



HAL
open science

Data-driven Mobility Analysis and Modeling: Typical and Confined Life of a Metropolitan Population

Haron Fanticelli, Solohaja Rabenjamina, Aline Carneiro Viana, Razvan Stanica, Lucas Santos de Oliveira, Artur Ziviani

► **To cite this version:**

Haron Fanticelli, Solohaja Rabenjamina, Aline Carneiro Viana, Razvan Stanica, Lucas Santos de Oliveira, et al.. Data-driven Mobility Analysis and Modeling: Typical and Confined Life of a Metropolitan Population. ACM Transactions on Spatial Algorithms and Systems, 2022, 8 (3), pp.1-33. 10.1145/3517222 . hal-03569051

HAL Id: hal-03569051

<https://inria.hal.science/hal-03569051v1>

Submitted on 12 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Data-driven mobility analysis and modeling: Typical and confined life of a metropolitan population

HARON C. FANTICELLI, National Laboratory for Scientific Computing - LNCC, Brazil and Inria, France

SOLOHAJA RABENJAMINA, Univ Lyon, Inria, INSA Lyon, CITI, France

ALINE CARNEIRO VIANA, Inria, France

RAZVAN STANICA, Univ Lyon, INSA Lyon, Inria, CITI, France

LUCAS SANTOS DE OLIVEIRA, State University of Southwestern of Bahia - UESB, Brazil

ARTUR ZIVIANI, National Laboratory for Scientific Computing - LNCC, Brazil

The idea of using mobile phone data to understand the impact of the Covid-19 pandemic and that of the sanitary constraints associated with it on human mobility imposed itself as evidence in most countries. This work uses spatiotemporal aggregated mobile phone data provided by a major French telecom operator, covering a geographical region centered on Paris for early 2020, i.e., periods before and during the first French lockdown. An essential property of this data is its fine-grained spatial resolution, which, to the best of our knowledge, is unique in the COVID-related mobility literature. Contrarily to regions or country-wide resolution, it describes population mobility flows among zones ranging from 0.025 km^2 to 5.40 km^2 , corresponding to 326 aggregated zones over the total area of 93.76 km^2 of the city of Paris. We perform a data-driven mobility investigation and modeling to quantify (in space and time) the population attendance and visiting flows in different urban areas. Second, when looking at periods both before and during the lockdown, we quantify the consequences of mobility restrictions and decisions on an urban scale. For this, per zone, we define a so-called *signature*, which captures behaviors in terms of population attendance in the corresponding geographical region (i.e., their land use) and allows us automatically detect activity, residential, and outlier areas. We then study three different types of *graph centrality*, quantifying the importance of each zone in a time-dependent weighted graph according to the habits in the mobility of the population. Combining the three centrality measures, we compute per zone of the city, its *impact-factor*, and employ it to quantify the global importance of zones according to the population mobility. Our results firstly reveal the population's daily zone preferences in terms of attendance and mobility, with a high concentration on business and touristic zones. Second, results show that the lockdown mobility restrictions significantly reduced visitation and attendance patterns on zones, mainly in central Paris, and considerably changed the mobility habits of the population. As a side effect, most zones identified as mainly having activity-related population attendance in typical periods became residential-related zones during the lockdown, turning the entire city into a residential-like area. Shorter distance displacement restrictions imposed by the lockdown increased visitation to more "local" zones, i.e., close to the population's primary residence. Decentralization was also favored by the paths preferences of the still-moving population. On the other side, "jogging activities" allowing people to be outside their residences impacted parks visitation, increasing their visitation during the lockdown. By combining the impact factor and the signatures of the zones, we notice that areas with a higher impact factor are more likely to maintain regular land use during the lockdown.

CCS Concepts: • **Information systems** → **Spatial-temporal systems**.

Additional Key Words and Phrases: human mobility, lockdown restrictions, land use detection, graph centrality

1 Introduction

Mobile phone data is nowadays acknowledged as a common tool to study human mobility [26, 27, 31]. With mobile phones becoming proxies for human presence, datasets collected by mobile operators outlined many interesting properties regarding human mobility and activity patterns, such as high periodicity [14, 45], shortest path routing [10, 47], or travel patterns that follow preferred locations [32, 33]. While these properties seem to be universal and have been observed in numerous studies over different geographical areas [19, 21, 44], mobile phone data have also proven useful in the study of specific events and situations. A few representative examples are the study of mobility patterns in urban regions during major cultural events [51], the analysis of migration patterns following important natural disasters [56], or the link between human mobility and disease propagation [49].

Authors' addresses: Haron C. Fanticelli, National Laboratory for Scientific Computing - LNCC, Petropolis, Brazil, Inria, Palaiseau, France, haroncf@hotmail.com; Solohaja Rabenjamina, solohaja.rabenjamina@inria.fr, Univ Lyon, Inria, INSA Lyon, CITI, Villeurbanne, France; Aline Carneiro Viana, aline.viana@inria.fr, Inria, Palaiseau, France; Razvan Stanica, ravan.stanica@inria.fr, Univ Lyon, INSA Lyon, Inria, CITI, Villeurbanne, France; Lucas Santos de Oliveira, lsoliveira@uesb.edu.br, State University of Southwestern of Bahia - UESB, Jequié, Brazil; Artur Ziviani, National Laboratory for Scientific Computing - LNCC, Petropolis, Brazil.

As expected, mobile phone data was extensively used in the fight against the Covid-19 pandemic. Some governments, mainly in Asian countries, included operator data in their contact tracing strategy [39, 53, 57] and most countries tried to assess the impact of sanitary measures, such as lockdowns and travel restrictions [18]. Many research teams specialized in epidemics also made extensive use of mobile phone data to parameterize their models [59]. However, most of the available studies focused on mobility at a large spatial scale, between different regions or different cities in a country.

In this study, we report a data-driven analysis focused on human mobility during the Covid-19 pandemic on a metropolitan scale. More precisely, we study the Paris metropolitan area and investigate how mobility patterns at an urban scale were affected by the pandemic, and especially by the harsh lockdown conditions enforced from March 17, 2020, to May 11, 2020, in France. These sanitary measures forced drastic changes on population mobility and the dataset we explore in this work allows us to quantify their impact. The mobility data we use are unique in the Covid-19 related literature, to the best of our knowledge, distinguishing themselves through two important properties. First of all, the dataset presents a spatial resolution describing population attendance and mobility over small geographical zones (varying from 0.025 km^2 to 5.40 km^2 , this later representing very few zones) and covering overall a significant area (i.e., 93.76 km^2) in a big European metropolis.

Second, the dataset, collected by SFR, a major French mobile operator, temporally covers two different contexts, describing a "usual" and a "radically locked down" urban life. From a human mobility study perspective, this represents a rare opportunity, as a radical change in population mobility habits of an important EU capital can not be planned for research purposes. Note that the literature on human mobility focused on very big planned events, such as the Olympic Games or the Football World Cup, which induce flow increase in some particular areas of the city (e.g. airports, touristic spots, stadiums, and sports centers), while the large part of the local population pursues a "normal" urban life, with maybe an adaptation of some transit zones. On the other hand, when crisis situations are studied, it is usually too late to retrieve "normal" mobility information about the area of interest. In our case, the entire population is impacted and the dataset also covers the period immediately before the lockdown, allowing for an original case of study.

In this work, we perform a data-driven mobility investigation and modeling to quantify (in space and time) the population attendance and visiting flows in different urban areas. Second, when looking at periods both before and during the lockdown, we quantify the consequences of mobility restrictions and decisions on an urban scale. The contributions of this study are:

- We conduct a detailed spatio-temporal analysis of human mobility during the lockdown, showing that mobility is divided by a factor 3-4 once the lockdown is established. However, we notice that hotspots still exist during the lockdown and some regions in Paris even observe an increased attendance (cf. Section 4).
- By adapting a well-known framework for automatic land use inference from mobile phone data [20], we classify each region in Paris as either residential, activity, or outlier area. This allows us to observe how the usage of any given area changes once the lockdown is established. We notice that most of the city transforms in a big residential area, with a few specific activity and outlier areas that we study in-depth (cf. Section 5).
- We model mobility flows using a time-varying weighted mobility graph and we study three different types of graph centrality, which quantify the importance of each zone in the city according to the habits in mobility of the population: (1) the betweenness centrality captures the tendency of people to follow shortest paths, (2) the closeness centrality captures the locality of people movement, and (3) the degree centrality captures topologically central hubs. By observing how these graph centralities change, we outline the fact that mobility became more local during the lockdown (cf. Section 6).
- We combine the three centralities and the general human density of an area into one metric, which we denote as impact factor. We use this metric to quantify the global importance of zones according to the mobility behavior of the population. We notice that the zones with a high impact factor in normal conditions are mainly located in the city center and the lockdown restrictions have a significant effect on them. Interestingly, many of these zones still appear among the most important ones in the city of Paris in terms of impact factor, even during the lockdown. Besides, our results also show a visitation increase in parks and in more local hotspots during the lockdown (cf. Section 3).

- We study the correlation between the impact factor and the type of area (residential, activity, other). We show that areas with an outlying behavior are more biased towards low impact factor and that stable areas (i.e. areas that do not change their usage once the lockdown is established) are more likely to have a high impact factor (cf. Section 7).

While the methodology proposed in this paper is used to obtain quantitative results regarding the impact of the Covid-19 pandemic on urban mobility, the approach we describe, combining time series analysis and graph theory, can be applied to any contextual scenario with an impact on human mobility, which brings the generality of our contributions. Generally speaking, mobility analytics such as those we propose bring an understanding of the urban space usability, which can indeed be used for epidemic prevention [12], but also for disaster response [41] and urban traffic management [52]. Moreover, these analytics also have the potential to improve mobility anticipation algorithms [4, 47] and quality of services offered by pervasive computing applications [1].

To cite just a few examples we are currently exploring, our contributions can feed Anti-Covid applications (i.e. spots to avoid in a journey, as already done for pollution peaks), as well as draw the authorities attention to particular areas. Another application is the provision of energy-efficient and cost-effective network infrastructures [17, 36], adapted to the modified population mobility patterns during these special situations. Indeed, telecommunications operators planned and deployed their infrastructure with certain attendance and mobility patterns in mind, so understanding how these patterns change in special situations is a first step in the proposal of flexible and adaptive approaches. As a final potential impact, our results are a valid indication of activity loss, and consequently, of the imposed socio-economic impact activity-labeled areas suffered during the lockdown.

The rest of the paper is organized as follows. In Section 2 we discuss related work on the use of mobile phone data in the context of the Covid-19 pandemic. Section 3 presents the dataset and our modeling approach. The main spatio-temporal insights are presented in Section 4, before studying the Paris land use in Section 5. The centrality-based analysis and the impact factor metric are discussed in Section 6. Finally, the correlation between impact factor and centrality makes the object of Section 7.

2 Related Work

Digital technologies are one of the main tools in the fight against the Covid-19 pandemic. This is especially true in the contact tracing step, a task for which many countries proposed specific smartphone applications [2]. Most of these solutions are based on Bluetooth ranging [28], with a lot of attention being given to privacy problems [11]. Because of these privacy issues, these applications can not actually be used to understand the impact of the pandemic on human mobility.

Mobile phone data collected by network operators has been used in the past to understand the propagation and the impact of mobility on infectious diseases, such as malaria [54] or HIV [9]. With respect to the Covid-19 pandemic, several countries, especially in Asia, included data collected by mobile operators in the official tracing and monitoring operations [39, 53, 57]. Numerous academic studies also included mobile phone data, with different objectives. Oliver et al. [38] provide a detailed discussion of these works and the way they can be used by governments to estimate the effectiveness of sanitary control measures.

A number of studies focused on the impact of Covid-19 and associated measures on human mobility. Hara and Yamaguchi [25] observed a mobility reduction of 20% in Japan, as well as noting that people avoided traveling to densely populated areas. A significant drop in inter-metropolitan mobility was also observed in Finland [55]. A relation between inter-metropolitan mobility and Covid-19 propagation is observed by a study in Brazil [40]. As a matter of fact, mobility changes during the pandemic were observed globally, as shown by a study on 124 countries [16], and the intensity of these changes is related to the development index of the countries, with a more significant mobility reduction in highly developed countries.

Similarly, in France, G. Pullano et al. [42] showed that 9 out of 10 symptomatic cases of COVID-19 were undetected by the surveillance system in the 7 weeks following the end of the first lockdown in France. This analysis was performed at the regional scale and using mobile phone data provided by Google Mobility Trends. In [18], the authors studied the impact and effectiveness of the first lockdown, when exploring different types and duration of social distancing interventions (e.g., home-office). Mobile phone data in the Ile de France region is used to deduce mobility reductions

(one of the studied impacted factors) during the lockdown (70% of reduction resulted from the analysis) and to model the effect of sanitary measures. However, all these studies focus on macroscopic mobility data, at a regional or even national level. Badr and Gardner [6] discuss at length the limitations of this country-wide analysis, showing that mobility is not always correlated with virus propagation at this large scale.

Most studies using mobile phone data are based on data coming from Apple and Google, the developers of the two mobile operating systems dominating the market [23]. However, the data provided through these means have a much lower resolution than the mobile phone data collected by the operators. For example, in [23], the authors observe a decrease of 80% in human mobility in the United Kingdom during Spring 2020 using data provided by Apple, but the resolution is countrywide. Moreover, the gradual sanitary restrictions imposed in the United Kingdom over the studied period were very different from the harsh and prompt lockdown applied in France.

A few studies used fine-grained spatial mobile phone data but provided valid scientific contributions with different objectives than our work. For example, Li et al. [30] study the impact of the pandemics on smartphone usage, showing a decreased user engagement during the outbreak. Zhou et al. [59] use mobile phone data from the metropolitan area of Shenzhen to build a virus transmission model and simulate the impact of different scenarios of mobility restrictions. Finally, Ayan et al. [5] use both aggregated and individual mobile phone data covering the city of Rio de Janeiro, in Brazil. Their spatial results focus on the information at the level of the district, which is much larger than the resolution we use for Paris. The authors mostly centered their analysis on the unique individual mobility dataset at their disposal, allowing them to distinguish that 45% of the users entirely stopped their mobility during the Brazil lockdown, while 4% of the users still present a very high mobility index.

