Spatio-temporal usage pattern analysis of the Paris Shared Bicycle Scheme: a data mining approach

Etienne Côme, Andry Randriamanamihaga, Latifa Oukhellou *

Université Paris-Est-IFSTTAR-COSYS-GRETTIA, F-77426 Champs Sur Marne, France

Abstract

In recent years, public authorities play an important role in boosting sustainable mobility practices in urban areas. One urban mobility measures adopted by major cities as an additional mean of sustainable intermodal transport is the implementation of Bike Sharing Systems. Within this context, the Vélib' Bike Sharing System (BSS) of Paris has been launched in July 2007. In this paper we investigate the analysis of BSS systems through the clustering of their stations and the origin-destination flow-data with respect to their trips data. This approach offers the possibility to highlight mobility patterns in the BSS usage. The analysis of the results shows that the clustering found is closely related to the city functions (transportation, leisure, employment) which can be helpful for a variety of applications, including urban planning and the choice of business location. Crossing the results of the model with sociological and economic data is carried out to show the close links between these two aspects and the use of a bike sharing transport scheme, which may be useful for bike redistribution planning and for designing new BSSs.

Keywords: Bike Sharing System; urban mobility; usage patterns; data mining.

Résumé

Les récents enjeux écologiques, la congestion routière grandissante et l'explosion démographique sont autant de facteurs qui ont motivé l'émergence de nouvelles politiques de mobilité durable, qui promeuvent l'usage des transports en commun et des modes doux de mobilité. Dans ce contexte, un système de Vélos en Libre-Service (VLS) a été mis en place par la ville de Paris (Vélib) depuis Juillet 2007. Dans cet article, nous proposons une approche statistique en vue d'extraire des données de déplacement générées par un tel système, des groupes de stations ou des couples Origine-Destination ayant des motifs d'usage similaires. L'analyse des résultats obtenus permet de mettre en évidence les liens existants avec d'autres données de la ville (Transport, loisirs, emploi). De plus, le croisement des résultats fournis par ce modèle avec des données socio-économiques telles que la densité de population, d'emploi et de commerces permet in fine de mieux appréhender l'usage qui est fait de ce système, ce qui pourrait être utile à la fois pour optimiser les politiques de redistribution des vélos et pour implémenter de nouveaux systèmes VLS.

Mots-clé: Vélos en Libre Service, mobilité urbaine, formes d'usage, fouilles de données.

* Corresponding author: Latifa Oukhellou, Senior Reseacher, Université Paris-Est, French Institute of Sciences and Technology for Transport, Development and Networks (IFSTTAR), COSYS, GRETTIA, 14-20 Boulevard Newton, Cité Descartes, Champs sur Marne, F-77447 Marne la Vallée Cedex 2, Tel.: +33 1 81 66 87 19; fax: +33 1 81 66 80 01.
E-mail address: latifa.oukhellou@ifsttar.fr.

1. Introduction

In recent years, public authorities play an important role in boosting sustainable mobility practices in urban areas aiming at reducing congestion and environmental issues related to pollution and noise. They encourage people to use public transport, soft modes of transport such as cycling, walking ... , or a combination of them, thus lessening the need for private cars usage (the car is the dominant mode of transport in all European countries). From a users's perspective, the attractiveness of a mode of transportation as an alternative to private car would be linked to its flexibility, its facilities and also its schedule organization. One urban mobility measures adopted by major cities as an additional mean of sustainable intermodal transport is the implementation of Bike Sharing Systems (BSSs).

