The limits of human mobility traces to predict the spread of COVID-19: A transfer entropy approach

Federico Delussu, Michele Tizzoni, Laetitia Gauvin

PNAS Nexus, 2023, 2, 1–12
https://doi.org/10.1093/pnasnexus/pgad302
Advance access publication 14 September 2023
Research Report

The limits of human mobility traces to predict the spread of COVID-19: A transfer entropy approach

Federico Delussu[a,b], Michele Tizzoni[a,c] and Laetitia Gauvin[a,d,*]

[a]ISI Foundation, via Chisola 5, 10126 Torino, Italy
[b]Department of Applied Mathematics and Computer Science, DTU, Richard Petersens Plads, DK-2800 Copenhagen, Denmark
[c]Department of Sociology and Social Research, University of Trento, via Verdi 26, I-38122 Trento, Italy
[d]UMR 215 PRODIG, Institute for Research on Sustainable Development - IRD, 5 cours des Humanités, F-93 322 Aubervilliers Cedex, France
*To whom correspondence should be addressed: Email: laetitia.gauvin@ird.fr
Edited By: Rui Reis

Abstract

Mobile phone data have been widely used to model the spread of COVID-19; however, quantifying and comparing their predictive value across different settings is challenging. Their quality is affected by various factors and their relationship with epidemiological indicators varies over time. Here, we adopt a model-free approach based on transfer entropy to quantify the relationship between mobile phone-derived mobility metrics and COVID-19 cases and deaths in more than 200 European subnational regions. Using multiple data sources over a one-year period, we found that past knowledge of mobility does not systematically provide statistically significant information on COVID-19 spread. Our approach allows us to determine the best metric for predicting disease incidence in a particular location, at different spatial scales. Additionally, we identify geographic and demographic factors, such as users' coverage and commuting patterns, that explain the (non)observed relationship between mobility and epidemic patterns. Our work provides epidemiologists and public health officials with a general—not limited to COVID-19—framework to evaluate the usefulness of human mobility data in responding to epidemics.

Keywords: human mobility, COVID-19, mobile phone data, transfer entropy

Significance Statement

Mobile phone data are considered a key ingredient of realistic disease transmission models. However, it is hard to gauge their usefulness in epidemic forecasting because their added value often depends on the specific definition of mobility and the modeling approach. We develop a general and model-free framework to quantify the information transfer between mobile phone-derived mobility indicators and epidemic time series. By measuring the relative information added by different types of mobility traces to predict the spread of COVID-19 in four European countries, we find that in 2020–2021 cell phone data provided limited information to forecast COVID-19. Our results provide guidance on the effective use of mobility metrics in response to epidemic outbreaks.

Introduction

The relationship between human movements and the spatial spread of infectious diseases has been recognized for a long time (1–3). Human movement has been shown to play a key role in the dynamics of several pathogens, through two basic mechanisms: traveling infectious individuals may introduce a pathogen in a susceptible population, and, at the same time, human movement increases the contact rate between individuals, creating new opportunities for infection. In the past 15 years, the increasing availability of mobility data derived from mobile phones has fueled a large body of work aimed at identifying opportunities to use them for infectious disease modeling and surveillance (4–10).

More recently, during the COVID-19 pandemic, mobile phone-derived data have been extensively harnessed to monitor the effect of nonpharmaceutical interventions (NPIs) across countries, understand the early dynamics of COVID-19 diffusion, and forecast its spread at different spatial scales, from countries to cities (11–17). By measuring human movements and combining them with phylogeography methods (18, 19), several studies shed light on the cryptic spread of new variants, their persistence over time and resurgence after the relaxation of NPIs (20–22).

Human mobility has been shown to strongly correlate with the spread of COVID-19 during the early phase of the outbreak in China and in many other countries (23–28). However, once COVID-19 established a foothold in a population, the relative importance of mobile phone-derived data to predict the epidemic dynamics on a local scale has been generally less understood and several studies have shown conflicting evidence about the use of mobility traces to model the spread of COVID-19 at later stages of the outbreak. For instance, it has been shown that the

Competing Interest: The authors declare no competing interest.
Received: February 17, 2023. **Accepted:** August 17, 2023

explanatory power of mobility metrics in relation to the case growth rate in the United States significantly declined in spring 2020, especially in rural areas (29–31). Similar trends have been observed in Europe (32). In parallel, mobile phone-derived data have been proven beneficial to model COVID-19 dynamics in largely populated urban areas of Western countries (33, 34), but less so in countries of the Global South (35).

Several reasons have been proposed to explain the varying relationship between mobility metrics and epidemic indicators (29). Mobility metrics are generally derived from raw mobile positioning data through complex and customized processing pipelines that can significantly vary across data providers (36). How raw data are processed, and the specific definitions of mobility metrics can significantly impact their interpretation with respect to epidemic variables (37). Moreover, the relationship between mobility and epidemic patterns often relies on modeling assumptions, typically considering linear dependencies, that may not capture the complex interplay of these quantities (30, 32). Finally, mobile phone-derived metrics are generated from a sample of users that is generally not representative of the whole population. It is therefore of paramount importance to define standardized approaches that can quantify the added value of mobility metrics for epidemiological analysis and make different metrics, across settings, directly comparable.

Here, we extensively quantify the relationship between cell phone-derived mobility metrics and COVID-19 epidemiological indicators through a model-free approach, based on an information-theoretic measure, transfer entropy (38), adapted for small sample sizes. Leveraging granular data provided by Meta that capture users' movements and colocation at a fine spatial scale (39) and Google Community Mobility Reports (40), we measure the information flow between mobility metrics and time series of COVID-19 incidence and deaths in four European countries, at a subnational scale, over a one-year period. We find that the relative information added by the past knowledge of mobility metrics to the knowledge of the current state of COVID-19 time series is often not statistically significant, and that its significance also depends on the spatial resolution considered. At the finest resolution, in statistically significant cases, we show that the relative information added by past knowledge of COVID-19 cases to the knowledge of current deaths is twice the information flow between past knowledge of mobility metrics and current deaths. We also show that the information flow of a given mobility metric to predict future COVID-19 incidence or deaths can be significant in one country but not in another, even if derived from the same original data source.

Being a general framework, our approach provides a quantitative measure of the relative added explanation brought by mobile phone data to the prediction of epidemiological time series that does not depend on the choice of a specific forecasting model. It thus helps to better identify the most appropriate mobility metrics to use among those available. Our results can guide epidemiologists and public health practitioners in the evaluation of mobile phone-derived mobility metrics when they are interpreted as a precursor of epidemic activity.

Results

Here, we first describe and then apply our framework to measure the information flow between human mobility traces and the time evolution of COVID-19 in four European countries.

A transfer entropy approach to link mobility behavior and COVID-19 epidemiology

With the aim of quantifying the information flow from mobility-derived data to COVID-19 data, we first gathered a set of mobility and epidemiological indicators. Figure 1 provides an overview of the datasets used in the study. In the Materials and methods section, we provide a full description of all data sources and the data processing steps. We considered four European countries—Austria, France, Italy, and Spain—and their administrative subdivisions at NUTS3 level (41) which is the lowest, i.e. the most granular, level of the standard hierarchy of administrative regions in Europe (Fig. 1, leftmost column).

In all administrative regions, we collected indicators of the COVID-19 epidemic dynamics, namely, the weekly and daily numbers of new COVID-19 cases and deaths over the period, from September 2020 until July 2021. During this period, the dynamics of COVID-19, exemplified by the incidence of new cases (Fig. 1, rightmost column), displayed subsequent waves, as a result of the complex interaction between the spread of new variants, the adoption of nonpharmaceutical interventions, the introduction of vaccines.

