A Dynamical Network View of Lyon's Vélo'v Shared Bicycle System

Pierre Borgnat*[1], Céline Robardet*[2], Patrice Abry[1], Patrick Flandrin[1], Jean-Baptiste Rouquier[3], and Nicolas Tremblay[1]

[1] Laboratoire de Physique, ENS de Lyon, CNRS, 46 allée d'Italie 69007 Lyon, France. {Pierre.Borgnat; Patrice.Abry; Patrick.Flandrin; Nicolas.Tremblay}@ens-lyon.fr
[2] LIRIS, INSA-Lyon, CNRS, Bâtiment Blaise Pascal, 69621 Villeurbanne, France. Celine.Robardet@insa-lyon.fr
[3] Eonos, 53 rue de la Boétie, 75008 Paris, France. jrouquie@gmail.com
* Corresponding Authors

Summary. Community shared bicycle systems are an instance of public transportation systems that provide digital footprints of all the movements made using this system. The completeness of such dataset allows for their study using a complex system point of view. We discuss in this chapter how Lyon's shared bicycle system, called Vélo'v, can be seen as a dynamical complex network, and how using community detection methods gives interesting results thanks to the aggregation in space and/or time that communities propose.

1 Introduction

A current challenge in the study of complex networks is to devise and validate methods that are not limited to static or growing complex networks, but that are adaptable to dynamical aspects of complex networks. Among all the complex networks offered by human activities, the transportation networks raise some challenge by themselves. The classical methods in transportation research rely on household surveys of movements, or on the direct observation of movements by sampling some places in the city. However, thanks to the developments of Information Technologies, more and more systems of transportation offer digital footprints of population movements and enable their study as complex systems: public transportations in subway (thanks to individual digital subscription cards) [1, 2], railways [3, 4], or air transportation thanks to database of flights [5], with applications ranging from urban planning to epidemiology [6]. The issues we now face are to understand how to cope with these new, large-scale datasets, and which methods are useful so as to obtain some insight on the dynamics of people's moves.

We will review in this chapter the progresses made in this direction in the particular case of shared bicycle systems. As modern cities are more and more overcrowded by cars in their centers, alternative means of transportation have been developed. Among them, shared bicycle programs have gained a renewal in interest in the past 10 years, thanks to the possibility of using fully automated rental systems that are available 24 hours a day, 7 days a week. This innovation led to the growth of different systems in many of the major cities in Europe, e.g., Vélib' in Paris [7, 8] and Vélo'v in Lyon [9, 10], Bicing in Barcelona [11, 12], OYBike in London [13], Bicikelj in Ljubljana [14] to cite only a few that have attracted quantitive analyses of their data. Indeed, the full automation of the systems creates digital footprints which provide a complete view of all the trips made with these bicycles. Such datasets would have been impossible to obtain before, although it gives, by definition, only a partial view of all the movements done in a city. Despite that, these data are interesting to test ideas related to the study of dynamical networks.

The contribution of this chapter will be first to review previous studies on shared bicycle systems from the point-of-view of information sciences. We discuss some of the general features that are displayed by these systems. Then, we focus on Vélo'v – the shared bicycle system that is deployed in the city center of Lyon, France's second largest urban community, and that was the largest scale system of this type when launched in May, 2005. Our main purpose is to discuss an adaptation to the case of dynamical networks, of the methods of aggregation in complex networks that rely on the notion of communities. Here, the Vélo'v network is inherently dynamical, in that the network appears only because there are bicycle trips connecting the stations.

The chapter is organized as follows. Section 2 recalls general facts about shared bicycle systems, about their global dynamics (cycle over the week and nonstationarity over several months). Section 3 shows how such a system can be seen as an instance of a dynamical complex network. In Section 4 is discussed how a spatial aggregation of the Vélo'v complex network can be obtained by adapting methods for community detection. Then, Section 5 proposes a typology of the dynamics in the Vélo'v network by estimating a similarity graph and detecting communities. We conclude in Section 6.

2 General dynamics of Vélo'v system

The Vélo'v system is basically a bicycle rental system consisting of 343 automated stations each composed of several stands from where a Vélo'v bicycle can be taken or put back at the end of the trip. The stations are spread out in the city with the objective that, in the center, the distance to a station is no more than 300m. There are around 4000 Vélo'v bicycles available at these stands, and, as the system is automated, a bicycle can be returned at any free stand, usually in a different station that the one it was taken from. The functioning is all automated and works 24 hours a day, all year long. People use long-term or short-term (obtained with a credit card) subscription card to rent a bicycle. This made possible the collection of all the data pertaining to the trips made with these bicycles, without sampling as was the rule previously in transportation studies. This also made possible the display of the state of each station through the web [15].

Thanks to JCDecaux – Cyclocity and the Grand Lyon, we had access to the anonymized version of the trips made with Vélo'v. The dataset consists of a log of all the rentals, with the station and the time of departure, and the destination station and the time of arrival. For privacy concerns, there is no information about users. Nevertheless, about the trips, the data is complete.

This dataset of the trips was instrumental in studying the general dynamics of the system. A first point was to understand how the size of the Vélo'v system and its popularity increased along the first years of the program. As shown in [10], there has been a clear nonstationary growth of its popularity, in term of long-term subscribers and in term of number of trips made (currently the average is more than 20 000 trips per day). Moreover the system is still planned to grow [16]. Also, the number of rentals reveal a general multiplicative trend at the scale of several months that can be estimated using recent data-driven statistical methods [17].