Within this context, the Vélib' Bike Sharing System (BSS) of Paris has been launched in July 2007. It has quickly become one of the largest BSS in the world: 20 000 bikes distributed on 1208 fixed stations now generate 224 000 annual subscribers and an average number of 110 000 travels per day. The availability of such sizeable dataset contributes to leverage the development of novel approaches based upon engineering and computer sciences and aiming to recover the underlying urban phenomena linked to city dynamics. This kind of analysis would surely help sociologists and planners in their task of apprehending the bikers' mobility patterns within large megacities. This knowledge can then be transferred to other cities wishing to introduce BSSs. Several studies (Froehlich et al., 2009) (Kaltenbrunner et al., 2010) (Borgnat et al., 2011) (Vogel et al., 2011) (Lathia et al., 2012) (Nair et al., 2012), have shown the usefulness of analyzing the data collected by BSSs operators and city authorities. A statistical analysis of such data helps in the development of new and innovative approaches for a better understanding of both urban mobility and BSS use and performance. The design of BSSs, the adjustment of pricing policies, the improvement of service level of the system (redistribution of bikes over stations) can all benefit from this kind of analysis (Dell'Olio et al., 2011) (Lin et al., 2011) (Buttner et al., 2011) (Apur, 2006) (De Maio, 2009), which also helps sociologists and planners to understand user mobility patterns within the cities.

In this paper we investigate the analysis of BSS systems through the clustering of their stations and origin-destination flow-data with respect to their usage data. A dedicated data mining approach based on poisson mixture models is designed to this end, it offers the possibility to highlight mobility patterns in the BSS usage from trips data collected on the system. The proposed methodology was tested to mine one month of usage data from the Paris Vélib' system. The analysis of the results shows that such an approach may provide latent factors that reveal how regions of different usage interact over the time. The obtained clusters are closely related to the city functions (transportation, leisure, employment) which can be helpful for sociologists and planners in their task of apprehending the bikers' mobility patterns withing large megacities. Additionnal analysis of the results are carried out to give insights into the relations between the kind of neighborhood of the stations (the type of amenities it offers, its demographics, etc.) and their associated usage profiles. Crossing the results of the model with sociological and economic data is carried out to this end, and shows the close links between these two aspects and the use of the Vélib' bike sharing transport scheme, which may be useful for bike redistribution planning and for designing new BSSs.

This paper is organized as follows, Section 2 presents the key characteristics of the vélib' Bike Sharing System of Paris. Section 3 introduces the proposed statistical model based upon count series clustering. The results are given and discussed in Section 4, followed by a general discussion and conclusion in Section 5.

2. The Vélib' Bike Sharing System (BSS)

2.1. General description of the Vélib' network

The Vélib' BSS has been deployed since July 2007 and is operated as a concession by Cyclocity, a subsidiary of the French outdoor advertising company JCDecaux. At its debut in 2007, 700 bicycles were spread across 750 fixed stations. In six years, it has expanded to 1208 stations which hire out more than 18,000 bikes throughout the city, with 224,000 annual subscribers making on average 110,000 trips each day. Vélib' is a large-scale scheme, one of the largest Bike Sharing Systems in the world and the biggest Bike Sharing System in Europe. Vélib' is available mainly in Paris *intramuros*, some stations being located in the suburbs.

Each Vélib' station is equipped with a non-stop and an automatic rental terminal. The whole network includes 40,000 docking points (between 8 and 70 per station). The bikes are locked to the electronically controlled docking points. Users can purchase a short-term daily or weekly subscription, or a long-term annual subscription. The subscription allows an unlimited number of rentals, the first half hour (or the first 45 minutes for a long-term subscription) of every individual trip being free. Registration of users is required. The bicycles can be hired at any of the stations and at any time and returned back at any other station and at any time.

2.2. Global Statistics of the Vélib' network

Global trends in Vélib' usage can be highlighted through some general statistics presented below. The dataset used to estimate these global statistics was provided by JCDecaux and Ville de Paris and corresponds to one month of trips data recorded in April 2011. The system recorded respectively around 2,500,000 trips after data cleansing. In preprocessing the data, trips with a duration of less than one minute and which looped at the same station were excluded. These trips correspond to user misoperation and not to real trips.

Figure 1 displays the total number of recorded Vélib' trips per hour and day of the week during a month, with respect to the type of subscription (annual plotted in blue and one day plotted in red). It shows that Vélib' usage is closely linked not only to the hour and the day but also to the kind of day (weekday or weekend) and to the type of subscription. A first significant difference in Vélib' usage between short-term and long-term subscribers can be seen. This difference is reflected in terms of the usage volume: most of the Vélib' trips are generated by long-term subscribers, even if the difference between the two subscriptions is smaller during the weekend. This can be linked to the fact that short-term subscriptions are mainly associated with leisure, while long-term subscriptions tend to cover the users'daily mobility routines.