In each country, we also collected weekly and daily time series describing movements and colocation patterns made available by Meta (42). We computed contact rates from colocation maps (see Material and methods section and online supplementary material for details), which measure the probability that two users from two locations are found in the same location at the same time (39). Colocation maps were generated by Meta on a weekly basis, only. To study human movement patterns, we considered movement range maps provided by Meta, which report the number of users who moved between any two 16-level Bing tiles with an 8 h frequency (43). To make colocation and movement patterns comparable in terms of scale, we focused on short-range movements, i.e. movements that occurred within the same tile, and we separately considered the mid-range movements, i.e. movements that occur between two different tiles in the same province. We then processed the three datasets, starting from their raw form, to aggregate them at the NUTS3 resolution and create the time series: $M^s(t)$ for the short-range movements, $M(t)$ for the mid-range movements and $\mathrm{CR}(t)$ for the contact rates. We also gathered daily mobility data that captures the relative change in mobility with respect to a baseline from two different data sources: the relative change in time spent at home, provided by Google, and the relative change in total movements, provided by Meta (see Materials and methods section for more details). The first dataset was available at NUTS3 resolution, while the second was only available at NUTS2 level. In the following, we refer to the residential time series as $M^r(t)$ and to the relative change in movement provided by Meta as $\mathrm{MRC}(t)$. We further aggregated the mobility metrics $M^s(t)$ and $M(t)$ at the NUTS2 level, to explore the effect of spatial resolution on our results.

Mobile phone-derived time series were then used as source variables in the information-theoretic analysis. In the remainder of the paper, we focus on the analysis of the $\mathrm{CR}(t)$, $M(t)$, and $M^s(t)$ time series at the finest spatial resolution, generally referring to NUTS3 units as provinces, although their nomenclature varies across countries.

Figure 2 illustrates our study design based on the transfer entropy (38). Transfer entropy is a metric that measures the directed statistical dependence between a source and a target time series and it has been applied to a wide range of research domains (44). Here, our approach consists, first, in computing the transfer entropy between mobility time series, $M^s(t)$, $M(t)$, and $\mathrm{CR}(t)$, and

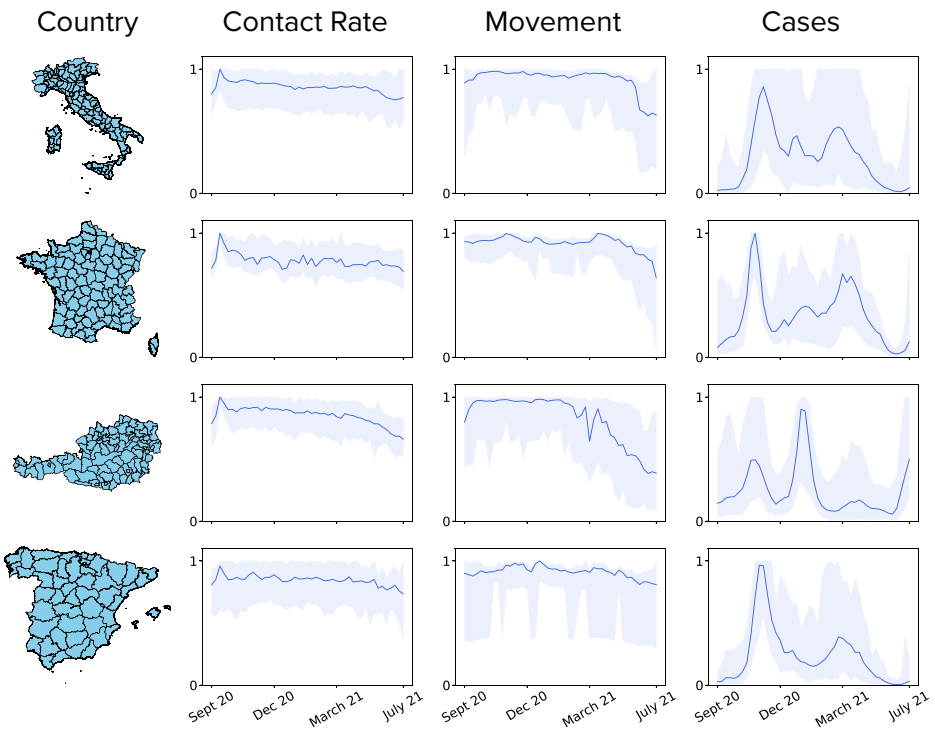

Fig. 1. Summary of behavioral and epidemiological indicators. In each country under study (from top to bottom: Italy, France, Austria, and Spain), we consider three different types of indicators: contact rates, movements (here for the sake of simplicity we only show the short-range movements), and COVID-19 cases. In each plot, the blue shaded area highlights the within-country variability, corresponding to time series in every administrative subdivision. The solid line represents the average value. All curves are normalized between 0 and 1, corresponding to their maximum value.

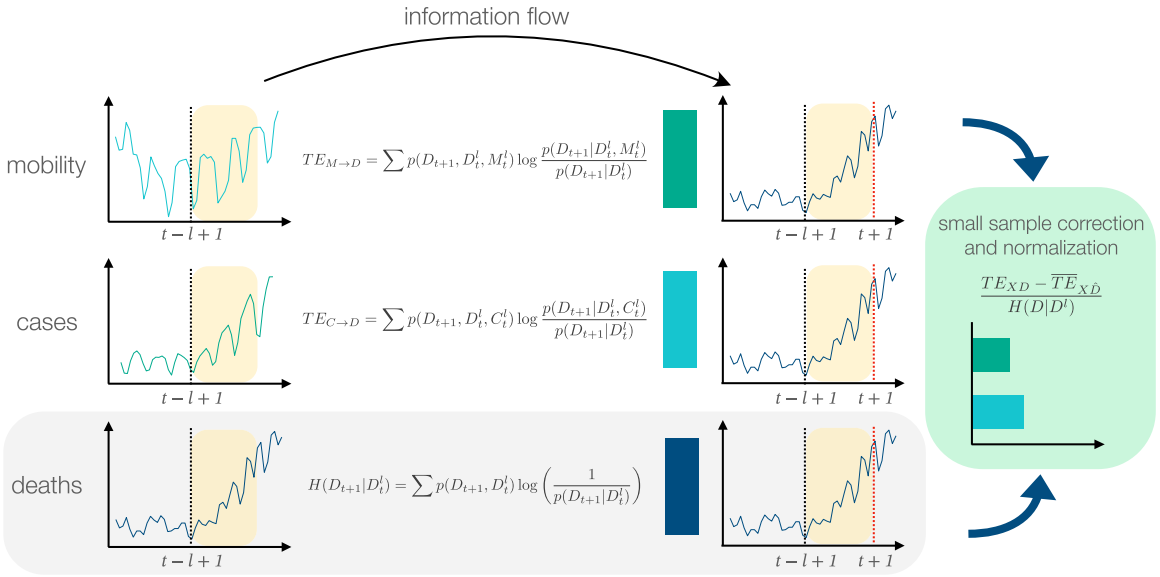

Fig. 2. Illustration of study design. We computed the transfer entropy $TE_{X\to Y}$ to measure the information flow between source X (on the left) and target time series Y (right), for a given time lag l. In the figure example, as target time series we consider the number of COVID-19 deaths, $D(t)$. As source time series, we consider either mobility indicators, $M^s(t)$, $M(t)$, $CR(t)$, or COVID-19 cases $C(t)$. Transfer entropy quantifies the amount of information that is added by past knowledge of mobility or cases (green and cyan bars, respectively) to current knowledge of deaths, with respect to the knowledge of past deaths only (blue bar). After correcting the TE for small sample sizes, and normalizing by the reference value represented by the blue bar, we finally compare the normalized effective transfer entropy of mobility and cases (rightmost box).

epidemiological time series such as the reported number of COVID-19 attributed deaths $D(t)$ and cases $C(t)$, in each administrative unit, and for different temporal lags l, using the definition of Shannon entropy, as described by the equations in Fig. 2. Intuitively, the transfer entropy between mobility and deaths, $TE_{M^s\to D}$ (resp. $TE_{M\to D}$ and $TE_{M^r\to D}$), can be interpreted as the degree of uncertainty of the reported deaths, D, at time t that is solved jointly by the time series of deaths and mobility trends M^s (resp. M and M^r) and exceeds the current degree of uncertainty of D, which can be solved by D's own past.