Our study presents considerable differences with respect to the existing literature. Contrarily to previous works, we investigate mobility behavior of population on a relatively small scale, in the city of Paris, in two different temporal contexts: (1) when urban mobility follows daily routine-like mobility habits of its population and (2) when rigid movement restrictions were firstly enforced by the government to slow down the virus spread. More important, both period investigations are based on mobile phone data describing population mobility flows among small spatial scale zones in Paris. Our analysis is spatially oriented, distinguishing the land use of different urban areas before and during the lockdown and using a series of metrics from graph theory to compute an epidemic-related impact factor for each urban zone.

3 Dataset, modeling, and methodology

For business (e.g., billing, network management purposes) and legal reasons, mobile operators log events generated during mobile communications. Practically, for each voice call, text message, and mobile data traffic session, as well as in some user mobility situations (known as location area updates), the operator collects a series of network signaling events [35]. These events are associated with individual mobile devices and are time-stamped and geo-referenced. In their most basic form, the locations of the geo-referenced events correspond to the cell towers where the signaling event was generated, which makes mobile network data an obvious proxy for the trajectories of the mobile network subscribers.

In this study, we use a dataset built from call detail records (CDR) [29] and network signaling data (NSD) [19], provided by the French telecom operator SFR. This dataset is already pre-processed by the operator, representing an aggregation of mobile network data of all the mobile phone subscribers of SFR and covering 20% of the entire population of the studied area. Note that we look for a relative importance comparison among zones, which can be perfectly obtained with this available 20% of the entire population. A more complete population coverage would require data availability from the four French cellular operators.

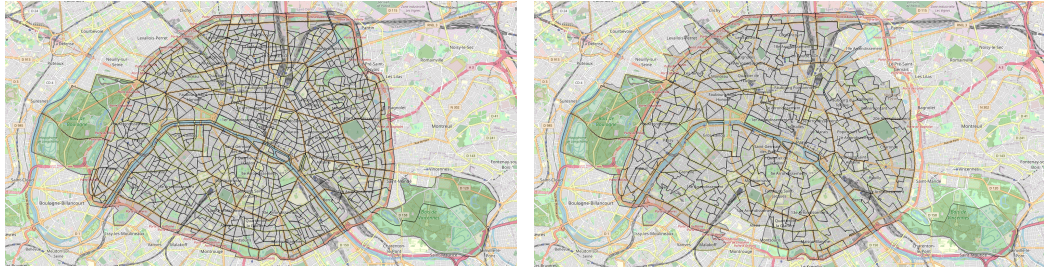
The original network data collected by the operator contain sensitive information regarding the users. Therefore, to avoid subscriber identification and to ensure a proper level of privacy, the provided dataset consists of user counters, aggregated both in time and space. We note that the European privacy regulations consider as personal data any individual trajectory data, such as those used in [5], even when the users' identifiers are pseudonymized.

More formally, we observe a large metropolitan area \mathcal{A} in the French region Ile de France, which basically comprises the Paris large metropolitan area, divided into a number of geographical zones $a \in \mathcal{A}$. For privacy reasons, the mobile operator cells, which represent the finest granularity achievable by the operator, were aggregated based on the official

units for statistical information (also known as IRIS zones), defined by the French National Institute of Statistics and Economic Studies (INSEE). The IRIS zones must respect geographic and demographic criteria and have borders that are clearly identifiable and stable in the long term¹. IRIS zones offer the most sophisticated tool to date to describe the internal structure of more than 1,900 municipalities in France with at least 5,000 inhabitants. Even this level of granularity is too fine in some cases, as some IRIS contain very few active users who could be identified this way. Therefore, some IRIS are grouped together by the operator to ensure a proper level of anonymity.

Practically, the dataset covers the city of Paris (i.e., the French department with code 75), for a total area of 93.76 km². There are 992 INSEE IRIS zones in the city of Paris, which are further aggregated into 326 zones in the SFR dataset. To achieve this level of aggregation, 8% of the INSEE IRIS were removed by SFR, about 11% of the INSEE IRIS have been kept the same, and the remaining ones have been merged into groups with size from 2 to 16 INSEE IRIS (an average merging of 2.9 INSEE IRIS).

Figure 1 shows the original INSEE IRIS fragmentation of Paris and the division provided by SFR denoted hereafter as *SFR-IRIS*, which appears in the studied dataset. In the central metropolitan area, the SFR-IRIS sizes are smaller, due to their high user attendance, when compared with SFR-IRIS covering parks or more humanly sparse zones.



(a) Original IRIS division given by INSEE/IGN.

(b) IRIS division given by SFR.

Fig. 1. IRIS division of Paris (a) given by INSEE and (b) given by SFR in the studied dataset.

The provided dataset contains aggregated counter or flow information of several types, per time interval t :

- (1) Continuous user presence – denoted as $MoveInside(t)$: for each of the SFR-IRIS zones, we have the number of users observed in the corresponding zone during time interval t which were also present in the zone during $t - 1$. These users can be static or moving inside the zone, but they are not logged outside the corresponding area during time interval t .
- (2) Incoming trips – denoted as $InComing(t)$: for each SFR-IRIS zone, we have the number of users that arrive in the zone during time interval t . These users have been observed in a different area during time interval $t - 1$ and are now logged in the area of interest during t .
- (3) Outgoing trips – denoted as $OutGoing(t)$: for each SFR-IRIS zone, we have the number of users that leave from the zone during time interval t . The last known location of these users was in the area of interest during $t - 1$, and they are now observed in a different area in time interval t .

Regarding the incoming and outgoing trips, these are divided into two categories: flows coming/going from/to one of the other SFR-IRIS zones in Paris (with the corresponding zone identifier) and flows coming/going from/to any other French commune (with its corresponding identifier). The dataset does not contain any information on individual trajectories, we only know the number of trips, aggregated per source (corresponding to $OutGoing$ flows) and destination (corresponding to $InComing$ flows).

For each time interval t , we compute the attendance p_a^t for every zone a , which represents an indicator of human presence in that zone: $p_a^t = MoveInside(t) + \frac{InComing(t) + OutGoing(t)}{2}$. This is simply a consequence of the fact that

¹<https://www.insee.fr/en/metadonnees/definition/c1523>

users in $MoveInside(t)$ are observed inside area s during time interval t , while users in $InComing(t)$ and $OutGoing(t)$ are logged as entering or leaving a during the time interval, but without any precise information in the exact time of these events. For each time interval t , we also know the flow f_{ab}^t (representing an aggregated number of users) between any two zones a and b , which represents an indicator of human mobility. We insist on the fact that attendance and flows are aggregated values provided by SFR, and we do not manipulate any individual user data to obtain them.

Besides, for each zone, we compute its *density* in terms of its attendance and spatial coverage (i.e., area in square kilometers). At time t , the *density* of a zone a is defined as: $d_a^t = \frac{p_a^t}{S_a}$, where S_a represents the area (in km^2) covered by a . The intuition behind using d_a^t is to take into account the size of the considered zone, i.e., two zones having the same attendance but different geographical size do not have the same density.

The considered zones vary in size, according to the spatial features of the dataset. While moving, aggregated flows interact with the infrastructure and other flows in the visited zones. The sequence of zones in such flows describe more than mere physical areas in the map but they reflect a social aspect: flows from a residential zone to zones with other purposes, e.g. business, commercial, or leisure. Indeed, the flows we observe mainly represent the activity-oriented mobility behaviors of the population in the considered metropolitan area, involving not only people displacements, but also their interaction with the frequented zones in a city. This brings our approach close to the rich literature on urban functions and urban functionality [58]. However, our approach is fundamentally different from classical studies in urban planning and geographical information science. While these latter are reliant on surveys conducted by experts and sociological interpretation, our approach is entirely bottom-up and data driven. This allows us to have an understanding of the lockdown impact on human mobility and urban land use without waiting for mobility-oriented surveys, proposing, to the best of our knowledge, the most precise view and investigation of the population mobility in a dense urban area in the context of Covid-19.

Concerning the temporal side, the aggregated dataset covers two weeks before the first lockdown in France and two weeks during this strict lockdown: i.e., from January 26th to February 8th and from March 22nd to April 4th. The dataset has a temporal granularity of 1 hour.

Beyond quantifying the impact of the lockdown, which is by itself an important information, our overall objective is to provide public and health authorities with a mobile data-driven solution to quickly detect the zones in a city that are most prone to gather and mix up the population. The methodology we use in this sense consists of three steps, summarized below and detailed in the following sections:

- (1) We adapt an automatic land use detection algorithm from the literature [20], which allows us to label each zone as *residential*, *activity* or *other*, both before and during the lockdown.
- (2) By taking a graph theory approach, we measure three different centrality metrics for each zone. These metrics provide complementary information regarding user mobility in the studied urban area, and we combine them in a general metric, denoted as impact factor. The impact factor is representative of the role each zone plays in terms of mobility during the lockdown period.
- (3) We combine these two previous contributions in order to understand whether the impact factor of a zone is correlated with its land use.

4 Spatio-temporal properties of population mobility

We first begin our study through a detailed spatio-temporal analysis of the dataset. We plot in Figure 2(a) the total number of users in the dataset for the 2-week periods before and during the lockdown, for all SFR-IRIS. We can see that, in both these periods, the number of users follows a weekly periodic pattern. In particular, for the period before lockdown, the number of users at SFR-IRIS increases on weekdays, reaches a weekly maximum on Thursday or Friday, and declines quickly at weekends. Nevertheless, the number of users observed in the dataset decreases significantly during the lockdown period, though a weekly periodic pattern is still present.

On each day, the average number of users also varies hourly. As shown in Figure 2(b), before the lockdown, the majority of the population is detected in the SFR-IRIS zones from 8am to 11pm, while less than 25% of the events are logged from 12am to 7am. This is a combination of two phenomena: (1) a decreased user activity on the mobile network during nights and early morning, which reduces the population observed by the mobile operator, but also (2) a reduced human mobility during these periods. During the lockdown, when the home-office practice became

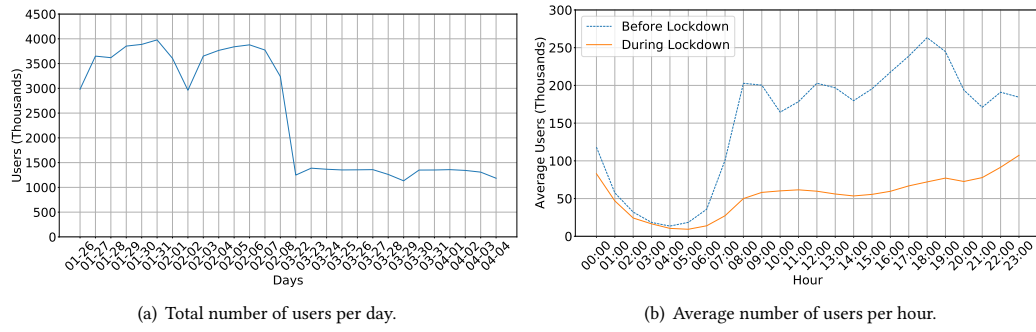


Fig. 2. Number of flows observed (1) per day and (2) per hour, before and during lockdown.

the rule and transportation time is removed from the daily routine of most of the population, the observed number of users per hour radically decreases from 8:00 to 23:00, though a similar activity is still present from 00:00 to 6:00. Interestingly, contrarily to the period before the lockdown, the activity here continues to increase after 6pm, when the remote working day should be usually over. This could be explained by the population choice in performing after 6pm the few activities allowed during the lockdown, such as jogging activities, as well as by the use of the mobile network for leisure activities during the evening.

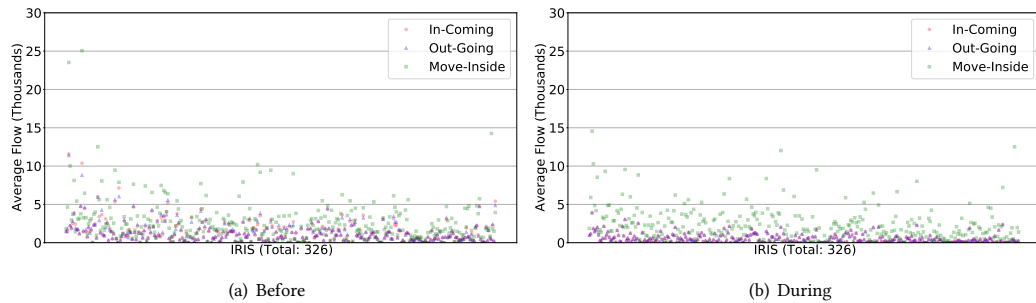


Fig. 3. Average number of *MoveInside*, *Outgoing*, and *Incoming* flows, before and during lockdown.

In order to better understand the reduction in terms of number of users during the lockdown, we decompose p_a^t in Figure 3, showing the flows *MoveInside*, *InComing*, and *OutGoing* individually. Figure 3(a) plots these flows before the lockdown, while Figure 3(b) shows them during the lockdown. From these figures, it is clear that the drop in user attendance is mainly a result of reduced mobility, as *InComing* and *OutGoing* flows significantly decrease, while the *MoveInside* volume is quite similar during the two periods, thus attesting for the maintenance of a local restricted movement during the lockdown.

