Nonnegative Matrix Factorisation of Bike Sharing System Temporal Networks

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

In recent years, bike sharing systems have become very popular in many major cities. Thanks to the data they generate, their activity can be tracked down, giving an overall view of how human activities are spread over time and space. We propose in the present article a novel method to extract mobility patterns that occur in such large-scale transportation systems. The trips made by the users are first represented as flows between the different stations of the system, describing a network whose structure evolves over time. A decomposition technique is then proposed using non-negative matrix factorisation, to express the resulting temporal networks as a mixture of sub-networks, each of them characterising the different behaviours of users over time and space. This method is applied on the Lyon's bike sharing system, and it is emphasised that key spatio-temporal elements of urban activity are retrieved, capturing known phenomena such as commuting. This approach could be easily extended to large-scale transportation systems exhibiting a network structure, paving the way to an unsupervised modelling of mobility patterns.

1 Introduction

Many cities have developed bike sharing system (BSS) program over recent years [6], promoting environmentally friendly and healthy means of transportation to reduce traffic congestion in urban areas. These systems offer bikes that can be hired in any of the fully automated stations spread over the city, and returned at any other station. This flexibility, allowing an easy integration along traditional means of public transportation, has been decisive in the success of BSS systems in world's top major cities, e.g., in Paris with the Vélib system [15]. As any recent systems, a large amount of data is collected and is used for the exploitation and the maintenance. Their use in research context has led to new insights concerning spatial and temporal distribution of the activity [8, 1, 12, 4], as well as to study the users of such systems and their practice of shared bikes [18, 14]. In these works, traditional clustering techniques, such as K-means, are applied on extracted features, designed for a specific task, for instance to establish a typology of stations or users. Therefore, these approaches do not allow for the intrinsic nature of movements, lying on a network of stations and distributed over space and over time. In [2], a temporal network model has been proposed to take into account this spatio-temporal structure. The resulting analysis is nonetheless limited to snapshots of the network, upon which static network tools are applied to cluster stations. In the present article, we propose to use a similar temporal network representation of bike sharing systems, with the aim to to automatically extract relevant spatio-temporal components. A method to consider jointly time and space, whose preliminary results have been discussed in [10], is introduced, relying on non-negative matrix factorisation (NMF) [13]. A formal representation of bike sharing systems by temporal networks is first given, followed by a presentation of the decomposition technique used to express the resulting temporal networks as a mixture of sub-networks, characterising different spatio-temporal components. An application on the Lyon's bike sharing system, called Vélo'v, is then presented, highlighting the ability of this framework to capture significant mobility patterns occurring in a large-scale transportation system.

2 Proposed method

2.1 Representation of BSS as temporal network

A temporal network representation is used to model the flows of bikes between stations of a bike sharing system. Each station of the system is considered as a vertex of the network, and connections between stations appear over time according to the individual trips made by the users. Unlike other transportation networks, such as road or subway networks, users have the freedom to start from any station of the network, and travel to any other station (including the departure station). It implies that the network is complete, i.e., each vertex is connected to all other vertices. In practice however, some trips seldom occur, in particular between stations at the periphery who are too distant to be travelled by bike. To take into account this phenomenon, weighted edges have to be introduced, to denote the number of trips made between each pair of stations.

A formal definition is introduced in the following. Let $\mathcal{S}$ be the set of stations of the network, and $\mathcal{T}$ the continuous time interval of the experiment. $\mathcal{T}$ is first divided between I intervals of length Δ_t, which forms the set of intervals $\mathcal{I}$:

$$\mathcal{I} = \{\mathcal{I}_i\}_{k\in\{0,\dots,I-1\}} = \left\{[kL, kL+\Delta_t[\right\}_{k\in\{0,\dots,I-1\}} \text{ such that } \bigcup_{k=0}^{I-1} \mathcal{I}_k = \mathcal{I} \tag{1}$$

