

A Markov-chain Based Model for a Bike-Sharing System

Emanuele Crisostomi
Department of Energy, Systems
Territory and Constructions Engineerings
University of Pisa
Italy
Email: emanuele.crisostomi@unipi.it

Mahsa Faizrahnemoon
and Arie Schlote
Hamilton Institute
National University of Ireland Maynooth
Maynooth, Ireland

Robert Shorten
University College Dublin
and IBM Research, Dublin
Ireland

Abstract—Bike-sharing systems are recently becoming ubiquitous in most cities, as an environmentally friendly alternative to other means of transportation. An optimal management of the bike-sharing service would in principle benefit from the availability of a mathematical model underlying the system. Accordingly, in this paper we propose a Markov-chain based approach to model the bike-sharing system, which we believe has a potential to develop alternative methods to implement classic control actions in a bike-sharing system (e.g., in terms of implementing alternative relocation strategies or planning advertising campaigns). The proposed methodology is validated on real data from the bike-sharing system in Boston, USA, and a first application of the proposed model is preliminarily illustrated in the paper.

I. INTRODUCTION

Bike sharing schemes are now available in most cities around the world as described among other in [1], [2], [3], [4]; see for example *Vélib'* in Paris, *Bicing* in Barcelona, *Capital Bikeshare* in Washington, D.C. and *Forever Bicycle* in Shanghai. A good overview can be gained from [5] and references therein. They are not only seen as an integral component in the public transportation networks for alleviating congestion and obesity problems, see [6], [7], [8], [9], but also as a fundamental tool for combating urban pollution, see [10], [11], [12]. They are also being suggested as a step towards solving the last mile problem¹ in intelligent transportation systems, see [13]. See also [14] for an overview over IT based bike sharing schemes in North America and [15] on a simulation study on the related field of electric bike sharing.

Given such a recent diffusion, there is also an interest in developing optimal ways of managing a bike-sharing system. In fact, as pointed out in [2], most bike-sharing systems have to rely on public fundings from municipalities or other public sources, since customer fees are not enough to match the capital costs plus the annual operational costs. Accordingly, there is a strong interest in devising smart ways of reducing operational costs, e.g., to minimise the costs of operating staffed trucks for manual relocation of bicycles in order to balance the difference between supply and demand at various stations (see [16]); or to maximise the revenues obtained by using the bikes for advertising.

Our interest here is to develop a model to capture the macroscopic aggregated properties of the bike-sharing system. From this perspective, we are mainly interested in answering high-level questions that pertain the identification of the main flows in the bike network, or how important are some bike stations with respect to others. Our ultimate goal is finally to support bike-sharing managers in implementing planning and control actions aiming at improving the efficiency of the bike network, in a proactive manner (i.e., anticipating the expected outcome of a strategic change in the bike network). The overall problem is particularly challenging if one intends to embed in the model the big quantity of data that become available in quasi real-time from a bike network. As an example, in this paper we use data collected from the bike-sharing network in Boston, USA².

For this purpose, our starting point is a recently proposed Markov-chain based framework for capturing macroscopic urban dynamics, namely [17]. That paper employed graph theory and Markov chain ideas to reveal some non-trivial patterns of urban mobility and to support engineers with practical tools to solve a number of mobility applications, such as routing, traffic light regulation and road works planning. Such a work has been recently extended along a number of lines (see among other [18] and [19]) and further validated over the small district of De Uithof, in Utrecht (in [20]) and for real data of the city of Beijing ([21]).

The objectives of this paper are two-fold: (i) first, we explain how the original idea of [17] can be adapted to the particular application of interest here, i.e., the bike-sharing network; (ii) second, we show some preliminary results that have been obtained by adopting the proposed modelling framework. The development of practical control applications based on the proposed model are currently under investigation and will be illustrated in a later work.