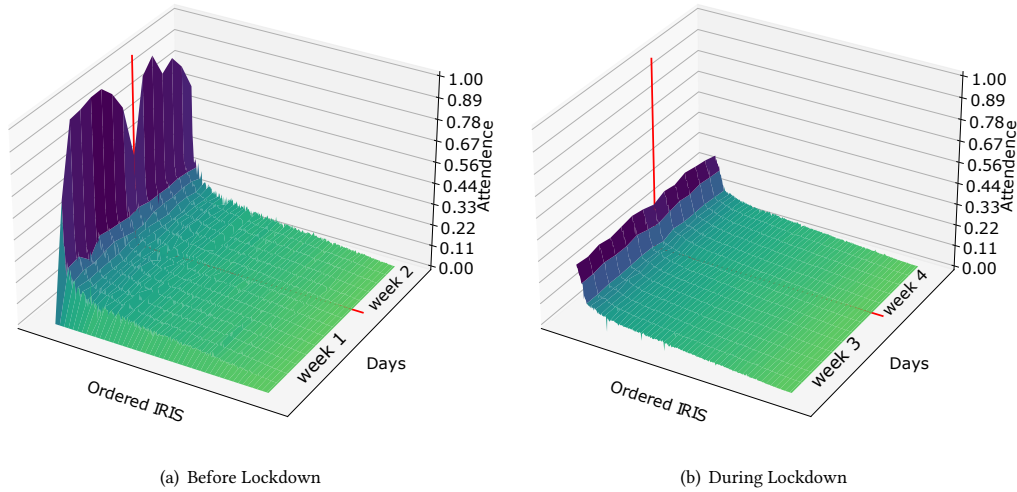


Fig. 4. Normalized average density per SFR-IRIS, ordered from high to low, before and during lockdown periods.

It is noteworthy that the user attendance is not only varying in time, but also in space. To show this, a spatiotemporal representation of the normalized density (with respect to the maximum attendance value of 44.000) is presented in Figure 4. We can see that, both before the lockdown (Figure 4(a)) and during the lockdown (Figure 4(b)), some SFR-IRIS zones are significantly more popular than others. However, the difference is much less important during the lockdown, with a much lower density in the popular IRIS. On the temporal side, we notice that the difference between week days and week-end practically disappears completely, as demonstrated by Figure 4(b).

In order to further investigate the spatial distribution of the zones attendance, we plot this metric for each SFR-IRIS on the map of the city of Paris in Figure 5. As expected, before the lockdown, the population attendance is mostly concentrated in the city center, as shown in Figure 5(a). As a matter of fact, the two most popular areas, clearly visible on the map, correspond to the areas where two main Parisian train-stations are located: Gare du Nord and Chat let. Both SFR-IRIS have an average attendance of approximately 35.400. During the lockdown, Figure 5(b) shows that the attendance is not only lower but also more uniformly distributed, with the city center no longer appearing as a hotspot, though the areas underlined above still remain visible on the map. The two most popular areas during the lockdown correspond to Gare du Nord and Vincennes park.

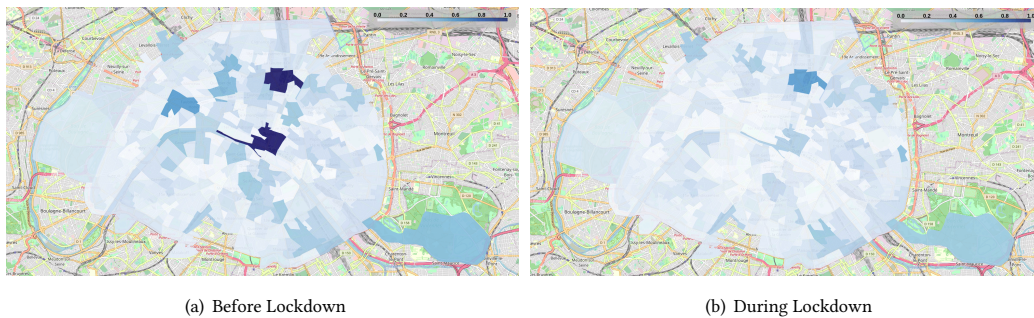


Fig. 5. Spatial distribution of the average attendance before and during lockdown periods.

Since the attendance does not consider the area covered by the SFR-IRIS, we also use the density as a complementary metric. However, to better depict this evolution in terms of population density, we choose a different representation in Figure 6, where we show the relative variation for each SFR-IRIS. Figure 6(a) is focused on the areas where the density is reduced during the lockdown, while Figure 6(b) outlines the zones where the density increases.

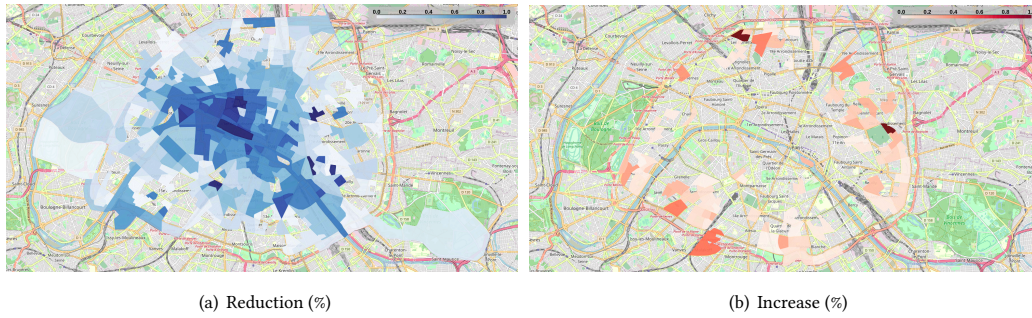


Fig. 6. Density (a) reduction and (b) increase when comparing periods before and during lockdown.

In Figure 6(a), the SFR-IRIS with the highest density reduction lost 19,000 flows in the lockdown period, equivalent to a 98% reduction. Interestingly, some zones actually showed an increased average density during the lockdown. From Figure 6(b), it is obvious that all these IRIS zones are situated at the periphery of the city, where they can welcome incoming flows from outside the city, becoming local hotspots. All the IRIS in the Paris city center presents a significant density reduction during the lockdown. For example, the SFR-IRIS with the highest density before the lockdown is located in the Opera Garnier area. It has an average attendance of 6624 users before lockdown, for a total area of 0.024 km², which results in an average density of approximately 270,000 user/km². The density of this area reduces by 80% during the lockdown, reaching an average attendance of 1370 users.

5 Attendance-based land use inference

An urban area is not used by its population in a uniform way. Depending on the settlements, arrangements, and inputs the population brings to a given area, its land use is different. Typically, the most basic division of an urban area is in *residential areas* and *activity areas*. Activity areas can be further divided in *office areas*, *leisure areas*, *educational areas*, etc. Mobile phone data have successfully been used in the past as a proxy for urban land use [46, 50]. Therefore, in this section, we exploit the available dataset to characterize the land use before and during the lockdown period, trying to assess the impact of these early sanitary measures. It is important to note that other methods (e.g. official surveys) could be used to characterize the land use during a normal period (i.e. before the lockdown). However, we are not aware of any other method that would allow us to characterize the land use in an exceptional period, such as during the lockdown. As a matter of fact, the municipality surveys imply significant costs and mobilize numerous people. For this reason, the Paris municipality only conducts mobility surveys once per decade, and this is in fact one of the highest frequencies at a global scale.

In order to detect residential and activity areas in a given urban region \mathcal{A} , we first need to classify its different subdivisions, trying to find recurrent and atypical patterns. In order to do this, we associate to each zone a in \mathcal{A} a *signature*, defined as the time series of the user attendance p_a^t . Several examples of signatures used in our study are given below in this section.

By classifying the different signatures, we can detect the major types of land-use in \mathcal{A} , mapped to the most significant classes obtained, as well as zones with a unique and specific land-use, which remain outside the major classes obtained by the classification algorithm. This methodology is further detailed in [20], where it shows excellent performance in terms of accuracy when compared with ground truth information. Our methodology can be summarized through the following steps, further detailed below:

- We associate two signatures to each SFR-IRIS, one signature for the period before the lockdown and a second period for the period during the lockdown.
- We use the Pearson correlation coefficient to measure the similarity between any two signatures in the dataset.
- By using an unsupervised machine learning approach, we classify the signatures in the dataset. Each class has a *characteristic signature*, obtained as the mean of all the signatures in that class.

- We label each class of signatures obtained in the previous step using information from municipality surveys produced before the lockdown.

5.1 Signature definition

Similarly to [20], we build a *typical week* signature. However, a particularity of our study is that each SFR-IRIS has two associated signatures one for the typical week before the lockdown and one for the typical week during the lockdown. Practically, the typical week signatures we use represent the mean of the two weeks of data we have before and during the lockdown, respectively. The signatures are normalized using a standard-score approach, in order to tackle the fact that the actual value of the user attendance p_i^a depends on multiple factors, such as the size of the area a or the smartphone adoption rate in the given area. This also allows comparing the signatures obtained before and during the lockdown, despite the significant difference in terms of user attendance observed in Section 4. Practically, the normalization allows us to take into account only the trend of each signature (i.e., the normalized typical median week), which is actually a more relevant feature for signature classification.

5.2 Hierarchical clustering

To detect the major types of land use in the studied area, we automatically classify the signatures by using an agglomerative hierarchical clustering algorithm [34], which is a non-parametric and unsupervised method. As an unsupervised method, the classification is done directly using the test data, on which we have no prior knowledge of classes. More precisely, we use the average linkage method, also known as Unweighted Pair Group Method Arithmetic Mean (UPGMA) algorithm [22], which means that each newly formed class will be represented by the mean of its elements.

As well known in the literature, the UPGMA algorithm requires the use of a *stopping rule* to select the best level of aggregation in the clustering process to be used, i.e., the best clusters and their composition. Practically, the hierarchical algorithm builds all possibilities of clusters by fusing similar classes in an iterative manner, from one element per class until one single class with all elements. A dendrogram usually shows a graphical information of all clustering possibilities.

With respect to the stopping rule, [20] proposed a composed metric mixing six different indices from the literature. However, these classical stopping indices were designed for totally different applications than ours, and it was difficult to find the intuition behind them. Therefore, we adapted the framework designed in [20], by using a more intuitive stopping rule, based on the Pearson correlation coefficient. Since our objective is to cluster together similar signatures, represented as time series, the Pearson correlation coefficient of two time series is a metric easy to understand and interpret. Therefore, at each step in the UPGMA algorithm, we group together the two clusters with the largest Pearson correlation coefficient and we stop the clustering once the correlation coefficient drops below a certain threshold. Since a correlation coefficient above 0.7 is usually considered to indicate highly correlated variables, and after conducting a series of tests that demonstrate that the obtained clusters are well balanced (i.e. not one single highly dominant class), we decided to use a correlation coefficient threshold of 0.7 in the rest of the study.

5.3 Labeling of zones

Using the UPGMA clustering algorithm, we obtain a total of 52 classes. Nonetheless, only a few of these classes contain the majority of elements. For instance, before the lockdown, three classes cover 80.5% of the total signatures and, during the lockdown, two classes contain 91.3% of the signatures, the rest being small classes. The two most important classes are present both before and during the lockdown. However, the importance of the classes is reversed between the two periods. The third most important class, colored in purple, is mostly present before the lockdown.

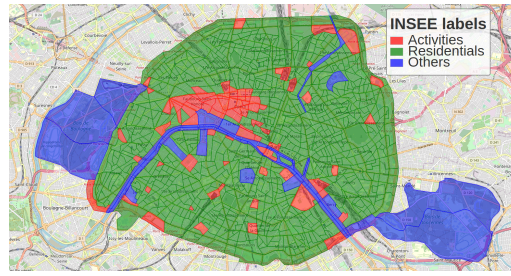


Fig. 7. Map visualization of INSEE class labels.

To label the classes using known land-use categories, we use the INSEE class labels as a reference. Based on census data, INSEE defines three different types of labels: activity, residential and diverse. IRIS labeled as residential have a population that is generally between 1,800 and 5,000. In terms of habitat types, they are homogeneous and their limits are based on the major cuts in the urban fabric (main roads, railways, waterways, etc.). Activity IRIS have more than 1,000 employees and have at least twice as many salaried jobs as the resident population. Finally, the diverse category covers large, specific areas that are sparsely inhabited and have a large surface area (leisure parks, port areas, forests, etc.).

In Figure 7, we can see that residential IRIS in green are predominant in the INSEE labeling. This corresponds to about 87% of IRIS in Paris. Activity IRIS, in red, are less prevalent, about 9%. They are mainly in the west center and occupied by main train stations, while some other small IRIS are scattered all over Paris. Then, diverse IRISs (denoted as *others* in the Figure), in blue, cover two large Parisian parks, the Seine river and other smaller parks.

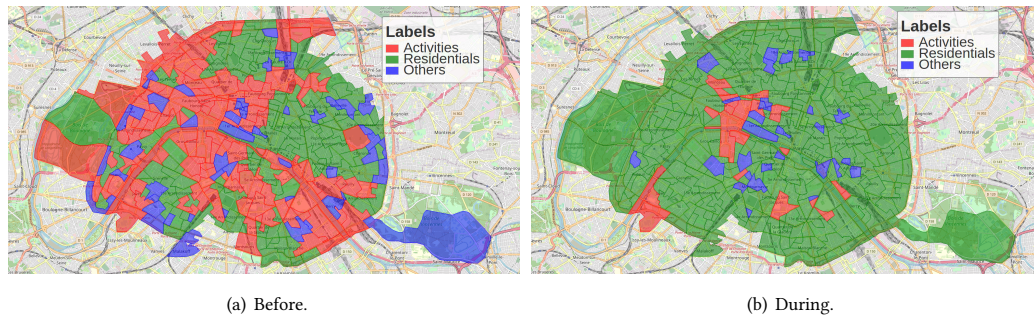


Fig. 8. SFR-IRIS labels before and during lockdown.

In order to label the SFR-IRIS in our study, we superimpose the classes obtained with the INSEE labels to those obtained by our clustering approach. A class will be labeled as *activity* or *residential* if its spatial occupation mainly corresponds to the corresponding INSEE label. For the classes covering very few elements, with an outlier behavior, we label them together as *others*. It is worth mentioning that the methodology developed in [20], which we adopt in our study, has been shown to provide better land use information than official census data. We also recall that, for privacy compliance, some SFR-IRIS zones are the aggregation of several INSEE IRIS. Therefore, we are not looking for a perfect mapping between INSEE data and our automatically obtained classes. As a matter of fact, as discussed below, the *activity* and *others* classes are over-represented in our classification method with respect to INSEE data. To summarize, following this methodology on our dataset containing in total 326 SFR-IRIS zones, we obtain three main labels, as shown in Figure 8, which are in line with the three classes defined by INSEE.