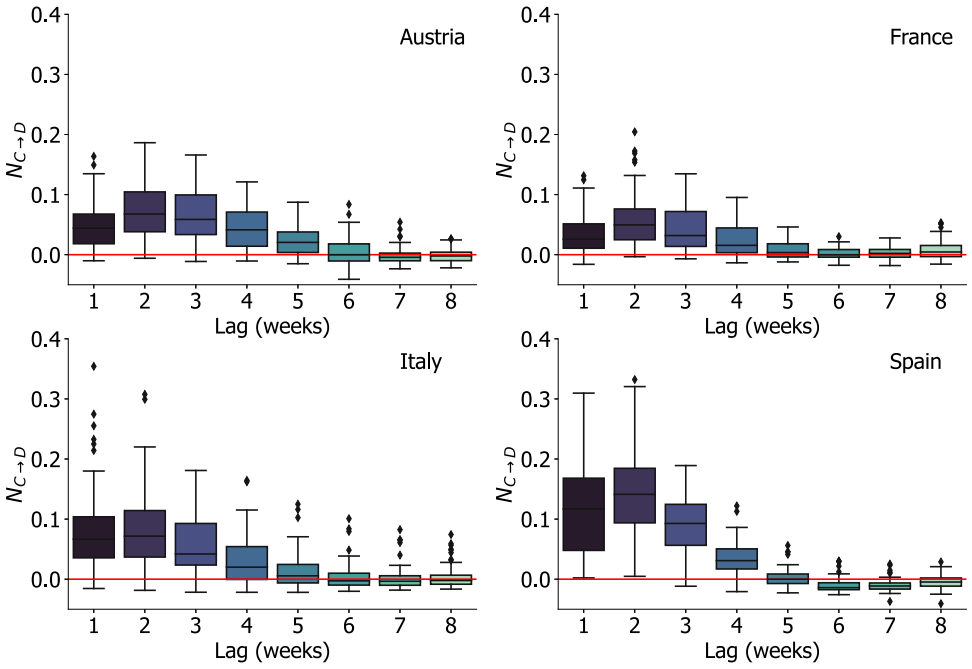

Fig. 3. Information flow between COVID-19 incidence and deaths. NETE between COVID-19 weekly reported cases and deaths in the NUTS3 administrative subdivisions (provinces) of Austria, France, Italy, and Spain. NETE is computed for lags ranging from 1 to 8 weeks, on the *x*-axis. Boxplots are computed on the distribution of NETE values of all the administrative subdivisions in each country. The horizontal red line marks the value $N_{C\to D} = 0$.

It is known that transfer entropy estimates suffer in case of small sample sizes and nonstationarity of the source and target time series (45). Moreover, due to the nonparametric nature of the transfer entropy, values computed between different source-target time series are not directly comparable. To address these issues, we first adopted the definition of effective transfer entropy (ETE) (45). ETE is obtained by subtracting from the original definition of TE a reference TE value using a shuffled version of the target time series (see Materials and methods section for details), thus removing spurious contributions to TE due to fluctuations observed in small sample sizes. Also, to address biases due to small sample sizes, we applied a Kernel Density Estimation, before the time series discretization that is necessary to compute the transfer entropy. Second, we normalized the effective transfer entropy by the Shannon entropy of the target variable, defining a normalized effective transfer entropy (NETE) (46). We obtain a metric that is always positive when it is statistically significant and whose zero value indicates the absence of information transfer between time series. In the remainder of the article, we thus refer to the NETE between source X and target Y as our main quantity of interest, using the symbol $N_{X\to Y}$ to denote it.

To better understand the cause–effect relationship between mobility and COVID-19 deaths, which are encoded in the value of $N_{M\to D}$, $N_{M^s\to D}$, $N_{M^r\to D}$ and $N_{CR\to D}$, we compared them against the transfer entropy $N_{C\to D}$, where C is the time series of new COVID-19 cases. As the causal relationship between the number of cases and deaths is established by definition, we used the transfer entropy $N_{C\to D}$ as a benchmark to evaluate the added value of mobility indicators to predict COVID-19 deaths. As an example, similar values of $N_{M\to D}$ and $N_{C\to D}$ would suggest knowledge of past COVID-19 incidence encodes a similar amount of information as knowledge of past mobility when it comes to predicting future deaths.

The information flow between COVID-19 incidence and deaths

As previously mentioned, to gauge our transfer entropy analysis framework, we first looked at the causal relationship between the incidence of COVID-19 cases and reported death counts. It is clearly expected that a major source of information that provides knowledge on future deaths is encoded in the time series of past case counts. We used NETE to quantify such information flow.

Figure 3 shows the NETE between the weekly time series of COVID-19 cases and deaths in the four countries under study. In all countries, median values of $N_{C\to D}$ increase from lags equal to 1 week up to a maximum of around 2–3 weeks, and then decline rapidly beyond the 3 weeks time lag. This is in line with early estimates of the median time delay between case reporting and fatality, which was estimated to range between 7 and 20 days in different countries (47, 48). At lag equal to 2 weeks, the mean relative explanation added by time series of cases with respect to deaths—that is how much of $D(t)$ can be explained only by the past knowledge $C(t-l)$—is 14% (SD = 8) in Spain, 8% (SD = 6) in Italy, 7% (SD = 5) in Austria, and 6% (SD = 5) in France. Boxplots computed on the distribution of administrative units in each country show a substantial heterogeneity of NETE across regions for lags shorter than 4 weeks. This may be partially explained by spatial heterogeneities of case and death reporting, and of testing strategies. Also, $N_{C\to D}$ values appear to be higher in Spain, with respect to the other countries. A transfer entropy analysis of daily time series of COVID-19 cases and deaths displays consistent results (see Fig. S1 in the online supplementary material), with NETE values that fall within the same range measured on a weekly time scale. These results suggest NETE estimates are robust with respect to the time scale at which source and target time series are compared. Moreover, it provides a reference value for NETE, in terms of orders of magnitude, when the existence of a causal relationship between time series is known.

The information flow between mobility traces and COVID-19 dynamics

Having defined a benchmark of information transfer using $N_{C\to D}$, we measured the information flow between behavioral time series of mobility indicators and COVID-19 cases and deaths. Figure 4 summarizes the main results of our analysis of the weekly time series. Values of $N_{X\to D}$, with X being either short-range movements, mid-range movements, or contact rates, were substantially smaller than $N_{C\to D}$ in all countries, for any given time lag l. In particular, Fig. 4a allows comparing the distributions of $N_{C\to D}$, $N_{CR\to D}$, $N_{M^s\to D}$, and $N_{M\to D}$, at the time lag l that maximized the median NETE for weekly time series, for all indicators. We found the largest median values of the normalized transfer entropy (NTE) at $l = 7$ weeks for both contact rates and movements (short-range and mid-range). The upper quartile of the NETE distributions derived from the mobility traces generally fell below 5%, in all countries, while the lower quartile of $N_{C\to D}$ was always above 5%. Also, the distributions of NTE computed from movements were much narrower and often included the value $N = 0$ within their interquartile range. Values of $N_{M\to C}$, shown in Fig. 4b, display a pattern similar to the NTE from the mobility time series to the death time series, with generally low values of NETE in all countries. Compared to movement time series, contact rates led generally to relatively higher values of NETE with both targets, cases and deaths, as shown in Fig. 4. Our result confirms the additional value of measuring contact rates from mobile phone data, with respect to other movement metrics (49). Besides, it shows that short-range mobility within a province had often limited predictive power to capture time trends of COVID-19 spread.

To obtain a more detailed picture of the predictive power of different mobility metrics in terms of NETE, we computed the percentage of provinces for which mobility time series provided significant relative information added, with respect to the past knowledge of epidemiological indicators only (see Table 1). On the one hand, our framework effectively captured the existing causal relationship between the time evolution of case counts and the number of deaths, as the NETE between these indicators was statistically significant ($P < 0.01$) in about 80% of the provinces, at 2 weeks lag. On the other hand, we observed a statistically significant information transfer from mobility time series to epidemiological ones in a much smaller fraction of provinces. Short-range movements NETE was significant in less than 20% of provinces when considered as a predictor of both cases and deaths. Mid-range movement time series and contact rates were significant in at most 27 and 40% of provinces. This means that in most provinces, mobility traces did not provide any additional information to predict future COVID-19 cases or deaths, at any lag between 2 and 8 weeks. Measures of contact rate extracted from colocation maps were more suitable than movement data to capture behavioral patterns relevant to predict COVID-19 spread.

By focusing only on those provinces where we could identify a significant information flow between mobility traces and COVID-19 indicators, we observe that the averaged relative explanation added by mobility data with respect to the epidemiological data ranges between 4 and 6%, which is about half of the averaged relative explanation added by past knowledge of cases to the prediction of future deaths (see Table 2 and Figs. S4–S11 in the online supplementary material).