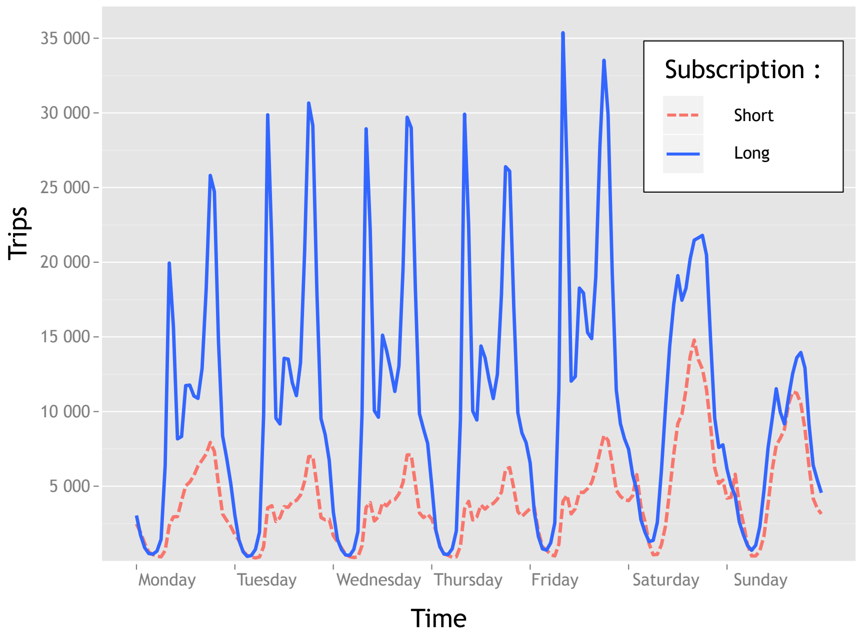

Fig. 1. Total number of trips during April 2011, summed over each hour of each day of the week, from Sunday to Saturday. The blue line (resp. red line) corresponds to the displacements carried out by one-year (resp. one-day) subscribers.

The blue curve shows a repetitive but distinct pattern depending on the type of day. Weekdays (Monday to Friday) are marked with peaks at the commutes (8am and 6pm) and during the lunch break, whereas the highest volume usage at weekends (Saturday and Sunday) is evenly distributed throughout the afternoons. The red curve depicts a totally different pattern with higher activity early morning and late afternoon. In addition, considering the volume of displacements during Saturdays, Sundays and Mondays, the typical weekend pattern for the one-day users lasts until Tuesday. It is reasonable to assume that these trips occuring at weekends are more leisure and recreational oriented, and the ones occurring at weekdays are more utilitarian oriented. These temporal trends are not peculiarly French feature. Froehlich et al. (2008) and Froehlich et al. (2009) who studied Bicing, the Bike Sharing System of Barcelona, similarly identified a distribution of trips marked by peaks at key moments of the day. Indeed these temporal trends of BSS usage can provide information on the sociological characteristics of the city. Considering the study carried out on the Barcelona Bicing system, some sociological differences between the two cities can indeed be highlighted. The lunch peak occurring at 2pm in Barcelona Bicing data occurs at 12 noon in Vélib' data, reflecting thus the late lunch culture of Spain (resp. the earlier lunch culture of France). Secondly, Friday is the least active day in Barcelona Bicing usage (resp. the most active one in Vélib usage).

In addition to these temporal trends in the use of Vélib bicycles, the duration and distance of trips can also be used as indicators of Vélib' usage. The trips are mostly short, one in two being less than 1.6 kilometres and four out of five being less than three kilometres, as shown in Figure 2(a), where the histogram represents the trip distance for the observed period of April 2011. This can be linked to the Vélib' pricing policy (free for half an hour). Figure 2(b), shows the histogram of trip duration: 91% of the trips last less than 30 minutes, which is the free usage for one-day subscriptions and 96% of them last less than 45 minutes, which is the free usage for one-year subscriptions. It should be noted that the trips recorded with null distances correspond to round trips, i.e. users rent bikes from and return them to the same station.

