Explaining the Initial Spread of COVID-19 using Mobile Positioning Data: a Case Study

Stefano Maria Iacus Carlos Santamaria Francesco Sermi Spyridon Spyratos
Dario Tarchi Michele Vespe
European Commission, Joint Research Centre
June 9, 2020

Abstract
This work provides an analysis on the potential of aggregate and anonymised mobility data from mobile phones to explain the recent COVID-19 outbreak in Europe. The data were processed by the European Commission in collaboration with EU Mobile Network Operators (MNOs) to improve the quality of modelling and forecasting for the pandemic at EU level. Connectivity matrices derived from MNOs at EU scale can help the understanding of the initial dynamics of the spread. This work is expected to provide input to epidemiological modelling in informing policies on de-escalation of control measures and recovery. Connectivity data between administrative areas can also be used to estimate impacts and spread of future outbreaks. A case study of France shows that mobility alone can explain from 52% up to 92% of the initial spread of the virus, while there is no effect before the outbreak, and has a slow decay effect after lock-down measures are effective. It also emerges that internal mobility is more important than mobility across geographical departments.

Highlights
• this work shows the importance of mobility data in understanding the initial spread of COVID-19;
• through a case study, it is shown that mobility alone can explain from 52% up to 92% of the observed phenomenon of excess cumulative deaths;
• in the analysed data, internal mobility within a department is more important than mobility across departments;
• the reduction mobility and the flattening of the cumulative excess deaths are highly correlated with a lag of about 14-20 days;
• the case study also shows that, when containment measures are in place, the impact of mobility on the spread of the virus reduces.

Keywords — COVID-19, coronavirus, mobile positioning data, excess deaths
1 Introduction

By means of a letter sent on 8 April 2020, the European Commission asked European Mobile Network Operators (MNOs) to share fully anonymised aggregate mobility data. The aim of the initiative is further specified by the Recommendation to support exit strategies through mobile data and apps. \[\ldots\] for research to combat the virus, modelling to understand how the virus will spread and modelling of the economic effects of the crisis. In particular, the data will help to understand and model the spatial dynamics of the epidemic and to assess the impact of social distancing measures (travel limitations, non-essential activities closures, total lock-down etc.) on mobility. This is essential \[\ldots\] to support the exit strategy with data-driven models that indicate the potential effects of the relaxation of the social distancing measures”. This aim is inline with calls from other researchers Kraemer et al. (2020) who conclude that “much further work is required to determine how to balance optimally the expected positive effect on public health with the negative impact on freedom of movement, the economy, and society at large\[1\].

The Joint European Roadmap towards lifting COVID-19 containment measures\[2\] introduces how the data can be beneficial to epidemiological modelling: “\[\ldots\] mobile network operators can offer a wealth of data on mobility, social interactions \[\ldots\] Such data, if pooled and used in anonymised, aggregated format in compliance with EU data protection and privacy rules, could contribute to improve the quality of modelling and forecasting for the pandemic at EU level.” While the value of mobile positioning data to describe human mobility has been explored (see e.g. [Csaji et al.] (2013)), its potential for epidemiology demonstrated (see e.g. [Wesolowski et al.] (2012)) also recently in the specific context of COVID-19 in China by Jia et al. (2020), in this paper we aim at exploring first insights using real data on the possible uses for policy in the EU.

The letter sent on 8 April by the European Commission sets out the purpose of the data processing, which aims to:

- “understand the spatial dynamics of the epidemics using historical matrices of mobility national and international flows;
- quantify the impact of physical distancing measures (travel limitations, non-essential activities closures, total lock-down etc.) on mobility, including the phasing out of such measures as relevant;
- feed epidemiological models, contributing to the evaluation of the effects of physical distancing measures on the reduction of the rate of virus spread in terms of reproduction number (expected number of secondary cases generated by one case);
- feed models to estimate the economic costs of the different interventions, as well as the impact of specific control measures on intra-EU cross border flows due to the epidemic.”

An example of the use of the data is the following: analysing the cumulative number of excess deaths in France, it appears that the COVID-19 had initially spread out of the Haut-Rhin Department. JRC analysed the data provided by Orange regarding aggregate and anonymised movements from Haut-Rhin to other Departments to investigate their role in the epidemiological dynamics of the virus. These movements have been found to largely explain the excess mortality with respect to previous year, a much more reliable indicator of the virus spread than number of cases. The study finds significantly higher mortality in those Departments that are more connected to the department of Haut-Rhin. This relationship is valid until mobility is severely reduced following to the national lock-down. This confirms once more the importance of mobility for understanding and forecasting the dynamics of the epidemic.