As a sensitivity analysis, we computed the NETE on a shorter time window, between September 2020 and January 2021, to exclude the confounding effect of introducing nationwide vaccination programs. Since in those months, all countries adopted

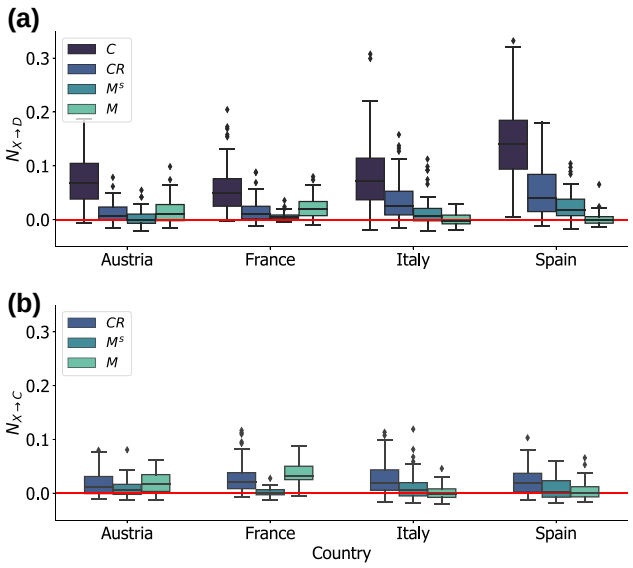

Fig. 4. Information flow from mobility data to COVID-19 incidence and deaths. Comparison between the NETE computed from source time series X and target time series of reported COVID-19 deaths D a) and cases C b). Source time series are COVID-19 cases (only for deaths), contact rates, short-range and mid-range movements. Boxplots are computed from the distribution of NETE values for a given time delay, l. In panel a: $l = 2$ weeks for cases, 7 weeks for contact rates and movement. In panel b: $l = 6$ weeks for short-range and mid-range movements. The horizontal red line marks the value $N_{X\to D} = 0$.

Table 1. Percentage of statistically significant NETE values across provinces in all the countries studied.

	→ C(t) (%)			→ D(t) (%)			
l (weeks)	CR (t)	M (t)	M^s (t)	CR (t)	M (t)	M^s (t)	C (t)
2	9	19	3	10	7	7	79
3	20	23	5	21	8	13	69
4	27	22	9	29	9	16	46
5	33	23	10	36	8	17	18
6	35	27	10	38	14	17	7
7	29	25	11	40	12	14	4
8	27	20	11	38	15	12	8

This table shows the percentage of provinces, in all countries, in which the NETE is statistically significant ($P < 0.01$) for lags (l) from 2 to 8 weeks.

Table 2. NETE results across provinces in all the countries studied.

	→ C(t) (%)			→ D(t) (%)			
l (weeks)	CR (t)	M (t)	M^s (t)	CR (t)	M (t)	M^s (t)	C (t)
2	4 (1)	4 (1)	4 (0)	4 (1)	5 (2)	4 (1)	11 (6)
3	4 (2)	4 (2)	4 (1)	5 (2)	4 (1)	4 (1)	9 (4)
4	5 (2)	4 (1)	4 (2)	5 (2)	4 (1)	5 (2)	6 (3)
5	5 (2)	4 (1)	5 (2)	6 (3)	4 (1)	5 (2)	5 (2)
6	6 (2)	4 (1)	5 (2)	6 (3)	4 (2)	5 (2)	5 (2)
7	5 (2)	5 (1)	5 (2)	6 (3)	5 (2)	6 (3)	5 (2)
8	5 (3)	5 (1)	5 (2)	6 (3)	5 (2)	6 (3)	4 (1)

The table shows the average relative explanation added by source time series, with respect to past knowledge of the target only. Only provinces having a statistically significant NETE are considered. Numbers in parenthesis report the standard deviation computed over all provinces for which the NETE was statistically significant.

mobility restrictions to mitigate the fall COVID-19 wave, we expect a stronger relationship between mobility and COVID-19 cases. Indeed, during this time frame, the information flow between movement time series and COVID-19 cases was consistently higher than in the full study period (see Fig. S12 in the online supplementary material). This result indicates that, provided with time series of adequate size, the NETE can effectively capture the time-varying relationship between human mobility time trends and COVID-19 dynamics. As an additional sensitivity analysis, we quantified the information transfer between mobility indicators and the time-varying reproductive number, R_t, which is a key measure of transmissibility during an outbreak (50). We considered the case of Italy, where estimates of R_t in 20 regions (NUTS2 level) were published by the National Institute of Public Health, every week. Results (see online supplementary Fig. S17) confirm that short-range mobility provides more information than mid-range mobility in predicting COVID-19 transmissibility, as we could measure by considering raw case counts, with similar values of NETE (online supplementary Table S5). However, the relationship between short-range mobility and R_t is statistically significant in more regions than we find with cases or deaths.

To gain a deeper insight into the effects of using different data sources and considering different spatial resolutions for our analysis, we performed a range of additional experiments, reported in the online supplementary material. First, we computed, on a daily resolution, the NETE between the residential time series, M^r, estimated by Google, and the epidemiological variables (see online supplementary Figs. S2 and S3). In this case, a very different measure of mobility, the change in residential time, showed very similar values of information transfer to cases and deaths, as that of the short-range mobility, confirming the robustness of our results against changes in data sources. Next, we measured the NETE between mobility indicators (M^s and M) aggregated at a lower spatial resolution, i.e. NUTS2 level, and COVID-19 cases and deaths in Austria, France, and Spain. The results are summarized in online supplementary Figs. S20–S22. At the regional level, we observe that the relative predictive value of short-range mobility and mid-range mobility is confirmed, with better performances of M in Austria and France, and of M^s in Spain. The regional aggregation, in general, leads to a larger number of statistically significant values in all countries. As an example, the values of $N_{M\to C}$ in the 12 French regions are always statistically significant, although the average NETE is not much higher than we observed at the province level. These results suggest the existence of a tradeoff between the information transfer of mobility indicators and the spatial scale at which time series are analyzed. In general, the higher the spatial resolution, the higher the noise, which may hide the existing relationship between mobility and disease dynamics.

To conclude, we considered as a source an additional metric of mobility provided by Meta at the NUTS2 scale on a daily basis, the change in movement, which is a relative measure of global change in population mobility (see Materials and methods section). The results of this analysis (online supplementary Figs. S18 and S19) show that such aggregate metric has predictive power in the same countries where mid-range mobility was evidenced as the best mobility metric but with smaller values of NETE.

Identifying the determinants of mobility data predictive power for COVID-19

In this section, we try to identify exogenous or endogenous factors that could explain the limited predictive power of mobility traces we observed in several provinces. We focus on Meta's data because this is the only source of data for which we have additional information related to the coverage. It also provides a way to compute the most predictive metrics among those tested here, i.e. the contact rate.

Maps of Fig. 5 highlight the spatial heterogeneity of $N_{X\to D}$ values observed within the same country, Spain, for a given time lag and different source time series (see online supplementary Figs. S13–S15 for the maps of Austria, France, and Italy). As previously mentioned, $N_{C\to D}$ displays higher and significant values in most of the country (Fig. 5a), with very few exceptions, while statistically significant values of $N_{M^s\to D}$ are found only in 16 provinces out of 42 (Fig. 5c).

To better understand the observed heterogeneity in NETE, and identify those features that can predict the likelihood to observe a statistically significant information transfer from mobility to COVID-19 death counts, we resorted to a classification model. Namely, we used a random forest classifier to predict when the value $N_{X\to D}$ is more likely to be statistically significant, using short-range movement and contact rate as source time series. We focused on these two metrics as they are quantities measured at the same spatial scale. Moreover, short-range movements represent on average 90% or more of all movements within a province (see online supplementary Table S1). As input features to the model, we considered a set of attributes of the provinces in each country. In particular, we investigated the effects of population size, province area in square kilometers, the density of Facebook users, the number of total cumulative deaths, the ratio between the number of commuters traveling from or to the province, and those who live and work there, as reported by the census (commuting flow), and the coverage consistency, that is the correlation over time between the number of Facebook users sharing their location and the number of Facebook users taken into account to compute the colocation maps.