A basic feature of transportation system studies is to discover its time pattern of use: When is it really used? What are the peak hours? Is it a means of transportation for the ordinary week-days or the week-ends? Using periodic averaging over the week combined to detrending of the nonstationary behavior, we were able to estimate the mean pattern of total rentals along the week, as shown in Figure 1 [10, 18, 19]. It reveals that Vélo'v is used first for ordinary transport on working days. Its activity peaks are during the morning, the lunch time and the evening and this is characteristic of a system used to go to work and then to come back, with the lunch break in the middle. In a French city such as Lyon, usual schedule for work is to begin around 9am, and finish between 5pm to 7pm. The lunch break is spread out between 11:30am and 2pm. Students at universities take classes between (at most) 8am and 6pm with a lunch break at the same time as workers. Finally, shops are usually open from 9am or 10am to around 7pm. All the peaks of activity detected are thus compatible with this lifestyle. Also all the week-days are similar in that respect, as shown on Figure 1(b). However, a second type of use exists during the week-ends, as seen on this second plot. The peaks are less sharp and more spread out around noon and during the afternoon. This is compatible with leisure activities. Finally, one can see that just after midnight each day, a small bump that can be related to the closure of the public transports right after midnight (this even causes local maxima on Friday and Saturday nights). This type of pattern is reminiscent to uses of ordinary public transports. An exception is the nocturnal activity when public transports are closed. Also, it is compatible with similar patterns for shared bicycle systems in other cities: Paris [7, 13], Barcelona [11, 12], Ljubljana [14].

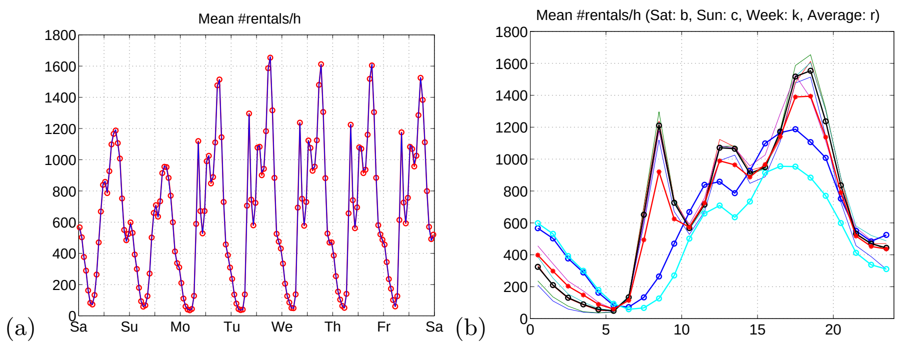

Fig. 1. Weekly pattern of the number of rentals made per hour, summed over all the Vélo'v stations: (a) pattern for the whole week; (b) superposition of individual day patterns in the week. The average over the 5 ordinary week-days from Monday to Friday is the thick black curve with circles. The five curves in thin lines that look alike are each for one week-day: Monday in blue; Tuesday in dark green; Wednesday in red; Thursday in cyan; Friday in purple. Saturday (thick blue curve with circles) and Sunday (thick cyan curve with circles) reveal a different structure, especially for the peak hours. The average over all days is the thick red curves and is dominated by the week-days. That is a reason to study separately week-days and week-ends.

The prediction of the number of rentals at a given day and time was addressed in [10, 18, 19] using statistical time-series analysis. The global number of rentals, summed over the city, can be predicted on an hourly basis if one takes into account several important factors: the weather (temperature and rain), the holiday periods, and the existence of a correlation over one hour. The first two features (temperature and rain) account for most of the non-stationary evolution over the year of the average rentals. When zooming in at finer time scales, the one-hour correlation reflects that one decides to use a bicycle depending on the conditions seen during the previous minutes, and a rain condition affects the decision on this hourly scale – as it could be expected.

Finally, several empirical studies of data of community shared bicycles revealed various features such as the most frequent paths taken by these bicycles or the advantages of using a bicycle as compared to a car [14, 20], the distribution of durations and lengths of the trips [10], or the usual speed of the bikers depending on the time of the day [10, 20]. All these features could help to design an agent-based model of shared bicycle systems. Note that their properties should be heterogeneous because the distributions found are usually with long tails. For instance, though the median duration of a trip with Vélo'v is 11 minutes and the average is little bit less than 30 minutes, there are rentals lasting more than 2 hours, and the distribution has a tail that is roughly a power-law [10].

All these studies give a good empirical description of the global features of these systems along time, or along space. However, the challenge is to design a joint analysis in space and time.

3 Vélo'v as a complex network

From individual trips to a network of stations.

As already proposed in [10, 21], it is worthwhile to study the Vélo'v system from a complex network perspective. Complex networks are usually employed when studying real-world dataset including some relational properties. In the context of shared bicycle systems, a network arises when looking at stations as nodes of a complex network. Let us define $\mathcal{N}$ the set of stations. Each node $n \in \mathcal{N}$ is at a specific geographical place in the city. Going from one station to any other is theoretically possible and around half of all theoretically possible trips have been done at least once. However some preferred trips appear in the data, and the they are not necessarily local in space. That is the reason why a representation of the Vélo'v system as a weighted network makes sense.

Our proposition is that the relation between the nodes of the Vélo'v network –the stations– is created by the trips made from one station to another. The greater the number of trips made, the more linked the stations are. Let us define $\mathcal{D} = \{(n, m, \tau)\}$ as the set of individual trips going from station $n \in \mathcal{N}$ to station $m \in \mathcal{N}$ at time τ. The Vélo'v network is defined as $\mathcal{G} = (\mathcal{N}, \mathcal{E}, T)$ where the set of possible edges is $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ and T is a function defining a weighted adjacency matrix varying in time. Let us define $\mathcal{T}$ as a set containing the times t of interest, and $\mathcal{S}$ a set of timescales Δ for aggregation. The function $T : \mathcal{E} \times \mathcal{T} \times \mathcal{S} \to \mathbb{N}$ is obtained by the following equation

$$T[n, m](t, \Delta) = \# \{(n, m, \tau) \in \mathcal{D} \mid t \leq \tau < t + \Delta\} \tag{1}$$

where $\#$ is the cardinal of the set. The result $T[n, m](t, \Delta)$ can be seen as an adjacency matrix of the weighted, directed graph, that represents a snapshot of the Vélo'v network. On each edge, the weight is then the number of bicycles going from the station n to station m between times t and $t + \Delta$. More generally, the network $\mathcal{G}$ is a dynamical network and issues arise when we need to deal both with its spatial nature (the nodes and the edges) and with its time evolution as obtained when varying t or the timescale Δ.

In the following of the article, we follow the same ideas as in [10, 21], but we propose a more unified and streamlined approach to study both space and time aggregation for the Vélo'v complex network. Other general ideas on how to cope with time evolving networks can be found in other chapters of this book, and in the review [22].

Timescales and aggregation in time of Vélo'v networks.