This work is organised as follows: Section 2 describes the mobility network operator raw data at hand. Section 3 introduces the concept of connectivity matrix and how it is obtained from the raw MNO data. Section 4 shows the main findings about the impact of mobility on the spread of the COVID-19 through a case study. Section 5 summarises the conclusions.

2 Mobile Positioning Data

The letter defines the basic characteristics of fully anonymised and aggregate data to be shared with the Joint Research Centre. Nevertheless, in order to act rapidly, it was agreed with MNOs to share available products with no additional developments on their side. This means that the products received by the Joint Research Centre from the different MNOs would be extremely heterogeneous, differing in terms of spatial granularity, refresh rate, latency, definitions and extrapolation. The Joint Research Centre would process this heterogeneous set of data in order to create a set of mobility indicators and maps (the “common denominator”), which are comparable across countries.

Data from MNOs are provided to JRC in the form of Origin-Destination-Matrices (ODMs). Each cell \([i − j]\) of the ODM shows the overall number of “movements” (also referred to as “trips” or “visits”) that have been recorded from the origin geographical reference area \(i\) to the destination geographical reference area \(j\) over the reference period. The nature of the geographical reference area is defined by the spatial granularity adopted by the MNO to build the ODM; it can be based on provinces, municipalities, districts, postal-code areas or even ad hoc grids defined by the MNO. “Origin” and “Destination” areas are defined in different ways by each MNO. Typically, a “dwelling time” (varying across MNOs from a few minutes to a few hours) is defined as the minimum period that a user needs to spend in a given reference area for that to be considered as origin or destination. In other cases, origin and destination are defined as the geographical areas where most of the time is spent during the reference period. Also the reference period of the ODM is different across MNOs, ranging from one hour to one day.

---

1 Commission Recommendation (EU) on a common Union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data, 2020/518.

2 Joint European Roadmap towards lifting COVID-19 containment measures.
of “movements” between couples of geographical reference areas. In order to define what a “movement” is, we need to
describe the data processing that takes place at the MNO. As part of the normal service operations, each MNO records
the unique identifier of all users (customers or inbound roamers) connecting to the network’s antennas. This process is
based on both Call Detail Records (CDRs) and eXtended Detail Records (XDRs), which are generated with the normal
use of the mobile phone (calls, SMSs, internet and a long list of background activities that our smartphones automatically
carry out) see e.g., [Bonato et al., 2020]. The position of the users is generally associated to that of the tower hosting the
antenna which the user is connected to (or alternatively, it is assessed through triangulation techniques). As a first form
of privacy protection, the MNO deletes the actual user’s unique identifier by replacing it with a dummy identifier that is
randomly renewed every 24 hours. This process is referred to as “pseudo-anonymisation” (Article 6(4) and Article 25(1)
of the GDPR [European Union, 2016]). Such procedure allows the MNO to track a user only within the 24 hours. Some
MNOs do not even keep track of a user and proceed to aggregation on the fly to ensure short term anonymisation, and
then intersect these aggregates to determine mobility. Through the above described methodology, the MNO aggregates
all users over the reference period in a mobile-cells based ODM. Because of the dwelling time, not all movements are
accounted for; for instance, a bike trip across the neighbour might not be included. The following step for the MNO
is to aggregate all movements between mobile cells within the same origin reference area and mobile cells within the
same destination area, according to the chosen spatial granularity. Thanks to this aggregation, any track of the dummy
identifier is lost and a primitive form of the ODM is obtained. As a further form of privacy protection, the MNO sets a
“confidentiality threshold”, discarding all aggregated number of movements falling below the threshold. The threshold
generally ranges between 4 and 20 and should be set in adequate proportion to the density of inhabitants and the size of
the geographical area of reference. At this stage, some of the MNOs introduce noise in the ODM, as a further measure
to ensure that individuals cannot be re-identified; other MNOs extrapolate the number of movements in the ODM to
the total population according to their share of the mobile network market in the country. It is straightforward that
the type of movements (in terms of range, duration and nature) that are captured in the ODMs highly depend on the
MNO’s choices about methodology and spatial/temporal granularity.

The nature of the ODMs therefore varies across MNOs and countries. Nevertheless, each ODM is consistent over
time and relative changes are possible to be estimated. This allows defining common indicators for “mobility” that can
be used, with all their caveats, to the purpose of this research.