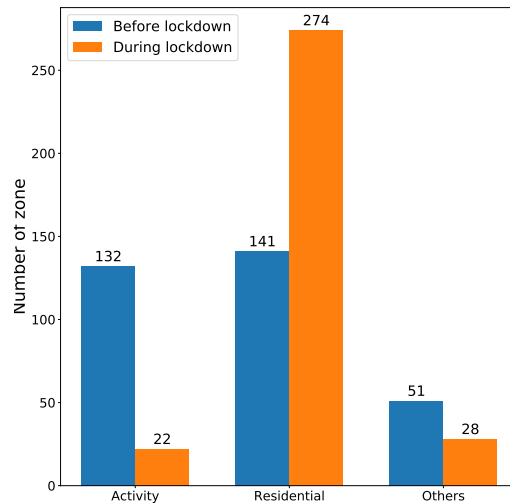


Fig. 9. Distribution of the number of SFR-IRIS zones per label before and during lockdown.

In more detail, we proceed as follows. First, all the classes with less than 5 elements (i.e., signatures) are grouped together as *others*. Next, using the classes identified before the lockdown as a reference, we label the *activity* and *residential* areas in Figure 8 using the INSEE labels as reference. The red cluster in Figure 8(a) covers 68% of the INSEE activity areas, therefore we label it as *activity*. It is important to note that the activity cluster detected by our method is larger than the one defined by INSEE (which is based on simple demographic statistics) and also covers some zones denoted as residential or diverse by INSEE. This can be explained by the mixed land-usage of many zones in Paris, which combine housing and commercial and office activities. The second important class before the lockdown, depicted in green in Figure 8(a), represents for 96% zones considered as residential by the INSEE classification. We therefore label this class as *residential*. Figure 9 shows the distribution of SFR-IRIS among the three identified main labels in the period before the lockdown, where we can notice the number of activity and residential SFR-IRIS is comparable.

For the period during the lockdown, shown in Figure 8(b), we apply the same labels previously defined. It is important to note that we do not have any ground truth information regarding the land use behavior during the lockdown. This underlines the interest in our methodology, where we cluster together with the signatures before and during the lockdown, which allows us to directly observe the change in land use once the lockdown is established. Figure 9 shows that the number of SFR-IRIS labeled as residential increases drastically, from 141 to 274. All of this is due to the decrease in activity caused by the lockdown. The number of zones labeled as *others* also decreased, some of them becoming residential. This allows a quantification of the expected impact of the strict lockdown imposed by the French governments on the population: most activities stopped and the population stayed at home, transforming the entire city of Paris into a big residential area. However, it is interesting to notice that some activity or diverse areas still appear, even during the lockdown. We provide a detailed analysis of this phenomenon below.

5.4 Signature analysis

As explained, each class has an associated characteristic signature, defining the general trend of user attendance for the zones in that class. Figure 10 shows the main characteristic signatures in the dataset, namely those associated with the activity and residential label for both the period before the lockdown and the one during the lockdown. Figure 10(a) shows the characteristic signature for the activity class. We can see that, during the weekend, user attendance decreases with respect to weekdays. Figure 10(b) shows the characteristic signature for the residential class. User attendance is fairly the same all the week even though there is a small decrease on Sunday. The 24h pattern, with increased communication during the day and reduced communication at night is clearly visible in both characteristic signatures. However, when comparing the activity and residential signatures, we notice that the low level of the time series is reached later in residential areas (the drop happens when people go to sleep) than in activity areas (the drop happens when people leave the activity area).

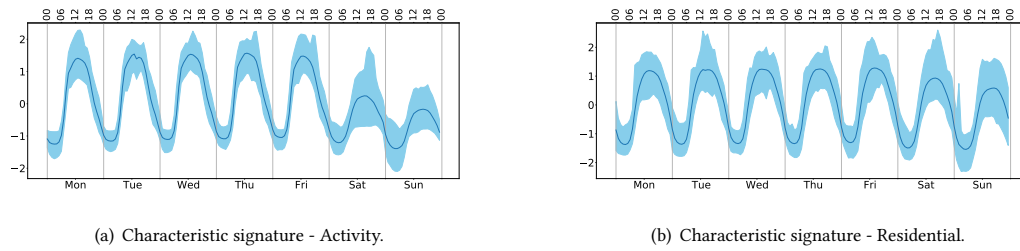


Fig. 10. Characteristic signatures for activity and residential zones.

As shown in Figure 9, almost all SFR-IRIS labels turned into residential during the lockdown. This is indeed the expected behavior following the sanitary measures taken by the government. However, some of the zones are still classified as activity, some join the others class, and we even notice some cases where former residential zones become activity areas. In this section, we analyze the evolution of labels before and during the lockdown. Please note that to better understand the results, we show below the non-normalized signatures, while the normalized versions are used in the clustering process.

5.4.1 *Evolution of the others label.* Figure 11 shows the evolution of the others label. The SFR-IRIS in yellow represent zones that were classified as others before the lockdown and became residential areas during the lockdown. In line with the goal of the lockdown, we consider this to be normal behavior and we do not investigate this further. However, we can also distinguish one SFR-IRIS, in black, labeled as others before the lockdown and that became an activity area during the lockdown. Also, we observe a few blue SFR-IRIS, which are zones labeled as others before the lockdown and that kept this label during the lockdown.

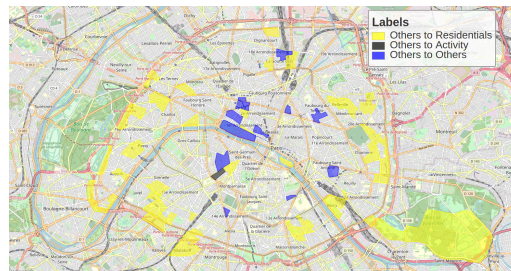


Fig. 11. Evolution of the others label when the lockdown is established.

Figure 12 shows an example of a SFR-IRIS signature labeled as others that became a residential area during the lockdown. This IRIS is located in the southwest of Paris, covering among others a sport complex that hosts the matches of the Paris Saint-Germain football team before the lockdown. We can see in Figure 12(a) that on Saturday, before the lockdown, there is an outstanding peak in the signature in the late afternoon. This corresponds to a match day. During the lockdown, in Figure 12(b), the signature changes, becoming a residential zone with similar trends to the characteristic signature of residential IRIS zones, shown in Figure 10(b). This can be explained by the interruption of all sports competitions in France during the lockdown.

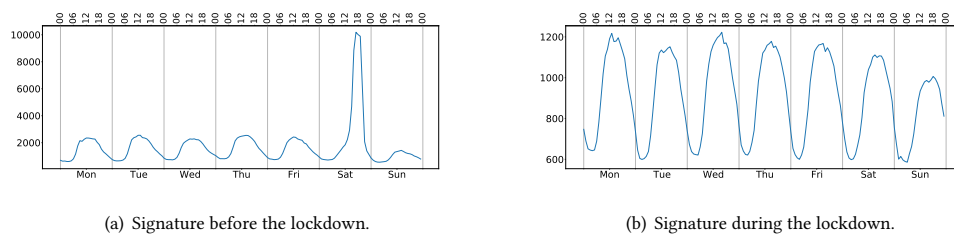


Fig. 12. Signature of an SFR-IRIS zone labeled as others before lockdown and residential during the lockdown.

For the only black SFR-IRIS in Figure 11, which evolved to an activity area, Figure 13 shows its signature before and during the lockdown. We can see in Figure 13(a) that, before the lockdown, the signature trend is different from the main characteristic signatures in the dataset (Figure 10). The signature shows that, for the first two days of the week, user attendance is lower than for the rest of weekdays, more similar to a Saturday. Then, for the rest of the weekday, user attendance increases, with a specific shape on Wednesday. This day user attendance increases until a peak reached towards the late afternoon, followed by a sharp decrease. This particular trend is the reason why, before the lockdown, this IRIS was labeled as others. The signature during the lockdown, in Figure 13(b), becomes similar to the characteristic signature of activity SFR-IRIS in Figure 10(a), where the evolution of user attendance during the weekday is stable, and it decreases during the weekend. After detailed verification, we identify that the corresponding area actually covers two major residential facilities for dependent elderly people (known as Ehpads in French). They are considered as medical-enhanced retirement homes adapted to the reception of residents with limited autonomy. The specific shape on Wednesday, before the lockdown, can be explained by the visits of families to their relatives who are residing in these Ehpads. Indeed, on Wednesday, French schoolchildren, junior high school, and high school students do not have classes. This day is therefore privileged for these types of visits to elderly family members. The reason that can explain why this SFR-IRIS turns into an activity area instead of a residential one during the lockdown is that, as medical-enhanced retirement places, Ehpads maintained their activities during the lockdown, even though visitation was interrupted. As hospitals, health care staff was present on-site during this period to provide residents care, resulting in this specific activity signature.

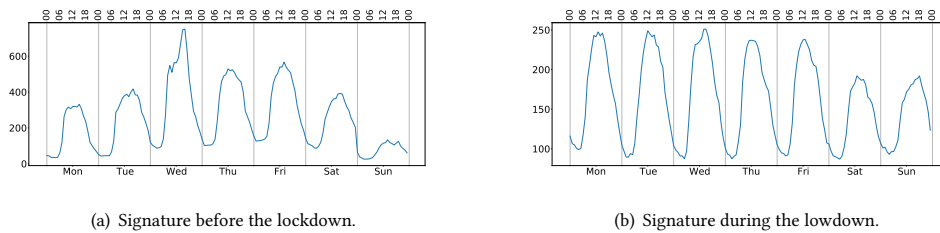


Fig. 13. Signature of an SFR-IRIS zone labeled as other before lockdown and activity during the lockdown.

For the blue SFR-IRIS in Figure 11, labeled as others both before and during the lockdown, their signature is still atypical compared to the main activity and residential signatures. However, the signature can change, even drastically, from one period to the other. An example of this is shown in Figure 14. Before the lockdown, in Figure 13(a), there is no clear difference between weekday (business day) and weekend. The Sunday user attendance is similar to that of Tuesday and Friday. As for Monday, Wednesday, Thursday, and Saturday, their user attendance is also similar and higher than the remaining days of the week. During the lockdown, in Figure 13(b), the signature is unique, noisy, and with a very low amplitude compared with the normal behavior. We can see that, for weekdays, user attendance (although generally low) is more important in the morning than in the afternoon. We can also see that some days, between 12pm and 6am, there are practically no users. On Saturday, almost all day, the attendance varies little, while on Sunday there are times when there is no user. In fact, these signatures correspond to an SFR-IRIS in the center of Paris, which covers a very popular mid-size park - Jardin des Tuileries - and a famous museum - the Louvre. A highly touristic area in regular times, this area was almost abandoned during the lockdown.

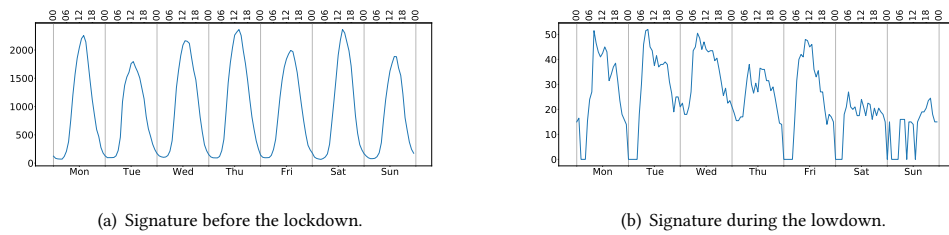


Fig. 14. Signature of an SFR-IRIS zone labeled as other both before and during the lockdown.

5.4.2 *The others label during the lockdown.* On the other hand, some SFR-IRIS labeled as activity or residential before the lockdown are labeled as others during the lockdown. Figure 15 shows these zones on a map view. Green areas represent residential SFR-IRIS that became outliers, while red areas are former activity areas with an outlying behavior during the lockdown.

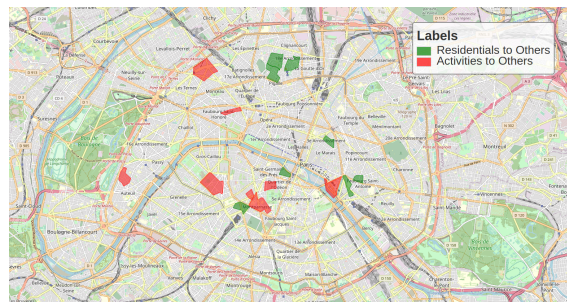


Fig. 15. Zones labeled as activity and residential that become others.

Figure 16 is a good example of this, showing a SFR-IRIS with a residential signature that is labeled as others during the lockdown. Before the lockdown, in Figure 16(a), the signature is similar to the residential characteristic signature in Figure 10(b). However, some particularities can be observed, namely the fact that the signature presents multiple peaks in some days, in the early afternoon, and in the evening. During the lockdown, in Figure 16(b), the signature shows that almost all activity disappears, with the exception of some sporadic peaks, one on Tuesday between 12am and 1pm and on Thursday at 10am. This SFR-IRIS corresponds to an area that covers mainly the Opera Bastille. The fact that this area is labeled as residential before the lockdown is simply a side effect of evening events at the opera, that attract the public at late hours, a behavior associated with residential areas by our classification method. However, the signature during the lockdown, when the Opera is closed, clearly signals the fact that this is not a typical residential zone.

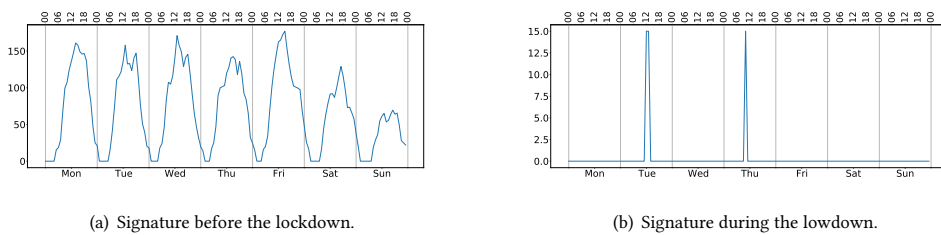


Fig. 16. Signature of an SFR-IRIS zone labeled as residential before lockdown and others during the lockdown.