First we take into account the cyclic nature of the network: the same pattern repeats itself each week and we estimate T using periodical averaging. For that, let us decide on a timescale Δ and then define a period $P = p\Delta,\ p \in \mathbb{N}$. Let us set $\mathcal{T}_P = \{k\Delta \mod P; k \in \{0, ..., p-1\}\}$ so that when applying eq. (1), the first interval in which the trips are counted is: $[0 \ \ \Delta] \bmod P$. The periodical estimation of the Vélo'v network is obtained by dividing T applied on this choice of $\mathcal{E} \times \mathcal{T}_P \times \mathcal{S}$, by the number of periods P in the data. As the main period in the data is the week [19], we let P be equal to 1 week in the following.

Usually, a specific value for time aggregation Δ is used. However, varying the timescales Δ for the analysis would be of potential interest, e.g. as it was done for computer networks analysis [23]. For complex networks, the importance of considering different time window lengths for aggregating networks, or the simpler idea that the analyses depend on the observation time, have already been put forward in various contexts: to study cattle mobility [24], communication network [25], or for information spreading in such network [26] to cite a few relevant references. For Vélo'v, given that the average duration of a rental is less than half an hour, aggregating in time over duration Δ larger than that smooths out most of the fluctuations due to individual trips. We classically use $\Delta = 1$h (as in Fig. 1) or 2h. For simplicity, $P = 1$ week and Δ=2h for all the results displayed hereafter.

First, it is possible to reduce the dimension in time of this evolving network by using a Principle Component Analysis in time, as in [10, 21]. It turns out that the principal components display peaks in their time evolution that correspond exactly to the different peaks already commented for the global number of rentals, as shown in Fig. 1. In the following, we will keep in the the dynamical adjacency matrix $T[n,m](t,\Delta)$, only the 19 peaks of activity in time, as given by the global behavior as well as the principle components: every ordinary days around 8am, 12am and 5pm, and each of the two week-end days around 12am and 4pm. We note $\mathcal{T}_P^*$ this set of peak moments and $\#\mathcal{T}_P^* = 19$. The representation of the series of snaphots of the Vélo'v network at peak activity is then $T[n,m](t,\Delta)$ with $t \in \mathcal{T}_P^*$. Another reason for keeping these 19 peaks is that classical approaches in transportation research are often focusing on the activity peaks and doing the same way will help future comparison of resuls.

Second, at large time scales, there are several manners to aggregate the network. Classical aggregation is to sum over time the different snapshots and this focuses on the strong exchanges between stations at the scale of the aggregation time (typically the week here). In a sense, it can be viewed as going from one timescale to a larger one (hence giving a crude version of multiresolution analysis). Using the 19 peak activity times $\mathcal{T}_P^*$, an aggregated view over the whole week is obtained as $\sum_{t\in\mathcal{T}_P^*} T[n,m](t,\Delta) \stackrel{\wedge}{=} < T[n,m] >_{\mathcal{T}_P^*}$ where $< \cdots >_{\mathcal{T}_P^*}$ stands for this particular aggregation in time. A question in

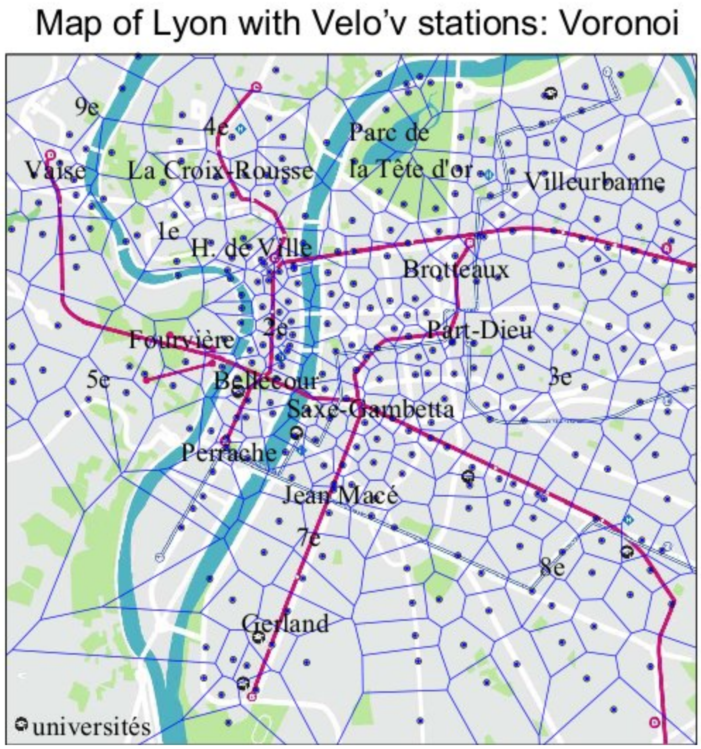

Fig. 2. Map of the cities of Lyon and Villeurbanne with the Vélo'v stations and their Voronoi diagram [27]. Each dot is a Vél'ov station surrounded by its own Voronoi cell. One sees that the city is well covered, with a higher density of stations in the center. Major roads are in white; main public transport lines are in red (subways) and grey (tramways). Parks are in green and the two rivers are in blue (the Saône on the west side, the Rhône on the east side). Names of the different parts of the city are given, as well as names of main hubs of transportations: Part-Dieu, Perrache for the main train stations; Vaise and Jean Macé for secondary train stations; Bellecour, Hôtel de Ville, Brotteaux (including Charpennes), Saxe-Gambetta for other important hubs of the public transport system. Gerland, Croix-Rousse and Villeurbanne are other parts of the city that will be discussed afterwards. Finally, the locations of the downtown university campuses are shown.

the following is how the different snapshots are similar to, or different from the aggregated network over the week.

4 Aggregation in space for the Vélo'v network

The series of snapshots of graphs convey detailed information, yet this is too much information for modeling. However, aggregating over all the nodes as done in Section 2 does not give enough details. We would like to aggregate on intermediate scales for the nodes in the network. There are two classical approaches to aggregation in a network: find clusters of nodes that are strongly linked together, also called communities [28], or use multiscale harmonic decomposition over a graph (e.g., [29]). Here, we explore the spatial aggregation that is obtained by looking at communities of stations. So as to compare with the urban organization of the city, a map of the city is in Fig. 2 that shows the main lines of transportations and provides the places and names of the most important hubs for public transportation. By referring to this map, the reader will follow with greater ease the comments about the spatial aggregation proposed by community aggregation in this section.