This paper presents an exploratory analysis for a specific outbreak in France, showing and quantifying the relative
importance of mobility in the early stages of the virus outbreak in the country.

3 Connectivity Matrix

An instance of “common denominator” to guarantee comparable indicators across countries is a measure of connectivity
at less granular level, using the common statistical classification of territorial units (NUTS). The NUTS3 level was
selected in this framework, where the average population size of administrative units in Member States lie between
150,000 and 800,000.

The connectivity matrix at time $t$ is represented by the Origin Destination Matrix $ODM_{i,j}(t)$ as extracted from
normalised and aggregated data received from MNOs. With reference to Figure 1, considering the NUTS3 area $A$ (rows
and columns $a_1$, $a_2$ and $a_3$) and the destination area $B$ (rows and columns $b_1$, $b_2$), then the relevant connectivity is
calculated as follows:

$$ODM_{A,B}(t) = \sum_{i \in \{a_1, a_2, a_3\}} \sum_{j \in \{b_1, b_2\}} ODM_{i,j}(t)$$

![Figure 1: Connectivity matrix construction starting from an ODM at higher level of granularity.](image)

---

3 Regulation (EC) 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common
classification of territorial units for statistics (NUTS)
The resulting connectivity matrix is obtained by averaging over one week of data. In the remainder we will analyse outward and internal movements along one row of the matrix (Section 4), and to movements internal to departments, i.e. on the diagonal of the matrix (Section 5).

4 Case study

The analysis here conducted demonstrates how mobility in and from the department (NUTS3 area) that is believed to be one of the main hot-spots for the outbreak of COVID-19 in France explains the excess of mortality in the whole country during the following weeks (see Figure 4).

Aiming to understand if there is a defined relation between mobility and COVID-19 spreading, we need to select a reliable indicator of the contagion. Since statistics on confirmed cases are by their nature heavily affected by the type and volume of the testing procedure, which generally varies both across departments and in time as the contagion evolves, we decided to consider the death toll. Given that the precise number of deaths due to COVID-19 is still uncertain, we rely, as many other authors (Bartoszek et al., 2020), on the total number of (officially confirmed) deaths by departments over the period 1 March 2020 - 27 April 2020 (INSEE, 2020). We consider the daily and cumulative number of excess deaths with respect to the same period of 2019, assuming that most of these excess cases are due to COVID-19 (Wu and McCann, 2020; Giles, 2020; The Economist, 2020).

To study the impact of mobility on the spread of COVID-19 in France, we fit a simple statistical model that aims at explaining the excess deaths of Haut-Rhin in terms of human mobility, as measured by the connectivity matrices derived from data provided by Orange (2020), as well as the distance between the department of Haut-Rhin and all other departments. We want to show that the connectivity (or human mobility) is more important than the geographical distance for the spread of the virus.

Then we consider the distance between the department of Haut-Rhin and all other departments an finally the ODM derived mobility indicator for the week of 23-29 February 2020. Although mobility can only explain part of the spread of COVID-19, in order to show the relative importance of the mobility component, we run the simple quadratic regression model of equation (2)

\[ y_i^d = \text{const} + \alpha_1 \log(\text{mobility}_i) + \alpha_2 + [\log(\text{mobility}_i)]^2 + \alpha_3 \text{distance}_i, \]

where \( y_i^d \) is the number of cumulative death excess for department \( i \) from 1 March 2020 till date \( d \) (\( d=2020-03-01, \ldots, 2020-04-27 \)). The variable distance\( i \) is the geographical distance between the department Haut-Rhin and the department \( i \), and mobility\( i \) is the normalised Mobility Index from Haut-Rhin to the department \( i \) for the week of 23-29 February 2020. As we consider log-scaled mobility, we drop the departments for which the outbound mobility from Haut-Rhin is zero, i.e., we only fit this and the other models for which mobility\( i > 0 \). The idea behind this data driven approach is to let the data tell us under which conditions and when the mobility data do matter in studying the contagion, and when not. The choice of the very simple structure of model (2) is to highlight clearly the impact of the mobility variable as well as to avoid any over-fitting effect due to a high number of covariates which can be of course included in more sophisticated models. The scope of the model is not to forecast cases, fatalities or else.