As for SFR-IRIS that were labeled as activity zones before the lockdown and that have become others, in Figure 17(a), the signature is similar to the activity characteristic signature (Figure 10(a)) where during the weekend the user attendance is reduced. However, we observe a peculiarity here, the user density presence on Monday being rather high compared to the rest of the weekdays. During the lockdown, in Figure 17(b), the signature is peculiar and unique. For

almost all days of the week, the user attendance is stable, even though it remains higher on Mondays. Nevertheless, on Tuesday and even more on Wednesday, the user attendance drops. There are several educational institutions in this IRIS, like middle and high school, and even a university (University Paris 2), which can partially explain this behavior, but we do not have any clear explanation for the highly atypical signature observed during the lockdown.

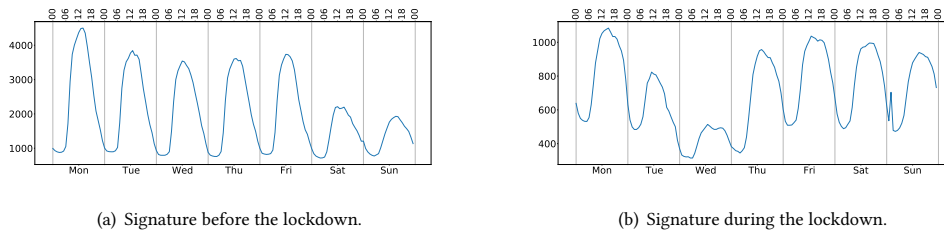


Fig. 17. Signature of an SFR-IRIS zone labeled as activity before lockdown and others during the lockdown.

5.4.3 Evolution of the activity label. Figure 18 shows the evolution of the SFR-IRIS zones labeled as an activity before or during the lockdown. Green IRIS are activity IRIS that transformed into residential IRIS during the lockdown. Yellow refers to SFR-IRIS which remain labeled as activity during the lockdown as well. Finally, SFR-IRIS colored in red were not labeled as activity areas before the lockdown, but became so during the lockdown.

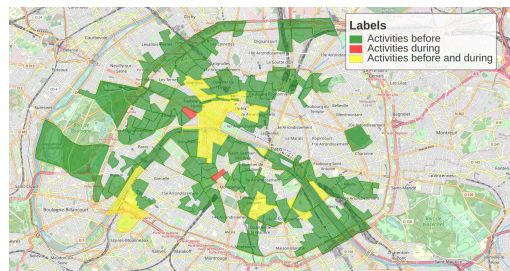


Fig. 18. Evolution of the activity label.

We consider residential IRIS to be the standard during the lockdown, therefore SFR-IRIS which became residential are normal. A good example of this type of IRIS is represented by major train station areas. Figure 19 shows the evolution of the signature of one of the biggest train stations in Paris, Gare du Nord. Before the lockdown, the signature is labeled as activity, due to the net decrease of user attendance during the weekend. However, we can notice that the behavior during weekdays is a little different from the activity characteristic signature shown in Figure 10(a). Indeed, before the lockdown, in Figure 19(a), the user attendance increases throughout the day until 5 p.m., when the peak is reached. On some days, there is also a second peak before the *rush hour* at 5 p.m. During the lockdown, in Figure 19(b), the signature became similar to the residential characteristic signature in Figure 10(b). It is also important to note the amplitude of the signature, which is much lower during the lockdown, indicating that it is in fact the weekday attendance that is highly reduced. This type of evolution is similar for most major train stations in Paris, which is expected considering that these train stations are surrounded by residential buildings, which become dominant during the lockdown.

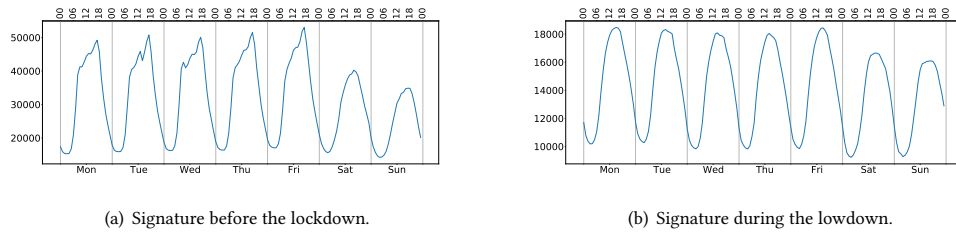


Fig. 19. Signature of a SFR-IRIS zone labeled as activity before lockdown and as residential during the lockdown.

In spite of the lockdown, some activity IRIS remain labeled as activity areas. Most of these IRIS are places of power like ministries, parliament, presidential palace, or embassies, but we also detect some hospitals. For example, before the lockdown, the signature of the SFR-IRIS covering the Val-de-Grace hospital, shown in Figure 20(a) corresponds well to the characteristic signature of activity areas (Figure 10(a)). Some particularities can be noted, such as the higher attendance on Wednesday (when children do not attend school and many parents do not work). During the lockdown, the hospital (Figure 20(b)) signature is classical, very similar to the activity characteristic signature. We can simply note a much lower signature amplitude during the lockdown, related to a lower general user attendance.

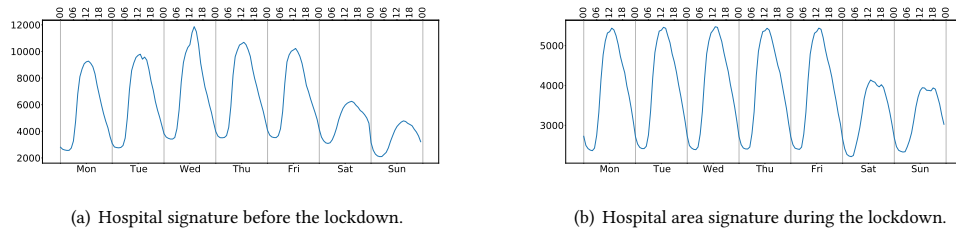


Fig. 20. Signatures of a SFR-IRIS zone labeled as activity both before and during the lockdown.

Finally, we distinguish two SFR-IRIS that became activity areas during the lockdown. One of the IRIS transforms from residential to activity and the other from others to activity (already discussed in Sec. 5.4.1). The signatures of the former, before and during the lockdown, are shown in Figure 21. Before the lockdown, in Figure 21(a), the signature is typical to the residential characteristic signature. During the lockdown, in Figure 21(b), we observe a typical signature for an activity IRIS, where there is a net decrease in user attendance during the weekend compared to weekdays. This SFR-IRIS corresponds to a part of Champs Elysees, labeled residential before the lockdown because it is a famous touristic zone where there are many people during evenings and during the weekend. It is interesting to notice that this pattern changes during the lockdown, with much lower amplitude and an activity-like pattern, which is probably a result of some businesses and offices still active in the area during the lockdown.

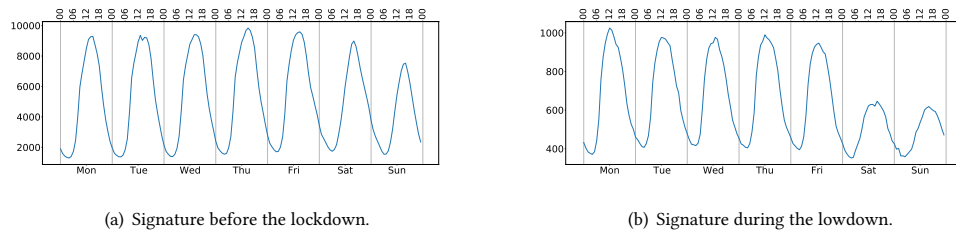


Fig. 21. Signature of a zone labeled as residential before the lockdown and as activity during the lockdown.

It is important to note that a number of touristic and cultural areas, such as Opera Bastille and Champs Elysees in our examples, are labeled as residential areas before the lockdown by our framework. As explained, this is due to

increased weekend attendance and a late evening drop-in user presence, similar to residential areas. As shown in [20], with more fine-grained spatial and temporal data (a point every 15 minutes at a cell level in [20]), the signature-based approach can distinguish these touristic areas from residential ones.

5.4.4 Evolution of the residential label. Becoming residential is the norm for most SFR-IRIS during the lockdown. Figure 22 shows that, in addition to SFR-IRIS which remain residential during the lockdown, depicted in yellow, many zones turn to residential during the lockdown, in red. A small number of IRIS, in green, were labeled as residential before the lockdown and switched to activity or others during the lockdown. Sections 5.4.2 and 5.4.3 give a more detailed discussion on these evolution-to-residential cases.

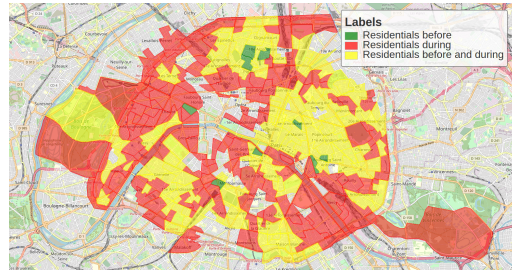


Fig. 22. Evolution of the residential label.

6 Zones preferences in population mobility

People's movements are typically characterized by routine behavior, having *regular and daily circadian rhythm* (e.g., from home to work) of actions guided by their decisions. This brings a certain *repetitiveness* [13, 24] in people mobility relative to a few locations. Additionally, when choosing an itinerary, people mobility shows a *shortest-path tendency* to reach their destination, also known as *desire line* [48]. The desire line is the shortest line (in terms of geographic distance or time) between origin and destination, and expresses the way a person would like to go if such a way was available [48]. The usual human itinerary is also characterized by a *maximum travel distance* (i.e., maximum displacement) and a *limited radius of gyration* (i.e., the total area of movement), both following a truncated power-law when aggregated over the whole user population [10]. In summary, people tend to perform most displacements around a limited number of locations; they roam close by their main locations (e.g., home or work); and their travel distances are restricted to the area containing their principal life-related locations. [3, 10, 32, 33, 47].

Although the privacy concerns which guided the collection of the dataset used in this work do not allow us to identify repetitiveness, shortest-path tendency, and locality in mobility on a per individual basis as in [33], we can still capture mobility-related decisions and trends of the concerned population on an aggregated basis. In general terms, we want to identify the SFR-IRIS as having important participation in the daily routine of people according to their habits of movements and visits. For this, we study the daily mobility flows of a population. Our strategy is methodologically structured as follows and was inspired by [32, 33, 37]:

- First, we create a time-dependent weighted graph to represent the interaction between flows and SFR-IRIS (cf. Section 6.1);
- Second, we use graph centrality measures to quantify the importance of each SFR-IRIS in the dataset in terms of different perspectives related to the way they are visited or frequented by the population. Indeed, some strategically positioned SFR-IRIS are likely to be more popular, to attract or to disseminate more people to neighboring areas (cf. Section 6.2);
- Finally, we measure the impact-factor of each SFR-IRIS by combining results of their centrality metrics and attendance, thus defining a specific mobility-related metric for each area in the city of Paris (cf. Section 6.4).

The intuition behind the impact-factor assigned to each SFR-IRIS is the identification of two main zone features: (1) To spot zones having a high population concentration due to the way people attend and visit them (e.g., transit IRIS); and (2) To point out the population variation and impact in the SFR-IRIS before and during the lockdown periods. We believe the computed impact-factor may in future analysis be correlated to the spatiotemporal risk of disease contamination and propagation in the corresponding zone. In other words, high population concentration

at certain hours and geographical areas intuitively increases the probability of agglomeration and, consequently, the contamination and propagation risks. As we will show hereafter, such risks are amplified by the high density of some zones, i.e., small geographical zones having a high population attendance. Of course, this correlation with the virus evolution is much more complex and should also account for sanitary measures and population contagiousness level. However, it is clear that social distancing is more complicated to achieve in dense areas, while transit areas, situated at the crossroad of different population flows, provide more opportunities for a virus to cross between different communities.

6.1 Time-varying weighted graph construction

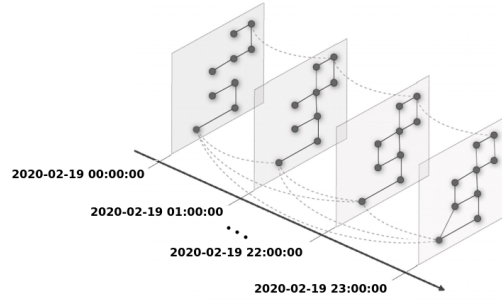


Fig. 23. Time-dependent graph example with time bins of one hour.

People move, build their trajectories, and interact with places and other people inside the sequence of urban areas composing their trajectories (in our case, population aggregated flows). In order to analyze and to capture the interaction between people and the urban scenario during the time, we built a *time-dependent weighted graph* $G(V, E)$. First, according to the dataset temporal aggregation into bins of constant time intervals, one graph is built per time interval, describing the flows of users, i.e., the *edges* E , between neighboring SFR-IRIS zones, which represent the *vertices* V . Next, the time-dependent weighted graph is composed of the set of temporal discretized graphs, where each vertex is connected to itself in the future. Figure 23 shows an example of a time-dependent graph composed of graphs generated for intervals of one hour (please note that this time interval is just given as an example, and it is not the one we use in our study, as explained below). The *time-dependent weighted graph* construction is the most important step in the impact-factor assignment procedure since all composing metrics capturing population visitation behavior are extracted from this graph. Therefore, we pay special attention to the definition of vertices and edges, as well as to the computation of edge weights.

Each SFR-IRIS is uniquely identified and its identifier can be mapped to its corresponding coordinates in each dataset department, resulting in a straightforward neighborhood extraction. From this point, the following rules have to be respected:

- A vertex $v \in V = \{v_1, v_2, \dots, v_w\}$ in the graph represents a SFR-IRIS. A vertex is created for a SFR-IRIS **if and only if** the latter observes **at least** one attending user. Note that a SFR-IRIS not respecting this condition (i.e., with no observed user) will not appear as a vertex in the graph of the considered time period. Hence, neighbor SFR-IRIS in the physical geographic area may not necessarily be neighbors in the time-dependent graph.
- Similarly, an edge $e \in E = \{e_1, e_2, \dots, e_y\}$ is created between two vertices **if and only if** their corresponding SFR-IRIS sequentially cover **at least** one human mobility flow in the considered time period.
- Additionally, we also represent in the graph the flows having as the source and destination the same SFR-IRIS, also denoted as *self-flows*, meaning stationary users or mobile flows inside a given SFR-IRIS for the considered time period.
- To capture the difference between input and output flows in a SFR-IRIS, we build a directed graph. The direction of edges represents the direction of flows, i.e., from which IRIS the flows leave and to which neighbor IRIS they arrive.
- The weight of the edges is defined by the inverse of the number of flows between each pair of SFR-IRIS, when a flow exists. For incoming (respectively outgoing) edges, we make the number of flows equal to the *InComing*

(respectively *OutGoing*) field in the dataset. The number of self-flows is computed as the number of *MoveInside* flows from the dataset.

- We discretized the whole dataset time period into constant time intervals of 24 hours. We then average each type of user flows (inside, incoming, outgoing) per day per SFR-IRIS. These average values are used to build *daily weighted graphs* (as detailed in the previous steps) that, when aggregated, resulting in the final temporal-dependent weighted graph.

Overall, we obtain a temporal-dependent weighted graph with 324 vertices and 2,187 edges for the period before lockdown and a second graph with 321 vertices and 2,021 edges for the period during the lockdown.

6.2 Capturing behaviors with centrality

6.2.1 Desire lines: Betweenness centrality. The concept of desire lines states that people tend to choose the shortest paths to arrive at their destinations. We argue that, in our graph model, the tendency of an SFR-IRIS to be present along desire lines can be captured by the betweenness centrality metric [8]. In our case, as we are using the inverse of the flow as the weight of the edges, the concept of the shortest path on the graph (i.e. path of minimum weight) can be understood as the most used path in practice.

Considering the graph $G(V, E)$, let $\sigma_{(u,x)}$ be the number of shortest paths between the vertices u and x . Now let $\sigma_{(u,x)}(w)$ be the number of shortest paths between u and x that pass through the vertex w . Then, the betweenness value C_{bt} of a vertex w is given by:

$$C_{bt}(w) = \sum_{u,x \in V} \frac{\sigma_{(u,x)}(w)}{\sigma_{(u,x)}}. \quad (1)$$

Vertices with high betweenness centrality are those that lay on most of the shortest routes they represent highly frequented SFR-IRIS, situated at the intersection of multiple urban mobility flows. Intuitively, these areas should receive increased attention with respect to sanitary measures enforcement compared to SFR-IRIS with a low betweenness centrality.

6.2.2 Locality of people movement: Closeness centrality. Human mobility is generally confined to a limited area. Even if people are not using the shortest routes, they are at least not moving far from their home location. We consider the closeness centrality metric [43] as a way to measure the local importance of an SFR-IRIS with respect to human mobility. Closeness centrality is computed based on the distance between all pairs of vertices in the network. In this work, we compute the closeness centrality based on the mobility flows, meaning that two SFR-IRIS are considered close if there is a high number of flows between them, regardless of the actual physical distance between the two areas.

Considering the $G(V, E)$ graph, the closeness centrality C_{cl} for a certain vertex $w \in V$ is defined as follows:

$$C_{cl}(w) = \frac{n-1}{\sum_{v \in V} d(v, w)}, \quad (2)$$

where $d(v, w)$ is the shortest-path distance between vertices v and w (i.e. the inverse of the mobility flow, in our case), and n is the number of nodes that are reachable by w . This assigns higher values for vertices closer to the remaining vertices of the network. In a city, we expect these to be highly central zones in the graph, major public transportation hubs, connected with many other areas of the city. Once again, SFR-IRIS with high closeness centrality intuitively present a higher risk from a virus propagation point of view.

6.2.3 Number of frontiers: Degree centrality. Not only mobility flows, but simple geography also plays a role in the importance of an urban area. For example, the higher is the number of frontiers a certain SFR-IRIS zone has with other zones, the higher is its disease-spreading potential, since it can be easily reached by different mobility flows. We consider the degree centrality metric [7] to capture this property.

The degree centrality shows how many connections a SFR-IRIS has in the weighted graph. It is calculated based on the number of neighbor vertices a certain vertex w has in the $G(V, E)$ graph. Remember that, because of the two strict rules used at the weighted graph construction (cf. Section 6.1), geographical SFR-IRIS neighbors may not necessarily be graph neighbors. Therefore, only visited SFR-IRIS appear in the graph and thus have a frontier with other SFR-IRIS. The degree centrality C_{dg} for a certain vertex $w \in V$ is defined as follows:

$$C_{dg}(w) = \sum_{v \in V} ngh(v, w), \quad (3)$$

where $ngh(v, w)$ identifies the neighboring SFR-IRIS of vertex w . This assigns higher values for highly central vertices in terms of number of frontiers.

6.3 Centrality results

In this section, we calculate the centralities previously defined, for both time periods (before and during lockdown). The results in terms of centralities show a ranking of the importance of each SFR-IRIS in the presented period according to the way they are frequented and located. The results also (1) show the impact the lockdown had on each SFR-IRIS, and (2) allow us to understand some factors explaining this lockdown impact, through the different mobility properties captured by each centrality.

Betweenness centrality. In Figure 24, the betweenness centrality is represented for all the SFR-IRIS in Paris, before and during the lockdown. The centrality values in the figure are normalized with respect to the maximum value of each period, so they should not be directly compared. We remind that, in this case, the way we define our edge weights produces a SFR-IRIS ranking that reflects the most-used paths in the city. Indeed, while the SFR-IRIS with a high betweenness before the lockdown preserve their relative importance, we observe a much more uniform distribution of the metric, especially in more peripheral areas. Thus, in addition to the reduction in the flow of people due to the lockdown period, already outlined in Section 4, we also observe a change in the distribution of the most used paths, with some alternative routes that are much better represented during the lockdown.

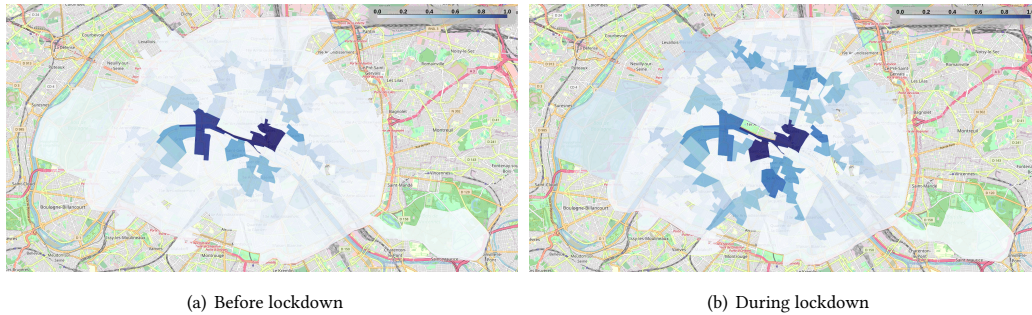


Fig. 24. Betweenness centrality normalized by the maximum value of each period in each dataset.

Closeness centrality. The darker regions of Figure 25 indicate the most central regions considering the flow of people, that is, the regions with high closeness centrality represent the most visited SFR-IRIS in our dataset in each period. Once again, we normalize the centrality results with respect to the maximum value observed in each period. This allows us to detect the areas most used by people who continue moving, even though we had a significant flow reduction in general during the lockdown period.

Similarly to betweenness results, we perceive the regions in the center of Paris to rank the highest in terms of closeness centrality before the lockdown period (Figure 25(a)). During the lockdown period (Figure 25(b)), we can also observe that the Eastern and Northern regions of Paris rank similarly to the central region. This is because the highest rates of flow reduction occurred mainly in central Paris. In this way, the flows of people existing during the lockdown are more uniformly distributed across Paris and not concentrated in the center.



Fig. 25. Closeness centrality normalized by the maximum value of each period.

Degree centrality. The degree centrality shows how connected is an SFR-IRIS in the weighted graph, i.e., how many neighbor zones of a SFR-IRIS disseminate (respectively attract) population flows to (respectively from) the given area. Regardless of the intensity of transiting flows, two SFR-IRIS are considered as neighbors only if there is at least one transiting flow between them. In other words, all neighboring SFR-IRIS in the weighted graph are also neighbors in the geographical space, but not all geographically close SFR-IRIS are neighbors in the graph.

Note that Figure 26 only shows the degree results for the period before the lockdown, as the city does not change shape during the lockdown. This lack of difference is due to the fact that the lockdown does not affect how the SFR-IRIS are connected in the graph (since the condition of “at least one transiting flow” does not change the graph connectivity), although the intensity of flows among SFR-IRIS reduces significantly during the lockdown.

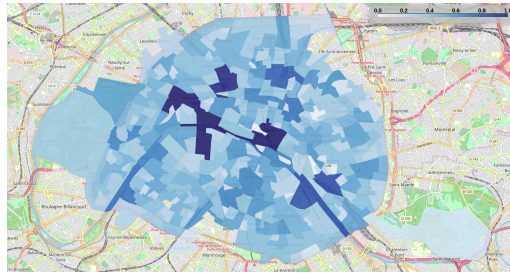


Fig. 26. Degree centrality normalized by the maximum value.

Figure 26 shows the degree centrality results normalized by the maximum observed value. The highest SFR-IRIS degree in the dataset is connected to 16 other SFR-IRIS zones and it is located in the central areas of Paris (Figure 26). We also verified that the degree is not a metric that is proportional to the actual area of the IRIS (i.e., covering large zones is not a condition for having many neighbor zones), which can be explained by the graph construction methodology used in our study, as well as by the geographical location of the concerned IRIS zones.

6.4 Impact-factor of zones

For the calculation of the impact-factor, we use a linear combination [15] using the three centralities: betweenness (C_{bt}), closeness (C_{cl}), degree (C_{dg}), as well as the density (D) (combining user attendance and SFR-IRIS area, as defined in Sec. 3). We analyze the Pearson correlation among all metrics considered in the impact-factor computation. A certain correlation exists between the metrics, but it is not significant, as shown in Table 1, confirming that each metric captures a complementary movement/activity behavior of the city population. Such a weak correlation justifies considering the four metrics in the following linear combination model:

$$RF(i) = \alpha C_{bt}(i) + \beta C_{cl}(i) + \gamma C_{dg}(i) + \delta D(i), \quad (4)$$

where i is the given SFR-IRIS and $RF(i)$ is the impact-factor of i . $C_{bt}(i)$, $C_{cl}(i)$, $C_{dg}(i)$ are the normalized values of centrality for SFR-IRIS i , and $D(i)$ is the normalized value of density of i . The coefficients α , β , γ , and δ represent the

proportion of each metric to be considered in the model, i.e., $\alpha + \beta + \gamma + \delta = 1$. The values of these coefficients are defined as follows.

		C_{bt}	C_{cl}	C_{dg}	D
Before	C_{bt}		0.44	0.55	0.07
	C_{cl}			0.46	0.33
	C_{dg}				-0.02
	D				
During	C_{bt}	0.40	0.54	0.13	
	C_{cl}			0.51	0.28
	C_{dg}				0.05
	D				

Table 1. Pearson correlation between centrality metrics and density.

First, we compute the maximum value per day of each metric normalized by the Z-Score. This computation is done separately for the periods before and during lockdown; this is done for comparison reasons among the two periods. Second, we compute the arithmetic mean among all days of the maximum Z-Score values. This mean is denoted k_x , where $x = \{C_{bt}, C_{cl}, C_{dg}, C_D\}$, and it refers to the centrality or density metrics in the impact-factor model, i.e., $k_{C_{bt}}$, $k_{C_{cl}}$, $k_{C_{dg}}$, and k_D . Finally, we compute the α , β , γ , and δ coefficients as follows:

$$\alpha = \frac{k_{C_{bt}}}{k_{C_{bt}} + k_{C_{cl}} + k_{C_{dg}} + k_D} \quad \beta = \frac{k_{C_{cl}}}{k_{C_{bt}} + k_{C_{cl}} + k_{C_{dg}} + k_D}$$

$$\gamma = \frac{k_{C_{dg}}}{k_{C_{bt}} + k_{C_{cl}} + k_{C_{dg}} + k_D} \quad \delta = \frac{k_D}{k_{C_{bt}} + k_{C_{cl}} + k_{C_{dg}} + k_D}$$

Note that the computation of coefficients allows us to weigh the difference in the maximum values among the different adopted metrics. For example, the betweenness centrality discovers only a few important SFR-IRIS zones, which results in a high α coefficient. On the other hand, in the closeness centrality, the vast majority of SFR-IRIS have high values, resulting in a low β coefficient. The difference in centrality values are clearly visualized in the centrality results of each dataset (e.g., Figures 24 and 25), as well as mirrored in the computation of the maximum Z-Score values.

The resulting arithmetic means are $k_{C_{bt}} = 6.85$, $k_{C_{cl}} = 2.21$, $k_{C_{dg}} = 4.32$, and $k_D = 7.48$ as well as the related coefficients are $\alpha \approx 0.33$, $\beta \approx 0.11$, $\gamma \approx 0.21$, $\delta \approx 0.35$. The impact-factor of SFR-IRIS i is then computed using the following linear combination:

$$RF(ir) = 0.33 \cdot C_{bt}(i) + 0.11 \cdot C_{cl}(i) + 0.21 \cdot C_{dg}(i) + 0.35 \cdot D(i). \quad (5)$$

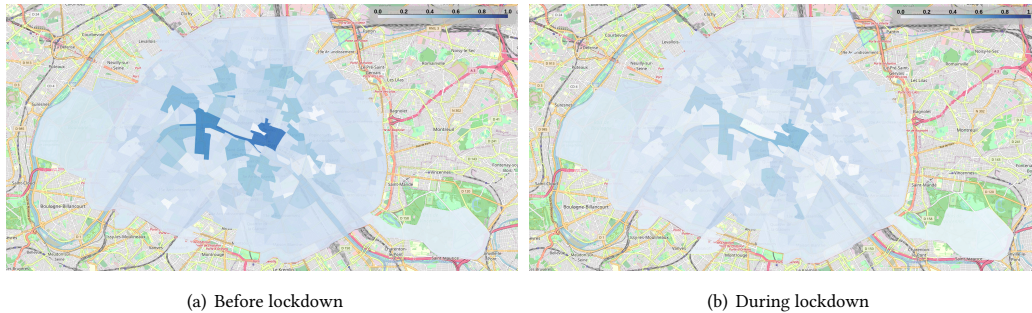


Fig. 27. Impact-factor of the different SFR-IRIS.

Figure 27 shows the result of the impact-factor, before and during the lockdown. For the calculation of the impact-factor, we consider the value of centralities and density normalized by the maximum value obtained when considering

both periods (i.e. before and during the lockdown), which allow us (1) to capture the flow reduction during the lockdown and (2) to compare the impact-factor results between the two periods.

In particular, some SFR-IRIS in central Paris have the highest impact-factor values during both periods. The maximum computed impact-factor value is approximately 0.70 before the lockdown and of 0.40 during the lockdown. This lower impact-factor value during the lockdown is due to the reduction of the number of flows, as well as to changes in the centrality and density ranking.

Finally, Figure 27 shows that during the lockdown: (1) IRIS zones that were top-ranked before the lockdown have their impact-factor reduced, which is explained by the reduction in attendance and visitation of zones, and (2) some new top-rank impact-factor IRISs appear (discussed below), reflecting the density increase in certain zones (e.g., parks).

7 Correlation between impact-factor and land use

After investigating separately the land-use dynamics (using the signature-based approach in Section 5) and the impact-factor of the SFR-IRIS (using the time-varying graph approach in Section 6), we check whether a certain correlation exists between the two. For this, Table 2 summarizes the results obtained in Section 5. Our automatic land use labeling approach detects three classes of SFR-IRIS: activity, residential, and others. Some of these SFR-IRIS keep the same label before and during the lockdown, and we denote them as *stable* in the following. Other SFR-IRIS, denoted as *unstable*, change their label once the lockdown is imposed. As shown in Table 2, most SFR-IRIS become residential during the lockdown, while the number of stable and unstable zones is quite similar (164 vs. 160).

Classes	Before	During	%	Unstable		Stable
				Only Before	Only During	Before and During
Activity	132	22	↓ 83%	112	2	20
Residential	141	274	↑ 94%	9	142	132
Others	51	28	↓ 45%	39	16	12
Total	324	324		160	160	164

Table 2. Number of SFR-IRIS in each class.

Figure 28 shows the normalized values of the three centralities, the density, and the impact-factor for the three classes of SFR-IRIS. Two important general remarks can be made from this figure:

- The SFR-IRIS labeled as others constantly score lower than the two other classes, for all the metrics and for both time periods.
- The average values for all the metrics and the classes decrease during the lockdown period. We note that, in this figure, the metrics were normalized with respect to the maximum value over the two periods, so they can be directly compared.

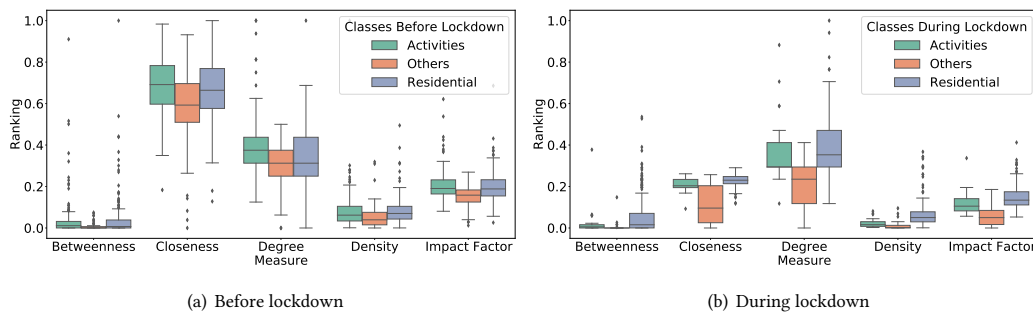


Fig. 28. Normalized values of centrality, density and impact-factor per land use class.

Considering the different metrics individually, the following observations can be made:

- Regarding the betweenness centrality, the top SFR-IRIS before the lockdown are rather uniformly distributed between the activity and residential classes. However, during the lockdown, almost all the top SFR-IRIS in terms of betweenness centrality belong to the residential class. This confirms the fact that activity areas, even if some of them are preserved, lose their central role in terms of transit during the lockdown.
- The closeness centrality is the metric with the most significant drop during the lockdown. This is a result of the significant drop in flow intensity, indicating that people are more confined and travel less during the lockdown. There is practically no SFR-IRIS with a high closeness centrality during the lockdown.
- The degree centrality values are identical before and during the lockdown. Nevertheless, their distribution slightly changes, with the most central SFR-IRIS becoming residential, in line with what happens in the entire city.
- In terms of density, we notice that before the lockdown the activity and residential classes are very similar. However, during the lockdown, the residential SFR-IRIS are by far the most important in terms of density, confirming that the population is mostly confined in these areas.
- The impact-factor is a composite metric of the four others, so it follows the same trend, with a significant reduction during the lockdown. Residential zones also become the leading class in terms of impact-factor during the lockdown.

	IRIS top 25%	Probability	Unstable IRIS	Probability	Stable IRIS	Probability
Activities	36	0.27	30	0.26	6	0.3
Residential	40	0.28	1	0.11	39	0.3
Others	5	0.1	5	0.12	0	0

Table 3. Analysis of the top 25% SFR-IRIS in terms of impact-factor before the lockdown.

As discussed above, SFR-IRIS zones with a high impact-factor are more likely to play an important role in the virus propagation and they should be given more attention in terms of sanitary measures. We, therefore, focus on these areas in the following, extracting the top 25% SFR-IRIS in terms of impact-factor, before and during the lockdown, respectively.

Table 3 shows the results for the period before the lockdown. We divide the results per land-use class, while also considering stable and unstable SFR-IRIS, as defined above. We show the number of top SFR-IRIS in each class. However, we note that the distribution of SFR-IRIS per class is not uniform, as shown in Table 2. Therefore, we also show the probability of an SFR-IRIS belonging to a certain class to be ranked in the top 25% impact factor. If no correlation would exist between the two events, we would expect a probability of 0.25 for each class. This means that classes with a probability higher than 0.25 in Table 3 are more likely to have a high impact-factor than expected in a random process, while values lower than 0.25 signal a probability lower than expected.

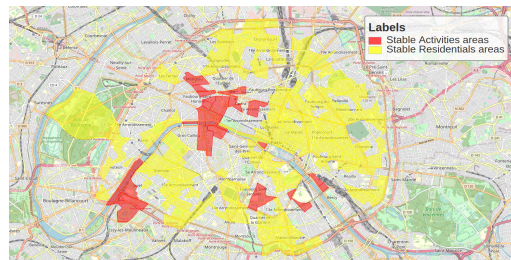


Fig. 29. Stable activity and residential SFR-IRIS

We observe that SFR-IRIS labeled as activity or residential before the lockdown are slightly more likely to have a high impact-factor, while zones labeled as others have a much lower probability than expected to be in the top 25%. When we divide the SFR-IRIS into stable and unstable, we notice that stable SFR-IRIS, both activity and residential, are those that are the most likely to appear in the top 25% before the lockdown. We show these stable SFR-IRIS on

the Paris map in Figure 29. We observe that they mostly cover peripheral parts of the city, especially on the Eastside, where we already observed a significant increase in closeness centrality in Figure 25.

	IRIS top 25%	Probability	Unstable IRIS	Probability	Stable IRIS	Probability
Activities	3	0.14	0	0	3	0.15
Residential	77	0.28	35	0.24	42	0.32
Others	1	0.04	1	0.06	0	0

Table 4. Analysis of the top 25% SFR-IRIS in terms of impact-factor during the lockdown.

Table 4 shows the same type of results, regarding the top 25% in terms of impact-factor, for the period during the lockdown. Residential SFR-IRIS, which become highly dominant in the city, are also more likely to be in the top 25% impact-factor, while activity and other areas seem less likely to score high during the lockdown. Once again, stable SFR-IRIS are the reason behind the high impact-factor of residential areas.

As noticed in all the results in this section, the SFR-IRIS labeled as others do not score very high in terms of impact-factor. Therefore, in Figure 30, we focus on the 6 SFR-IRIS labeled as others that present a high impact-factor (5 areas in the period before the lockdown and 1 area during the lockdown). By inspecting the map, we notice that, before the lockdown, these SFR-IRIS correspond to important transportation hubs (e.g. train stations such as Gare du Nord, Gare de Bercy, Gare de Neuilly). The red SFR-IRIS on the map, the only one labeled as others presenting a high impact-factor during the lockdown, covers the area of the Montmartre neighborhood, well known for its animated lifetime, indicating this behavior did not entirely stop during the lockdown, despite the mobility restriction policies. All these examples represent very well the fact that the impact-factor metric really manages to detect SFR-IRIS with the announced properties: areas where relatively important flows of people temporarily meet.

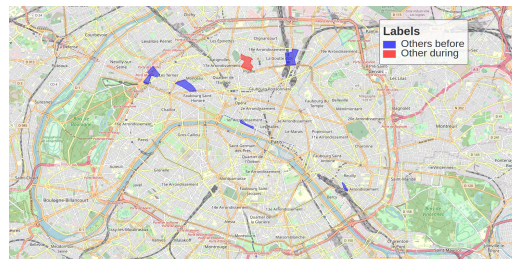


Fig. 30. SFR-IRIS labeled as others and ranked in the top 25% in terms of impact-factor

8 Conclusion

Urban mobility has been thoroughly investigated using mobile phone data, outlining important properties. For example, people's movements are typically characterized by routine behavior, having regular and daily circadian rhythm (e.g., from home to work) of actions guided by their decisions. This brings a strong repetitive behavior in people's mobility relative to a few locations, it reveals path preferences on the way people daily move, and it gives indications to the most representative attendance classes of urban areas.

Covid-19 related lockdown restrictions, however, highly perturbed our mobility patterns and use of urban spaces. In this paper, we investigated, under different perspectives, the impacts caused by harsh lockdown conditions in Spring 2020 in the city of Paris. As such, we proposed a data-driven methodology toward the tracking of the evolution in space and time of population mobility habits. Our results reveal that the lockdown mobility restrictions significantly reduced the number of observed users in most parts of the city, mainly in central Paris. However, some zones located at the periphery of the city observed an increase in human attendance during the lockdown. As a side effect of the lockdown, urban areas where specific activities were observed in normal periods became residential zones (i.e. showing residential-like attendance patterns) during the lockdown. Although this was somehow expected, it is striking to observe how an entire diverse and lively city practically becomes a huge residential-like area.

The other impacting effect of lockdown is represented by the changes imposed on the mobility habits of the population. In particular, our study reveals that top-ranked visited zones located in central Paris suffered from a

significant reduction in population flows and lost their popularity during the lockdown. Hence, we observed a decentralization in the popularity of zones, with the Eastern region of Paris becoming a major hub during the lockdown. In particular, (1) shorter distance displacement restrictions imposed by the lockdown increased visitation to more local zones, i.e., the ones close to population main residence; (2) decentralization was also favored by a shift in the path preferences of the still-moving population; and (3) jogging activities, allowing people to be outside their residences, impacted parks visitation, which had an increase in visitation during the lockdown.

We combine several graph centrality and attendance metrics in an impact factor, computed for each zone and indicating its importance in the urban area with respect to human mobility habits. We argue that zones with a high impact factor have the potential to play a more important role in the virus propagation and that they should be given increased attention when adopting sanitary measures. Our analysis shows that the zones with the highest impact factor, both before and during the lockdown, are located in central Paris. However, during the lockdown, peripheral zones, especially in the North and East sides, also present an increased impact factor. We show that stable zones, which keep the same land use before and during the lockdown, are more likely to have a high impact factor, while areas with an outlying behavior are rarely represented in the top-ranked zones in terms of impact factor.

We believe that the methodology we propose in this study can produce output directly usable by anti-Covid navigation applications, which propose routes avoiding high-risk spots, as already done for pollution hotspots. Our methodology can be used in real-time (or close to real-time), in order to draw the authorities' attention to particular areas, where sanitary measures need to be enforced. It is worth mentioning here that the impact factor, which we compute over a 2 weeks period, can be computed at a finer granularity, daily or even hourly, providing a detailed spatio-temporal evolution of the impact factor of zones in a metropolitan area. Finally, our results could also be a valid indication of activity loss, and consequently, of the economic impact (e.g., for activity-labeled areas) that different city zones suffered during the lockdown.

Acknowledgments

This study was funded by Inria in the context of the COVID-19 mission, launched in 2020. We wish to thank Loic Lelievre and the entire SFR Geostatistics team for providing us with the valuable dataset used in this study and for patiently answering our numerous questions.

References

- [1] Lauri Aalto, Nicklas Göthlin, Jani Korhonen, and Timo Ojala. 2004. Bluetooth and WAP push based location-aware mobile advertising system. In *MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*. ACM Press, New York, NY, USA, 49–58.
- [2] Nadeem Ahmed, Regio A. Michelin, Wanli Xue, Sushmita Ruj, Robert Malaney, Salil S. Kanhere, Aruna Seneviratne, Helge Janicke, and Sanjay K. Jha. 2020. A Survey of COVID-19 Contact Tracing Apps. *IEEE Access* 8 (July 2020), 134577–134601.
- [3] Licia Amichi, Aline Carneiro Viana, Mark Crovella, and Antonio A F Loureiro. [n.d.]. Understanding individuals' proclivity for novelty seeking. <https://hal.inria.fr/hal-02944150>
- [4] Licia Amichi, Aline Carneiro Viana, Mark Crovella, and Antonio A.F. Loureiro. 2020. Understanding Individuals' Proclivity for Novelty Seeking. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems* (Seattle, WA, USA) (*SIGSPATIAL '20*). Association for Computing Machinery, New York, NY, USA, 314–324.
- [5] Necati Ayan, Nilson Damasceno, Sushil Chaskar, Peron de Sousa, Arti Ramesh, Anand Seetharam, and Antonio de A. Rocha. 2021. Characterizing Human Mobility Patterns During COVID-19 using Cellular Network Data. In *LCN - IEEE 46th Conference on Local Computer Networks*.
- [6] Hamada S. Badr and Lauren M. Gardner. 2020. Limitations of Using Mobile Phone Data to Model COVID-19 Transmission in the USA. *The Lancet* 21, 5 (November 2020).
- [7] Albert-László Barabási and Márton Pósfai. 2016. *Network science*. Cambridge University Press, Cambridge. <http://barabasi.com/networksciencebook/>
- [8] Ulrik Brandes. 2008. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks* 30, 2 (2008), 136–145. <https://doi.org/10.1016/j.socnet.2007.11.001>
- [9] Sanja Brdar, Katarina Gavric, Dubravko Culibrk, and Vladimir Crnojevic. 2016. Unveiling Spatial Epidemiology of HIV with Mobile Phone Data. *Scientific Reports* 6 (January 2016).
- [10] D Brockmann, L Hufnagel, and T Geisel. 2006. The scaling laws of human travel. *Nature* 439, 7075 (Jan. 2006), 462–465.
- [11] Claude Castelluccia, Natalia Bielova, Antoine Boutet, Mathieu Cunche, Cedric Lauradoux, Daniel Le Metayer, and Vincent Roca. 2020. *ROBERT: ROBUst and privacy-presERving proximity Tracing*. Technical Report. Inria.
- [12] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2021. Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening. *Nature* 589, 7840 (2021), 82–87.
- [13] Guangshuo Chen, Aline Carneiro Viana, Marco Fiore, and Carlos Sarraute. 2019. Complete Trajectory Reconstruction from Sparse Mobile Phone Data. *EPJ Data Science* (Oct. 2019). <https://doi.org/10.1140/epjds/s13688-019-0206-8>
- [14] Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and Mobility: User Movement in Location-based Social Networks. In *Proc. International Conference on Knowledge Discovery and Data Mining*.

- [15] Mehmet Şimşek and Henning Meyerhenke. 2020. Combined Centrality Measures for an Improved Characterization of Influence Spread in Social Networks.
- [16] Katarzyna Czech, Anna Davy, and Michal Wielekhowski. 2021. Does the COVID-19 Pandemic Change Human Mobility Equally Worldwide? Cross-Country Cluster Analysis. *Economies* 9, 4 (November 2021).
- [17] F. De Rango, P. Fazio, and S. Marano. 2009. Utility-Based Predictive Services for Adaptive Wireless Networks With Mobile Hosts. *IEEE Transactions on Vehicular Technology* 58, 3 (2009), 1415–1428.
- [18] Laura Di Domenico, Giulia Pullano, Chiara E. Sabbatini, Pierre-Yves Boëlle, and Vittoria Colizza. 2020. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. *BMC Medicine* 18, 1 (30 Jul 2020), 240. <https://doi.org/10.1186/s12916-020-01698-4>
- [19] Mariem Fekih, Tom Bellemans, Zbigniew Smoerda, Patrick Bonnel, Angelo Furno, and Stephane Galland. 2020. A Data-driven Approach for Origin-Destination Matrix Construction from Cellular Network Signalling Data: A Case Study of Lyon Region France. *Transportation* (2020).
- [20] A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, and Z. Smoreda. 2017. A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas. *IEEE Transactions on Mobile Computing* 16, 10 (2017), 2682–2696. <https://doi.org/10.1109/TMC.2016.2637901>
- [21] Rahul Goel, Rajesh Sharma, and Anto Aasa. 2021. Understanding Gender Segregation through Call Data Records: An Estonian Case Study. *PLoS ONE* 16, 3 (March 2021).
- [22] Ilan Gronau and Shlomo Moran. 2007. Optimal Implementation of UPGMA and Other Common Clustering Algorithms. *Inform. Process. Lett.* 104, 6 (2007), 205–210.
- [23] Georgios Hadjidemetriou, Manu Sasidharan, Georgia Kouyialis, and Ajith Parlikad. 2020. The Impact of Government Measures and Human Mobility Trend on COVID-19 Related Deaths in the UK. *Transportation Research Interdisciplinary Perspectives* 6, 100167 (July 2020).
- [24] Daniel S. Hamermesh. 2003. *Routine*. Working Paper 9440. National Bureau of Economic Research.
- [25] Yusuke Hara and Hiromichi Yamaguchi. 2021. Japanese Travel Behavior Trends and Change under COVID-19 State-of-Emergency Declaration: Nationwide Observation by Mobile Phone Location Data. *Transportation Research Interdisciplinary Perspectives* 9 (March 2021).
- [26] Corina Iovan, Ana-Maria Olteanu-Raimond, Thomas Couronné, and Zbigniew Smoreda. 2013. Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies. In *Geographic Information Science at the Heart of Europe*. Springer, 247–265.
- [27] Shan Jiang, Gaston A Fiore, Yingxiang Yang, Joseph Ferreira Jr, Emilio Frazzoli, and Marta C González. 2013. A review of urban computing for mobile phone traces: Current methods, challenges and opportunities. In *Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing*. ACM, 2.
- [28] Douglas J. Leith and Stephen Farrell. 2020. Coronavirus Contact Tracing: Evaluating the Potential of using Bluetooth Received Signal Strength for Proximity Detection. *ACM SIGCOMM Computer Communication Review* 50, 4 (October 2020).
- [29] Yannick Leo, Anthony Busson, Carlos Sarraute, and Eric Fleury. 2016. Call Detail Records to Characterize Usages and Mobility Events of Phone Users. *Computer Communications* 95 (2016), 43–53.
- [30] Tong Li, Mingyang Zhang, Yong Li, Emil Lagerspetz, Sasu Tarkoma, and Pan Hui. 2021. The Impact of Covid-19 on Smartphone Usage. *IEEE Internet of Things Journal* (April 2021).
- [31] Xin Lu, Erik Wetter, Nita Bharti, Andrew J. Tatem, and Linus Bengtsson. 2013. Approaching the Limit of Predictability in Human Mobility. *Scientific Reports* 3, 1 (Oct. 2013), 2923. <https://doi.org/10.1038/srep02923>
- [32] Eduardo Mucelli Rezende Oliveira, Aline Carneiro Viana, K.P. Naveen, and Carlos Sarraute. 2017. Mobile data traffic modeling: Revealing temporal facets. *Computer Networks* 112 (Jan. 2017), 176–193. <https://doi.org/10.1016/j.comnet.2016.10.016>
- [33] Eduardo Mucelli Rezende Oliveira, Aline Carneiro Viana, Carlos Sarraute, Jorge Brea, and Ignacio Alvarez-Hamelin. 2016. On the regularity of human mobility. *Pervasive and Mobile Computing* 33 (Dec. 2016), 73–90. <https://doi.org/10.1016/j.pmcj.2016.04.005>
- [34] Fionn Murtagh and Pedro Contreras. 2011. Algorithms for Hierarchical Clustering: An Overview. *Data Mining and Knowledge Discovery* 2, 1 (2011), 86–97.
- [35] Diala Naboulsi, Marco Fiore, Stephane Ribot, and Razvan Stanica. 2016. Large-Scale Mobile Traffic Analysis: A Survey. *IEEE Communications Surveys & Tutorials* 18, 1 (2016), 124–161. <https://doi.org/10.1109/comst.2015.2491361>
- [36] Apollinaire Nadembega, Abdelhakim Hafid, and Tarik Taleb. 2015. Mobility-Prediction-Aware Bandwidth Reservation Scheme for Mobile Networks. *IEEE Transactions on Vehicular Technology* 64, 6 (2015), 2561–2576.
- [37] Eduardo Mucelli Rezende Oliveira and Aline Carneiro Viana. 2014. From routine to network deployment for data offloading in metropolitan areas. In *2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 126–134. <https://doi.org/10.1109/sahcn.2014.6990335>
- [38] Nuria Oliver, Bruno Lepri, Harald Sterly, Renaud Lambiotte, Sebastien Deletaille, Marco De Nadai, Emmanuel Letouze, Albert Ali Salah, Richard Benjamins, Ciro Cattuto, Vittoria Colizza, Nicolas de Cordes, Samuel P. Fraiberger, Till Koebe, Sune Lehmann, Juan Murillo, Alex Pentland, Phuong N. Pham, Frederic Pivetta, Jari Saramaki, Scarpino Samuel V., Michele Tizzoni, Stefaan Verhulst, and Patrick Vinck. 2020. Mobile Phone Data for Informing Public Health Actions Across the COVID-19 Pandemic Life Cycle. *Science Advances* 6, 23 (June 2020).
- [39] Sangchul Park, Gina Jeehyun Choi, and Haksoo Ko. 2020. Information Technology–Based Tracing Strategy in Response to COVID-19 in South Korea—Privacy Controversies. *JAMA* 323, 21 (April 2020).
- [40] Pedro S. Peixoto, Diego Marcondes, Claudia Peixoto, and Sergio M. Oliva. 2020. Modeling Future Spread of Infections via Mobile Geolocation Data and Population Dynamics. An Application to COVID-19 in Brazil. *PLoS ONE* 15, 7 (July 2020).
- [41] Lu Xin; Bengtsson Linus; Holme Petter. 2012. Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences* 109, 29 (2012), 11576–11581.
- [42] Giulia Pullano, Laura Di Domenico, Chiara E. Sabbatini, Eugenio Valdano, Clément Turbelin, Marion Debin, Caroline Guerrisi, Charly Kengne-Kuetche, Cécile Souty, Thomas Hanslik, Thierry Blanchon, Pierre-Yves Boëlle, Julie Figoni, Sophie Vaux, Christine Campèse, Sibylle Bernard-Stoecklin, and Vittoria Colizza. 2021. Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature* 590, 7844 (01 Feb 2021), 134–139. <https://doi.org/10.1038/s41586-020-03095-6>
- [43] Gert Sabidussi. 1966. The centrality index of a graph. *Psychometrika* 31, 4 (1966), 581–603. <https://doi.org/10.1007/BF02289527>
- [44] Carlos Sarraute, Pablo Blanc, and Javier Burrioni. 2014. A Study of Age and Gender Seen through Mobile Phone Usage Patterns in Mexico. In *ASONAM - IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.
- [45] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of Predictability in Human Mobility. *Science* 327, 5968 (2010).

- [46] Victor Soto and Enrique Frias-Martinez. 2011. Automated Land Use Identification using Cell-Phone Records. In *Proceedings of the 3rd ACM Workshop on Hot Topics in Planet-scale Measurement*. 17–22.
- [47] Douglas Do Couto Teixeira, Aline Carneiro Viana, Mário S. Alvim, and Jussara M. Almeida. 2019. Deciphering Predictability Limits in Human Mobility. In *ACM SIGSPATIAL 2019 - 27th International Conference on Advances in Geographic Information Systems*. Chicago, United States. <https://doi.org/10.1145/3347146.3359093>
- [48] James A. Throgmorton and Barbara. Eckstein. 2000. Desire Lines: The Chicago Area Transportation Study and the Paradox of Self in Post-War America. <http://www.nottingham.ac.uk/3cities/throgeck.htm>
- [49] Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M. Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C GONZALEZ, and Vittoria Colizza. 2014. On the use of human mobility proxies for modeling epidemics. *PLoS Computational Biology* 10, 7 (2014), 1–15. <https://doi.org/10.1371/journal.pcbi.1003716>
- [50] Jameson L. Toole, Michael Ulm, Marta C. Gonzalez, and Dietmar Bauer. 2012. Inferring Land Use from Mobile Phone Activity. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. 1–8.
- [51] V.A. Traag, A. Browet, F. Calabrese, and F. Morlot. 2011. Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Inference. In *SocialCom - IEEE 3rd International Conference on Social Computing*.
- [52] Bengtsson L.; Lu X.; Thorson A.; Garfield R.; von Schreeb J. 2011. Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data. *PLOS Medicine* 8, 8 (2011), 1–9.
- [53] C. Jason Wang, Chun Y. Ng, and Robert H. Brook. 2020. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. *JAMA* 323, 14 (March 2020).
- [54] Amy Wesolowski, Nathan Eagle, Andrew J. Tatem, David L. Smith, Abdisalan M. Noor, Robert W. Snow, and Caroline O. Buckee. 2012. Quantifying the Impact of Human Mobility on Malaria. *Science* 338, 6104 (October 2012), 267–270.
- [55] Elias Willberg, Olle Jarv, Tuomas Vaisanen, and Tuuli Toivonen. 2021. Escaping from Cities during the COVID-19 Crisis: Using Mobile Phone Data to Trace Mobility in Finland. *International Journal of Geo-Information* 10, 2 (February 2021).
- [56] Robin Wilson, Elisabeth Zu Erbach-Schoenberg, Maximilian Albert, Daniel Power, Simon Tudge, Miguel Gonzalez, Sam Guthrie, Heather Chamberlain, Christopher Brooks, Christopher Hughes, Lenka Pitonakova, Caroline Buckee, Xin Lu, Erik Wetter, Andrew Tatem, and Linus Bengtsson. 2016. Rapid and Near Real-Time Assessments of Population Displacement Using Mobile Phone Data Following Disasters: The 2015 Nepal Earthquake. *PLoS currents* 8 (24 Feb 2016), ecurrents.dis.d073fbeece328e4c39087bc086d694b5c. <https://doi.org/10.1371/currents.dis.d073fbeece328e4c39087bc086d694b5c> 26981327[pmid].
- [57] Huso Yi, Shu Tian Ng, Aysha Farwin, Amanda Pei Ting Low, Cheng Mun Chang, and Jeremy Lim. 2021. Health Equity Considerations in COVID-19: Geospatial Network Analysis of the COVID-19 Outbreak in the Migrant Population in Singapore. *Journal of Travel Medicine* 28, 2 (March 2021).
- [58] Guolei Zhou, Chenggu Li, and Jing Zhang. 2020. Identification of urban functions enhancement and weakening based on urban land use conversion: A case study of Changchun, China. *PLOS ONE* 15, 6 (06 2020).
- [59] Ying Zhou, Renzhe Xu, Dongsheng Hu, Yang Yue, Qingquan LI, and Jizhe Xia. 2020. Effects of Human Mobility Restrictions on the Spread of COVID-19 in Shenzhen, China: A Modelling Study using Mobile Phone Data. *The Lancet Digital Health* 2, 8 (August 2020), 417–424.