The results summarized in Table 3 show that the model achieves a good overall performance in terms of precision and recall, as indicated by f1-scores generally higher than 0.6. In particular, of all provinces that are classified by the model as characterized by a statistically significant value of NETE, 90% or more display a significant transfer of information, as shown by precision values. On the other hand, the model's recall is close to 0.95 when it comes to identifying provinces characterized by a not statistically significant NETE, therefore the model correctly identifies 95% of those provinces where there is no actual transfer of information between mobility and deaths.

To explore the importance of province features in our classification model, we examined the SHAP (SHapley Additive exPlanations) values associated with each, as shown in Fig. 6. SHAP is a method based on a game theoretic approach to explaining the output of classification models (51). As expected, the choice of the time lag to compute the NETE is crucial in determining the presence of a significant information transfer between mobility metrics and epidemiological indicators. Indeed, *lag* is ranked as the most and second most important feature explaining the classification, for contact rate and short-range movement, respectively. Commuting flow is the most important predictor of the statistical significance of NETE between short-range movements and deaths: when the number of commuters leaving or entering a province represents an important fraction with respect to those who remain within the province, the relationship between short-range mobility and COVID-19 dynamics gets weaker. However, the same feature has only a marginal impact on the NETE between contact rates and deaths, which suggests contact

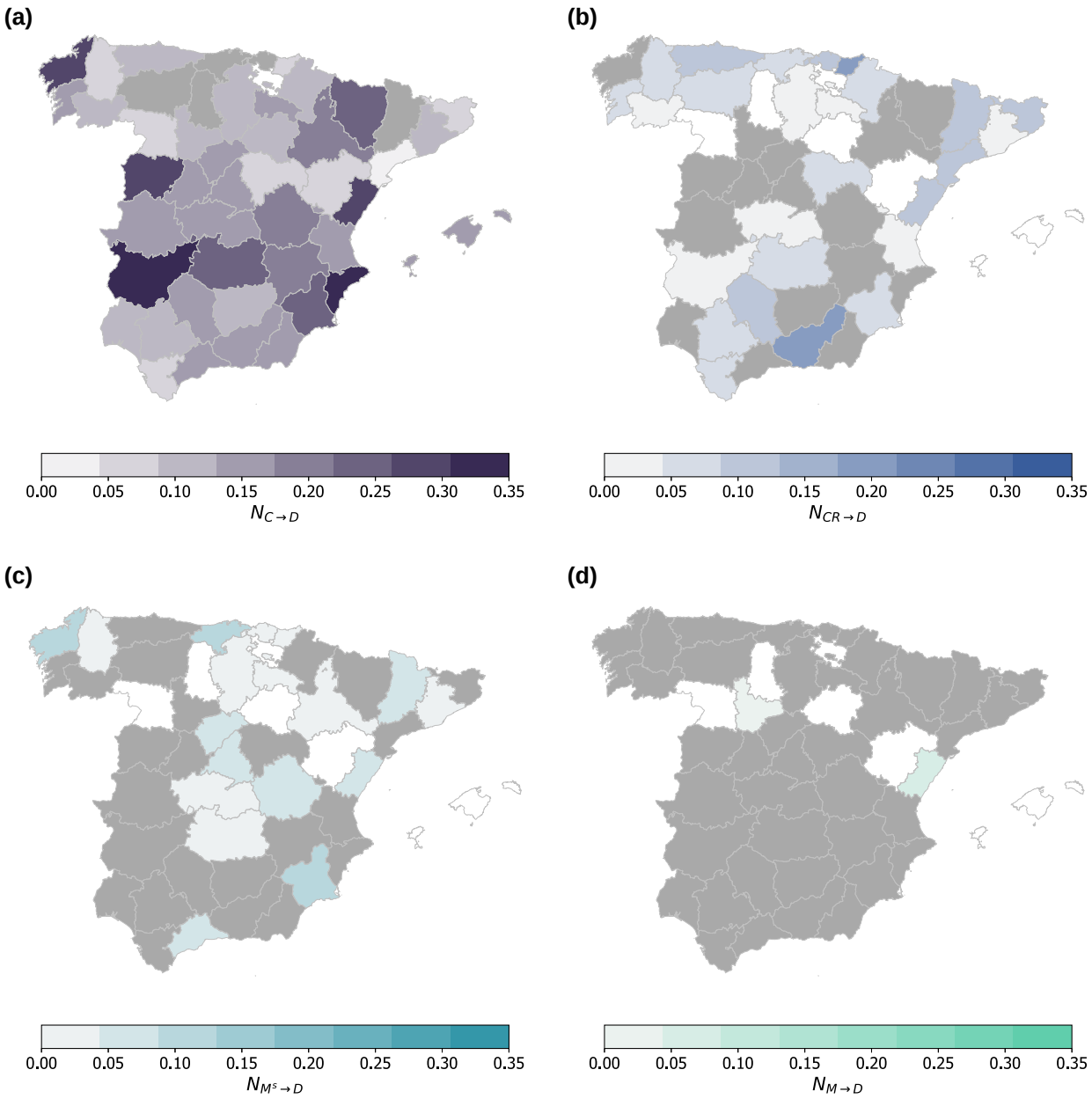

Fig. 5. Spatial variations of NETE. Maps of NETE values computed for different source time series and weekly COVID-19 deaths, in the provinces of Spain: a) source is COVID-19 cases at lag $l = 2$ weeks, b) source is contact rate at lag $l = 7$ weeks, c) source is short-range movement at lag $l = 7$ weeks. d) source is mid-range movement at lag $l = 7$ weeks. Dark gray indicates provinces with nonsignificant values of NETE ($P > 0.01$). Provinces in white are excluded from our sample.

Table 3. Classification performance metrics.

	Movement		Contact rate	
	$P \geq 0.01$	$P < 0.01$	$P \geq 0.01$	$P < 0.01$
Precision	0.64	0.90	0.71	0.92
Recall	0.95	0.47	0.95	0.61
f1-score	0.77	0.62	0.81	0.74

Summary of model's classification performance to predict the statistical significance of NETE at the $P < 0.01$ threshold when the input source is short-range movement (1) or contact rate (2) and target variable are COVID-19 deaths.

rate should be preferred over short-range movements to predict epidemic outcomes when a province is characterized by large population inflows/outflows. Province area and population size have also a significant impact on the information transfer between short-range movement and COVID-19 deaths. Indeed, a larger area and population size correspond to a higher likelihood of NETE significance for short-range movements. This effect may partly explain why we observed NETE values that were statistically significant only in a few provinces of Austria, where spatial units were particularly small. When looking at the information flow between contact rates and time series of deaths, the total

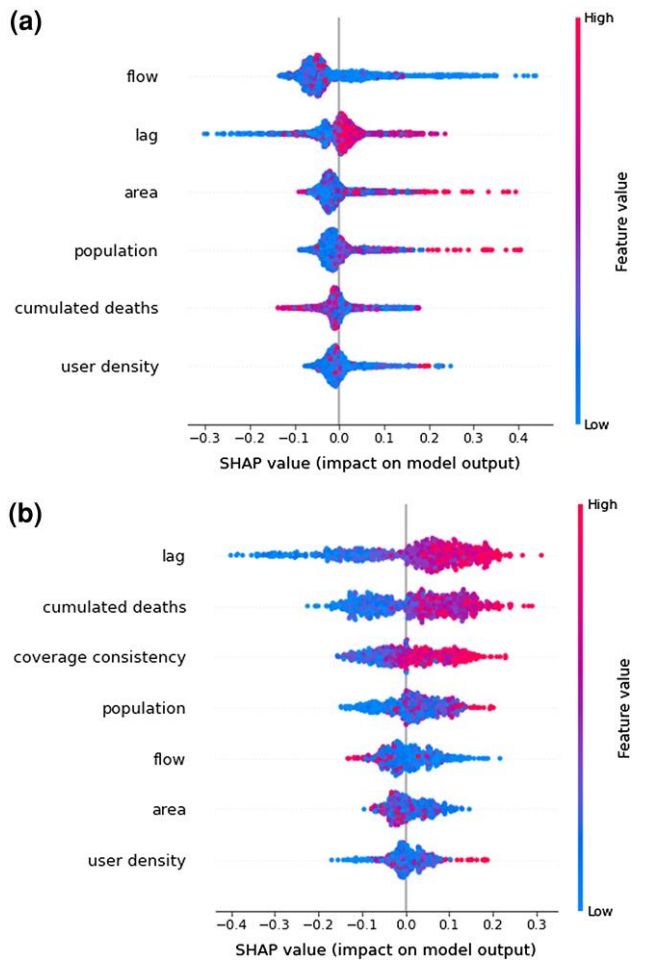

Fig. 6. SHAP (SHapley Additive exPlanations) plots of feature importance to predict the statistical significance of the NETE for all selected provinces. Color represents the feature value (blue is low and red is high). a) Describes the results for $N_{M^s \to D}$, b) for $N_{CR \to D}$. The SHAP value, on the horizontal axis, indicates the feature importance on the model output, with larger values corresponding to higher relevance. Each dot represents a single observation. Features are ranked by importance.