Aggregation of network by communities.

At a given timescale and instant, we propose to aggregate the network in space over its communities of nodes. A community is often defined as a subset of nodes that are strongly linked together inside the network (see the review in [28]). We adopt the modularity as a metrics to find communities. Modularity was first proposed in [30] and extended in [31] to the case of directed networks. Assume that t and Δ are set to specific values and use the adjacency matrix $T[n,m](t,\Delta)$ obtained that way. Modularity is defined as:

$$Q = \frac{1}{2W} \sum_{\{n,m\} \in \mathcal{N} \times \mathcal{N}} \left[T[n,m] - \frac{\sum_{j \neq n} T[j,n] \cdot \sum_{k \neq m} T[m,k]}{2W} \right] \delta_{c_n, c_m} \quad (2)$$

where $W = \sum_{n,m} T[n,m]$ is the total weight of the network and c_n is the partition index of the group in which n is. Modularity is a number between -1 and $+1$. If there is a community structure in the graph and the index c_n reflects this structure (by taking a different value for each community), Q should be large, typically larger than 0.4. Conversely, if one finds a partition index c_n for which Q is large enough, it tells that there are communities of nodes. As a consequence, finding communities is possible by maximizing Q over the set of $\{c_n, n \in \mathcal{N}\}$ of possible partitions of nodes. However, this task is hard: the complete maximization is NP-complete [32] and many approximations such as the one in [33], have a tendency to propose too big communities. In this work, we use the fast, hierarchical and greedy algorithm proposed in [34] (called the Louvain algorithm), as a simple way to find relevant communities. It is reviewed in [28] that modularity is a good metrics to find communities and that this algorithm works correctly as compared to other methods. Nevertheless, the results shown hereafter are not specific of this choice of algorithm to find communities in a network, and for instance the *infomap* algorithm of [35] would work as well.

Communities in Vélo'v networks.

First, the method is applied to the time-aggregated network $< T[n,m] >_{\mathcal{T}_P^*}$. Figure 3 shows the community structure obtained by approximate maximiza-

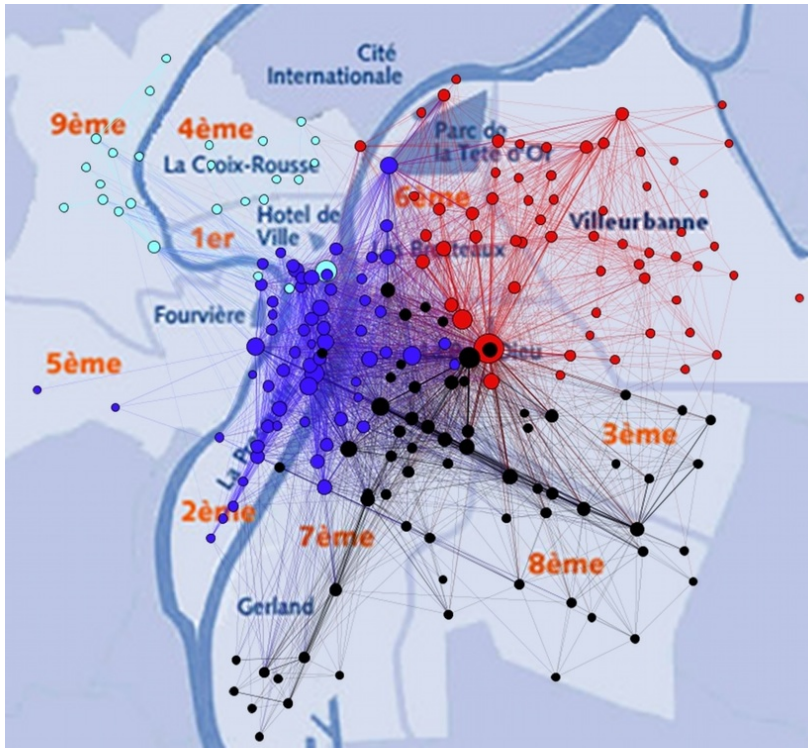

Fig. 3. Aggregation in space and time of the Vélo'v network: Communities of the average network $< T[n,m] >_{\mathcal{T}_P^*}$ are shown on top of a simplified map of Lyon. Each community has its own color. The size of one node is proportional to the number of trips made to and from this station; the width of each edge is proportional to the number of trips made between two stations. For the sake of clarity, the undirected version of the graph is shown and stations with small connectivity (degree at most 2) and edges with low activity (at most one trip per week) are not shown.

tion of the modularity Q by the Louvain algorithm. Four communities appear in this network and they are displayed on the Figure. The main feature is that the obtained communities, when shown on a map using the GPS coordinates of each Vélo'v station, are easily grouped on a geographical basis. This can be surprising as the partitioning in communities is blind to any geographical consideration. Anyway, one recognizes in the proposed communities a partition of Lyon city that reflects its general organization. The center of the city is spread out between the Presqu'île (between the two rivers, the Saône on the west, the Rhône on the east) and Part-Dieu (transport hub comprising the main railway station and a subway station) (blue community); the north-east part contains the 6th district and Villeurbanne which are well connected together with a major science university campus in the north (red community); the south and south-east parts are organized along two major roads (one from Gerland to Saxe-Gambetta and then Part-Dieu, a second along the limit between the 3rd and the 8th district to Saxe-Gambetta) (black community on the map). Finally, the north (Croix-Rousse) and north-west (Vaise, 9th district) are separate from other parts because Croix-Rousse is on the top of a high hill, and Vaise accessible only along the Saône river between Croix-Rousse hill and Fourvière hill (5th district); this creates a fourth separate community. This good matching of communities found by modularity and of the geographical partition of the city in large zones was first discussed in [10]. It even holds if one uses a smaller scale for communities, by keeping the hierarchical structure that is obtained thanks to the Louvain algorithm.

Communities in Vélo'v networks during the week-days.