This paper is organised as follows: Section II explains how we build a Markov-chain transition matrix from bike trips data, and motivates why Markov chains are expected to be useful to handle big-data analyses. Section III shortly describes the used database of data and validates the transition

¹see C. MacKernie, *The Last Mile Problem* at url http://publictransport.about.com/od/Transit_Planning/a/The-Last-Mile-Problem.htm.

²<http://hubwaydatachallenge.org/>

matrix built on such data. Section IV shows clustering results obtained using the second eigenvector of the transition matrix, and provides an interpretation of the result in terms of the bike network. Finally, Section V concludes the paper and outlines our current lines of research.

II. A MARKOV-CHAIN MODEL FOR A BIKE-SHARING SYSTEM

In this paper we are interested in discrete-time, finite-state, homogeneous Markov chain. A Markov chain with n states is completely described by the $n \times n$ transition probability matrix \mathbb{P} , whose entry \mathbb{P}_{ij} denotes the probability of passing from state S_i to state S_j in exactly one step. \mathbb{P} is a row-stochastic non-negative matrix, as the elements in each row are probabilities and they sum up to 1. Within Markov chain theory, there is a close relationship between the transition matrix \mathbb{P} and a corresponding graph. The graph consists of a set of nodes that are connected through edges. The graph associated with the matrix \mathbb{P} is a directed graph, whose nodes are given by the states $S_i, i = 1, \dots, n$, and there is a directed edge leading from S_i to S_j if and only if $\mathbb{P}_{ij} \neq 0$. A graph is strongly connected if for each pair of nodes there is a sequence of directed edges leading from the first node to the second one. The matrix \mathbb{P} is irreducible if and only if its directed graph is strongly connected. We do not describe here other properties of Markov chains, and the interested reader can refer to conventional books, e.g., [22] for more details.

In our model of a bike-sharing network, we assume that the i 'th bike station is associated with two states: state BS_i refers to a bike that is currently parked at the i 'th Bike Station; state TB_i refers to a bike that is Travelling from the i 'th Bike station to any other destination. Thus, at every time step (in our simulation, one second), a bike staying in the BS_i state can either remain in the same state, or move to state TB_i (i.e., if somebody picks the bike to reach his destination). Similarly, a bike staying in the TB_i state can either remain in the same state (i.e., if it keeps travelling because the destination has not been reached yet) or can move to another state corresponding to any bike station BS_j , where the j 'th bike station corresponds to the destination. Note that j can be equal to i (i.e., some users may pick up a bike at the i 'th bike station, and then return the bike to the same place after having completed their own business). Also, note that it is not possible to pass directly from state BS_i to BS_j if $i \neq j$ or from state TB_i to TB_j if $i \neq j$. Our model is exemplified in the graph shown in Figure 1.

According to the previous discussion, the transition matrix of the Markov chain has size $2N \times 2N$ where N is the number of bike stations, and it is very sparse (i.e., all but two entries for the states BS_i are equal to zero, $\forall i = 1, \dots, N$, and $n - 1$ entries are equal to zero for the states TB_i , $\forall i = 1, \dots, N$). Then, the diagonal entries pertaining the BS states should be built proportional to the average time spent (parked) at a given bike station (i.e., by denoting such time with tp_i , then the entry should be $(tp_i - 1)/tp_i$ as explained in the Appendix in [17]), while the other non-zero entry of the same row is easily determined by enforcing the row to be stochastic (i.e., sum equal to 1). The diagonal entries pertaining the TB_i states are similarly computed proportionally to the average length