These general statistics show the global dynamic of the Vélib' system. Let us now present the proposed statistical model build to automatically extract patterns from BSS data.

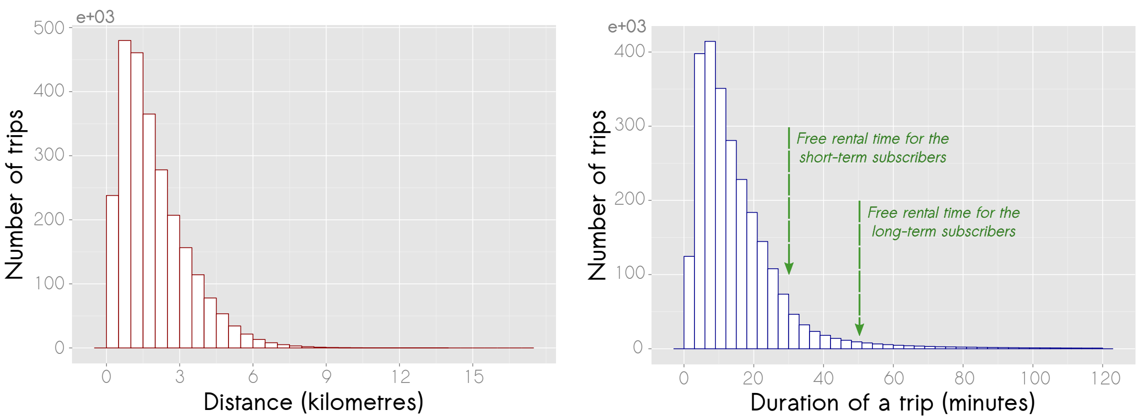

Fig. 2. (a) Histograms of trip distance; (b) Histograms of trip duration computed on the Vélib' April 2011 data.

3. Statistical methodology

As in the previous studies related to BSS station clustering (Froehlich et al.,2009), (Vogel en al. 2011), (Borgnat et al. 2012) (Lathia et al., 2012), the method investigated here aims to identify groups of stations with similar usage profile, but it differs from these studies on a number of points:

- First, it differs from the work of (Froehlich et al., 2009) and (Lathia et al., 2012) since it does not use the station occupancy data but arrivals/departures count time series derived from trips data to describe in a more detailed way the stations.
- The proposed model can also directly deal with the differences in behavior observed during weekdays and at weekends whereas it was handled through data preprocessing and feature construction in the two studies (Borgnat et al., 2012) (Vogel et al., 2011).
- The station usage count for each day will be processed by the model and not a summary over a long period which may be affected by seasonal or meteorological factors. Also, the proposed model can handle the specific nature of the observations, *i.e.* that they are counts whereas previous solutions do not use this particularity.

To achieve these goals, we propose a generative mixture model and derive the associated EM algorithm to estimate the parameters of the model and the clustering. This work adopts, therefore, the model-based clustering framework (Lachlan 2000), (Fraley et al., 2002}, with specific hypotheses related to the phenomena under analysis as discussed in the following paragraphs. We begin with a more formal description of the count time series construction, and then introduce the notations used in the rest of the paper.

3.1. Representation of the trips data

The target dataset of the proposed method corresponds to classical trips data recorded on BSS systems. It contains the following information for each trip: station of departure, time of departure, station of arrival, time of arrival. These raw data can be used to derive the following counts statistics to describe station usage:

- X_{sdh}^{out}: Departure count for station $s \in \{1, \cdots, S\}$ during day $d \in \{1, \cdots, \mathrm{D}\}$ and at hour $h \in \{1, \cdots, 24\}$;
- X_{sdh}^{in}: Arrival count for station $s \in \{1, \cdots, S\}$ during day $d \in \{1, \cdots, \mathrm{D}\}$ and at hour $h \in \{1, \cdots, 24\}$.