Communities can also be looked at for individual snapshots $T[n,m](t,\Delta)$, with $t \in \mathcal{T}_P^*$, of the Vélo'v network. Figure 4 displays the community structure obtained by the Louvain algorithm for two snapshots taken during ordinary job days (here, on Mondays). The community structure at a given time in the week still matches a geographical partitioning of the city. Because the timescale is finer, with less aggregation in time, more details are apparent and some specific Vélo'v stations are not in the same community as their surrounding stations. For instance, the northern community (Villeurbanne) which includes a major university campus, includes also a station on the banks of the Rhône where there is also a university. Also, there are more communities (6 here) than in the average network. One new community groups together the Croix-Rousse hill with the Hôtel de Ville and the 6th district which contain the closest downhill subway stations. The comparison of the maps in the morning (on the left) and at the end of the afternoon (on the right) is interesting: it shows that most of the communities are left unchanged. This indicates that Vélo'v, like other ordinary transportation means, is used for commuting (from home to work in the morning, and back in the late afternoon). However, the community grouping the Croix-Rousse hill and the 6th district is not present anymore: it is not hard to figure out that people living on the Croix-Rousse will not use Vélo'v bicycles (which are heavy bicycles) to go up the hill back to their home. Apart from that, more than 90% of the remaining nodes did not change of community. Note that the same results are obtained for other ordinary week-days.

Communities in Vélo'v networks during the week-ends.

When turning to the analysis of the week-end uses of Vélo'v, the features change a little bit as shown in Figure 5. A first point is that the most active stations during the week-ends are not always the same than during the ordinary days. Major transportation hubs are unchanged (Part-Dieu, down-hill of Fourvière, Hôtel de Ville,...) yet new active stations appear near places for

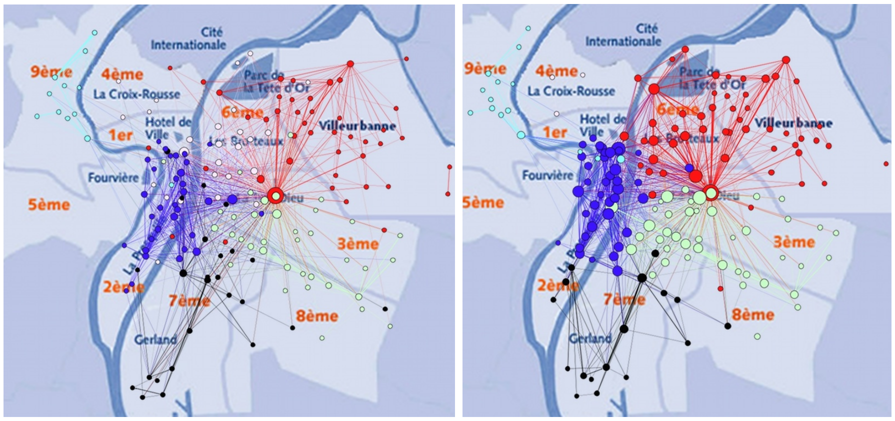

Fig. 4. Aggregation in space of snapshots of the Vélo'v network. Left: Monday am (7am-9am); Right: Monday pm (4pm-6pm). Each community has an arbitrary color. The size of each node is proportional to its incoming flow added to its outgoing flow. Note that the number of trips is usually larger during the afternoon (as was already seen on Fig.1, the peak for two hours in the afternoon is 1.5 times higher than the one in the morning).

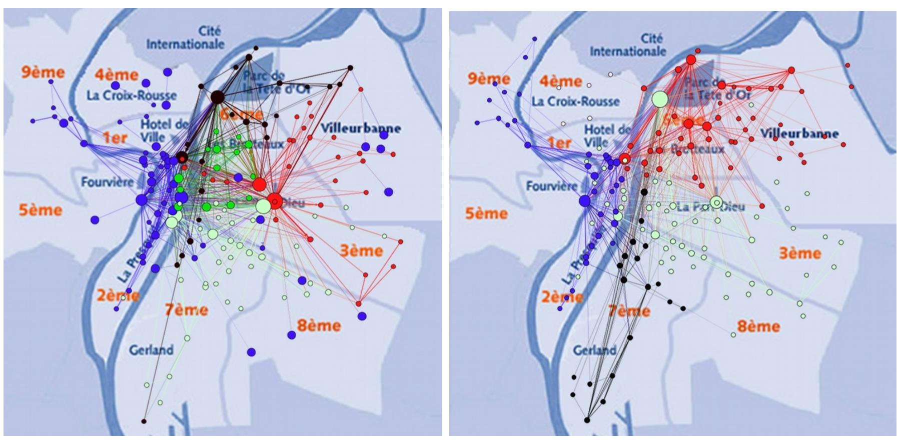

Fig. 5. Aggregation in space of snapshots of the Vélo'v network. Left: Saturday pm (3pm-5pm); Right: Sunday pm (3pm-5pm). Each community has an arbitrary color.

shopping (in the Presqu'île) and all around the large and green city park of la Tête d'Or in the north.

On Sunday (on the right), the communities could be reminiscent of the time-averaged ones excepted on two points. First, Vaise (9th district) is grouped with the Presqu'île community, possibly because the paths along the river are a pleasant leisure trip. Second, the community ranging from

Part-Dieu to the 3rd and 8th districts contains a station which is the most active during the week-ends: the station at the entrance of the city park of la Tête d'Or. Also some stations along the Rhône river are grouped in the same community. Here again, this is not a great surprise as the park of la Tête d'Or is a main destination for Sunday's outdoor activities. This community connects this park to other places that are either hubs of transportation (like Part-Dieu and Saxe-Gambetta) or other subway stations.

For Saturdays (on the left), the situation is more complex and does not reflect easily a simple geographical partition of the city. A community organized around the park of la Tête d'Or and grouping many stations around the park and on the river banks (having an easy access to the park thanks to bicycle paths on the river banks) is clearly visible. A surprising feature is that the periphery (Vaise, Croix-Rousse but also the most eastern parts of the city) is grouped in the community of the city center (in blue). This is a clue showing that some people uses Vélo'v for longer distance trips on Saturdays than on ordinary days.

This last aspect is one example of the fine scale analyses that are made possible by aggregating the network in a meaningful manner. It helps finding some unexpected structure that could be probed with more details in the complete dataset.