where L is the hop size, i.e., the shift in time between two consecutive intervals, and is comprised between 0 (excluded) and Δ_t. If L is equal to Δ_t, there is no overlap between intervals. Conversely, decreasing L increasing the overlap and then the redundancy of information. We will consider in the following $L = \Delta_t$. From there, $t \in \mathcal{T}$ belongs to the interval $\mathcal{I}_k$ if $t \in [kL, kL + \Delta_t[$. The value of Δ_t determines the temporal resolution of the resulting temporal network: if Δ_t is small (in the order of the minute), almost each trip separately will be considered, at the cost of a great variability. Conversely, a high value for Δ_t (more than 2 hours) will allow smoother transitions between each snapshot of the temporal network, but will also remove relevant information. As described in [1], $\Delta_t = 1h$ is shown to be a satisfactory compromise ensuring smoothness while preserving enough information. By convenience, $\mathcal{I}$ will refer to the set of indices of intervals $\{0, \dots, I-1\}$, instead of the intervals themselves.

A trip, defined as a user leaving a station m at time t_m and arriving at a station n at time t_n, can be formally defined as an element of the set $\mathcal{S}^2 \times \mathcal{I}^2$; $(m, n) \in \mathcal{S}^2$ is an ordered pair of stations, where m (respectively n) is the departure (respectively arrival) stations, while $(i_m, i_n) \in \mathcal{I}^2$ is an ordered pair of time intervals, such as the time of departure t_m (respectively of arrival t_n) belongs to the interval $\mathcal{I}_{i_m}$ (respectively $\mathcal{I}_{i_n}$). It is also possible to consider undirected trip by considered unordered pairs $\{i_m, i_n\}$, but we will restrain our study to the directed case.

A preliminary temporal network is then defined by $\mathcal{G}_0 = (\mathcal{V}, \mathcal{E})$, where the set of vertices $\mathcal{V}$ is the set of stations $\mathcal{S}$, and the set of edges $\mathcal{E}$ is a subset of $\mathcal{S}^2 \times I$. From a trip (m, n, t_m, t_n), the corresponding edge is obtained using either the departure or the arrival time, such that $e = (m, n, i_m) \in \mathcal{E}$ if $t_m \in \mathcal{I}_{i_m}$. It is also possible to consider an edge during all time intervals in which the movement occurs. The effects of this choice is not studied in the following, where only the departure time is used. This graph can be extended by adding weights on directed edges, by counting the number of repetitions of a given element $\mathcal{E}$. A weighted graph $\mathcal{G}_W = (\mathcal{S}, \mathcal{E}, \mathcal{W})$ is then introduced, where $\mathcal{W} = {\mathbb{N}_+}^{S^2 \times I}$, and $w_e \in \mathcal{W}$ gives the number of occurrences of trip leading to edge e in the dataset. From this definition, the resulting temporal network is defined over a potentially long period of time. As discussed in [1], it is relevant to average these intervals over a shorter period of time, typically chosen as the week, in order to reduce the fluctuations. We then introduced a temporal network $\tilde{\mathcal{G}}_{\text{Week}}$ exploiting this idea: Let $D : \mathcal{I} \to \{1, \dots, 7\}$ a mapping from an interval $i \in \mathcal{I}$ to the day of the week (1: Sunday, 2: Monday, etc.). For practical reasons, we will assume in the following that an interval does not overlap between two days. We have then $\tilde{\mathcal{G}}_{\text{Week}} = (\mathcal{S}, \mathcal{E}, \tilde{\mathcal{W}})$, where $\tilde{\mathcal{W}} = \{\tilde{w}_e\}_{e\in\mathcal{E}}$ is defined for an edge e defined on the interval i by

$$\tilde{w}_e = \frac{1}{N_i} \sum_{j\in\mathcal{I}} \mathbf{1}_{D(i)=D(j)} w_e \tag{2}$$

with $N_j = \sum_{j \in \mathcal{I}} \mathbf{1}_{D(i)=D(j)}$.