The aggregation at 1 hour was used to produce the counts since it gives a good trade-off between resolution of details and fluctuations (Borgnat et al., 2011}. These two time series of counts are then concatenated in a vector

$$\mathbf{X}_{sd} = \left(X_{sd1}^{in}, X_{sd2}^{in}, \cdots, X_{sd24}^{in}, X_{sd1}^{out}, X_{sd2}^{out}, \cdots, X_{sd24}^{out}\right) \tag{1}$$

These activity vectors can then be arranged in a tensor (or three-way array) of size N x D x T, with N the number of stations, D the number of available days in the dataset and T the length of the description vector, here 48 since non-overlapping windows of one hour are used to compute the arrivals and departures counts.

3.2. Vélib' Stations Clustering with Poissonian mixture model

Since the observed data are counts, we propose to use the following generative Poisson mixtures that discerns K clusters of stations:

$$\begin{aligned} &Z_s \sim M(1, \pi) \\ &\mathrm{X}_{sd1} \sqcup \cdots \sqcup \mathrm{X}_{sdT} \mid \{Z_{sk} = 1,\ W_{dl} = 1\} \\ &X_{sdt} \mid \{Z_{sk} = 1, W_{dl} = 1\} \sim P(\alpha_s\ \lambda_{klt}) \end{aligned} \tag{2}$$

Where Z_s is the indicator variable that encodes the cluster membership of the stations, $M(1, \pi)$ the multinomial distribution of parameter π (π are the prior proportions of each cluster of the mixture), W_d the indicator variables which are attached to the days and encode the differences between weekdays and weekends (which present very different usage profiles) and $P(\alpha_s\ \lambda_{klt})$ is the Poisson distribution of parameter $\alpha_s\ \lambda_{klt}$. The parameter α_s is a scaling factor specific to station s and will capture the global activity of the station. The parameters λ_{klt} describe the temporal variations of departures/arrivals and are specific to each station cluster and day type (weekday/weekend). These scaling factors are necessary to produce useful results since stations may share a common usage profile but differ strongly in terms of departure/arrival volume. For the parameters to be identifiable we must ensure the following constraints:

$$\textstyle\sum_{l,t} D_l \lambda_{klt} = \mathrm{DT}\,,\ \forall\, k\ \in \{1, \cdots, K\} \tag{3}$$

Where $D_l = \sum_d W_{dl}$ is the number of days in cluster *l*. Under some asumptions, the likelihood of the parameters expresses as :

$$L(\theta; X|W) = \textstyle\sum_{s=1}^{S} \log\left(\sum_{k=1}^{K} \pi_k\ \prod_{d,t,l}^{b} p(X_{sdt}; \alpha_s \lambda_{klt})^{W_{dl}}\right. \tag{4}$$

The estimate $\widehat{\Theta} = \left(\widehat{\alpha}, \widehat{\lambda}, \widehat{\pi}\right)$ that locally maximises the log-likelihood is computed with an Expectation – Maximisation (EM) type algorithm (Dempster et al., 1977), a well fitted two-step iterative approach for maximum likelihood estimation in statistical problems involving latent variables. EM also gives the partitions of stations by returning the a posteriori probabilities t_{sk} of the station *s* to be in cluster *k*. The total number of cluster K, can be determined by means of model evaluation criteria. A detailed description of the algorithm as well as the choice of the number of cluster K can be found in Randriamanamihaga et al. (2013). Next section will reveal the efficiency of the proposed model through the Vélib' usage data analysis.

When performing the clustering of origin-destination flow-data instead of stations, the variable *s* could be substituted by a couple *(u,v)* of OD and the statistical methodology remains valid (Randriamanamihaga et al., 2013).

4. Results and discussion

A first way to investigate the nature of the different clusters found is to look at their temporal profiles given by the parameters λ of the model. The number of the clusers was chosen equal to 8 (Randriamanamihaga et al., 2013). The results for five specific clusters are presented in Figure 3 and Figure 4. We name the first cluster "Railway stations", the second "Parks", the third "Housing" and the latter two "Employment 1" and "Employment 2" because these clusters correspond to stations close to these five kinds of amenities. This can easily be seen from the corresponding maps also presented in Figure 3 and Figure 4, which show the cluster station positions above a map background which presents the metro and railway lines, the parks and the Seine. The relationship between each of the clusters and the corresponding amenities is clearly shown.

