations, extensions and implications

In the analysis above, we explored how a policy change resounded throughout the entire anatomy of London's shared bicycle system. In doing so, we focused on what data mining techniques, combined with station capacity readings throughout the city, can tell us about the *changes* to urban life's footprint as decisions (e.g., access policy) are modified. In this section, we first expand on the limitations of this study, which we first addressed in Section 4; we then discuss the implications of this work, by considering *who* is affected by these changes and how.

A clear direction to improve work of this kind would be to have access to finer grained data, as well as to match it with other sources of information (e.g., demographics). Access to the transport authority's central database would allow us to view the system from a trip-basis: we would know each user-bicycle-origin–destination tuple, as well as the journey start and end times. This perspective would not only allow us to better the study performed above (and be able to explicitly differentiate between registered and casual users), but also to investigate how the system is utilised from the users' perspective (Lathia et al., 2010), including the trips, trip times, and stations that become linked by cyclists travelling between them. We note that Transport for London has released one such data set,[8] albeit without the user identifiers in order to protect the citizen's privacy. Matching this dataset with daily readings of precipitation, humidity, and temperature data would allow us to contribute to the discussion on the effect of weather in determining shared bicycle usage; we did not do so in this work since our focus was on the policy shift, we did not examine the data at the daily (non-aggregate) level, and the literature on the influence of weather (as reviewed in Section 2.1) remains inconclusive.

Access to a longer period of observation would also enable us to view the long-term effects and changes that the system underwent; in particular, further analysis (using, for example, sensitivity analysis) would allow us to observe how both seasonal and designed changes affect the system. However, we explicitly limited this by pruning data in the analysis above; this was in order to hone in on the time where the symptoms of the access policy change would be first exhibited.

Given that we did not have prior knowledge of the imminent policy change, we were unable to accompany our analysis with qualitative surveys, in order to measure the pre-change perception of both users and transport operators. However, the overarching limitation of such surveys would be the inescapable selection bias which was similarly experienced by Shaheen et al. (2011); our case studies were similarly lacking in survey and demographics due to time and budget constraints.

The results above show that the aggregate changes that we measured are not huge. This can be explained using another press release from Transport for London[9]: by February 2011, the system had nearly 110,000 members, while casual users had purchased 28,000 access periods for over 70,000 trips; if each access period represented a different cyclist (which gives us the maximum possible estimate), this would amount to 20% of the users falling into the casual category. It is therefore likely that the actual figure is much lower than this. More importantly, however, the changes that we measure are not spatially uniform: the stations that changed behaviour reside towards the perimeter of the system, rather than the centre. In the following, we discuss how these changes impact the different parties that are involved in the system's operation.

6.1. Travel operators

As discussed in Section 2.1, research into load-balancing shared-bicycle systems focuses on the *static* problem: optimising how a fleet of trucks should move throughout the city given that the station contents are, for the most part, not themselves

[8] http://www.tfl.gov.uk/tfl/syndication/feeds/BarclaysCycleHireUsageStats.zip.
[9] http://www.tfl.gov.uk/corporate/media/newscentre/archive/18060.aspx.

on the move (i.e., at night time). However, as we have seen above, the dynamics of day-time travel are such that the problem of empty (or full) stations is more prominent during daylight hours, particularly during commuting times. Real-time monitoring and forecasting stations' state—both from the aggregate and local perspective—may thus be of high value to transport operators, who may re-balance the system in a dynamic manner, transparently adapting to changes in the observed usage patterns over time (e.g., across policy chance).

6.2. End users

There exists a growing number of smartphone applications and online maps that aim to help travellers both find and return bicycles; nonetheless, users still report lack of available bikes/parking spots as the main problem with the London cycle hire scheme. It would thus seem that such applications would benefit from being augmented with forecasting techniques to not only allow travellers to see the system's current state, but also the likely state when they reach the nearest station. Once again, accurate real-time monitoring of stations' state, coupled with adaptive forecasting across (un)planned usage pattern changes, may attract more users to the system.

6.3. Urban planners

Lastly, an overriding theme from urban sensing work (as discussed in Section 2.3) is that both mobile and embedded sensors can highlight facets of urban life that would otherwise remain hidden or be too difficult to measure. In the above, we show how changes in urban mobility and usage change with policy: this latter dimension should thus play an important role in all urban sensing experiments, whose findings can be used by urban planners when designing shared spaces.

7. Conclusion and future work

As various forms of sensors continue to be embedded inside urban spaces, large-scale data becomes easily accessible that can be used to measure the interplay between the policy, design, and usage of transport systems. The work presented in this paper has focused on what data analysis and mining techniques (e.g., clustering) can tell us about how the spatio-temporal signature of London's bicycle hire scheme changed when the user access policy was modified. While at the macro-level similar usage patterns emerged, with concentric circles forming around the centre of London, and with variations across the north/south of the river divide, various micro (station) level changes could be detected, some of which of major intensity.