cumulative deaths represent an important explanatory variable for the classification model. Besides the analysis presented in Fig. 6 suggests that the coverage consistency needs to be sufficiently high in order to get a statistically significant transfer entropy from contact rate to deaths. In France, where in most provinces the coverage consistency is low and the commuting inflow and outflow are higher than in other countries (see online supplementary Table S2), mid-range movements seem to provide a better alternative to contact rates and short-range movements to partially explain time trends of COVID-19 cases and deaths (see Fig. S16 of the online supplementary material). From our analysis, we thus conclude that NETE values computed using contact rates as source time series are less sensitive to the province's geographic or demographic features, rather than to the noise of the target time series. Given good coverage, and consistency over time, contact rates thus represent a better epidemiological predictor of future COVID-19 deaths than short-range movements.

Discussion

In this work, we have introduced a general framework based on transfer entropy to quantify the amount of information that is transferred from mobile phone-derived mobility metrics to epidemiological time series. Given the important role that mobility indicators have played in the COVID-19 pandemic, we tested our approach on mobility and epidemic time series collected in four European countries, between 2020 and 2021, at a subnational scale. We found that, in general, the relative explanation added by mobility time series to predict future epidemic trends, whether new cases or deaths, was relatively small, ranging between 4 and 6% on average, and not statistically significant in the large majority of the provinces we considered, for any mobility metric. As a comparison, these values were about half of the relative explanation added by past knowledge of COVID-19 incidence to predict future deaths. Our method allowed us to directly compare the relative explanation added by different mobile phone-derived metrics of mobility from different data providers: change in residential mobility, short- and mid-range mobility, and contact rates. We generally found a higher information transfer from contact rates than movement, in line with previous studies (49), however, we also observed significant heterogeneities within the same country and between countries. We identified spatial features that may explain such heterogeneities. In provinces characterized by large populations, good coverage consistency over time, and small commuting in- and outflows, short-range movements can represent a useful metric to predict disease dynamics. Where commuting flows are large, such as in France, and Austria, mid-range movements, which represent less than 10% of the total movements, provided a better alternative to short-range ones. We also observed that the statistical significance of the information transfer depends on the spatial resolution considered. Aggregating mobility and epidemiological indicators at less granular spatial scales can help identify a clearer statistical signal for some mobility metrics. Our results suggest the choice of the best mobility metric to inform epidemic predictions can depend on a number of different factors, even when using one single data provider. Moreover, our findings show that cell phone mobility metrics do not always capture epidemiologically relevant behaviors and alternative data sources could be more effective for this aim, as, for instance, the collection of survey data (52).

There is an emerging common understanding that mobility indicators measured from mobile phone data present significant gaps and do not provide a consistent picture of mobility across countries, and data providers (53, 54). Previous studies have also highlighted the fact that coupling between mobility indicators and COVID-19 epidemiology is often weak, and it changes over time (29). The approach we introduced here addresses the above challenges by providing a general framework to evaluate the quality of metrics derived from passively collected mobility traces as a predictor of epidemic outcomes. Our framework has the advantage of being model-free, meaning that it does not depend on modeling assumptions regarding the expected relationship between mobility and epidemic dynamics, nor it requires any parametrization. The NETE we adopted is a general method that can be applied beyond COVID-19. It allows us to rigorously compare different mobility indicators, across epidemiological settings, by measuring the relative information added by mobility time series to the prediction of future disease incidence. To this end, we release the code to reproduce our analysis between any two source and target time series (see Data availability section). Researchers can use this tool in any epidemiological context to gauge the added value of a specific mobile phone-derived behavioral measure for epidemic intelligence.

Our study comes with a number of limitations and opens new directions for future work. We only focused on European

countries for which epidemiological data were shared with sufficient spatial and temporal granularity.

Such data are not free from biases. It is well known that time series of cases and deaths typically suffer from underreporting and they do not capture the true extent of the COVID-19 burden, while serosurvey studies may provide a more accurate description of the epidemic spread (55). However, COVID-19 cases and deaths represented the main target of forecasting modeling efforts during the pandemic (15) and they are key indicators for rapid public health response during epidemic outbreaks, while serosurveys, although more accurate, become available only at later stages of the epidemic.

On the mobility side, we considered data providers that, while being present in several countries and extensively studied in the literature (56–60), might overlook an important part of the population. Overall, it will be important to assess our findings on mobility data from other providers, and, most importantly, in countries of the non-Western world. Finally, it is important to note that transfer entropy measurements become more accurate as the length of the source and target time series increases (45). We worked with a relatively short time series, addressing the bias due to the small sample by adopting the effective transfer entropy. However, we could not systematically investigate how the information transfer changed over time, performing our analysis over different time windows and comparing them. Future work could benefit from longer epidemic time series, over several years, to identify temporal changes in the information flow between human movements and COVID-19 dynamics.

Measures of human mobility inferred from mobile phone data have been a critical ingredient to inform the public health response during the COVID-19 pandemic (61) and they will be an important asset in the fight against future pandemics. At the same time, their widespread use raises some relevant ethical concerns due to re-identification risks (62), therefore, it is fundamental to assess the added value of using cell phone mobility data in a given epidemic scenario and whether the benefits outweigh the risks. Our work provides a practical guide to identifying when and where mobile phone mobility metrics truly capture behavioral patterns that are relevant to predict disease dynamics.

Materials and methods

Epidemiological indicators

We collected epidemiological time series in the four countries under study from two data sources. Daily reported cumulative COVID-19 cases were collected from the COVID-19 Data Hub (63), an open source aggregator of up-to-date COVID-19 statistics, at the NUTS3 level in Austria, France, Italy, and Spain. Daily reported cumulative deaths in Austria, France, and Spain were also collected from the COVID-19 Data Hub. For Italy, death statistics were only available on a weekly time scale from the public platform CovidStat (https://covid19.infn.it/iss/). For Italy, we collected the weekly reproduction number R_t at a regional level from the National Institute of Public Health. Data are publicly available from: https://github.com/Biuni/rt-italy. The transfer entropy analysis on R_t is performed over the temporal intersection between the R_t dataset time-range and the full study period. For the analysis, we generated daily incidence time series from cumulative data by computing day-to-day differences. Then, we aggregated the daily time series of deaths and cases into weekly ones, to perform the transfer entropy analysis on a weekly scale. For additional transfer entropy analyses, we further spatially aggregated the deaths and cases time series from the NUTS3 to the NUTS2 level.

Mobility-derived indicators

In our study, we computed daily and weekly movement and contact rates from data provided by Meta through its Data for Good program (42) and by Google through its Community Mobility reports (40).

In the online supplementary material, we provide an extensive description of the raw data sources and the processing pipeline we adopted to generate our input time series for the transfer entropy analysis. More briefly, we collected the following datasets that were publicly released by Meta since the beginning of the COVID-19 pandemic, in Austria, France, Italy, and Spain:

- *Movement range maps*. It reports the number of users who moved between any two 16-level Bing tiles, with an 8-h frequency.
- *Colocation maps*. It estimates the probability, P, that, given any two administrative regions, p_1 and p_2, a randomly chosen user from p_1 and a randomly chosen user from p_2 are simultaneously located in the same place during a randomly chosen minute in a given week (39). The dataset also reports the *number of users* in p_1 and p_2.
- *Relative change in movement*. It reports the daily average number of 16-level Bing tiles visited by the users of a given region with respect to a baseline that predates the introduction of social distancing measures.