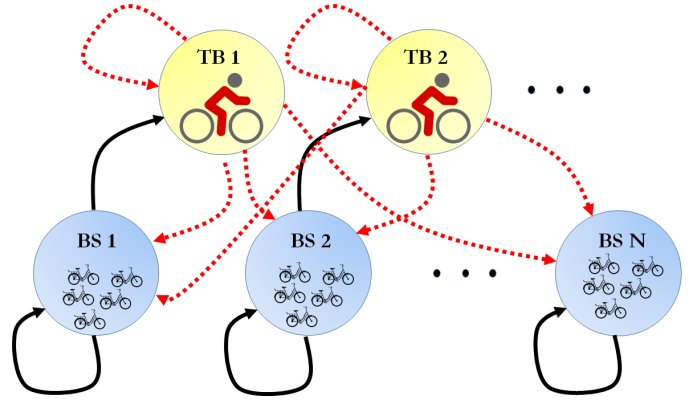


Fig. 1. Graph modelling a bike-sharing network. States BS_i 's are only connected to themselves or to states TB_i 's (black solid connections). States TB_i 's are connected to themselves and to every single state BS (red dashed connections).

of trips starting from the i 'th bike station. The other entries in the same row are chosen proportionally to the frequency with which a given bike station is chosen as a destination for a trip starting from the i 'th bike station, and scaled in order to make the row stochastic as well. Thus, the transition matrix can be built only if average parking times, average trip durations and a full list of trips are available. However, note that such a requirement is typically fulfilled, since such data are available anyway for other purposes as well, e.g., typically, GPS-like sensors are mounted on bikes to prevent people from stealing them; bike stations are monitored to choose optimal relocation strategies to avoid bike stations to remain empty; and finally information about the trips are available for pricing purposes. In particular, such data were publicly available for the city of Boston, USA, and in the next Section we explain how we built the transition matrix for that specific use case. The next subsection collects some known results to motivate the adoption of Markov chains for big-data applications.

A. Effects of simple updates of Markov chain transition matrices

Markov chains are particularly suited to big data applications for several reasons.

- First, microscopic behavior is embedded into the chain through aggregation; namely, in the form of probabilities. These probabilities are easily measured or calculated without need for large data processing or storage capabilities.
- The suitability of Markov chains for capturing and modelling complex dynamics is discussed and justified in [23].
- Third, the suitability of Markov chains for big data application is discussed, for example, in the context of Google's PageRank algorithm [24]. Here it is shown that well-established and robust algorithms are already available to handle data-sets of the size of thousands, if not millions, of webpages that might contain the relevant information pertaining the user's query.

- Finally, many of the key properties (e.g., Perron eigenvectors) of a Markov Chain can be calculated using simple update formulae [25]. This applies to updates of sparse transition matrices and simple rank-1 updates of the transition matrix (depending on application).

To gain insight into fast recalculation of Markov chain quantities for changed data, we now collect a number of theorems that can be used in this context. To set up notation, throughout this section let \mathbf{I} always denote the identity matrix of appropriate dimension, as well as e and e_i the all ones vector and the i 'th unit vector of appropriate dimension. Our starting point is the following well-known theorem from [25].

Theorem 1: Let \mathbb{P} and $\tilde{\mathbb{P}}$ be irreducible $n \times n$ Markov chain transition matrices and let π and $\tilde{\pi}$ be their respective left Perron eigenvectors. Let $\Pi = e\pi^\top$. Let further G be a 'one condition' generalised inverse of $\mathbf{I} - \mathbb{P}$, i.e. G satisfies $(\mathbf{I} - \mathbb{P})G(\mathbf{I} - \mathbb{P}) = \mathbf{I} - \mathbb{P}$. In this situation the following equality holds.

$$\tilde{\pi}^\top - \pi^\top = \tilde{\pi}^\top (\tilde{\mathbb{P}} - \mathbb{P})G(\mathbf{I} - \Pi). \quad (1)$$

The above theorem gives insight into the relationship between the stationary distributions of a Markov chain before and after a parameter update. In our situation we assume that $\tilde{\mathbb{P}}$ and \mathbb{P} as well as π along with Π are known and we wish to find an efficient way to compute $\tilde{\pi}$ (e.g., the update of the transition matrix after we receive information of one more trip). There are several candidates for a matrix G with the desired property. In particular, the group inverse matrix $Q^\#$ (see e.g. [26]) can be used. Ideally we wish to have an explicit formula for $\tilde{\pi}$. A sufficient condition for the existence of such a formula is that the matrix $\mathbf{I} - (\tilde{\mathbb{P}} - \mathbb{P})G(\mathbf{I} - \Pi)$ is invertible. This is simple to check for some forms of updates. For example, if $(\tilde{\mathbb{P}} - \mathbb{P})$ is rank-1 (which corresponds to the case where the transition matrix is updated due to a new trip), then one could use the following theorem from [27] addressing row updates to Markov chain transition matrices and their effect on the stationary distributions.