As introduced, the temporal network $\mathcal{G}$ is a sequence of networks, each of them giving a snapshot of movements made between stations during a specific time intervals. An alternative way to represent $\mathcal{G}$ is to define an adjacency tensor $\mathcal{A} \in \mathbb{N}_+^{S^2 \times I}$, where each slice $\mathbf{A_k} \in \mathbb{N}_+^{S^2}$ gives the weights of edges between all pairs of stations for the interval k. This representation allows for the use of technique such as non-negative matrix factorisation, introduced in the next section.

2.2 Decomposition of temporal adjacency matrix

Besides the wide collection of techniques to decompose networks into clusters, commonly called communities [7], only a few works have been focused on temporal networks. In [9], the temporal adjacency matrix is decomposed as a tensorial product of rank-one matrix, using non-negative tensor factorisation [3]. The low-rank decomposition favours the apparition of blocks of vertices, forming clusters In the case of BSS temporal network, this structure in communities may be relevant on static networks, as described in [2], but has less sense in the temporal case. We then propose to use non-negative matrix factorisation (NMF) for the decomposition, with the advantage that no hard constraints on extracted network structures, such as clusters, are implied.

NMF consists in approximating a non-negative data matrix $\mathbf{V} \in \mathbb{R}_+^{F \times N}$ as the product $\mathbf{WH}$ of two non-negative matrices $\mathbf{W} \in \mathbb{R}_+^{F \times K}$ and $\mathbf{H} \in \mathbb{R}_+^{K \times N}$. Often, $K < \min(F, N)$, such that $\mathbf{WH}$ is a low-rank approximation of $\mathbf{V}$. Every sample $\mathbf{v}_n$, the n-th column of $\mathbf{V}$, is thus decomposed as a linear combination of K elementary *components* or *patterns* $\mathbf{w}_1, \ldots, \mathbf{w}_K \in \mathbb{R}_+^F$, the columns of $\mathbf{W}$. The coefficients of the linear combination are given by the n-th column $\mathbf{h}_n$ of $\mathbf{H}$. The non-negativity constraint of $\mathbf{W}$ and $\mathbf{H}$ induces to express the data matrix $\mathbf{V}$ as an additive combination of K basis components, given by the columns of $\mathbf{W}$, whose mixture coefficients are given by the rows of $\mathbf{H}$. In [5] and [16], several algorithms have been proposed for the unsupervised estimation of $\mathbf{W}$ and $\mathbf{H}$ from $\mathbf{V}$, by minimising the following cost function

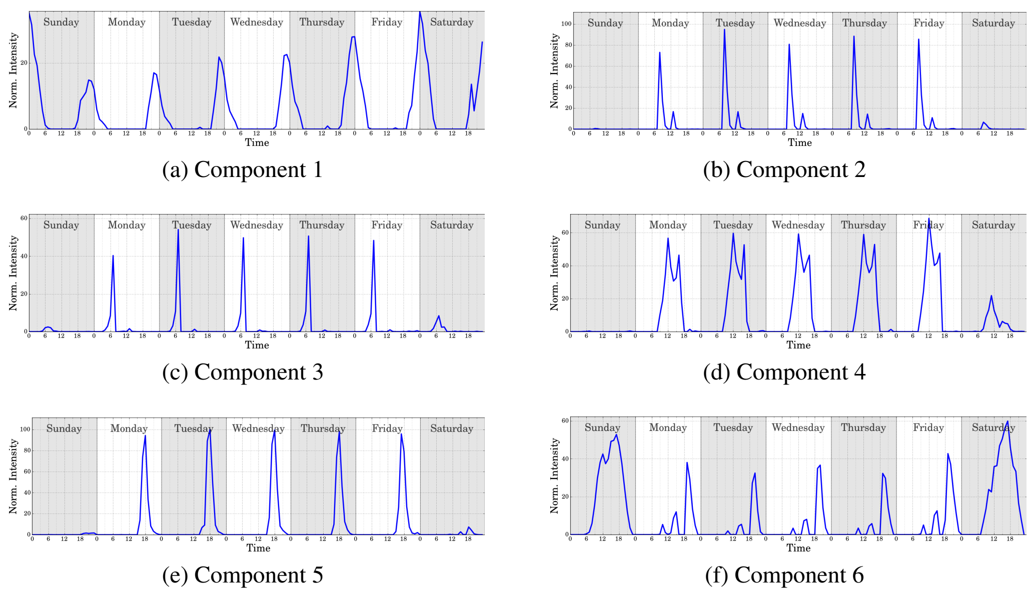
$$D_\beta(\mathbf{V} \,|\mathbf{WH}) = \sum_{fn} d_\beta(v_{fn} | [\mathbf{WH}]_{fn}) \tag{3}$$