One of the key strengths of the work we conducted lies in the *scale* of the quantitative analysis we have been reporting. By passively monitoring the system state via sensors embedded in station docks, we could collect and analyse an extremely fine-grained and accurate dataset, thus obtaining detailed insights (both spatially and temporally) that no user questionnaires could ever offer. While this work highlights the promise that embedded computing holds for the future evaluation of transport system, we must also acknowledge its limits: passive sensors can tell us *how*, *when*, and *what* changes occur, but fail to contain any information as to *why*. We observed important variations in bike usage across policy changes, and we can seek for hints by inspecting the places (e.g., venues, transport stations, work institutions, residential areas) that surround the changing stations, but even these observations offer little explanation as to why flow changes occur. The qualitative work on bicycle usage reviewed in Section 2.1 thus compliments the observations we have made here: research that aims to understand *why* travellers opt for bicycles may decorate our results with additional contextual information. For example, one of the stations that we considered (Jubilee Gardens), which is within a short distance of a Waterloo, is just one of many cycle hire stations that surround the rail station. Why was this station the one that was affected most, while the others remain (relatively) unchanged? As stated above, sensor readings alone cannot answer this question; note, however, that highlighting interesting areas in the system for which further investigation is required, they offer valuable insights to survey designers as to what questions to ask and, crucially, to whom. Passive sensing and active user reporting will thus have to be combined for a more thorough understanding of the system in use.

Within the scope of passive sensing and quantitative analysis, there is a further dimension we have not delved into: the dynamics of day-to-day changes in activity as bicycle stations are closed (due to technical failures or upgrades), permanently relocated, or given additional capacity. We intend to pursue this line of work in the future by performing a full-scale comparison of shared-bicycle data from various cities across the world.

Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-SST-2008-RTD-1) under Grant Agreement No. 234239.

References

Baltes, M., 1996. Factors influencing nondiscretionary work trips by bicycle determined from 1990 US census metropolitan statistical data. Journal of the Transportation Research Board (1538), 96–101.