5 Typology of dynamics of the Vélo'v network

Another method for aggregation of nodes is possible: we now want to group two nodes if they have the same usage pattern, whereas in the previous section we grouped nodes exchanging many bicycles. Such an aggregated view is different from summing up the individual snapshots. With the objective of proposing a streamlined methodology for aggregating networks in space and/or time, we will show how the notion of communities can be tailored to group nodes with similar behaviors. For that, the idea is to first build a new similarity network from the dynamical network, before finding communities of similar nodes.

Similarity graph for the dynamics.

The principle is to quantify the resemblance over time of the different flows between stations. For that, one considers each snapshot to be one observation of the network, and then builds a similarity matrix between stations based on these observations. Given a station $n \in \mathcal{N}$, two feature vectors characterize its activity: the incoming flows $F^{in}[n](t) = \sum_{i \in \mathcal{N}} T[i,n](t,\Delta)$ and the outgoing flows $F^{out}[n](t) = \sum_{j \in \mathcal{N}} T[n,j](t,\Delta)$ where $t \in \mathcal{T}_P$ and Δ is constant. For a given pair of stations $[n,m]$, quantities can be computed to quantify if these activity patterns look alike or not. A general approach relies on the choice of a distance d between features (see for instance [36] for many possible distances, or [28] for application on graphs), leading to distances between activities of stations n and m:

$$D^{in}[n,m] = d(F^{in}[n], F^{in}[m]) \quad \text{and} \quad D^{out}[n,m] = d(F^{out}[n], F^{out}[m]). \quad (3)$$

For dealing with observations at different times $t \in \mathcal{T}_P$, it is natural to use a correlation distance over the various observations. The empirical estimator of correlations reads as

$$D^{in}[n,m] = \frac{1}{\#\mathcal{T}_P} \sum_{t \in \mathcal{T}_P} \tilde{F}^{in}[n](t)\tilde{F}^{in}[m](t) \quad (4)$$

where $\tilde{F}^{in}[n]$ is the centered and normalized version for each n of $F^{in}[n]$ (and respectively for $\tilde{F}^{out}[n]$), and $\mathcal{T}_P$ is a set of times of interest.

When looking at individual snapshots of the Vélo'v network, we have commented that the behaviors during the week-days are roughly unchanged from one day to another. It makes sense to take as the set of relevant times $\mathcal{T}_P$ the 15 peaks of activity of the week-days that were already used previously: 8am, 12am and 5pm, with $\Delta = 2$h. We obtain two correlation matrices of size $\#\mathcal{N} \times \#\mathcal{N}$, that we note ρ^{in}_{week} and ρ^{out}_{week}. For the week-ends, the behavior of the stations is different and we compute separately a correlation for the features during the week-ends. Using for $\mathcal{T}_P$ the times 12am and 4pm of Saturday and Sunday and a timescale $\Delta = 2$h, two other correlations matrices are obtained: $\rho^{in}_{\text{w.-end}}$ and $\rho^{out}_{\text{w.-end}}$. Note that one could use not only the peak activities but the whole temporal features over these days (like the one reported in Figure 1 for the global system). However, it is less important if during the low activity parts of the days, two stations have the same behavior (which could be no activity at all, and that would have no real relevance).

Remind that the goal is to compare the behaviors of stations, hence the choice of looking at in-in or out-out correlations. An alternative would be to study in-out correlations between pairs of stations; this metrics would describe whether two stations are well connected in the meaning that bicycles leaving one station have a good chance to arrive at another station. However, this metrics appear to be less interesting: first, the mere study of the flows connecting two stations, as studied in Section 4, gives already a picture of how well two stations are connected in this acceptance and this new study would be somewhat redundant; second, the flows leaving a given station are usually really spread between many other stations: the statistical confidence on estimated in-out correlations is low and we leave their study to further work.

For the Vélo'v network, the situation is that of many nodes ($\#\mathcal{N} = 334$) but only a few observations on one dataset because $\#\mathcal{T}_P$ is 15 for week-days, and 4 for week-ends. Recent theoretical studies about *Correlation Screening*

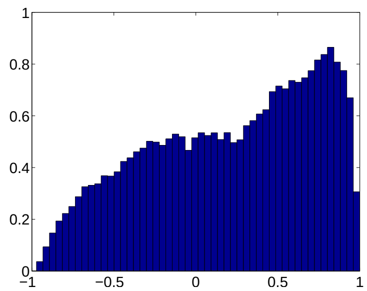

Fig. 6. Distribution of the empirical correlations ρ^{out}_{week} values.

[37] have shown that, even under the null hypothesis of no correlations between the node features, one should expect large estimated values if using the empirical correlation for large $\#\mathcal{N}$ and small $\#\mathcal{T}_P$. The number of false discovery of non-zero correlations can then be really large. In [37], expressions are given to estimate the threshold under which false discovery becomes dominant. As a consequence, if one wants to build a network of similarity between nodes based on correlation for a small number of observations (as it is often the case), it is expected that a thresholding operation is needed on the correlation to reduce the number of false discoveries. Using [37], and the specific values for the Vélo'v network, a threshold in correlation of 0.8 is reasonable to obtain some statistical confidence of discovering real correlations. Figure 6 shows the histogram of the values of the correlation matrix ρ^{out}_{week} outside the diagonal. As expected, the distribution is broad in $[-1, 1]$. Still, a maximum in the probability of finding large correlations occurs around 0.85 that is not predicted by a null hypothesis of uncorrelated node features. This is a sign of existing similarities in the nodes' activities in the network.

We build a similarity graph for the nodes by thresholding the 4 correlation matrices and summing up the thresholded correlations. The weight $S[n, m]$ on each edge of the similarity graph is then:

$$S[n, m] = \sum_{\substack{dir.=\{in,out\} \\ time=\{\text{week},\text{w.-end}\}}} \max(\rho^{dir.}_{time}, \eta) - \eta. \tag{5}$$

where η is the threshold. Based on the analysis before, we set $\eta = 0.8$ (though the results are not sensitive to small changes of η). An important remark is that the threshold is applied directly on ρ, not on its absolute value: for the Vélo'v network, negative correlations are discarded because stations with opposite activities would not be in a group of similar behavior. On the contrary, it could be used as to detect stations whose behavior are far remote, but this is not our objective here.