Similar findings are seen for the week before and the week after as will be shown at the end of this section.
Then we also fit two other alternative models (3) and (4) in which only one of the two explanatory variables are present

\[ y_i^d = const + \beta_1 \log(\text{mobility}_i) + \beta_2 [\log(\text{mobility}_i)]^2 \]  
\[ y_i^d = const + \gamma_d \text{distance}_i. \]  

For each model we evaluate the $R^2$ index as well as the significance of each coefficient. To further stress the relative importance of the variables in the full model of equation (2), we consider the evolution of the standardised coefficients in time. Figure 5 shows the goodness of fit of the three models of equations (2) (Full in the plots), (3) (DistanceOnly) and (4) (MobilityOnly) in terms of $R^2$ and the $p$-value of the estimated coefficients. The vertical lines set on 9 and 13 March represent two large gatherings mentioned in the above and that on 16 March the first lock-down measure (school closing). The next three lines are translated by 14 days, which is the supposed time of incubation of the virus. From Figure 5 it is clear that the full model of (2) dominates the other two but also that, up to the maximum value of the $R^2 = 0.45$ (around 24-25 March), the model which includes only the mobility component (3) is almost equally good ($R^2 = 0.42$). Then, on the long run, the distance become slightly more important but clearly the rest of the variability is explained by the mix of the three. This figure also confirms that the department of Haut-Rhin could have been the initial source of the outbreak of COVID-19 in France. It also shows the positive impact of the lock-down measures on the reduction of excess deaths. On the other hand, Figure 6 shows the standardised coefficients of the full model of (2) as a function of time. Standardised coefficients are all on the same scale and thus their values can be compared in terms of their impact on the outcome variable. Looking at the plot, then becomes clear that the mobility coefficient is the largest effect up to 27 March, then the excess deaths number is explained by the constant in the model, i.e., by other factors not included in the simple regression analysis.

The explanatory impact of mobility can be improved further if we select those cases for which the mobility index and the cumulative excess deaths are positive. This selection is natural both because what is interesting is the positive excess of deaths and because, due to the process of non-identification and anonymisation of the data, some entries of the original OD matrix are cut to zero by construction. Figure 7 shows the scatterplot of the two dimensions for all the data. The significance of the coefficients for the constant and the mobility variable are even improved while the for the variable distance is worsened. The standardised coefficients show a similar path to those of Figure 6, see Figure 9. Figure 9 put in evidence that, when statistically significant in the model, the impact of mobility is quite strong. Figure 10 also shows that this evidence does not change substantially if we choose a different mobility week. Indeed, the Figure shows that this evidence does not change substantially if we choose a different mobility week. Indeed, the Figure shows that this evidence does not change substantially if we choose a different mobility week.

Although, we are not trying to forecast the number of deaths with mobility due to the extreme simplicity of the models considered, we notice that the $R^2$ path shows an observed time lag between mobility and excess of deaths which we impute to COVID-19: according to Li et al. (2020) the mean incubation period was 5.2 days (95% confidence interval 4.1 to 7.0) while according to Wang et al. (2020) the median number of days from the occurrence of the first symptom
Figure 5: The $R^2$ (upper panel) is explained mostly by the mobility indicator, and reaches its maximum on 24 and 25 March 2020. When the lock-down measures are in force, then the spread is explained by the distance from the Haut-Rhin department, where the spreading of the virus has believed to have started, but also other factors captured by the constant effect. In the bottom panel the p-values of the coefficients are represented. The horizontal line represents the 0.05 significance level (coefficient below this threshold are statistically significant). The black vertical lines set on 9, 13 and 16 March represent two big gatherings and the first lock-down measure. The vertical red lines are translated by 14 days.

to death was 14 (range 6-41) days.
Figure 6: The upper panel is the same as Figure 5. The lower panel represents the standardised coefficients. The impact of mobility index is predominant when $R^2$ is maximal, then most of the effect is captured by the constant. In the legend: mobility and mobility2 are the coefficients for the log-mobility and the squared log-mobility in model (2).
**Figure 7:** The data for log-mobility and cumulative excess deaths on 16 March 2020. In red the selected subset of positive excess deaths and mobility. The other points are the residual data. Red and light blue together represent the whole data. The two curves correspond to the fitted model 2 for both data sets. The straight line represent simple linear regression of the two variables in the plot.

**Figure 8:** The $R^2$ for the whole data or the selected subset (cut) for which excess deaths and mobility index are both positive. $R^2$ goes from 0.52 to 0.92 for the selected data set.
Figure 9: The p-values and the standardised coefficients for the full model applied to the selected data set. In the legend: mobility and mobility2 are the coefficients for the log-mobility and the squared log-mobility in model 2.