where $d_\beta(x|y)$ is the β-divergence, whose special cases include the Euclidean distance ($\beta = 2$) and the Kullback-Leibler divergence ($\beta = 1$). Due to the non-convexity of this problem, resulting matrices are only an approximation of the best solution, and a better estimates is obtained by repeating the optimisation process 10 times and retaining the best achieved solution.

As defined in (3), NMF is defined for matrices. In our application, the temporal adjacency matrix of dimensions $S \times S \times I$ is transformed into a matrix by stacking end-to-end the columns of $\mathbf{A}_k$ for each interval k. The resulting matrix $\mathbf{V}$ is then a matrix of dimension $S^2 \times I$, and is used as input of the NMF. After minimisation of (3), the columns of matrix $\mathbf{W}$ regroups a collection of K spatial patterns, which can be unfold to form an adjacency matrix describing a static network structure between stations. For each snapshot of the temporal network, the rows of $\mathbf{H}$ give the mixture of these patterns, and then reveal the sub-structures of the activity. Before applying NMF, two parameters need to be set, namely the measure of dissimilarity, controlled by the value of β, and the number of components K. The choice of β is guided by the probabilistic model of the data we consider: entries v_{it} of the data matrix $\mathbf{V}$ describe an average number of bikes departing from a station to another one during a time interval, and may be described by a Poisson law: $v_{it} \sim \mathcal{P}(v_{it}, \sum_{k=1}^{K} w_{ik} h_{kt})$, where $\mathcal{P}(x, \lambda) = e^{-\lambda} \frac{\lambda^x}{\Gamma(x+1)}$ with $\Gamma(x+1)$ the Gamma function. As proven in [17], under the assumption that the entries of $\mathbf{V}$ are i.i.d., maximising the likelihood of this model is equivalent to minimising the Kullback-Leibler divergence, that is to say the β-divergence for $\beta = 1$. As for the number of components, there is no natural choice as it is in practice guided by the data as well as the required level of details. A low number of components is easier to interpret, but can fail to capture all the information. Conversely, a high number of components increases the diversity of spatio-temporal profiles, but makes the interpretation more complex. For this study, the number of components is arbitrarily set to 6, selected as a compromise value between complexity and interpretation.

3 Experiments on Lyon's bike sharing system

Thanks to a partnership with the "Grand Lyon" City Hall and the operator Cyclocity, all the records of the Vélo'v system[1] in Lyon, France, were made available to us for the year 2011. For this year, the system comprises 343 stations spread over the cities of Lyon and Villeurbanne. Previous studies [1] have highlighted that Vélo'v activity over the day and the week, captured through the measure of number of rentals, is not stationary over the year. We then focus our study on the spring season, as the good weather and the absence of holidays reduce the variability of data. A temporal network including trips from this period of time averaged over the week is built, and decomposed using the method described above.