As a consequence, the similarity weights $S[n, m]$ are between 0 and 4. If two nodes are never similar, neither during the week nor during the weekends, both for incoming and leaving flows, the weight is 0 and the nodes (i.e.,

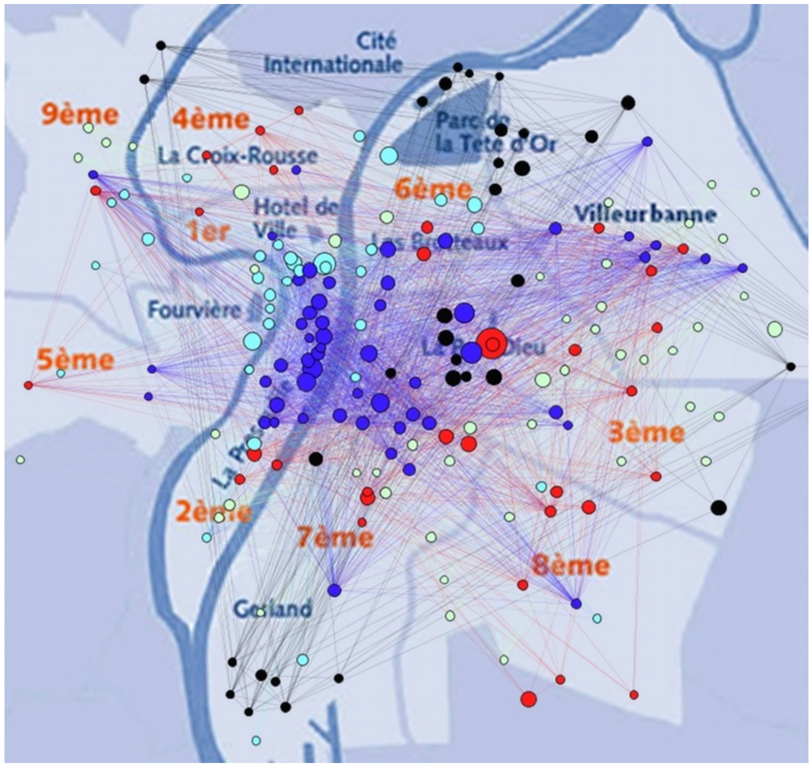

Fig. 7. Communities in the similarity graph $S[n,m]$ of the Vélo'v network. Each community (in an arbitrary color) reflects a specific pattern of usage of the station during the week. Edges are non-zero weights of similarity $S[n,m]$ and nodes have a size proportional to the number of trips made to and from this station.

Vélo'v stations) are not connected in the similarity graph. If the nodes have similar behavior along time for some of the features, the weight will increase by being higher than η, 2η, 3η or 4η if they are similar for one, two, three or all of the feature correlation matrices. Using thresholding before summing the correlations allows us to escape the poor estimation in correlation screening.

Communities of dynamical activities.

Given the similarity graph, quantifying if the activities of two stations look alike along time or not, it is possible to build a typology of the stations by grouping them according to these correlations. This is simply framed as a problem of detecting communities in the similarity graph of weights S. The same method of community detection, using modularity and the Louvain algorithm, is used on the weighted similarity graph. It provides communities of stations that share, by design of S, their pattern of activity in time. Each community is a type of dynamical activity in the Vélo'v network. The set of communities can be seen as a typology of the different dynamics at work in the network. Figure 7 shows the obtained typology on the Lyon map.

As compared to the previous communities obtained for space aggregation, the similarity communities can not be matched on a simple geographical partitioning of the city. However, it can be interpreted as a kind of segmentation of the city in various zones of activity. For instance, the community in black groups most of the nodes from the university campus (Villeurbanne and near the park in the north, Gerland in the south, the medicine university in the east and the university on the banks of the Rhône) and parts of the city with many companies – places to which people commute. The community of Part-Dieu (in red) includes many places of major subway or tramway hubs. Another (in dark blue) is spread out in the city center and has extensions along the stations of the subway lines crossing the center at Bellecour. Finally, the two remaining communities (in light colors) group parts of the cities that are in the east (mostly residential area) or near the Saône river banks in the center (where there are many shops, especially active during the week-ends).

All these communities, found through the similarity patterns in the dynamical Vélo'v network and thanks to information science methods, offer a typology of the different neighborhoods in Lyon that will be compared in the future to social science studies.

6 Conclusion

Studying the Vélo'v shared bicycle system as a dynamical network, we have discussed in this chapter how the methods for complex networks can be used to understand part of the dynamics of movements in a city. A key point of the study is the existence of digital footprints left by the use of such automated systems. The main point of this article, after a review of general results on community shared bicycle systems, is to show that community detection thanks to modularity maximization offers a way to create space and/or time aggregation of a network representation of the system. This method for aggregation in space would make possible the modeling of the Vélo'v systems in two levels: a first level inside each community (each having many trips inside the group); a second level between the communities. Separating the modeling in two levels like that is a good step to reduce the dimensionality of possible modeling of the number of trips made with bicycles in space and time. Also, the aggregated networks that were obtained are interesting as they enable future comparisons of results obtained by these complex network methods, to results obtained more traditionally by economical and social studies of the city and its transportation systems.

Acknowledgments JCDecaux – Cyclocity and the Grand Lyon Urban Community are gratefully acknowledged for having made Vélo'v data available to us. The authors thank Eric Fleury, Antoine Scherrer, Pablo Jensen, Luc Merchez, Charles

Raux, Alain Bonnafous and Marie Vogel for interesting discussions on the study of Vélo'v. This work has been partially supported by IXXI (Institut des Systèmes Complexes – Complex Systems Institute) Rhône-Alpes in Lyon.