Theorem 2: Let \mathbb{P} and $\tilde{\mathbb{P}}$ be irreducible $n \times n$ Markov chain transition matrices that satisfy the relationship $\tilde{\mathbb{P}} - \mathbb{P} = -e_i\delta^\top$ where $\delta \in \mathbb{R}^n$. Let π and $\tilde{\pi}$ be the respective left Perron eigenvectors of \mathbb{P} and $\tilde{\mathbb{P}}$. Let $Q = (\mathbf{I} - \mathbb{P})$ and let $Q^\#$ be its group inverse. Then

$$\tilde{\pi}^\top = \pi^\top - \epsilon^\top, \quad (2)$$

where $\epsilon^\top = \frac{\pi_i}{1 + \delta^\top Q^\# e_i} \delta^\top Q^\#$. Further

$$(\mathbf{I} - \tilde{\mathbb{P}})^\# = Q^\# + e\epsilon^\top(Q^\# - \frac{\epsilon^\top Q^\# e_i \mathbf{I}}{\pi_i} - \frac{Q^\# e_i \epsilon^\top}{\pi_i}). \quad (3)$$

The above theorem allows one to explicitly compute the stationary distribution of a Markov chain after updating a single row using the original stationary distribution and the group inverse of $(\mathbf{I} - \mathbb{P})$. It also allows to directly compute the group inverse of the updated Markov chain and this theorem can thus be used iteratively to obtain updated stationary distributions for arbitrary changes in the transition matrix by means of describing them as consecutive row updates. In some situations even simpler formulas can be obtained. For example, the following theorem was proved in [18].

Theorem 3: Let \mathbb{P} and $\tilde{\mathbb{P}}$ be irreducible $n \times n$ Markov chain transition matrices such that $\tilde{\mathbb{P}}$ is obtained from \mathbb{P} by multiplying the i 'th diagonal entry with a factor $w_i > 0$ for each $i = 1 \dots n$ and scaling the off diagonal entries in each row so that their ratios remain constant. Let π and $\tilde{\pi}$ be the respective left Perron eigenvectors of \mathbb{P} and $\tilde{\mathbb{P}}$. Then

$$\tilde{\pi} = \kappa W \pi, \quad (4)$$

where $W = \text{diag}(w_1, \dots, w_n)$ and $\kappa = \frac{1}{\|W\pi\|_1}$ is a scaling factor that ensures that the entries of $\tilde{\pi}$ sum to 1.

In the bike example, the previous theorem can be used as soon as updates regarding the average parking time of bikes at a given bike station, or regarding the average time length of trips from a given bike station are received.

III. BOSTON DATA

The data used in this work involve about 70000 bike trips occurring between July 28 and September 21 in year 2011. Overall, 53 bike stations and 569 bikes were effectively used in the period under exam. The transition matrix was built upon such available data, and thus corresponds to average values in the examined time range. The transition matrix has size 106×106 and, consistently with the discussion in Section II, not more than $2N + (N + 1) \cdot N$ non-zero entries (actually, there are even less non-zero entries, namely, 2654, due to the fact that in the period under exam some particular pairs of origin-destination were never chosen by the users).