(a) Component 1

(b) Component 2

(c) Component 3

(d) Component 4

(e) Component 5

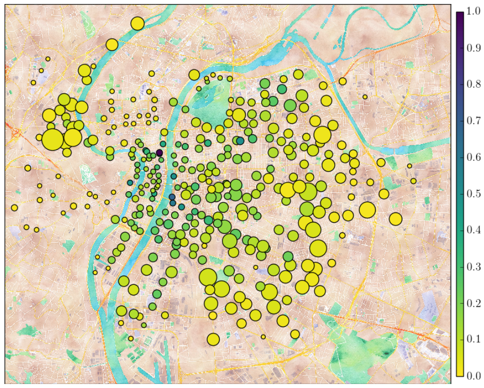
(f) Component 6

Figure 1: Activation coefficient for each component. The intensity is normalised by the highest intensity for all components.

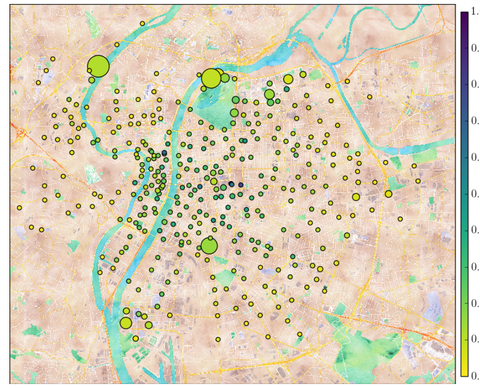
Figure 1 displays for each component the activation coefficients over the week. Temporal description of activation coefficients lead to a classification of components according two modalities, the type of day (weekdays or weekend) and the period of the day (morning, midday, afternoon, evening, night). Weekend activities is mainly caught by component 6, even if the mobility pattern represented by this component is also present in weekdays with a lower intensity. For weekdays, components 2 and 3 describe the activity in the morning, components 2 and 4 at midday, and components 4 and 5 for late afternoon / evening. It is interesting to note that the midday activity is composed of patterns present both in the morning and in the evening. We also observe small variations of the amplitude of peaks with respect to the day, and more particularly, a lower amplitude appears on Wednesday, that may be explained by the primary school schedule. This modulation is clearly visible in component 1, which spans over the whole week during the night, but with a different amplitude for each day. The intensity of peaks follows the expected intensity of the nightlife, much more stronger on Thursday, Friday and Saturday night than the other day. These extracted temporal components are consistent with the ones identified in various studies on Vélo'v, for instance in peak activities described in [1].

One of the advantages of a joint decomposition in time and space is that a mobility pattern is associated to each component, captured in the columns of the matrix $\mathbf{W}$. They are displayed in Figure 2 as static network embedded in the geographical area of Lyon and Villeurbanne. Each vertex (represented as dot) corresponds to a Vélo'v station. Colour of dots indicates the weighted degree of the corresponding vertex, i.e., the average weight of outgoing and in-going edges. This is then related to the activity of the station (number of rentals and of deposits). Size of dots depends on the ratio between in-going weighted degree (average weight of in-going edges) and outgoing weighted degree