We also collected the following data from Google:

- *Mobility trends for place of residence*. It estimates the daily relative change in the time spent at places of residence, with respect to a baseline, by all users in a given administrative region.

We derive from the above data sources, four different mobility time series. The short-range movement rate is defined as

$$M^{s}_{p,w} = \frac{M^{(within)}_{p,w}}{N^{(pop)}_{p,w}} \tag{1}$$

that is the proportion of users who moved within the same 16-level Bing tile in a given province, p, in a given week w. The mid-range movement rate is defined as

$$M_{p,w} = \frac{M^{(between)}_{p,w}}{N^{(pop)}_{p,w}} \tag{2}$$

representing the proportion of users who moved between different tiles in a given province, p, in a given week w. The contact rate is defined as

$$CR(t)_{p,w} = \hat{P}_{p,w} \cdot N^{(pop)}_{p,w} \tag{3}$$

where $\hat{P}$ denotes the colocation probability corrected by a factor that takes into account the overestimation of colocation probabilities due to the heterogeneous distribution of users across provinces and the presence of a significant fraction of static users in some periods of mobility restrictions (58) (see the online supplementary material for additional details).

Finally, the residential mobility, $M^{r}(t)$, is the relative change in time spent at home, as provided by Google. Although we use the same symbol M, as for mobility metrics derived by Meta, it is

important to note that $M^r(t)$ is a measure of duration, thus very different in nature from $M(t)$ and $M^s(t)$.

Province sample selection

The population of Facebook users who contribute to the generation of the movement and colocation time series varies across countries, and it changes over time. Moreover, the metrics of movement (short- and mid-range) and colocation, are computed from different users' samples of different sizes: $N_{p,w}^{(pop)}$ and $N_{p,w}^{(coloc)}$, respectively.

In our analysis, to limit bias that may be caused by the little representativeness of the underlying sample of users, we selected NUTS3 regions in the 4 countries, according to the following criteria. First, we considered only regions where the sample $N_{p,w}^{(pop)}$ represented at least 3% of the census population to guarantee we had at least 500 users in each province. Furthermore, we considered only those regions where the two sample sizes $N_{p,w}^{(pop)}$ and $N_{p,w}^{(coloc)}$ were always positively correlated over time, during the whole study period. We denote the Pearson's correlation of weekly values of $N_{p,w}^{(pop)}$ and $N_{p,w}^{(coloc)}$ as *coverage consistency*.

After the selection, our analysis includes 47 provinces in Austria, 51 provinces in France, 93 provinces in Italy, and 42 provinces in Spain, for a total of 233 spatial units.

Normalized effective transfer entropy

Given two discrete temporal signals represented as time series X and Y, the TE (38) is a measure of the amount of information delivered from X to Y, defined as

$$\mathrm{TE}_{XY} = H(\mathrm{Y}|\mathrm{Y}^{(l)}) - H(\mathrm{Y}|\mathrm{Y}^{(l)}, X^{(l)}), \tag{4}$$

where $X^{(l)}$, $Y^{(l)}$ are, respectively, the l-lagged time series of X and Y and TE_{XY} is formulated as a difference between two conditional entropy terms, where conditional entropy is expressed as $H(a|b) = H(a, b) - H(b)$, and $H(\cdot)$ is the Shannon Entropy. Given a discrete time series S, its observations can be expressed as the sample $\{s_i; i = 1, \ldots, n\}$, and we obtain the discrete probability distribution $p(s_j)$. We compute the Shannon Entropy as: $H(S) = \sum_j p(s_j) \cdot \log_2(p(s_j))$. Thus, TE_{XY} can be expressed as

$$\mathrm{TE}_{XY} = H(\mathrm{Y}, \mathrm{Y}^{(l)}) - H(\mathrm{Y}^{(l)}) - H(\mathrm{Y}, \mathrm{Y}^{(l)}, X^{(l)}) + H(\mathrm{Y}^{(l)}, X^{(l)}). \tag{5}$$

The time series that we consider in our experiments are continuous, therefore they need to be discretized before computing TE_{XY}. We employ the kernel density estimation (KDE) for TE estimation. KDE method evaluates the entropy terms of Eq. 5 from the discretized density estimated from each of the four features sets: $\{\mathrm{Y}^{(l)}, (\mathrm{Y}, \mathrm{Y}^{(l)}), (\mathrm{Y}, \mathrm{Y}^{(l)}, X^{(l)}), (\mathrm{Y}^{(l)}, X^{(l)})\}$. KDE employs a Gaussian kernel for density estimation. Performing tests on synthetic datasets of different sizes, we checked this was the method the most adapted to small samples. For the selection of the kernel's bandwidth, we use the Scott method (64). The continuous density is then discretized with a grid obtained by an equal-width discretization of each feature's density domain. We select 20 as the number of bins for each feature's domain discretization. The discretized density is computed with the integral of the continuous probability density functions over each grid cell. Concerning the implementation, for TE estimation we use the PyCausality Python package (https://github.com/ZacKeskin/PyCausality).

Effective transfer entropy. We introduce the effective transfer entropy (ETE) as a correction to TE for small sample time series, as originally proposed by Marschinski and Kantz (45):

$$\mathrm{ETE}_{XY} = \mathrm{TE}_{XY} - \frac{1}{N_s}\sum_{j=1}^{N_s} \mathrm{TE}_{X\hat{Y}_j}, \tag{6}$$

where the correction term is obtained by performing N_s iterations of Y shuffling, obtaining $\hat{Y}_j$ and computing the average of $\{\mathrm{TE}_{X\hat{Y}_j};\ j = 1, \ldots, N_s\}$. In our experiments, we performed 500 shuffling iterations.

Normalized transfer entropy. We would like to employ TE in order to compare a set of input signals $\{X_j;\ j = 1, \ldots, N\}$ in terms of their transfer entropy TE_{X_jY} towards a specific output Y. From Eq. 4, we have that TE_{X_jY} is evaluated as a difference of conditional entropy where the first term $H(\mathrm{Y}|\mathrm{Y}^{(l)})$ depends only on target Y. In order to ensure comparability over the set $\{\mathrm{TE}_{X_jY}; j = 1, \ldots, N\}$, we reformulate the difference as a relative difference dividing by $H(\mathrm{Y}|\mathrm{Y}^{(l)})$. Thus, the set of inputs are compared according to $\{\mathrm{TE}_{X_jY}/H(\mathrm{Y}|\mathrm{Y}^{(l)});\ j = 1, \ldots, N\}$ and we refer to $\mathrm{TE}_{XY}/H(\mathrm{Y}|\mathrm{Y}^{(l)})$ as NTE.

Normalized effective transfer entropy. By combining the ETE and the NTE, we can finally introduce the NETE, which is obtained by dividing the ETE by the first conditional entropy term $H(\mathrm{Y}|\mathrm{Y}^{(l)})$ as in Perilla and Woolf (65):

$$\mathrm{NETE}_{XY} = \frac{\mathrm{TE}_{XY} - \frac{1}{N_s}\sum_{j=1}^{N_s} \mathrm{TE}_{X\hat{Y}_j}}{H(\mathrm{Y}|\mathrm{Y}^{(l)})}. \tag{7}$$

In this way, the NETE accounts both for bias in small sample time series and ensures comparability between different input sources $\{X_j\}$ in terms of information transfer to different targets. Besides, it enables estimating the percentage of explanation value added with respect to only knowing the past of the time series used as a target.

Classification model

The introduction of the ETE allows associating a P-value, a metric of statistical significance, to each NETE value computed between any pair of time series.

In our study, we investigated a number of explanatory features to understand better why in some provinces the NETE could not identify a significant transfer of information between mobility time series and epidemiological indicators. Specifically, we trained a Random Forest classification model to predict the significance of $N_{X\to Y}$ at the threshold of $P < 0.01$, in each province under study. The random forest was performed with 100 decision tree classifiers on various sub-samples of the dataset and used averaging to improve the predictive accuracy and control for over-fitting. The function to measure the quality of a split was the Gini impurity. Before applying the random forest, the data were split between training and test sets (30%). To compensate for the imbalance of the datasets, we applied a synthetic minority oversampling technique (66) on the test set.