The temporal profiles of the two first clusters also present interesting points: the profile of the "Railway stations" cluster shows much activity around peakhours for both departures and arrivals. The "Parks" profiles give a totally different picture, with a rush of activity during the weekend afternoons and a low activity during the weekday peak hours. The maps (see Figure 3) which depict the positions of the cluster stations confirm the interpretation of these two usage clusters. All the railways stations of Paris are clearly visible on the first map, along with several important metro stations like Nation, Denfert-Rochereau, Porte d'Orléans and Vincennes. The map of the stations which belong to the "Parks" cluster also gives a clear view of the nature of this cluster: all the stations are close to parks like Vincennes, Buttes-Chaumont, Montsouris, La Villette, etc. A map showing the location of the metro stations is presented in Figure 5.

The main remaining clusters shown in Figure 4 can also be explained in terms of geography and demographics. The "Housing" cluster has an asymmetrical profile, with a lot of departures during the morning rush but few arrivals, and the reverse during the end of the work peak. These stations belong to a belt surrounding the center of Paris which presents a high population density. The next two clusters, "Employment 1" and "Employment 2", present a reverse asymmetry to that of the "Housing" clusters: a lot of arrivals during the morning rush but few departures, and the reverse during the end of work peak. These two clusters are correlated to the employment density as shown in Table 1. During the weekend the two clusters present differences, with more activity in stations from "Employment (2)".

The above observations concerning the relationships between the clusters and the nature of the neighborhoods of the stations that belong to them made through the analysis of the maps presented in Figure 3 and 4 can be quantitatively investigated. To this end the per cluster average of additional socio-economic variables (population density, employment density, services (restaurants, hairdressers, etc..) and shops density) have been computed (see Table 1). An analysis of variance confirms that the station clusters derived from BSS usage data are significantly different with respect to these four variables. As expected, the local density of inhabitants is particularly high for the "Housing" cluster, the density of employment being, at the opposite end, high for the "Employment 1" and "Employment 2" clusters. Finally, the shops and services densities are high for the "Spare-time" clusters.

Table 1. Mean of each cluster with respect to population density (number of inhabitants per hectare), employment density (number of jobs per hectare), services density (number of personal services such as restaurants, hairdressers, etc. per hectare) and shops density (number of shops per hectare). Sources "Recensement 2008", "Base permanente des équipements", Insee.

Cluster Name	inhabitants/ha	jobs/ha	services/ha	shops/ha
All	162	237	4.2	3.7
Parks	172	90	2	1.7
Railway Stations	209	206	2.4	1.8
Housing	**375**	108	3.8	2.7
Employment 1	138	**409**	4.5	2.8
Employment 2	157	**456**	5.7	5.6
Spare-Time 1	367	189	**6.3**	**4.4**
Spare-Time 2	261	322	**7.7**	**6.9**
Mixed	301	163	3.8	2.8

Cluster “Railway stations”

Cluster “Parks”

Fig. 3. Maps of station positions for “Railway stations” and “Parks” clusters. The map background presents the metro and railway lines, the parks and the Seine. The areas of the dots representing the stations are proportional to the station scaling factor α_s. Each cluster map is completed with the temporal profile of the cluster; to this end the parameters λ_{klt} are arranged according to departure/arrival and weekday/weekend. The quantiles 0.25 and 0.75 of the total population of station activity (scaled by their average activity) are also shown in order to highlight the temporal specificities of each cluster.

Cluster “Housing”

Cluster “Employment 1”

Cluster “Employment 2”

Fig. 4. Maps of station positions for “Housing”, “Employment 1” and “Employment 2” clusters. The map background presents the metro and railway lines, the parks and the Seine. The areas of the dots representing the stations are proportional to the station scaling factor α_s. Each cluster map is completed with the temporal profile of the cluster; to this end the parameters λ_{klt} are arranged according to departure/arrival and weekday/weekend. The quantiles 0.25 and 0.75 of the total population of station activity (scaled by their average activity) are also shown in order to highlight the temporal specificities of each cluster.

Fig. 5. A map showing the location of the metro stations.