Bawa-Cavia, A., 2011. Sensing the urban: using location-based social network data in urban analysis. In: Workshop on Pervasive and Urban Applications. San Francisco, CA.
Benchimol, M., Benchimol, P., Chappert, B., Taille, A.D.L., Laroche, F., Meunier, F., Robinet, L., 2011. Balancing the stations of a self-service "Bike Hire" system. RAIRO Operations Research 45, 37–61.
Borgnat, P., Abry, P., Flandrin, P., Rouquier, J.-B., 2009a. Studying Lyons VloV: a statistical cyclic model. In: European Conference on Complex Systems. Warwick University, Coventry, UK.
Borgnat, P., Fleury, E., Robardet, C., Scherrer, A., 2009b. Spatial analysis of dynamic movements of VloV, Lyon's shared bicycle program. In: European Conference on Complex Systems. Warwick University, Coventry, UK.
Borgnat, P., Robardet, C., Rouquier, J.-B., Fleury, E., Abry, P., Flandrin, P., 2010. Shared bicycles in a city: a signal processing and data analysis perspective. Advances in Complex Systems.
Bows, A., Anderson, K., 2007. Policy clash: can projected aviation growth be reconciled with the UK government's 60% carbon-reduction target? Transport Policy 14, 103–110.
Bryan, H., Blythe, P., 2007. Understanding behaviour through smartcard data analysis. Proceedings of the Institution of Civil Engineers: Transport 160 (4).
Calabrese, F., Ratti, C., 2006. Real time Rome. Networks and Communications Studies 20, 247–258.
boyd, D. , Crawford, K., 2011. Six provocations for big data. In: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society.
Dill, J., Carr, T., 2003. Bicycle commuting and facilities in major US cities. Journal of the Transportation Research Board (1828), 116–123.
Eliasson, J., Jonsson, L., 2011. The unexpected "yes": explanatory factors behind positive attitudes to congestion charges in Stockholm. Transport Policy 18, 636–647.
Etzioni, O., Knoblock, C.A., Tuchinda, R., Yates, A., 2003. To buy or not to buy: mining airfare data to minimize ticket purchase price. In: ACM SIGKDD 2003 Conference on Knowledge Discovery and Data Mining. Washington DC, USA.
Ferris, B., Watkins, K., Borning, A., 2010. OneBusAway: results from providing real-time arrival information for public transit. In: ACM Human Factors in Computing Systems (CHI). Atlanta, GA, USA.
Froehlich, J., Krumm, J., 2008. Route prediction from trip observations. In: Intelligent Vehicle Initiative, SAE World Congress. Detroit, Michigan.
Froehlich, J., Neumann, J., Oliver, N., 2008. Measuring the pulse of the city through shared bicycle programs. In: International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems (UrbanSense). Raleigh, North Carolina, USA.
Froehlich, J., Neumann, J., Oliver, N., 2009. Sensing and predicting the pulse of the city through shared bicycling. In: Proceedings of the 21st International Joint Conference on Artificial intelligence. Pasadena, California, USA, pp. 1420–1426.
Graham, D.J., Glaister, S., 2006. Spatial implications of transport pricing. Journal of Transport Economics and Policy (JTEP) 40 (2).
Hanson, S., Hanson, P., 1977. Evaluating the impact of weather on bicycle use. Journal of the Transportation Research Board (629), 43–48.
Hayes, J., Stephenson, M., 2011. Bridging the social and physical sensing worlds: detecting coverage gaps and improving sensor networks. In: Workshop on Pervasive and Urban Applications. San Francisco, CA.
Institute of Transportation Engineers, T.S.C., 2009. ITE TSC before-and-after study. In: Technical Brief. Washington, DC, USA.
Jensen, P., Rouquier, J.-B., Ovtracht, N., Robardet, C., 2010. Characterizing the speed and paths of shared bicycle use in lyon. Transportation Research Part D, 522–524.
Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., Banchs, R., 2010. Urban cycles and mobility patterns: exploring and predicting trends in a bicycle-based public transport system. IEEE Pervasive and Mobile Computing 6, 455–466, http://dx.doi.org/10.1016/j.pmcj.2010.07.002.
Lathia, N., Capra, L., 2011a. How smart is your smartcard? Measuring travel behaviours, perceptions, and incentives. In: ACM International Conference on Ubiquitous Computing. Beijing, China.
Lathia, N., Capra, L., 2011b. Mining mobility data to minimise travellers' spending on public transport. In: ACM SIGKDD 2011 Conference on Knowledge Discovery and Data Mining. San Diego, CA, USA.
Lathia, N., Froehlich, J., Capra, L., 2010. Mining public transport usage for personalised intelligent transport systems. In: IEEE International Conference on Data Mining. Sydney, Australia.
Miranda-Moreno, M., Fernando, L., Nosal, T., 2011. Weather or Not to Cycle; Whether or Not Cyclist Ridership has Grown: A Look at Weather's Impact on Cycling Facilities and Temporal Trends in Urban Environments. US Transportation Research Board Annual Conference.
Nair, R., Miller-Hooks, E., Hampshire, R., Busic, A., 2012. Large-scale bicycle sharing systems: analysis of V'Elib. International Journal of Sustainable Transportation.
Nankervis, N., 1999. The effect of weather and climate on bicycle commuting. Transportation Research Part A 33, 417–431.
Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J., White, P., 2006. The demand for public transport: the effect of fares, quality of service, income and car ownership. Transport Policy 13, 295–306.
Pucher, J., Buehler, R., 2006. Why Canadians cycle more than americans: a comparative analysis of bicycling trends and policies. Transport Policy 13, 265–279.
Quercia, D., Lorenzo, G.D., Calabrese, F., Ratti, C., 2011. Mobile phones and outdoor advertising: measurable advertising. IEEE Pervasive Computing, 28–36.
Rattenbury, T., Good, N., Naaman, M., 2007. Towards automatic extraction of event and place semantics from Flickr tags. In: ACM SIGIR. Amsterdamn, Netherlands.
Ratti, C., Frenchman, D., Pulselli, R., Williams, S., 2006. Mobile landscapes: using location data from cell phones for urban analysis. Environment and Planning B 33, 727–748.
Raviv, T., Tzur, M., Forma, I., 2010. The Static Repositioning Problem in Bike-Sharing Systems. Working Paper.
Replogle, M., 1992. Bicycle and Pedestrian Policies and Programs in Asia, Australia and New Zealand. US Department of Transportation.
Rietveld, P., Daniel, V., 2004. Determinants of bicycle use: do municipal policies matter? Transportation Research Part A 38, 531–550.
Rose, G., Ahmed, F., Figliozzi, M., Jacob, C., 2011. Quantifying and Comparing the Effects of Weather on Bicycle Demand in Melbourne (Australia) and Portland (USA). US Transportation Research Board Annual Conference.
Saneinejad, S., Roorda, M., Kennedy, C., 2012. Modelling the impact of weather conditions on active transportation travel behaviour. Transportation Research Part D 17, 129–137.
Shaheen, S., Guzman, S., Zhang, H., 2010. Bikesharing in Europe, the Americas, and Asia: past, present, and future. In: 2010 Transportation Research Board Annual Meeting. Washington, DC, USA.
Shaheen, S., Zhang, H., Martin, E., Guzman, S., 2011. Hangzhou public bicycle: understanding early adoption and behavioural response to Bikesharing in Hangzhou, China. In: 2011 Transportation Research Board Annual Meeting.
Smith, M., Kauermann, G., 2011. Bicycle commuting in Melbourne during the 2000s energy crisis: a semiparametric analysis of intraday volumes. Transportation Research Part B 45, 1846–1862.
Soto, V., Frias-Martinez, E., 2011. Robust land use characterization of urban landscapes using cell phone data. In: Workshop on Pervasive and Urban Applications. San Francisco, CA.
Thomas, T., Jaarsma, C., Tutert, S., 2009. Temporal variations of bicycle demand in the Netherlands: the influence of weather on cycling. In: Annual Transportation Research Board.
Vogel, P., Mattfeld, D., 2010. Modeling of repositioning activities in bike-sharing systems. In: World Conference on Transport Research (WCTR).
Zuckerman, E., 2011. Desperately seeking serendipity. In: Keynote at ACM Conference on Human Factors in Computing Systems. Vancouver, Canada.