All the information required to build the Markov chain transition matrix can be extracted from the list of trips: in particular, every trip stores the ID of the bike, the time at which the bike is taken, the bike station of origin, the time at which the bike is delivered back and the bike station of destination. Other information was also available, such as whether the user was a registered or a casual customer (i.e., without a membership to the bike-sharing service); the zip code of the user; the birth year of the user and the gender. Note that all of such information can be used to deliver particular services (e.g., advertising), but have not been used in this work and their analysis will be left for a future work.

Remark 1 : One might expect that the destination of one trip should coincide with the origin of the next trip (of the same bike). This is true unless the bike gets relocated to another bike station, some time during the two successive trips (e.g., because there is a shortage of bikes at one bike station). Accordingly, in the list of trips from the available data, most of the times the two stations coincide, but in some cases there is a discrepancy. We assumed that every discrepancy was due to relocations, and did not consider such inconsistent trips when computing parking times (because the time of relocation was not available, and thus it could not be established how long the bike had been parked before relocation, or after relocation before being used again).

Remark 2 : Note that the available data only reveal a part of the true story. For instance, some users might have wanted to use a bike to take a specific trip but did not find any available bike at the origin bike station; or some users might have wanted to choose a specific destination but found that it

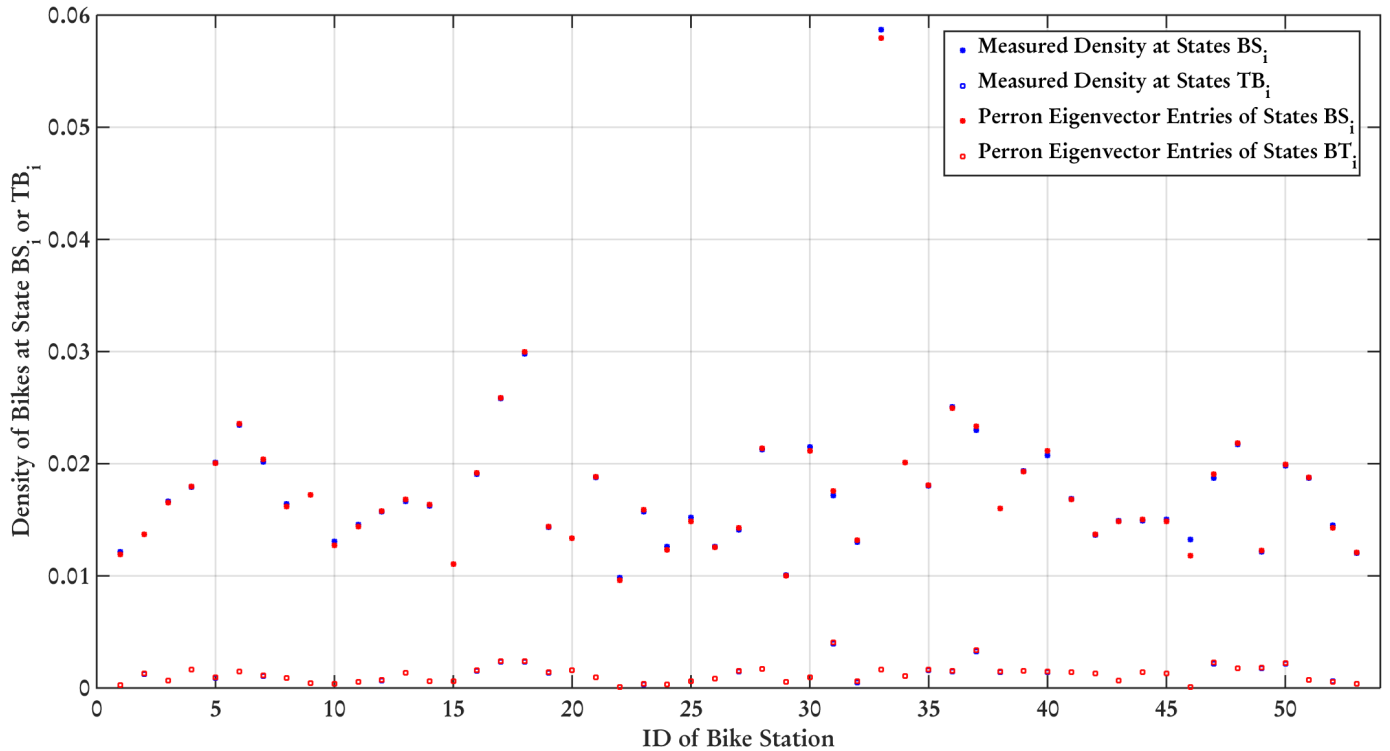


Fig. 2. Comparison between the Perron eigenvector of the Markov chain transition matrix (in red) and the same information extracted from the trip data (in blue).