As input to the classification model we used a set of features that characterize each province:

1. population size (as reported by the latest available census);
2. area (in km^2);
3. density of Facebook users (measured as $N_{p,w}$ divided by area);
4. total cumulative number of reported COVID-19 deaths during the study period;
5. commuting flow;
6. coverage consistency.

The commuting flow is defined as the ratio between the total number of daily commuters who travel from or to a province and the total number of commuters who work and live in that province. Commuting data were collected from the latest available census statistics in each country. The coverage consistency is the correlation over time between the users' populations $N_{p,w}^{(pop)}$ and $N_{p,w}^{(coloc)}$. Before including the above features into the model, we checked for multicollinearity using the variation inflation factor.

To quantify the importance of different features in our classification model, we used their SHAP values (51). SHAP is a method to explain model predictions based on Shapley values from game theory. In particular, we use TreeSHAP (67), an algorithm to compute SHAP values for tree ensemble models, such as the random forest classifier of our study.

Acknowledgments

The authors gratefully acknowledge Alex Pompe for his help to understand the details of mobility data from Meta.

Supplementary material

Supplementary material is available at *PNAS Nexus* online.

Funding

F.D. gratefully acknowledges support from the CRT Lagrange Fellowships in Data Science for Social Impact of the ISI Foundation, where this work was conducted. M.T. and L.G. acknowledge the Lagrange Project of the ISI Foundation funded by CRT Foundation. The funders had no role in the study design, decision to publish, or preparation of the manuscript.

Author contributions

F.D. collected data, conducted experiments, interpreted the results, made figures, and contributed to the writing of the paper. M.T. and L.G. conceived and designed the study, conducted the statistical analysis, interpreted the results, made figures, and wrote the paper. All authors read and approved the final version of the manuscript.

Preprints

A preprint of this article is published at: https://doi.org/10.48550/arXiv.2301.03960.

Data availability

The data and code to reproduce our analysis are available at: https://doi.org/10.5281/zenodo.7700401

References

1 Longini IM Jr. 1988. A mathematical model for predicting the geographic spread of new infectious agents. *Math Biosci*. 90(1–2): 367–383.
2 Findlater A, Bogoch II. 2018. Human mobility and the global spread of infectious diseases: a focus on air travel. *Trends Parasitol*. 34(9):772–783.
3 Balcan D, *et al*. 2010. Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. *J Comput Sci*. 1(3):132–145.
4 Wesolowski A, Buckee CO, Engø-Monsen K, Metcalf CJE. 2016. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *J Infect Dis*. 214(Suppl 4):S414–S420.
5 Wesolowski A, *et al*. 2012. Quantifying the impact of human mobility on malaria. *Science*. 338(6104):267–270.
6 Mari L, *et al*. 2012. Modelling cholera epidemics: the role of waterways, human mobility and sanitation. *J R Soc Interface*. 9(67): 376–388.
7 Buckee CO, Wesolowski A, Eagle NN, Hansen E, Snow RW. 2013. Mobile phones and malaria: modeling human and parasite travel. *Travel Med Infect Dis*. 11(1):15–22.
8 Charu V, *et al*. 2017. Human mobility and the spatial transmission of influenza in the United States. *PLoS Comput Biol*. 13(2): e1005382.
9 Tizzoni M, *et al*. 2014. On the use of human mobility proxies for modeling epidemics. *PLoS Comput Biol*. 10(7):e1003716.
10 Peak CM, *et al*. 2018. Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data. *Int J Epidemiol*. 47(5):1562–1570.
11 Zhang M, *et al*. 2022. Human mobility and COVID-19 transmission: a systematic review and future directions. *Ann GIS*. 28: 501–514.
12 Oliver N, *et al*. 2020. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci Adv*. 6(23):eabc0764.
13 Buckee CO, *et al*. 2020. Aggregated mobility data could help fight COVID-19. *Science*. 368(6487):145–146.
14 Gatto M, *et al*. 2020. Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc Natl Acad Sci USA*. 117(19):10484–10491.
15 Cramer EY, *et al*. 2022. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proc Natl Acad Sci USA*. 119(15):e2113561119.
16 Chang S, *et al*. 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*. 589(7840):82–87.
17 Lucchini L, *et al*. 2021. Living in a pandemic: changes in mobility routines, social activity and adherence to COVID-19 protective measures. *Sci Rep*. 11(1):1–12.
18 Lemey P, *et al*. 2014. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H_3N_2. *PLoS Pathog*. 10(2):e1003932.
19 Lemey P, *et al*. 2020. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nat Commun*. 11(1):1–14.
20 Kraemer MUG, *et al*. 2021. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. *Science*. 373(6557): 889–895.
21 Davis JT, *et al*. 2021. Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave. *Nature*. 600(7887):127–132.
22 Lemey P, *et al*. 2021. Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature*. 595(7869):713–717.
23 Chinazzi M, *et al*. 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. 368(6489):395–400.
24 Peixoto PS, Marcondes D, Peixoto C, Oliva SM. 2020. Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. *PLoS ONE*. 15(7):e0235732.
25 Kraemer MUG, *et al*. 2020. Mapping global variation in human mobility. *Nat Hum Behav*. 4(8):800–810.

26 Jia JS, *et al.* 2020. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*. 582(7812):389–394.
27 Persson J, Parie JF, Feuerriegel S. 2021. Monitoring the COVID-19 epidemic with nationwide telecommunication data. *Proc Natl Acad Sci USA*. 118(26):e2100664118.
28 Iacus SM, *et al.* 2020. Human mobility and COVID-19 initial dynamics. *Nonlinear Dyn*. 101(3):1901–1919.
29 Kishore N. 2021. Evaluating the reliability of mobility metrics from aggregated mobile phone data as proxies for SARS-CoV-2 transmission in the USA: a population-based study. *Lancet Digit Health*. 4(1):27.
30 Jewell S, *et al.* 2021. It's complicated: characterizing the time-varying relationship between cell phone mobility and COVID-19 spread in the US. *NPJ Digit Med*. 4(1):1–11.
31 Badr HS, Gardner LM. 2021. Limitations of using mobile phone data to model COVID-19 transmission in the USA. *Lancet Infect Dis*. 21(5):e113.
32 Nouvellet P, *et al.* 2021. Reduction in mobility and COVID-19 transmission. *Nat Commun*. 12(1):1–9.
33 Aleta A, *et al.* 2020. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav*. 4(9):964–971.
34 Aleta A, *et al.* 2022. Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas. *Proc Natl Acad Sci USA*. 119(26):e2112182119.
35 Ramiadantsoa T, *et al.* 2022. Existing human mobility data sources poorly predicted the spatial spread of SARS-CoV-2 in Madagascar. *Epidemics*. 38:100534.
36 Kishore N. 2021. Mobility data as a proxy for epidemic measures. *Nat Comput Sci*. 1(9):567–568.
37 Levin R, Chao DL, Wenger EA, Proctor JL. 2021. Insights into population behavior during the COVID-19 pandemic from cell phone mobility data and manifold learning. *Nat Comput Sci*. 1(9):588–597.
38 Schreiber T. 2000. Measuring information transfer. *Phys Rev Lett*. 85(2):461–464.
39 Iyer S, *et al.* 2023. Large-scale measurement of aggregate human colocation patterns for epidemiological modeling. *Epidemics*. 42:100663.
40 Google. 2020. Covid-19 community mobility reports [accessed 2021 Nov 6]. https://www.google.com/covid19/mobility/
41 Eurostat. Eurostat. Your key to European Statistics [accessed 2023 Sept 27]. https://ec.europa.eu/eurostat/web/nuts/background
42 Herdağdelen A, Dow A, State B, Mohassel P, Pompe A. 2020. Protecting privacy in Facebook mobility data during the COVID-19 response [accessed 2021 Nov 6]. https://research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/
43 Facebook Data for Good. 2020. Movement range maps [accessed 2021 Nov 6]. https://data.humdata.org/dataset/movement-range-maps
44 Bossomaier T, Barnett L, Harré M, Lizier JT. 2016. *An introduction to transfer entropy*. Cham: Springer. p. 1–190. https://doi.org/10.1007/978-3-319-43222-9
45 Marschinski R, Kantz H. 2002. Analysing the information flow between financial time series. *Eur Phys J B*. 30(2):275–281.
46 Zeng Z, *et al.* 2022. