Model-based count series clustering for Bike Sharing System usage mining, a case study with the Velib system of Paris

Etienne Come, Latifa Oukhellou

Model-based count series clustering for Bike Sharing System usage mining, a case study with the Vélib' system of Paris.

CÔME ETIENNE and OUKHELLOU LATIFA
Université Paris-Est, IFSTTAR, GRETTIA,
F-77447 Marne-la-Vallée, France

September 24, 2014

Abstract

Today, more and more bicycle sharing systems (BSS) are being introduced in big cities. These transportation systems generate sizable transportation data, the mining of which can reveal the underlying urban phenomena linked to city dynamics. This paper presents a statistical model to automatically analyze the trips data of a bike sharing system. The proposed solution partitions (*i.e.* cluster) the stations according to their usage profiles. To do so, count series describing the stations' usage through departure/arrival counts per hour throughout the day are built and analyzed. The model for processing these count series is based on Poisson mixtures and introduces a station scaling factor which handles the differences between the stations' global usage. Differences between weekday and weekend usage are also taken into account. This model identifies the latent factors which shape the geography of trips and the results may thus offer insights into the relationships between station neighborhood type (its amenities, its demographics, etc.) and the generated mobility patterns. In other words, the proposed method brings to light the different functions in different areas which induce specific patterns in BSS data. These potentials are demonstrated through an in-depth analysis of the results obtained on the Paris Vélib' large-scale bike sharing system.

1 Introduction

With a growing population and more and more people living in cities, as well as the increase in nuisance factors such as pollution, noise, congestion, and greenhouse gas emissions, the development of new sustainable mobility strategies in urban areas has become a necessity. Public authorities need to deploy urban mobility policies in order to organize passenger mobility differently and thus lessen the negative impact of mobility demands. One possible approach is the

promotion of soft modes of transport such as walking and cycling, which are economical, healthy, less pollutant and more equitable Pucher and Buehler [2008], Büttner et al. [2011], Dill [2009].

The implementation of Bike Sharing Systems (BSSs) is one of the urban mobility measures proposed in many cities all over the world as an additional means of sustainable intermodal transport. Over the last few years, different BSSs have been implemented in European cities, the main motivation being to provide users with free or rental bicycles especially suited for short-distance trips in urban areas, thus reducing traffic congestion, air pollution and noise. In Europe BSSs are most popular in southern European countries, where a cycling tradition does not exist. Thanks to their unquestionable success De Maio [2009], Büttner et al. [2011], more and more cities want to provide this mode of mobility in order to show that they are sustainable and modern. In France, since the implementation of the first BSS in Lyon in 2005 (it is called Vélo'v), BSSs have been launched in twenty French cities, including Paris, one of the most large-scale BSSs implemented in France (it is called Vélib').

The key to its success is having good knowledge of BSS usage and performance. This knowledge can then be transferred to other cities wishing to introduce BSSs. Several studies Froehlich et al. [2009], Borgnat et al. [2011], Vogel and Mattfeld [2011], Lathia 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 and Yang [2011], APUR [2006], which also helps sociologists and planners to understand user mobility patterns within the cities.

However, the amount of data collected on such systems is often very large. It is therefore difficult to acquire knowledge using it without the help of automatic algorithms which extract mobility patterns and give a synthetic view of the information. This paper presents one such automatic algorithm based on a new statistical model which will automatically cluster BSS stations according to their usage profile. The analysis will help us to understand the BSS station's attractiveness, with respect to city geography and demographics, by identifying different functions in different areas, which give rise to specific usage patterns in BSS data. The model proposed here therefore shares some of the objectives highlighted in Yuan et al. [2012], *i.e.*, finding functional areas in a city through the mining of mobility data (taxi journeys in this application). However, the specific nature of the transport mode analyzed here (which is mainly used for short distance trips) requires the development of a particular model more fitted to these data. The clustering of the BSS stations is closely related to the city's activities (transportation, leisure, employment) and can be helpful for a variety of applications, including urban planning and the choice of business location, as mentioned by the previous authors. In addition, the analysis of the results provided by the model provides 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 a bike sharing transport scheme, which may be useful for bike redistribution planning and for designing new BSSs.

This paper presents a dedicated model based upon count series clustering which has been developed in order to highlight mobility patterns in the BSS usage data. The model, which uses trips data to describe station usage, is a generative mixture model; an EM algorithm is derived to learn the model parameters and to perform the station clustering. The formalization of the model is general enough to take into account specific hypotheses related to the BSS case study. The proposed approach is validated through extensive investigations carried out on data collected on the Paris large-scale BSS (Vélib').

This paper is organized as follows, Section 2 presents a survey of related work in the relevant literature. Section 3 introduces the proposed statistical model based upon count series clustering. The Vélib' example is detailed in Section 4, the results are given and discussed in Section 5, followed by a general discussion and conclusion in Section 6.

2 Related work

Mobility patterns are traditionally analyzed using human and social science methodologies. The data used for these studies are collected either from sensing devices or through observational mechanisms, *e.g.* surveys. However, the emergence of information and communication technologies, as well as the advent of new observation and tracking capabilities, has boosted the availability of sizable mobility datasets. Indeed, most people carry passive urban mobile sensors including mobile phones, GPSs, etc., and leave digital traces through ticketing data of public transportation systems. All these datasets can be used to recover mobility footprints and the development of such new sensors therefore greatly benefits urban mobility studies. This leads to the emergence of a new field of research, namely urban computing. The type of dataset used and the applicative objectives addressed by this research are many and various: human mobility can be captured through GPS trajectories of vehicles or pedestrians Wangsheng et al. [2012], cell phone usage Ratti et al. [2006], as well as data related to bicycle sharing systems as is the case here. Such digital traces can then be used to extract mobility patterns as in Calabrese et al. [2013] and Yuan et al. [2012], to estimate traffic information Hofleitner et al. [2012] or to detect traffic problems Yuan et al. [2010]. Another example of such new usage of these data is the work of Zheng et al. [2011] where GPS trajectories of taxicabs have been used to detect flawed urban planning in a city. Finally, the aim of such research may also be to design new smart services, as in the work of Ma et al. [2013] who proposed a large-scale taxi ridesharing service, for example. In this instance, real-time requests sent by taxi users are taken into account in order

to generate ridesharing schedules that reduce the total travel distance significantly. More generally, the availability of these new data sources underlines the importance of the development of novel approaches based upon engineering and computer sciences. Tools for processing mobility data are needed for a better understanding of mobility patterns of travelers and goods, as well as of the use and performance of transportation systems.

For the specific case of BSSs, several requirements have motivated earlier studies from the urban computing field: improvement of existing systems, growth of knowledge on urban mobility and, more generally, developing the BSSs of tomorrow. The design of new BSSs can benefit from the experience gathered on existing systems, the analysis of which can help us to better understand their usage. The long-term goal is to be able, before and after BSS implementation, to optimize station planning in terms of urban planning, mobility needs and the redistribution capacities of the system.

One of the main issues raised by users in recent surveys is the availability of bikes: users find themselves with empty stations when renting or borrowing bikes, and full stations when returning them. Redistribution of bikes is indeed necessary in most BSSs to compensate for the uneven demand of users by relocating the bikes among the stations, thus ensuring a good quality of service of the system. This is generally performed by redistribution trucks driving around the city which move bikes between stations. Several studies address the issues related to the optimization of bike redistribution policies, including Benchimol et al. [2011], Chemla et al. [2011], Nair et al. [2012].

Other work from computer science or signal processing domains has involved the study of existing BSSs. These approaches vary according to the kind of data they use and the goal they set out to reach. The collected data on existing systems might correspond to station occupancy statistics, such as station occupancy over the day or over several time frames, but trips data could also be available *i.e.* for each trip made using the system, in which case the departure station and starting time, and the arrival station and stopping time are recorded. This last form of dataset is interesting since it provides information on the users' starting place and destination and therefore enables the construction of OD matrices indexed by time.

Using these types of datasets two main topics have been investigated by researchers, namely clustering and prediction. Whereas the aim of clustering is to identify mobility patterns in BSS usage by partitioning the stations into different clusters having a similar usage, prediction focuses on developing models able to predict the occupancy of the stations or more globally the state of the network over time.

The prediction problem has been studied by Froehlich et al. [2009], Borgnat et al. [2009], Kaltenbrunner et al. [2010], Michau et al. [2011] and Vogel et al. [2011]. In the first study, the problem of forecasting near-term station usage is addressed using Bayesian Networks, the performance of which is analyzed with respect to factors such as time of day and station activity. The same problem is addressed in Kaltenbrunner et al. [2010], Vogel et al. [2011] using a time series analysis. Borgnat et al. [2009] predicts the global rental volume using

the cyclostationarity of the temporal series while Michau et al. [2011], using a parsimonious statistical regression model, seeks to relate social, demographic and economic data of the various neighborhoods of the city with the actual number of trips made from and to the different parts of the city.

In almost all of the clustering studies carried out until now, the bicycle sharing stations are grouped according to their usage profiles, thus highlighting the relationships between time of day, location and usage. Depending on how the station usage is described and the clustering technique they use, different approaches have been proposed. The first attempt in this line of work is due to Froehlich et al. [2008], who analyzed a dataset from the Barcelona Bicing system by means of clustering techniques. The data correspond to station occupancy statistics in the form of free slots, available bikes over several time frames and other station activity statistics derived from station occupancy raw data collected every 5 minutes. The clustering is then performed using a Gaussian Mixture model estimated by an EM algorithm. In Froehlich et al. [2009] two clusterings are compared, both being performed by hierarchical aggregation. The first one uses activity statistics derived from the evolution of station occupancy while the second uses directly the number of available bicycles along the day. Other studies like Lathia et al. [2012] use similar clustering techniques and data to study the effect of changing the user-access policy in the London Barclays cycle hire scheme. The authors investigate how the change affected the system usage throughout the city via both spatial and temporal analysis of station occupancy data. As in Froehlich et al. [2009], each station is described by a time series vector which corresponds to the normalized available bicycle value of the station along the day. Each element of the feature vector is therefore equal to the number of available bicycles divided by the station size (the 95th percentile of the sums of free slots and available bikes). These time series are then smoothed using a moving average and clustered using a hierarchical agglomerative algorithm [Duda et al., 2001, see p. 552], with a cosine distance.

Other approaches closer to the one proposed here use trips data to build station usage profiles and perform the clustering on this basis. One such example is the recent study of Borgnat et al. [2013], who uses different graphs to describe the similarity of usage profiles (in terms of arrivals/departures count correlations) between pairs of stations for weekdays and weekends. The resulting four graphs (departures weekdays, arrivals weekdays, departures weekends, arrivals weekends) are then thresholded and summed together to provide a single graph that gives the global similarity in terms of usage between the stations. This final graph is then analyzed using a community detection algorithm, based on modularity optimization [see Newman, 2006, for details on modularity clustering], which provides clusters of stations with similar usage profiles. Another piece of research that uses the same type of data was proposed by Vogel and Mattfeld [2011], Vogel et al. [2011]; it aims to identify station clusters in order to better understand temporal and spatial causes of imbalances between BSS stations. The proposed methodology, based on the Geographical Business Intelligence process, was successfully applied to data collected from Vienna's BSS (Citybike Wien). It used feature vectors to describe the stations that come from

normalizing arrival and departure counts per hour, and also handled weekdays and weekends separately. Classical clustering algorithms, namely, K-means, Gaussian mixture model estimated through the EM algorithm and sequential Information-Bottleneck (sIB), were then compared.

Lastly, Borgnat et al. [2011, 2013] considers other approaches that do not partition stations with respect to their usage profiles but with respect to the way they exchange bikes. This dynamic complex network view of the problem uses graph clustering algorithms to identify communities of stations that exchange bikes in a preferential way. Using several temporal aggregation schemes enable the investigation of the dynamics of the system in particular contexts. Another line of work is also investigated in Borgnat et al. [2011], where clustering tools are used to partition the flows between stations which exhibit similar usage profiles and not the stations themselves.

In this paper we investigate the analysis of BSS systems through the clustering of their stations with respect to their usage data. Before detailing the proposed model and the estimation procedure, we position our work with respect to the related work presented above and describe the counts statistics used by the model to achieve the clustering.

3 Count series clustering

The approach undertaken here follows the line of research initiated by Froehlich et al. [2008] and pursued in Vogel et al. [2011], Borgnat et al. [2013], Lathia et al. [2012], but with a new tool tailored to fit the specificity of the data. As in the previous studies related to BSS station clustering, the method investigated here aims to identify groups of stations with similar usage profiles, but it differs from these studies on a number of points. First, it differs from the work of Froehlich et al. [2008] 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 the stations. This description of the stations is significantly more detailed than station occupancy statistics. In particular, it is able to differentiate between situations when a lot of bikes come to and leave from a station and the cases where there is no activity at the station, whereas the descriptions built from station occupancy data cannot account for such a difference. The proposed model can also deal with the differences in behavior observed during weekdays and at weekends. As noted by several authors Borgnat et al. [2013], Vogel et al. [2011], there are great differences between the two and this must be taken into account when performing the clustering. These differences are handled directly by the model proposed in this paper whereas it was handled through data preprocessing and feature construction in the two previous studies. Furthermore, the station usage count for each available day will be processed by the model and not a summary over a long period which may be affected by seasonal or meteorological factors. Lastly, the proposed model can handle the specific nature of the observations, *i.e.* that they are counts and therefore belong to $\mathbb{N}$, 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 Mclachlan and Peel [2000], Fraley and Raftery [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 Trips data and count time series

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 and possibly a type of user subscription (day/year). This last piece of information is not used in the proposed model since its nature may depend on the BSS implementation and is therefore not generic enough. These raw data can be used to derive the following counts statistics to describe station usage:

- X^{out}_{sdh}: departure count for station $s \in \{1, \ldots, S\}$ during day $d \in \{1, \ldots, D\}$ and at hour $h \in \{1, \ldots, 24\}$;
- X^{in}_{sdh}: arrival count for station $s \in \{1, \ldots, S\}$ during day $d \in \{1, \ldots, D\}$ and at hour $h \in \{1, \ldots, 24\}$.

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 $\mathbf{X}_{sd}$ describing the arrival and departure activity of station s during day d:

$$\mathbf{X}_{sd} = (X^{in}_{sd1}, \ldots, X^{in}_{sd24}, X^{out}_{sd1}, \ldots, X^{out}_{sd24}). \tag{1}$$

These activity vectors can then be arranged in a tensor (or three-way array) of size $N \times D \times T$, 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 Poisson mixture model

Since the observed data are counts, we propose to use Poisson mixtures to build the generative model. Poisson mixtures have already been used successfully in several applicative domains, and can take different forms depending on specific assumptions Rau et al. [2011], Thomas [2010], Karlis and Meligkosidou [2003], Govaert and Nadif [2010]. As in this earlier work we will consider that conditionally on the clusters the observed variables are drawn from Poisson distributions, but we will adapt the model to our needs by making further hypotheses on the model parametrization. The generative model which we propose uses two additional sets of variables. The first is a classic one corresponding to indicator

variables (denoted by Z_s) encoding the cluster membership of the stations and taking their values in $\mathcal{Z} = \left\{\{0,1\}^K : \sum_k Z_{sk} = 1\right\}$, with K the number of station clusters; these variables are not observed and must be obtained. The variables in the second set denoted by W_d are also indicator variables, but they are attached to the days and encode the differences between weekdays and weekends (which present very different usage profiles). These variables take their value in $\mathcal{W} = \left\{\{0,1\}^2 : \sum_l W_{dl} = 1\right\}$ and we consider that they are observed. Using these two sets of variables the following generative model is then assumed for the observed data:

$$\begin{aligned} Z_s &\sim \mathcal{M}(1,\pi) \\ X_{sd1} \perp\!\!\!\perp \ldots \perp\!\!\!\perp X_{sdT} &\mid \{Z_{sk}=1, W_{dl}=1\} \\ X_{sdt}|\{Z_{sk}=1,W_{dl}=1\} &\sim \mathcal{P}(\alpha_s\lambda_{klt}), \end{aligned}$$

with $\mathcal{P}(\lambda)$ the Poisson distribution of parameter λ and $\mathcal{M}(1,\pi)$ the multinomial distribution of parameter π. This generative model assumes therefore that, knowing the cluster of the station and the cluster of the day, the departure and arrival counts of each hour are independent, and that they follow a 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). For the parameters to be identifiable we must have constraints on the λ. The following constraints will ensure that the model is identifiable up to the permutation undetermination which is unavoidable in all mixture models:

$$s.t. \sum_{l,t} D_l \lambda_{klt} = DT, \forall k \in \{1,\ldots,K\}, \tag{2}$$

with $D_l = \sum_d W_{dl}$ the number of days in day cluster l. The conditional independence assumption relates this model to the naive Bayes model, and can be criticized; it is nonetheless a good first approximation. The Poisson hypothesis is natural for count data and furthermore it enables the introduction of the station scaling factor α_s Rau et al. [2011], Govaert and Nadif [2010]. 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. Using these assumptions the conditional density of an activity vector $\mathbf{x}_{sd}$ can be derived as:

$$\begin{aligned} f(\mathbf{x}_{sd}|\{Z_{sk}=1,W_{dl}=1\}) &= \prod_{t,l} p(\mathbf{x}_{sdt};\alpha_s\lambda_{klt})^{W_{dl}} \\ &= \prod_{t,l} \left(\frac{(\alpha_s\lambda_{klt})^{\mathbf{x}_{sdt}}}{\mathbf{x}_{sdt}!} \exp^{-\alpha_s\lambda_{klt}} \right)^{W_{dl}}, \end{aligned}$$

with $p(.,\lambda)$ the density of a Poisson distribution of mean λ. Parameter estimation and station clustering with such a model is performed in the classical

maximum likelihood framework. The log-likelihood must therefore be derived. It is given by:

$$L(\mathbf{\Theta}; \mathbf{X}|\mathbf{W}) = \sum_{s=1}^{S} \log \left(\sum_{k=1}^{K} \pi_k \prod_{d,t,l} p(X_{sdt}; \alpha_s \lambda_{klt})^{W_{dl}} \right) \tag{3}$$

The maximization of this quantity can be achieved by an EM-type algorithm described in the next section.

3.3 EM Algorithm

The EM algorithm Dempster et al. [1977], Mclachlan and Krishnan [1996] is a popular algorithm for maximum likelihood estimation in statistics when the problem involves missing values or latent variables. It is an iterative algorithm

ALGORITHM 1: EM algorithm to estimate the model's parameters and clustering

Input: Data $\mathbf{X}$: tensor of size $(N \times D \times T)$, W indicators of day clusters: matrix of size $(D \times 2)$, desired number of cluster K
Output: Estimated parameters $\mathbf{\Theta} = (\alpha, \lambda, \pi)$, posterior probabilities t_{sk}
Initialization;
for *each station* s *in* $\{1, \ldots, N\}$ **do**
    compute the station's scaling factor;
    $\alpha_s = \frac{1}{DT} \sum_{d,t} X_{sdt}$;
end
for *each cluster* $k \in \{1, \ldots, K\}$ **do**
    initialize $\hat{\pi}_k^{(0)}$;
end
for *each station* $s \in \{1, \ldots, N\}$, *cluster* $k \in \{1, \ldots, K\}$ *and day cluster* $l \in \{1, 2\}$ **do**
    initialize $\hat{\lambda}_{klt}^{(q)}$;
end
repeat
    E step: compute the a posteriori probabilities;
    for *each station* $s \in \{1, \ldots, N\}$ *and cluster* $k \in \{1, \ldots, K\}$ **do**
        $t_{sk} = \frac{\pi_k^{(q)} \prod_{d,t,l} p(X_{sdt}; \alpha_s \lambda_{klt}^{(q)})^{W_{dl}}}{\sum_k \pi_k^{(q)} \prod_{d,t,l} p(X_{sdt}; \alpha_s \lambda_{klt}^{(q)})^{W_{dl}}}$;
    end
    M step: update the parameters;
    for *each cluster* $k \in \{1, \ldots, K\}$ **do**
        $\hat{\pi}_k^{(q)} = \frac{1}{N} \sum_s t_{sk}$;
    end
    for *each station* $s \in \{1, \ldots, N\}$, *cluster* $k \in \{1, \ldots, K\}$ *and day cluster* $l \in \{1, 2\}$ **do**
        $\hat{\lambda}_{klt}^{(q)} = \frac{1}{\sum_s t_{sk} \alpha_s \sum_d W_{dl}} \sum_{s,d} t_{sk} W_{dl} X_{sdt}$;
    end
until *convergence*;

which alternates between maximizing a lower bound of the log-likelihood and updating the bound. This bound is classically obtained from the completed likelihood which introduces the latent variable Z:

$$Lc(\mathbf{\Theta}; \mathbf{X}, \mathbf{Z}) = \sum_{s,k} Z_{sk} \log \left(\pi_k \prod_{d,t,l} p(X_{sdt}; \alpha_s \lambda_{klt})^{W_{dl}} \right) \tag{4}$$

During the E step of the algorithm the conditional expectation of this function over Z with respect to the current parameter values is computed. This expectation will provide the lower bound of the log-likelihood which will be maximized during the M step. This expectation is given by:

$$\mathbb{E}[Lc(\mathbf{\Theta}; \mathbf{X}, \mathbf{Z}) | \mathbf{X}, \mathbf{\Theta}^{(q)}] = \sum_{s,k} t_{sk} \log \left(\pi_k \prod_{d,t,l} p(X_{sdt}; \alpha_s \lambda_{klt})^{W_{dl}} \right), \tag{5}$$

where the t_{sk} are the *a posteriori* probabilities (given the current parameters estimate $\mathbf{\Theta}^{(q)}$) of each cluster given by:

$$t_{sk} = \frac{\pi_k^{(q)} \prod_{d,t,l} p(X_{sdt}; \alpha_s^{(q)} \lambda_{klt}^{(q)})^{W_{dl}}}{\sum_k \pi_k^{(q)} \prod_{d,t,l} p(X_{sdt}; \alpha_s^{(q)} \lambda_{klt}^{(q)})^{W_{dl}}}. \tag{6}$$

These quantities are computed during the E step of the algorithm. During the M step, this expectation is maximized with respect to the parameters in order to increase the likelihood. This maximization, detailed in Appendixes A and B, leads to the following update rules:

$$\hat{\alpha}_s = \frac{1}{DT} \sum_{d,t} X_{sdt} \, , \, \hat{\pi}_k = \frac{1}{N} \sum_{s=1}^{S} t_{sk} \tag{7}$$

$$\hat{\lambda}_{klt} = \frac{1}{\sum_s t_{sk} \alpha_s \sum_d W_{dl}} \sum_{s,d} t_{sk} W_{dl} X_{sdt} \tag{8}$$

The update formulas have natural interpretations, the scale factor of station s, α_s is simply given by the average of its activity vectors along all the time frames and days. Since they do not depend on the t_{sk}, they can be computed only once. The proportions π_k are classically updated using the *a posteriori* probabilities of each cluster. The λ_{klt} are given by a weighted mean of the activity of cluster k stations in day cluster l and time frame t. Lastly the E and M steps are iterated to build an EM algorithm (see Algorithm 1) which will converge towards a local maxmimum of the log-likelihood. This algorithm was implemented in R R Core Team [2012] and the source code of this implementation is available upon request to the author.

Before exploring the results obtained with this model on a one-month trips dataset recorded in April 2011 on the Vélib' BSS, a general view of this system and some historical background are provided in the next section.

4 The Vélib' case

4.1 History

Since 2001, the city of Paris has deployed urban policies aiming to favor public transportation and soft modes of transport such as bicycles, walking, etc. Within this context, the Vélib' bike sharing system was launched in July 2007. Vélib' is operated as a concession by Cyclocity, a subsidiary company of the French advertising corporation JCDecaux. 7,000 bikes were initially distributed on 750 fixed stations. Five years ago, the Vélib' system was extended to 20,000 bikes spread out over 1,208 fixed stations, with 224,000 annual subscribers making on average 110,000 trips each day. Vélib' is a large-scale scheme, the second largest BSS in the world after a BSS launched in China. Vélib' is available mainly in Paris *intramuros*, some stations being located in the suburbs.

Vélib' offers a non stop service (24/7). Each Vélib' station is equipped with 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.

Despite the boost in bike use in Paris which followed the introduction of the BSS, the cycling modal share is still very low compared to other cities in Europe. Analyzing modal splits Büttner et al. [2011] in Paris can give hints about the local cycling culture. Cycling share is still very low in Paris (3%) but has been on the increase in the last few years. The modal share of BSS is about 2%. Public transport has an estimated modal share of 40% while car share is estimated at 21% APUR [2006]. Even if France does not have a strong cycling culture (the primary purpose of cycling is leisure), people seem to be very enthusiastic about public bike plans. Bikes are considered to be environmentally friendly by 62% of people in France APUR [2006].

4.2 General view of the system

The aim here is to obtain some general statistics to highlight global trends in Vélib' usage. The dataset used to estimate these global statistics and analyze the results of the proposed methodology corresponds to one month of trips data recorded in April 2011. This corresponds to roughly 2,500,000 trips after data cleansing. The data were cleansed by just removing trips with a duration of less than one minute and with the same station as point of departure and destination. These trips correspond to user misoperation and not to real trips.

Figure 1 shows the whole number of recorded Vélib' trips per hour and day of the week during a month, with respect to the type of subscription (day or year). It shows that Vélib' usage is closely linked not only to the hour and

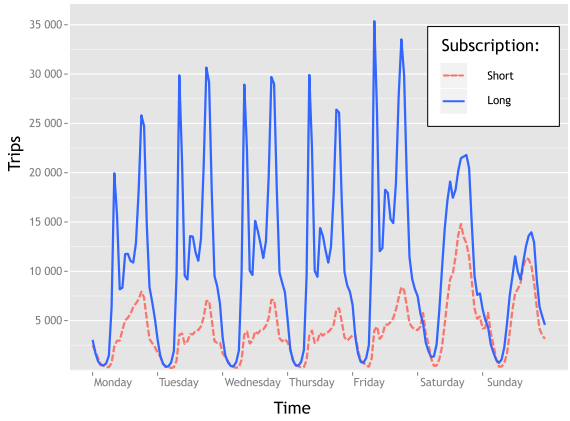

Figure 1: Number of recorded trips per hour and day of the week during April 2011 with respect to the type of subscription (long: one year, short: one day).

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.

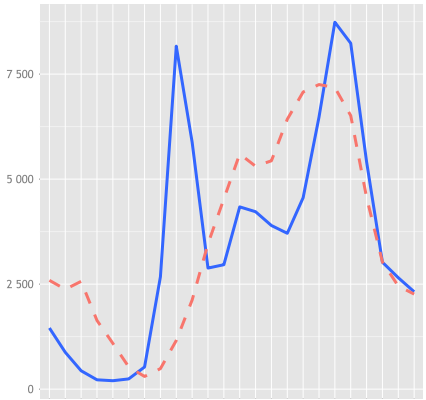

Figure 2: Average number of trips per hour during weekdays (continuous blue line) and at the weekend (dashed red line).

This figure also shows a difference in Vélib' usage during weekdays and at

the weekend. A cyclostationarity pattern can be seen in Vélib' usage during weekdays. Three peaks of weekday usage can be observed in Figure 2: the two most significant correspond to the commutes (8am and 6pm) while the third one at 12 noon corresponds to the lunch break. As can be expected, the morning peak usage disappears during the weekend, where Vélib' usage gradually increases to reach a maximum in the afternoon. It should be noted that Friday is the peak usage for weekdays. These temporal trends of BSS

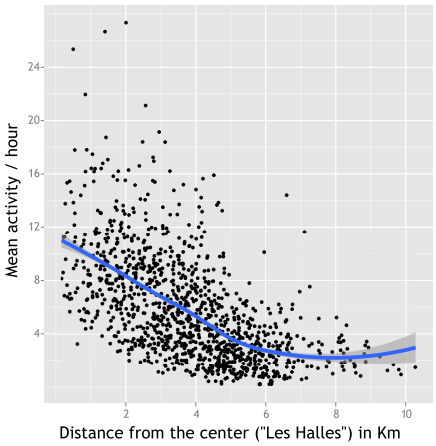

Figure 3: Average activity of stations (number of actions: departure or arrival) per hour with respect to the distance of the stations from the center of Paris ("Les Halles").

usage can provide information on the sociological characteristics of the city. Considering the study carried out by Froehlich et al. [2009] 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, spatial trends closely linked to geographical aspects of the city can also be identified. Figure 3 shows the average number of observed departures and arrivals per hour with respect to the distance from the station to the center of Paris. It is clear that the closer the station to the center of Paris, the greater this mean activity. Furthermore, the duration and distance of trips can also be used as indicators of Vélib' usage. As shown in Figure 4, half of the Vélib' trips last twelve minutes. This can be linked to the Vélib' pricing policy (free for half an hour). 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 first statistics show the global dynamic of the Vélib' system. Let us now examine the clustering results obtained using the proposed statistical model

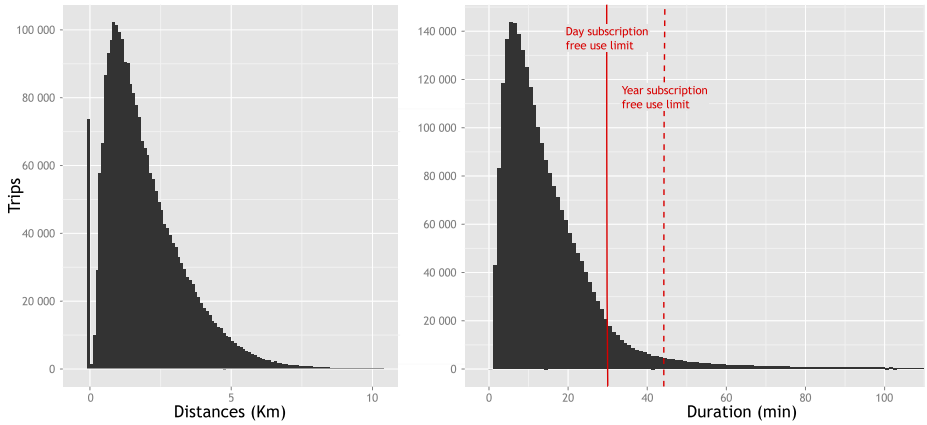

Figure 4: Histogram of trip length in kilometers (left) and of trip duration in minutes (right).

to automatically extract finer details from BSS data.

5 Clustering results and discussion

Count series for each station were derived from the cleansed raw trips data as described in section 3.1 and used for the algorithm. The final formated data correspond in the Vélib' case to a tensor of size 1185 (the number of stations) times 30 (the number of available days) times 48 (the number of counts in the station profiles). Before discussing in detail the results obtained with a particular number of clusters, it is necessary to look at the methodology and the results that we used to select this value.

5.1 Selection of the number of clusters

The proposed algorithm was tested on these data with a varying number of clusters from 1 to 25. To pick an appropriate value for the number of clusters the evolution of the model log-likelihood with respect to the number of clusters was analyzed. This curve is depicted in Figure 5 and presents an elbow around eight clusters; above this value the gain in log-likelihood is linear with respect to K. According to the elbow heuristic a value of eight seems therefore to be a good candidate for the number of clusters. We therefore choose to analyze in more detail the clustering found for $K = 8$ in this section.

5.2 Results

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. In order to give

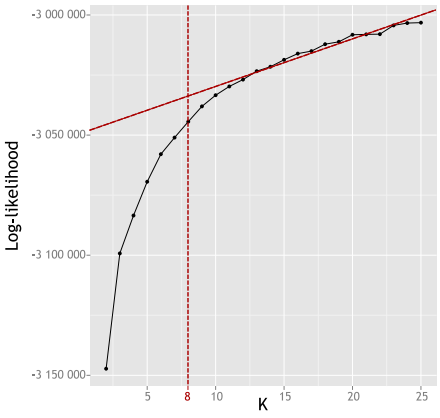

Figure 5: Evolution of the model's log-likelihood with respect to the number of clusters K.

a clear overview of these profiles we organize them according to the nature of the count (departures/arrivals), in rows, and to the day type (week/weekend), in columns.

The results for two specific clusters are presented in Figure 6. We name the first cluster "Railway stations" and the second "Parks", because these two clusters correspond to stations close to these two kinds of amenities. This can easily be seen from the corresponding maps also presented in Figure 6, 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 these two clusters and their corresponding amenity is clearly shown. The temporal profiles of these two clusters also present interesting points: the profile of the "Railway stations" cluster shows much activity around peak hours for both departures and arrivals, the other time frame being the average of the total station population. 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 6) 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.

The remaining clusters shown in Figures 9 and 10 can also be explained in terms of geography and demographics. The "Spare-time (1)" and "Spare-time (2)" clusters present high activity values during the night. The difference between these two sets of stations appears during the weekend, when the "Spare-time (2)" cluster has a higher activity. From a geographical point of view these

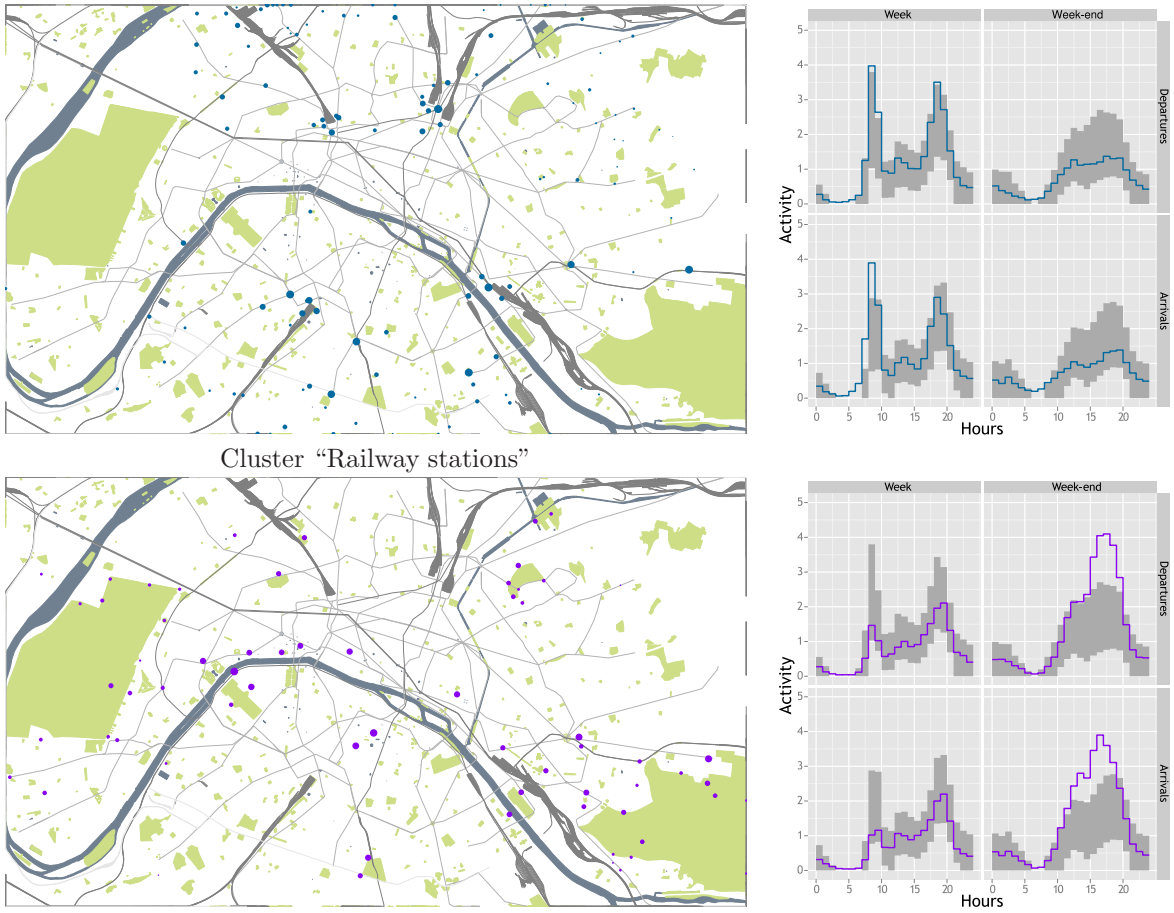

Figure 6: Maps of station positions for "Railway stations" and "Parks" clusters. The map background presents the metro and railway lines, the parks and the Seine. The areas of the dots representing the stations are proportional to the station scaling factor α_s. Each cluster map is completed with the 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.

stations also correspond to specific parts of the city that are close to important tourist places like the Eiffel Tower, the Cité des sciences et de l'industrie (science museum complex), the old historical center of Paris, etc. whereas the stations from the "Spare-time (1)" cluster are in neighborhoods with known night activities like Pigalle, Mouffetard, etc.

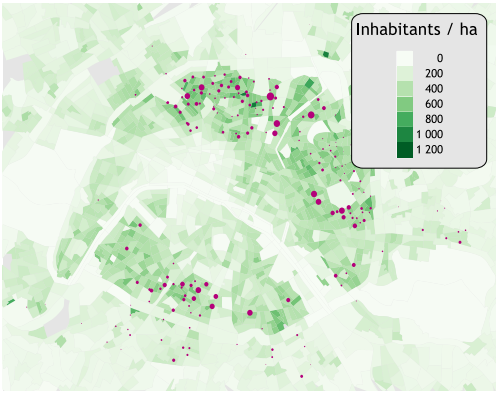

Figure 7: Map of station positions for the " Housing" cluster. The map background presents the density of inhabitants per hectare. The areas of the dots representing the stations are proportional to the station scaling factor α_s (Sources "Recensement 2008" (2008 Census), "Base permanente des équipements", Facilities Database) Insee).

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 the high population density visible in Figure 7.

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 Figure 8. During the weekend the two clusters present differences, with more activity in stations from "Employment (2)". Finally, the last "Mixed" cluster appears to be the stations having a mixed usage with an average temporal profile and no specific features.

5.3 Relations between clusters and additional socio-economic variables

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 7 and 8 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"

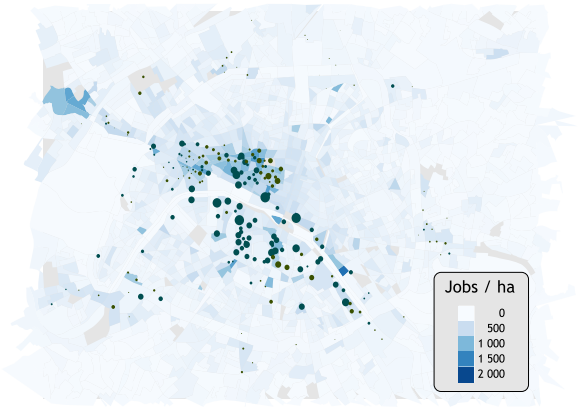

Figure 8: Map of station positions for the "Employment (1)" cluster, with light green dots, and the "Employment (2)" cluster, with light blue dots. The map background presents the density of jobs per hectare. The areas of the dots representing the stations are proportional to the station scaling factor α_s (Sources "Recensement 2008" (2008 Census), "Base permanente des équipements", Facilities Database) Insee).

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" (2008 Census), "Base permanente des équipements" (Facilities Database), Insee.

Cluster name	inhabitants/ha	jobs/ha	services/ha	shops/ha
All	162	237	4.2	3.7
"Spare-time (1)"	367	189	**6.3**	**4.4**
"Spare-time (2)"	261	322	**7.7**	**6.9**
"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
"Mixed"	301	163	3.8	2.8

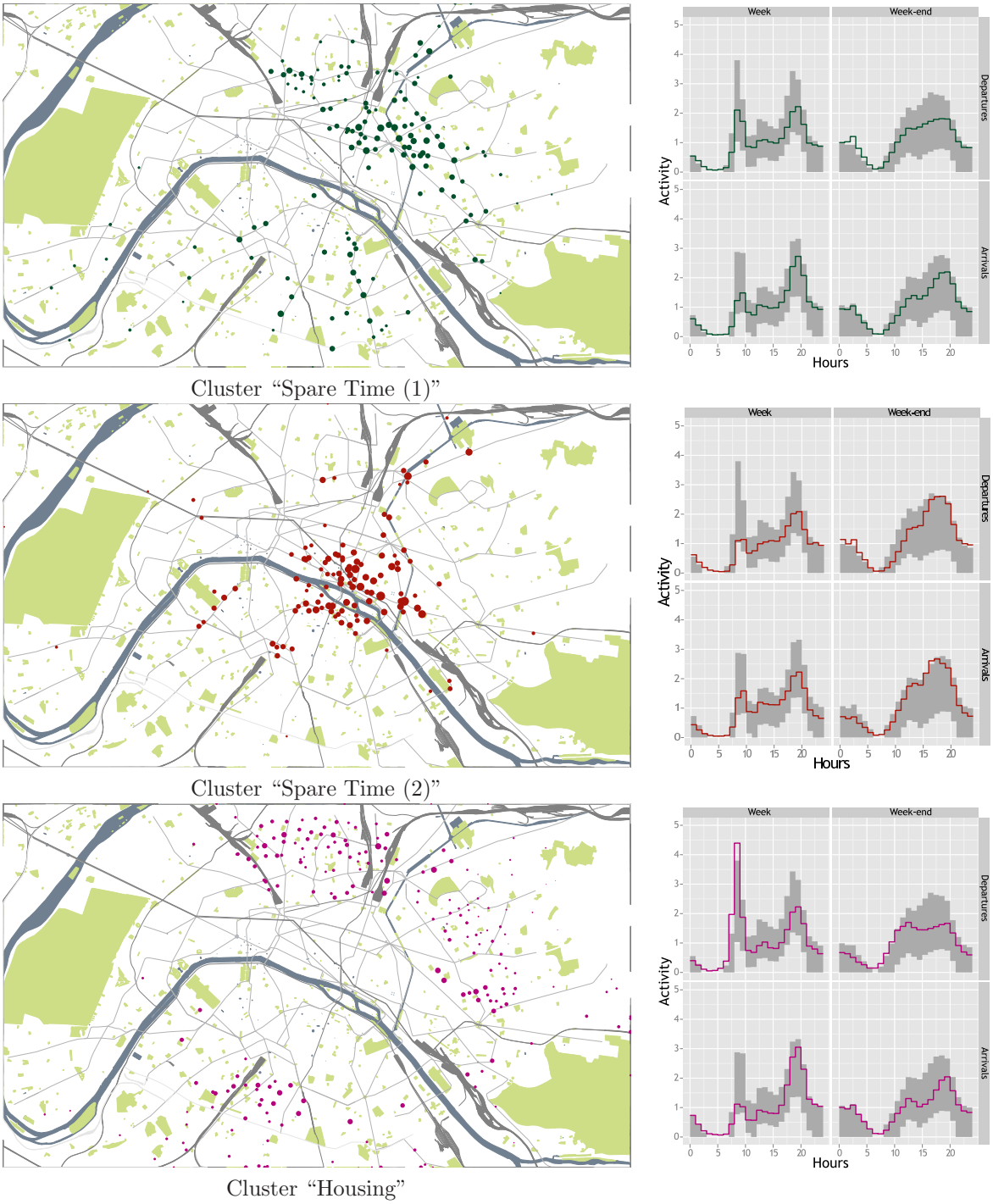

Figure 9: Maps of station positions for “Spare-time (1)”, “Spare-time (2)” and “ Housing” clusters. The map background presents the metro and railway lines, the parks and the Seine. The areas of the dots representing the stations are proportional to the station scaling factor α_s. Each cluster map is 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.