was already full of bikes. We do not have information of such (somewhat) unsatisfied customers, but note that they might introduce some noise in the data (i.e., due to unavailability of bikes, they might choose a different starting point than the desired one; or due to full bike stations, they might choose a different destination point than the desired one).

Figure 2 shows in blue the Perron eigenvector of the Markov chain transition matrix built according to the procedure described in Section II. The entries of the Perron eigenvector correspond to the long-run fraction of time that a bike spends in the corresponding state, i.e., either parked at the i 'th bike station, or on a trip starting from the i 'th bike station. Figure 2 also compares the Perron eigenvector with the same information extracted from the same data (and shown in red). As can be noticed from the figure, the two densities are very close, and discrepancies are due to the fact that the real data correspond to a possible realisation of the Markov chain transition matrix given a specific initial distribution of bikes. Another interesting information that can be extracted from the figure is that bikes are available more likely than busy.

IV. CLUSTERING

As illustrated in reference [17] in a different context, the advantage of aggregating data in a Markov chain, is that some quantities that can be easily computed in the context of Markov chains have an insightful interpretation in terms of the original data that in some cases can not be easily extracted in other ways. As an example, the “second eigenvector”

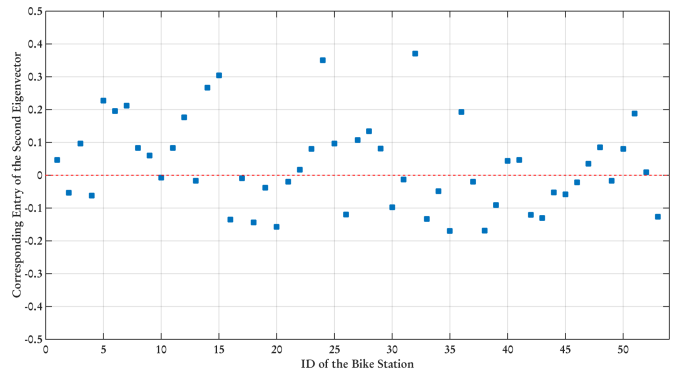


Fig. 3. Second eigenvector of the submatrix corresponding to the trips of vehicles.

(i.e., the eigenvector associated with the second eigenvalue of largest module, if real) is known to have clustering properties (see Appendix of the reference [17]). To obtain a meaningful interpretation of clusters in the bike network, we only considered the flows of bikes neglecting the times spent parked in the station, and the time spent travelling along a trip. Thus, we considered the submatrix of the original one whose states would correspond to TB_i , and renormalised its rows to have a unitary sum. Then, the second eigenvector of the matrix is shown in Figure 3.