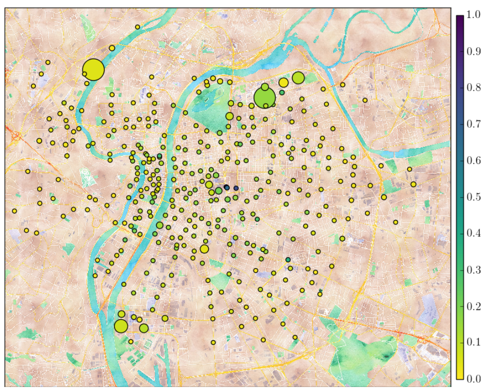
[1] http://www.velov.grandlyon.com



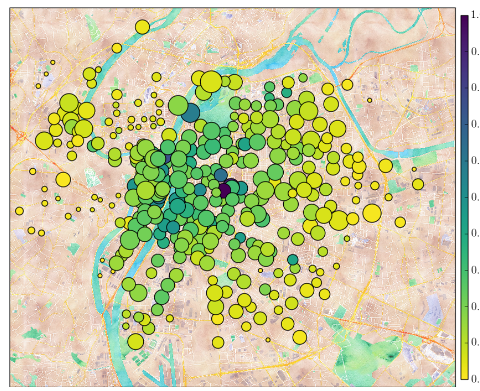
(a) Component 1



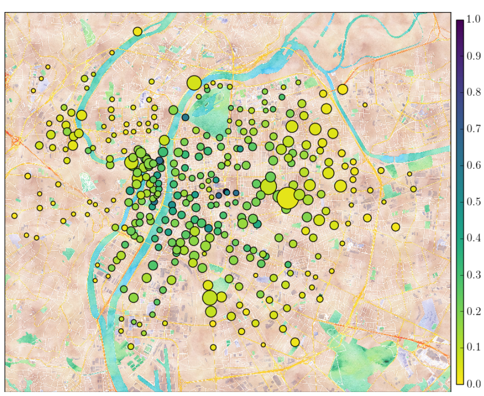
(b) Component 2



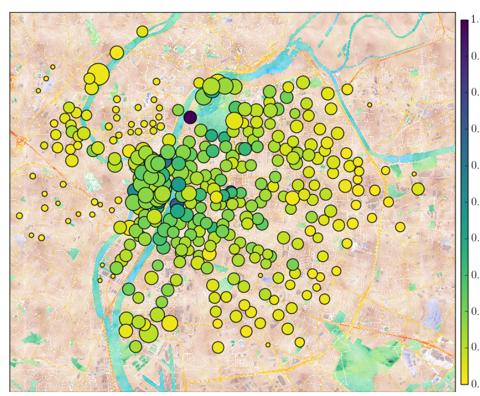
(c) Component 3



(d) Component 4



(e) Component 5



(f) Component 6

Figure 2: Representation of patterns (columns of $\mathbf{V}$ as network embedded in the geographical area of Lyon and Villeurbanne. Vertices correspond to stations. Their colour indicates their weighted degree, i.e., the average weight of outgoing and in-going edges, while their size indicates the ratio between in-going weighted degree (average weight of in-going edges) and outgoing weighted degree (average weight of outgoing edges). For sake of readability, directed edges are not displayed.

(average weight of outgoing edges). Large dots correspond to stations being filled with bikes, while small dots represent stations being emptied. For sake of readability, directed edges are not displayed.

Using this decomposition, it becomes much easier to explore the spatial distribution of flows of bikes over the city. A little knowledge of the metropolis is sufficient to interpret the structure of the extracted networks, and understand the key characteristics of the use of the Vélo'v system. First, we can note that for all components, the activity is mainly concentrated on a specific zones such as Part-Dieu district, regrouping the main railway station and business and shopping precincts, as well as the Presqu'île, the strip of land between the two rivers considered as the nerve centre of the city of Lyon. Second, stations located in hilly zones, such as the west side of the city or above the Presqu'île, are likely to be empty as going up is tiresome when cycling (this phenomenon is particularly visible in component 6, where only stations from these areas are represented by small dots).

An joint interpretation between the mobility patterns and their activation coefficients gives some hints on some behaviours: for the weekdays in the morning, movements are mainly from residential areas to employment or educational areas, as indicated by the small dots in the outskirts of the city and the city centre regrouping many residential areas, and larger dots close to universities and business precincts in components 2 and 3. Conversely, flows in the opposite direction appear in components 4 and 5, representing spatial distribution of the activity in the late afternoon. It is interesting to note that these patterns, highlighting the use of BSS for commuting, also appear at midday, albeit less strongly. Other types of behaviours arise at different period of time, outlining different uses of the system: in component 1, night activity is concentrated around around theatres, restaurants and bars. The stations close to these areas of interest tend to get empty, in favour of the ones at the outskirts of the city, located in residential areas. This component then suggests that bike-sharing system might be an alternative mean of transportation at night, when public transportation has closed.