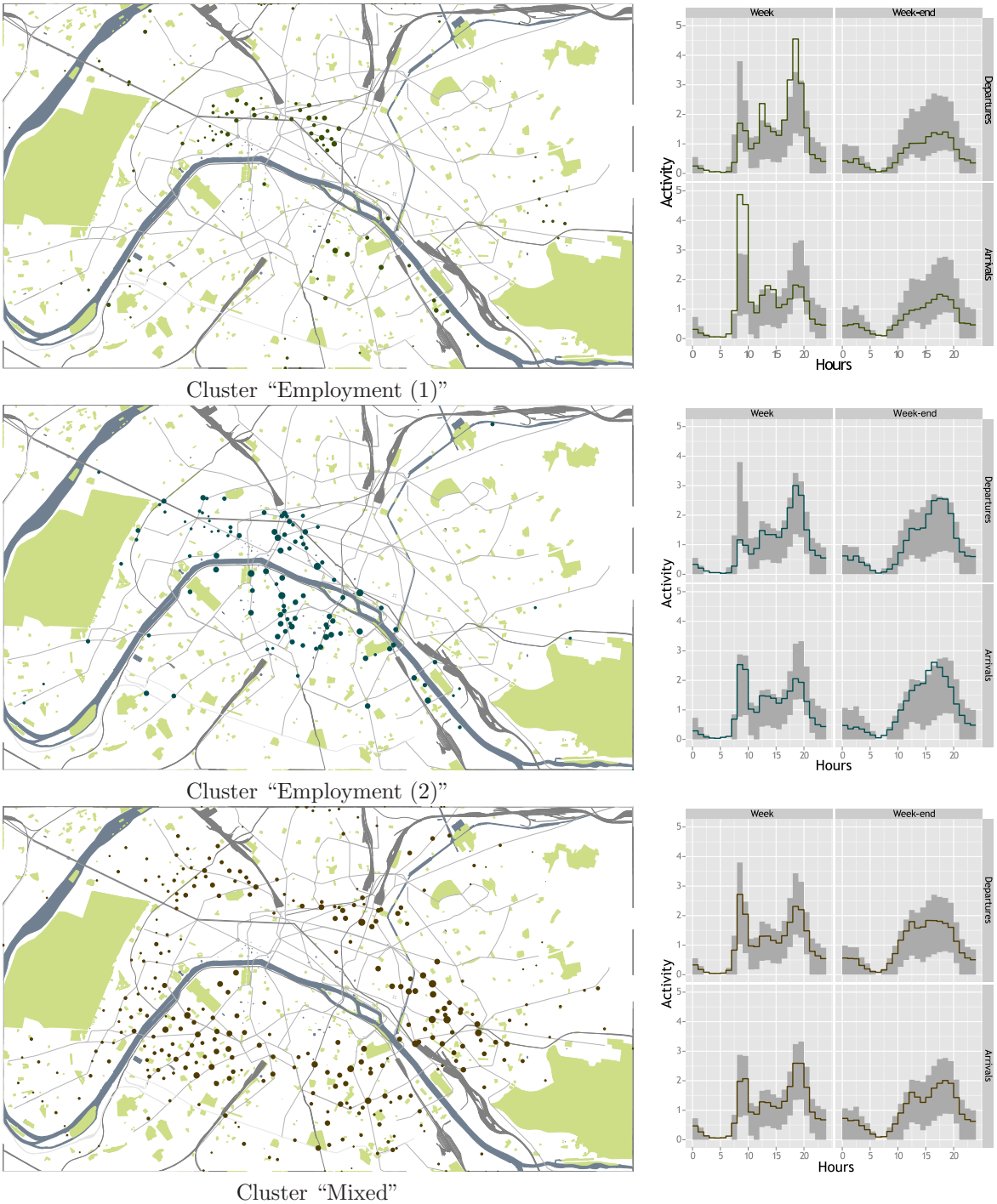

Figure 10: Maps of stations positions for "Employment (1)", "Employment (2)" and " Mixed" clusters. The map background presents the metro and railway lines, the parks and the Seine. The areas of the dots representing the stations are proportional to the station scaling factor α_s. Each cluster map is 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.

6 Conclusion

This paper has presented a new model-based clustering methodology to explore the usage statistics generated by BSSs. This model introduces a latent variable to encode station cluster membership, and an observed variable which deals with the difference in usage between weekdays and weekend. Conditionally on these variables the observed counts are assumed to be Poissonians and independent. Their intensities take into account a station scaling factor which handles the differences between the global station activities. An EM algorithm was then derived to estimate the parameters of the model. The methodology was tested to mine one month of usage data from the Paris Vélib' system. The clustering found is rich with interpretable clusters which are easily linked to the presence of certain types of amenities such as parks and railway stations, and to sociological variables like population, job and service densities.

Such a clustering tool, which is an exploratory technique may serve as a first step towards the initial applicative objectives: urban planning, firm location and BSS fleet management. The clustering model results can be particularly used to calibre simulation tools from the operational-research field with realistic values to optimize bike redistribution policies and plan new BSS systems (position, size of the stations, bike fleet size, ...). Furthermore, the crossing of the clustering results with socio-economical and geographical variables give clues on the important aspects of the city that explain the demand-variation and may therefore help to design new predictive demand models used by urban planners to position and dimension the BSS stations. Eventually, the spatial analysis of the discovered clusters may be helpful to understand a complex metropolitan and can benefit from a variety of applications, such as location choosing for a business, advertisement casting, and social recommendations Yuan et al. [2012].

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

Concerning the BSS application of the method, interesting open questions remain. For example, previous studies carried out on BSSs mainly use station occupancy statistics to describe station usage. It would be interesting to conduct a comparative study to quantify how the clustering results are affected by the type of data (occupancy statistics/trips data) we use to obtain the partition. This will require availability of both types of data collected over the same time period. Such a dataset will also be helpful to address the problem of demand estimation. This problem affects all the studies done with usage data since, by definition, they only contain information on the satisfied part of the demand. The unsatisfied part of the demand, which corresponds to the users that have left the system for bike unavailability reasons, do not leave numerical traces in the system records. Developing tools to handle this issue by trying to estimate the unsatisfied part of the demand is therefore a challenging and interesting

direction for future research.

From a methodological point of view there is also room for possible improvements. Because of the limited data size (one month) used during the experiments, further investigations involving a larger dataset (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 framework 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. Also, the naive assumption of conditional independence between the time frames could perhaps be removed with profit using approaches like Karlis and Meligkosidou [2003] and Thomas [2010]. The use of Zero inflated Poisson or Negative Binomial laws to model the observed counts would also deserve to be tested and compared with the approach proposed here. Lastly, the use of a mix-membership mixture model like LDA Blei et al. [2003] will be interesting in order to describe the mixed nature of the city neighborhoods.

Acknowledgements

The authors wish to thank François Prochasson (Ville de Paris) and Thomas Valeau (Cyclocity-JCDecaux) for providing Vélib' data. We also thank Isabelle Saint-Saens (Ifsttar) for valuable discussions. We are grateful to Samuel Sellam (Ifsttar) for his help at the beginning of the study.

References

APUR. Etude de localisation des stations de vélos en libre service. rapport. Technical Report 349, Atelier Parisien d'Urbanisme, December 2006.

M. Benchimol, P. Benchimol, B. Chappert, A. De La Taille, F. Laroche, F. Meunier, and L. Robinet. Balancing the stations of a self-service bike hire system. *RAIRO-Operations Research*, 45(1):37–61, January 2011.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

P. Borgnat, E. Fleury, C. Robardet, and A. Scherrer. Spatial analysis of dynamic movements of Vélo'v, Lyon's shared bicycle program. In Francois Kepes, editor, *European Conference on Complex Systems, ECCS'09*. Complex Systems Society, September 2009.

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

P. Borgnat, C. Robardet, P. Abry, P. Flandrin, J.B. Rouquier, and N. Tremblay. *Dynamics On and Of Complex Networks, Volume 2*, chapter A Dynam-

ical Network View of Lyon's Vélo'v Shared Bicycle System. Springer Berlin Heidelberg, 2013. URL http://liris.cnrs.fr/publis/?id=5713.

H. Büttner, J. Mlasowky, T. Birkholz, D. Groper, a.C. Fernandez, Emberger G., and M. Banfi. Optmising bike sharing in european cities, a handbook. Technical report, Intelligent Energy Europe Program (IEE, OBIS projext), August 2011.

F. Calabrese, M. Dia, G. Di Lorenzo, J. Ferreaira, and C. Ratti. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C*, 26:301–313, 2013.

D. Chemla, F. Meunier, and R. Wolfler Calvo. Balancing a bike-sharing system with multiple vehicles. In *Congrès annuel de la société Française de recherche opérationelle et d'aide à la décision, ROADEF2011*, Saint-Etienne, France, Mars 2011. Société Française de recherche opérationelle.

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

L. Dell'Olio, A. Ibeas, and J. L. Moura. Implementing bike-sharing systems. In *ICE - Municipal Engineer*, volume 164, pages 89–101. ICE publishing, 2011.

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

J. Dill. Bicycling for transportation and health: The role of infrastructure. *Journal of Public Health Policy*, 30:95–110, 2009.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, New York, 2. edition, 2001.

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

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

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

G. Govaert and M. Nadif. Latent Block Model for Contingency Table. *Communications in Statistics-Theory and Methods*, 39(3):416 – 425, January 2010.

A. Hofleitner, R. Herring, P. Abbeel, and A. M. Bayen. Learning the Dynamics of Arterial Traffic From Probe Data Using a Dynamic Bayesian Network. *IEEE Transactions on Intelligent Transportation Systems*, 13(4):1679–1693, 2012.

A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs. Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6(4):455–466, 2010.

D. Karlis and L. Meligkosidou. Model based clustering for multivariate count data. In *18th International Workshop on Statistical Modelling*, pages 211–216. Katholieke Universiteit Leuven, July 2003.