The interpretation of clusters is that people most likely travel within a single cluster, and more rarely travel from

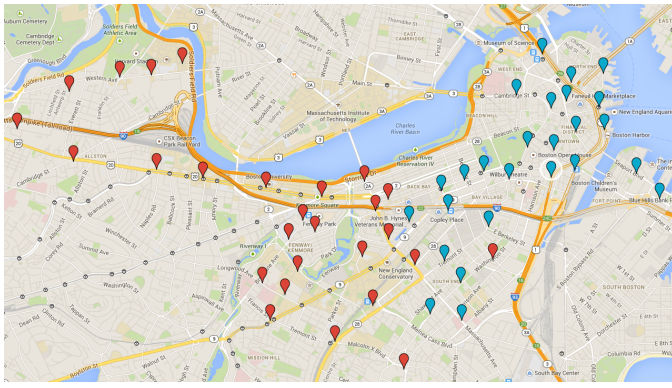


Fig. 4. Positive and negative entries of the second eigenvector do seem to provide a natural geographical characterisation of the bike stations in Boston.

one cluster to the other cluster. The possible membership to one cluster rather than to another one is given by the sign of the corresponding entry of the second eigenvector, and Figure 4 shows the positive and the negative entries of the second eigenvector with a different colour (the figure is taken from Google Map, using the information of the latitude and longitude of each single bike station, available from the data-set). The interesting property of the so-obtained clusters is that they do not exploit any geographical information about the position of the bike stations, but they only exploit the knowledge of the typical trips made by the customers, and thus, of the typical choice of origin-destination pairs. As shown in the figure, the clusters do seem to provide a correct information, as the locations are in fact consistently situated in two geographically separated regions.

Remark 3 : In principle, it could occur that different clusters arise at different times of the day. In this case, we further computed the second eigenvector taking into account only the trips occurring in the morning, those in the afternoon, and still we obtained the same results. This suggests that in this specific application clusters appear not to be time-varying.

The knowledge of the clusters can be used for a number of applications: since most trips occur within a cluster, and more rarely from one cluster to another cluster, the number of bikes within a cluster should be constant most of the time. Accordingly, one could plan two separate trucks performing relocation actions in the two clusters. Also, since there are two (quite) separate flows of trips within the two clusters, two advertising campaigns could be conducted in parallel in the two areas.

V. CONCLUSION

This paper proposed a new Markov chain based model to describe the functioning of a bike-sharing network. The model was inspired by a similar work pertaining vehicular flows, and adapted for the specific application which, among other things, required two states for each bike station (i.e., one to take into account the bikes parked at the bike station, and one to consider all trips starting from that given bike station). We used real data from the Boston bike-sharing network to build our Markov chain transition matrix, and to validate the

model with respect to the data.

The paper mainly focussed on the modelling part, while the only practical application described so far was the clustering one, which indeed did provide an interesting partitioning of the bike-sharing network, which could not be extracted from the same information in a straightforward manner. However, in our mind many other practical applications can be based on the proposed model, and in particular we are currently working towards designing novel relocation strategies and new ideas for advertising campaigns. For this purpose we plan to use other variables which can be easily extracted from the Markov chain (e.g., the Mean First Passage Times (MFPTs)), and other information from the database (e.g., the age and the gender of customers) which have not been used yet.

ACKNOWLEDGMENT

This work was supported in part by Science Foundation Ireland grant 11/PI/1177.