Figure 10: $R^2$ evolution for each mobility week using the model based on mobility alone on the selected data set of positive excess deaths and mobility index. A similar pattern is shown, explaining that mobility has no effect before the COVID-19 outbreaks and has a fast decay after the lock-down measures are in place (left and right tails of the figure). On the contrary, during the initial emergency phase mobility is a key element of the spread of the virus and therefore on fatalities.
5 The impact of internal mobility

Most of the correlation captured by the $R^2$ in the analysis of Section 4 is due to the internal mobility within the department and by the connectivity between the Haut-Rhin department and few others (Bas-Rhin, Vosges, Aisne but not, for example, Moselle, see also Figure 7). In fact, the distance alone cannot explain the excess deaths in the chosen simplified model (2). Starting from this evidence, which is common to most of the departments considered in the analysis, we now study the impact, if any, of internal mobility alone on the excess deaths of the same department through a time series approach. We consider the time series of excess deaths and of the reduction of mobility (100% = total lock-down, 0% no restriction to mobility) for each department through time. We normalise the reduction of mobility index, to the local maximum within the department to have a [0,1] measure, i.e., we define the normalised mobility as

$$nmob_{i,j} = 1 - \frac{mobility_{i,j} - \min_j(mobility_{i,j})}{\max_j(mobility_{i,j}) - \min_j(mobility_{i,j})},$$

for each origin department $i$ and all outbound departments $j$. We also normalise the cumulative sums of excess deaths in the same way

$$cumdeaths_i = \frac{cumdeaths_{i} - \min_j(cumdeaths_{i})}{\max_j(cumdeaths_{i}) - \min_j(cumdeaths_{i})},$$

for each department $i$. Figure 11 shows a plot of the two indicators for the Haut-Rhin department. The graphs suggest an evident lag could exist between the two curves. We now estimate the lag via the lead-lag estimator. Let $\theta \in (-\delta, \delta)$ be the time lag between the two nonlinear time series $X$ and $Y$. Roughly speaking, the idea is to construct a contrast function $U_n(\theta)$ which evaluates the Hayashi-Yoshida covariance estimator (Hayashi and Yoshida, 2008, 2005) for the times series $X_t$ and $Y_{t+\theta}$ and then to maximize it as a function of $\theta$. The lead-lag estimator $\hat{\theta}_n$ of $\theta$ is defined as (Hoffmann et al., 2013)

$$\hat{\theta}_n = \arg \max_{-\delta < \theta < \delta} |U_n(\theta)|.$$

(5)

When the value of $\hat{\theta}_n$ is positive it means that $X_t$ and $Y_{t+\hat{\theta}_n}$ (or $X_{t-\hat{\theta}_n}$ and $Y_t$) are strongly correlated, so we say “$X$ leads $Y$ by an amount of time $\hat{\theta}_n$”, so $X$ is the leader and $Y$ is the lagger. Viceversa for negative $\hat{\theta}_n$. The lead-lag estimator is provided by the yuima package (Iacus and Yoshida, 2018). The estimator (5) provides a value of 14 days.

5.1 The impact of local mobility on other departments

After analysing the internal mobility of the Haut-Rhin department, we consider the internal mobility of each department and estimate a lead-lag parameter against the $cumdeaths$ curves of all departments. This way of proceedings takes into account possible spin-offs from one department to another. After running the extensive 9216 (96x96) lead-lag analyses, we summarise the results through a network analysis. Therefore, we build a directed graph representing the lead-lag
relationships between the *lagger* cumulative deaths curves and the *leader* internal mobility. The nodes in the graph from \( A_i \) to \( B_j \) are such that \( B_j \) is the *cumdeath* curve for a target department \( j \) and \( A_i \) is the internal mobility of a department \( i \). Cutting those edges that form loops (internal mobility on the same department) and further selecting the edges \( i \rightarrow j \) which presents the maximal lagged-correlation for a given destination \( j \), we obtain a simplified graph. The edges on the graph are weighted according to the \( R^2 \) statistics. On this graph, a standard community detection algorithm for directed graphs is run in order to discover similar paths between mobility within a department and impact on other departments. Keeping in mind that lead-lag analysis does not involve any causal effect, the clusters that are obtained can be seen only as similar patterns of cross-correlation between internal mobility and excess deaths in outbound departments. Most of the lag discovered are around 14-20 days. Shorter lag estimates usually correspond to very flat excess cumulative deaths curves. Too large lags (i.e. 30 days) are mostly spurious correlation due to few observations. Figure 13 and Figure 14 are a graphical representation of the graph. It is important to stress that no causality effect should be interpreted from this graph. In fact, notice that the reason why the department of Hautes-Alpes is the origin of the edges of the graph towards many other departments is because the mobility there was reduced much earlier and therefore the correlation with the cumulative excess deaths is much anticipated by simple correlative effects. So what can be retained from this analysis? This network says that dynamics of both mobility reduction and cumulative excess deaths have similar patterns regardless of the region. Therefore this result, coupled with previous analysis on internal mobility, can inform policy makers on the expected spread containment in terms of human mobility reduction and the corresponding lag needed to achieve the flattening of the excess deaths curve.
Figure 13: A zoomed section of the graph of lead-lag relationships between internal mobility (origin) and cumulative excess deaths (destination). Full network is available in Figure 14. The graph is zoomed on Haut-Rhin. This cluster shows that the internal mobility of Haut-Rhin correlates maximally with the cumulative excess deaths of several other departments at different lags. The width of the edges is inversely proportional to the estimated lag: the smaller the lag, the ticker the edge. Red edges connect two different clusters. No causality effect should be read from this graph.