5. Conclusion

This paper has introduced a statistical clustering methodology to explore the usage statistics generated by BSSs. This model introduces a latent variable to encode the stations cluster membership, and an observed variable which deals with the difference of usage between weekdays and weekend. The methodology was tested to mine one month of usage data from the Paris V´elib' system. The clustering found is rich with interpretable clusters easily linked to the presence of certain type of amenities like parks and railway and railway stations, and to sociological variables like population, jobs and services densities.

The proposed methodology is obviously very general and can be applied to many other problems involving data that could be obtained by the urban activity. Data produced by a self-service vehicle system obsviously can be analyzed following the same methodology. Similarly, data collected by public transport fare collection systems in the origin-destination form can be analysed by the proposed methodology to extract mobility patterns.

There is nonetheless room for possible improvements, because of the limited data size (one month) used during the experiments, further investigations involving a larger data size (collected over the year for example) have to be made in order to take into account the influence of seasons and weather conditions. The proposed mixture model is flexible enough to easily take into account this season variability. Adding an observed variable linked to weather conditions could be an interesting way to handle this issue.

Acknowledgements

The authors wish to thank Ville de Paris and Cyclocity - JCDecaux for providing Vélib' data.

References

Apur. (2006). Etude de localisation des stations de vélos en libre service. Rapport. Tech. Rep. 349, *Atelier Parisien d'Urbanisme*. December.

Borgnat, P., & Robardet, C., & Abry, P., & Flandrin, P., & Rouquier, J., & Tremblay, N. (2013). *Dynamics On and Of Complex Networks*, Volume 2. Springer Berlin Heidelberg, Chapter A Dynamical Network View of Lyon's Vélo'v Shared Bicycle System.

Borgnat, P., & Robardet, C., & Rouquier, J.-B., & Parice, A., & Fleury, E., & Flandrin, P. (2011). Shared Bicycles in a City: A Signal processing and Data Analysis Perspective. *Advances in Complex Systems* 14, 3, 1–24.

Büttner, J. & Mlasowky, H., & Birkholz, T., & Groper, D., & Fernandez, A., G., E., & BANFI, M. (2011). Optimising bike sharing in european cities, a handbook. Tech. rep., *Intelligent Energy Europe Program (IEE, OBIS projext).* August.

De Maio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation* 12, 4, 41–56.

Dell'olio, L., & Ibeas, A., & Moura, J. L. (2011). Implementing bike-sharing systems. In *ICE - Municipal Engineer*. Vol. 164. ICE publishing, 89–101.

Dempster, A. P., & Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* B 39, 1–38.

Fraley, C. & Raftery, A. (2002). Model based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association* 97, 458, 611–631.

Froehlich, J., & Neumann, J., & Oliver, N. (2008). Measuring the pulse of the city through shared bicycle programs. In *UrbanSense08*. 16–20.

Froehlich, J., & Neumann, J., & Oliver, N. (2009). Sensing and predicting the pulse of the city through shared bicycling. In *21st International Joint Conference on Artificial Intelligence, IJCAI'09*. AAAI Press, 1420–1426.

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. *Pervasive and Mobile Computing* 6, 4, 455–466.

Lathia, N., & Saniul, A., & Capra, L. (2012). Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies* 22, 88–102.

Lin, J., & Yang, T., (2011). Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research Part E: Logistics and Transportation Review* 47, 2, 284–294.

Nair, R., & Miller-Hooks, E., & Hampshire, R. C., & Busic , A. (2012). Large-Scale Vehicle Sharing Systems: Analysis of Vélib'. *International Journal of Sustainable Transportation* 7, 1, 85–106.

Randraimanamihaga, A., & Côme, E., & Oukhellou, L., & Govaert, G., Clustering the Vélib' origin-destinations flows by means of Poisson mixture models. The 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN2013, Bruges, Belgique.

Vogel, P., & Greiser, T., & Mattfeld, D. (2011). Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia - Social and Behavioral Sciences* 20, 0, 514 – 523.

Vogel, P. & Mattfeld, D. (2011). Strategic and operational planning of bikesharing systems by data mining - a case study. In *ICCL*. Springer Berlin Heidelberg, 127–141.