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

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

S. Ma, Y. , Zheng, and O. Wolfson. T-share: A large-scale dynamic taxi ridesharing service. In *29th IEEE International Conference on Data Engineering, ICDE 2013*, Brisbanne, Australia, April 2013. IEEE.

G. J. Mclachlan and T. Krishnan. *The EM algorithm and Extension*. Wiley, 1996.

G. J. Mclachlan and D. Peel. *Finite Mixture Models*. Wiley, 2000.

G. Michau, C. Robardet, L. Merchez, P. Jensen, P. Abry, P. Flandrin, and P. Borgnat. Peut-on attraper les utilisateurs de vélo'v au lasso ? In *XXIIe Colloque GRETSI - Traitement du Signal et des Images*, pages 46–50. GRETSI, 2011.

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

M.E. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, June 2006.

P. Pucher and R. Buehler. Making cycling irresistible: Lessons from the netherlands, denmark and germany. *Transport Reviews*, 28(4):495–528, 2008.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

C. Ratti, R. M. Pulselli, S. Williams, and D. Frenchman. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, 33(5):727–748, 2006.

Andrea Rau, Gilles Celeux, Marie-Laure Martin-Magniette, and Cathy Maugis-Rabusseau. Clustering high-throughput sequencing data with Poisson mixture models. Technical Report RR-7786, INRIA, November 2011.

Sarah Julia Thomas. *Model-based clustering for multivariate time series of counts*. PhD thesis, Rice University, 2010.

P. Vogel and D.C. Mattfeld. Strategic and operational planning of bike-sharing systems by data mining - a case study. In *ICCL*, pages 127–141. Springer Berlin Heidelberg, 2011.

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

Z. Wangsheng, L. Shijian, and P. Gang. Mining the semantics of origin-destination flows using taxi traces. In *ACM Conference on Ubiquitous Computing, UbiComp'12*, pages 943–949. ACM, 2012.

J. Yuan, Y. Zheng, W. Xie, X. Xie, and Y. Huang. T-drive: Driving directions based on taxi trajectories. In *18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS'10*, pages 99–108, New York, USA, November 2010. ACM.

J. Yuan, Y. Zheng, and X. Xie. Discovering regions of different functions in a city using human mobility and pois. In *18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'12*, pages 186–194, Beijing, China, August 2012. ACM.

Y. Zheng, Y Liu, J. Yuan, and X. Xie. Urban computing with taxicabs. In *13th ACM Conference on Ubiquitous Computing, UbiComp 2011*, pages 89–98, Beijing, China, September 2011. ACM.

APPENDIX

A Maximization of the lower bound with respect to λ_{klt}

The optimization must take into account the constraints $\sum_{l,t} D_l \lambda_{klt} = DT, \forall k \in \{1, \ldots, K\}$, with $D_l = \sum_d W_{dl}$ the number of days belonging to cluster l. The Lagrangian associated with these K equality constraints is given by:

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \sum_{s,d,t} \sum_{k,l} t_{sk} W_{dl} \left(X_{sdt} \log(\alpha_s \lambda_{klt}) - \alpha_s \lambda_{klt} \right)) + \sum_k \gamma_k (DT - \sum_{l,t} D_l \lambda_{klt}), \tag{9}$$

with γ_k the Lagrange multiplier associated with the k^{th} constraints. The partial derivative of this lower bound with respect to λ_{klt} is given by :

$$\frac{\partial \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\lambda})}{\partial \lambda_{klt}} = \sum_{s,d} t_{sk} W_{dl} \left(\frac{X_{sdt}}{\lambda_{klt}} - \alpha_s \right) - \gamma_k D_l \tag{10}$$

The Lagrange multipliers can be obtained by setting these equations to zeros as follow:

$$\begin{aligned}
\frac{\partial \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\lambda})}{\partial \lambda_{klt}} &= \sum_{s,d} t_{sk} W_{dl} \left(\frac{X_{sdt}}{\lambda_{klt}} - \alpha_s \right) - \gamma_k D_l = 0 \qquad (11) \\
&\Rightarrow \sum_{s,d} t_{sk} W_{dl} X_{sdt} - \sum_{s} t_{sk} \alpha_s D_l \lambda_{klt} - \gamma_k D_l \lambda_{klt} = 0 \\
&\Rightarrow \sum_{s,d} t_{sk} W_{dl} X_{sdt} - D_l \lambda_{klt} \left(\sum_{s} t_{sk} \alpha_s + \gamma_k \right) = 0 \\
&\Rightarrow \sum_{l,t} \left(\sum_{s,d} t_{sk} W_{dl} X_{sdt} - D_l \lambda_{klt} \left(\sum_{s} t_{sk} \alpha_s + \gamma_k \right) \right) = 0 \\
&\Rightarrow \sum_{s,d,t} t_{sk} X_{sdt} - \sum_{l,t} D_l \lambda_{klt} \left(\sum_{s} t_{sk} \alpha_s + \gamma_k \right) = 0 \\
&\Rightarrow \sum_{s,d,t} t_{sk} X_{sdt} - DT \left(\sum_{s} t_{sk} \alpha_s + \gamma_k \right) = 0 \\
&\Rightarrow \gamma_k = \frac{1}{DT} \sum_{s,d,t} t_{sk} X_{sdt} - \sum_{s} t_{sk} \alpha_s. \qquad (12)
\end{aligned}$$

The update formulas for the λ_{klt} are then given by:

$$\begin{aligned}
&\Rightarrow \sum_{s,d} t_{sk} W_{dl} X_{sdt} - D_l \lambda_{klt} \frac{1}{DT} \sum_{s,d,t} t_{sk} X_{sdt} = 0 \\
&\Rightarrow \sum_{s,d} t_{sk} W_{dl} X_{sdt} - D_l \lambda_{klt} \frac{1}{DT} \sum_{s} t_{sk} \frac{1}{DT} \sum_{d,t} X_{sdt} = 0 \\
&\Rightarrow \sum_{s,d} t_{sk} W_{dl} X_{sdt} - D_l \lambda_{klt} \sum_{s} t_{sk} \hat{\alpha}_s = 0 \\
&\Rightarrow \hat{\lambda}_{klt} = \frac{1}{\sum_s t_{sk} \hat{\alpha}_s D_l} \sum_{s,d} t_{sk} W_{dl} X_{sdt} \qquad (13)
\end{aligned}$$

B Maximization of the lower bound with respect to α_s

The partial derivative of the lower bound with respect to α_s is given by:

$$\frac{\partial \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\lambda})}{\partial \alpha_s} = \sum_{d,t} \sum_{k,l} t_{sk} W_{dl} \left(\frac{X_{sdt}}{\alpha_s} - \lambda_{klt} \right) \tag{14}$$

The update rules for these parameters are obtained by setting these derivatives to zero, this leads to:

$$\begin{aligned}
&\Rightarrow \quad \sum_{d,t} \sum_{k,l} t_{sk} W_{dl} \left(X_{sdt} - \alpha_s \lambda_{klt} \right) = 0 \\
&\Rightarrow \quad \sum_{d,t} X_{sdt} - \alpha_s \sum_{k} t_{sk} \sum_{l,t} \sum_{d} W_{dl} \lambda_{klt} = 0 \\
&\Rightarrow \quad \sum_{d,t} X_{sdt} - \alpha_s \sum_{k} t_{sk} \sum_{l,t} D_l \sum_{t} \lambda_{klt} = 0 \\
&\Rightarrow \quad \sum_{d,t} X_{sdt} - \alpha_s DT = 0 \\
&\Rightarrow \quad \hat{\alpha}_s = \frac{1}{DT} \sum_{d,t} X_{sdt}
\end{aligned} \tag{15}$$