6 Caveats of the analysis

It is important to remark once again that the choice to use a very simple model of equation (2) is aimed at showing very rough effect of mobility on the initial spread of the virus. Clearly, such model do not and cannot produce any other result, like forecasting the number of deaths. In this work the excess deaths are treated as a proxy of virus outbreak. It is also clear and well known, that the excess deaths due to coronavirus are influenced by the age structure, the availability of ICU units, the readiness of the health care system, etc, of each department considered in the analysis. Part of the excess deaths after lock-down events, is probably also inflated by the lower number of deaths due to road accidents. There is also a technical issue due to the construction of the OD matrix that cuts movements below some threshold for the confidentiality requirements already mentioned in the above. And so forth. Some of the confounding effects are captured by the constant coefficient of the model, but clearly there could be many confounders. Still, these results looks quite stable and in line with similar findings like, e.g., in China (Jia et al., 2020).

7 Conclusions

The paper demonstrates that mobility across provinces can explain from 52% up to 92% of the dynamics of excess of deaths that can be linked to the virus during the early stages of COVID-19 outbreak in France. This information could be useful in forecasting scenarios for future waves of the virus in cases of limited additional protective measures are in place (e.g. wearing masks, physical distancing etc.). This is achieved by exploring the linkage between the geographic distribution of the excess of mortality with respect to 2019 and mobile positioning data from Mobile Network Operators, processed to derive connectivity matrices (a proxy of mobility) between territorial units. Besides predicting dynamics of future COVID-19 outbreaks, connectivity information can be used as basis to plan targeted control measures to contrast the spread of the virus, while minimising socio-economic impacts of interventions on mobility. Future data gathered in the context of this European Commission initiative will make possible an analysis at EU regional scale, providing a framework for sharing best practices and data-driven input to inform coherent strategy among EU Member States for de-escalation and recovery.
Acknowledgements

The authors would like to acknowledge colleagues from the JRC, DG CONNECT, Eurostat and ECDC for their valuable discussions in drafting the data request and for setting up a secure environment enabling to host and process the data. Finally, the authors acknowledge the support of Orange in providing access to aggregate and anonymised data.

Figure 14: Full graph of lead-lag relationships and cluster of patterns. The width of the edges is inversely proportional to the estimated lag. Red edges connect two different clusters. No causality effect should be read from this graph.

References

Bartoszek, K., E. Guidotti, S. M. Iacus, and M. Okrój (2020). Are official confirmed cases and fatalities counts good enough to study the covid-19 pandemic dynamics? A critical assessment through the case of Italy. pp. 1–40.

Bonato, P., P. Cintia, F. Fabbri, D. Fadda, F. Giannotti, P. L. Lopalco, S. Mazzilli, M. Nanni, L. Pappalardo, D. Pedreschi, F. Penone, S. Rinzivillo, G. Rossetti, M. Savarese, and L. Tavoschi (2020). Mobile phone data analytics against the COVID-19 epidemics in Italy. Flow diversity and local job markets during the national lockdown, 1–24.

Csáji, B. C., A. Browet, V. A. Traag, J.-C. Delvenne, E. Huens, P. Van Dooren, Z. Smoreda, and V. D. Blondel (2013). Exploring the mobility of mobile phone users. Physica A: Statistical Mechanics and its Applications 392(6), 1459–1473.