References

1. Latora, V. and Marchiori, M. "Is the Boston subway a small-world network?", Physica A 314, 109–113 (2002).
2. Roth, C., Soong Moon Kang, Batty, M., Barthelemy, M. "Structure of Urban Movements: Polycentric Activity and Entangled Hierarchy Flows", *PLoS One*, vol. 6 (1), p. e15923 (Januray 2011).
3. Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., and Manna, S. S. "Small-world properties of the Indian railway network", *Physical Review E*, vol. 67, p. 036106 (2003).
4. Ghosh, S., Banerjee, A., Sharma, N., Agarwal, S. and Ganguly, N. "Statistical Analysis of the Indian Railway Network: A Complex Network Approach", *Acta Physica Polonica B Proceedings Supplement*, vol. 4(2), pp. 123–137 (2011).
5. Barrat, A., Barthélemy, M. and Vespignani, A. "The effects of spatial constraints on the evolution of weighted complex networks", *J. Stat Mech.*, vol. 5, p. P05003 (2005).
6. Barrat, A., Barthélemy, M. and Vespignani, A. *Dynamical processes on complex networks.* Cambridge University Press, Cambridge (2008).
7. Girardin, F. "Revealing Paris Through Velib' Data" http://liftlab.com/think/fabien/2008/02/27/revealing-paris-through-velib-data/ (2008).
8. Nair, R., Miller-Hooks, E., Hampshire, R., Busic, A. "Large-scale bicycle sharing systems: analysis of Vélib", *International Journal of Sustainable Transportation* (2012).
9. http://www.velov.grandlyon.com/
10. Borgnat P., Robardet C., Rouquier, J.-B., Abry, P., Fleury, E., and Flandrin P. "Shared Bicycles in a City: A Signal Processing and Data Analysis Perspective", *Advances in Complex Systems*, Vol. 14, No. 3, p. 415-438 (2011).
11. Froehlich, J., Neumann, J., and Oliver, N. "Measuring the Pulse of the City through Shared Bicycle Programs", *International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems - UrbanSense08*, Raleigh, North Carolina, USA (November 4, 2008).
12. Froehlich, J., Neumann, J., and Oliver, N. "Sensing and Predicting the Pulse of the City through Shared Bicycling", *Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*, pp. 1420-1426, Pasadena, California, USA, (July 11 - 17, 2009).
13. Lathia, N., Ahmed, S., Capra, L. "Measuring the impact of opening the London shared bicycle scheme to casual users", *Transportation Research Part C: Emerging Technologies*, vol. 22, pp. 88–102 (2012).
14. "Bicikelj: Environmental Data Mining on the Bicyle", submitted to *ITI - Information Technology Interfaces* 2012.
15. http://www.velov.grandlyon.com/Plan-interactif.61.0.html
16. http://www.grandlyon.com/PDU.55.0.html

17. Moghtaderi, A., Flandrin, P. and Borgnat P. "Trend Filtering via Empirical Mode Decompositions", *Computational Statistics & Data Analysis*, In Press, (Available online 2 June 2011).
18. Borgnat, P., Abry, A., and Flandrin, P. "Modélisation statistique cyclique des locations de Vélo'v à Lyon", Symposium GRETSI-09, Dijon, FR (Sept., 2009).
19. Borgnat, P., Abry, A., Flandrin, P., and Rouquier, J.-B. "Studying Lyon's Velo'V: A Statistical Cyclic Model", *Proceedings of ECCS'09*, Warwick, UK (Sept., 2009).
20. Jensen, P., Rouquier, J.-B., Ovtracht, N. and Robardet, C. "Characterizing the speed and paths of shared bicycles in Lyon", *Transportation Research Part D: Transport and Environment*, vol. 15 (8), pp. 522–524 (2010).
21. Borgnat, P., Fleury, E., Robardet, C., and Scherrer, A. "Spatial analysis of dynamic movements of Vélo'v, Lyon's shared bicycle program", *Proceedings of ECCS'09*, Warwick, UK (Sept., 2009).
22. Holme, P., Saramäki, J. "Temporal Networks", *ArXiv* 1108.1780v2 (2011).
23. Abry, P., Baraniuk, R., Flandrin, P., Rieidi, R. and Veitch, D. "Wavelet and Multiscale Analysis of Network Traffic", *IEEE Signal Processing Magazine*, vol. 3 (19), pp. 28–46 (2002).
24. Bajardi, P. Barrat, A., Natala, F., Savini, L., Colizza, V. "Dynamical patterns of cattle trade movements", *PLoS ONE*, vol. 6 (5), p. e19869 (2011).
25. Krings, G., Karsai, M., Bernhardsson, S., Blondel, V., Saramäki, J. "Effects of time window size and placement on the structure of an aggregated communication network", *EPJ Data Science*, vol. 1 (4), P. 1-16 (2012).
26. Miritello, G. Moro, E., Lara, R. "Dynamical strength of social ties in information spreading", *Phys. Rev. E*; vol. 83 (4), p. 045102 (2011).
27. Preparata, F. R. and Shamos, M. I. "Computational Geometry: An Introduction." New York: Springer-Verlag (1985).
28. Fortunato S., "Community detection in graphs", *Physics Reports*, vol. 486, pp. 75–174 (2010).
29. Hammond, D., Vandergheynst, P., Gribonval, R. "Wavelets on graphs via spectral graph theory", *Applied and Computational Harmonic Analysis*, vol. 30 (2), pp. 129-150 (2011).
30. Newman, M. and Girvan, M. "Finding and evaluating community structure in networks", *Phys. Rev. E*, vol. 69, 0206113 (2004).
31. Leicht, E. and Newman M. "Community structure in directed networks", *Phys. Rev. Lett.*, vol. 100, 118703 (2008).
32. Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikoloski, Z. and Wagner, D. "On modularity clustering", *IEEE TKDE*, vol. 20 (2), pp.172–188 (2008).
33. Clauset, A., Newman, M., and Moore, C. "Finding community structure in very large networks", *Phys. Rev. E*, vol. 70, 066111 (2007).
34. Blondel, V. Guillaume, J.-L., Lambiotte, R. and Lefebvre E. "Fast unfolding of communities in large networks", *J. Stat. Mech.*, P10008 (2008).
35. Rosvall, M. and Bergstrom, C. T. "Maps of information flow reveal community structure in complex networks", *PNAS*, vol. 105, p.1118 (2008).
36. Basseville, M. "Distance measures for signal processing and pattern recognition", *Signal Processing*, vol.18 (4), pp.349-369 (1989).
37. Hero, A.O. and Rajaratnam, B. "Large Scale Correlation Screening," *Journal of the American Statistical Association*, vol.106(496), pp. 1540-1552 (2011).