REFERENCES

- [1] S. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in Europe, the Americas, and Asia: Past, present and future," *Transportation Research Record*, 2010.
- [2] P. Midgley, "Bicycle-sharing schemes: enhancing sustainable mobility in urban areas," *United Nations, Department of Economic and Social Affairs*, 2011.
- [3] J. E. Pucher and R. E. Buehler, *City cycling*. Cambridge, MA: MIT Press, 2012.
- [4] E. Fishman, S. Washington, and N. Haworth, "Bike share: A synthesis of the literature," *Transport Reviews: A Transnational Transdisciplinary Journal*, vol. 33, no. 2, pp. 148–165, 2013.
- [5] P. DeMaio, "Bike-sharing: History, impacts, models of provision, and future," *Transport Reviews: A Transnational Transdisciplinary Journal*, vol. 12, no. 4, pp. 41–56, 2009.
- [6] P. Midgley, "The role of smart bike-sharing systems in urban mobility," *Journeys*, pp. 23–31, 2009.
- [7] E. Fishman, "The impact of public bicycle share schemes on transport choice," in *Asia-Pacific Cycle Congress*, Berlin, Germany, 2011, pp. 18–21.
- [8] N. Lathia, S. Ahmed, and L. Capra, "Measuring the impact of opening the London shared bicycle scheme to casual users," *Transportation Research Part C*, vol. 22, pp. 88–102, 2012.
- [9] E. Murphy and J. Usher, "The role of bicycle-sharing in the city: Analysis of the Irish experience," *International Journal of Sustainable Transportation*, vol. 9, no. 2, pp. 116–125, 2015.
- [10] J. Pucher, J. Dill, and S. Handy, "Infrastructure, programs, and policies to increase bicycling: An international review," *Preventive Medicine*, vol. 50, no. Supplement, pp. S106–S125, 2010.
- [11] R. Buehler and J. Pucher, "Cycling to sustainability in Amsterdam," *Sustain: A Journal of Environmental and Sustainability issues*, vol. 21, pp. 36–40, 2010.
- [12] J. Pucher, R. Buehler, and M. Seinen, "Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies," *Transportation Research Part A*, vol. 45, no. 6, pp. 451–475, 2011.
- [13] J. Pucher and R. Buehler, "Making cycling irresistible: Lessons from the Netherlands, Denmark and Germany," *Transport Reviews: A Transnational Transdisciplinary Journal*, vol. 28, no. 4, pp. 495–528, 2008.
- [14] S. Shaheen, E. Martin, A. Cohen, and R. Finson, "Public bikesharing in North America: Early operator and user understanding," *Research Report 11-26, Mineta Transportation Institute*, 2012.
- [15] S. Ji, C. Cherry, L. Han, and D. Jordan, "Electric bike sharing: simulation of user demand and system availability," *Journal of Cleaner Production*, vol. 85, pp. 250–257, 2013.

- [16] J. Pfrommer, J. Warrington, G. Schildbach, and M. Morari, "Dynamic vehicle redistribution and online price incentives in shared mobility systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1567–1578, 2014.
- [17] E. Crisostomi, S. Kirkland, and R. Shorten, "A Google-like model of road network dynamics and its application to regulation and control," *International Journal of Control*, vol. 84, no. 3, pp. 633–651, 2011.
- [18] E. C. A. Schlote, S. Kirkland, and R. Shorten, "Traffic modelling framework for electric vehicles," *International Journal of Control*, vol. 85, no. 7, pp. 880–897, 2012.
- [19] T. Morimura, T. Osogami, and T. Idé, "Solving inverse problem of markov chain with partial observations," in *IBM Research Report RT0952*, 2013.
- [20] B. B. Klaassen, "Improving traffic mathematically: Expanding the work of Crisostomi et al. (2011)," in *Thesis - Faculty of Utrecht - Nederland*, 2015.
- [21] V. Moosavi and L. Hovestadt, "Modeling urban traffic dynamics in coexistence with urban data streams," in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, 2013.
- [22] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*. Princeton: Van Nostrand, 1960.
- [23] A. Schlote, *New Perspectives on Modelling and Control for Next Generation Intelligent Transport Systems*. Hamilton Institute, National University of Ireland Maynooth: PhD Thesis, 2014.
- [24] A. N. Langville and C. D. Meyer, *Google's PageRank and beyond: The science of search engine rankings*. Princeton: Princeton University Press, 2006.
- [25] J. Hunter, "Stationary distributions and mean first passage times of perturbed markov chains," *Linear Algebra and its Applications*, vol. 410, no. 15, pp. 217–243, 2005.
- [26] C. D. Meyer, "The role of group generalized inverse in the theory of finite markov chains," *SIAM Review*, vol. 17, no. 3, pp. 443–464, 1975.
- [27] A. Langville and C. D. Meyer, "Updating Markov chains with an eye on Google's PageRank," *SIAM Journal of Matrix Analysis and Applications*, vol. 27, no. 4, pp. 968–987, 2006.