Table 1: Classification of NMF components

#	Period of time	Space distribution	Direction
1	Evening and night	Nightlife districts	City outskirts
2	Weekdays - Morning and midday	Residential and business areas	Universities / schools
3	Weekdays - Early morning	Residential and business areas	Universities
4	Weekdays - Midday and afternoon	City centre / Railway station	City centre
5	Weekdays - Late afternoon	Residential and business areas	Residential areas
6	Weekend - Daytime	Recreational areas	City centre

A summary of the interpretation of components is given in Table 1. This example showed that even without prior information on bike-sharing systems, it is possible to achieve a relevant spatio-temporal characterisation on the main behaviours. Further investigations should be done to precisely characterise spatial components, with respect for instance to socio-economic variables.

4 Conclusion

In this paper, we proposed a novel framework to explore large datasets of trips made using bike sharing system, through a joint analysis of temporal and spatial dynamics. The method, based on a temporal network representation and a decomposition technique, automatically extracts the most significant spatio-temporal patterns. It has been successfully applied on real-world data from the bike-sharing system in Lyon: The extracted components are consistent with the current knowledge about the system studied through extensive socio-economical surveys, confirming the relevance of the proposed approach to easily access to a simplified view of how the system works, without any supervision. These components could be used to gain new insights about how transportation systems involving fleets of vehicles is spatially and temporally structured, and therefore help to build predictive model of the traffic flows over a transportation network.

References

[1] P. Borgnat, P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, and E. Fleury. Shared bicycles in a city: a signal processing and data analysis perspective. *Adv. Complex Syst.*, 14(03):415–438, 2011.

[2] P. Borgnat, C. Robardet, P. Abry, P. Flandrin, J.-B. Rouquier, and N. Tremblay. A dynamical network view of lyon's vélo'v shared bicycle system. In *Dynamics On and Of Complex Networks, Volume 2*, Modeling and Simulation in Science, Engineering and Technology, pages 267–284. Springer New York, 2013.

[3] A. Cichocki. *Nonnegative matrix and tensor factorizations applications to exploratory multi-way data analysis and blind source separation*. John Wiley, Chichester, U.K., 2009.

[4] E. Côme and L. Oukhellou. Model-based count series clustering for bike sharing system usage mining: A case study with the vélib'system of paris. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):39, 2014.

[5] C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the β-divergence. *Neural Computation*, 23(9):2421–2456, 2011.

[6] E. Fishman, S. Washington, and N. Haworth. Bike Share: A Synthesis of the Literature. *Transport Reviews*, (ahead-of-print):1–18, 2013.

[7] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010. bibtex: Fortunato2010.

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

[9] L. Gauvin, A. Panisson, and C. Cattuto. Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PLoS ONE*, 9(1):e86028, Jan. 2014.

[10] R. Hamon, P. Borgnat, C. Févotte, P. Flandrin, and C. Robardet. Factorisation de réseaux temporels : étude des rythmes hebdomadaires du système Vélo'v. In *Colloque GRETSI 2015*, Lyon, France, Sept. 2015.

[11] R. Hamon, P. Borgnat, P. Flandrin, and C. Robardet. Networks as signals, with an application to a bike sharing system. In *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, pages 611–614. IEEE, 2013.

[12] N. Lathia, S. Ahmed, 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.

[13] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, Oct. 1999.

[14] C. Morency, M. Trepanier, and F. Godefroy. Insight into montreal's bikesharing system. In *Transportation Research Board 90th Annual Meeting*, number 11-1238, 2011.

[15] 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, Jan. 2013.

[16] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama. Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with β-divergence. *Proc. of MLSP*, 10:1, 2010.

[17] T. Virtanen, A. T. Cemgil, and S. Godsill. Bayesian extensions to non-negative matrix factorisation for audio signal modelling. In *Proc. of ICASSP*, pages 1825–1828. IEEE, 2008.

[18] M. Vogel, R. Hamon, G. Lozenguez, L. Merchez, P. Abry, J. Barnier, P. Borgnat, P. Flandrin, I. Mallon, and C. Robardet. From bicycle sharing system movements to users: a typology of vélo'v cyclists in lyon based on large-scale behavioural dataset. *Journal of Transport Geography*, 41:280–291, 2014.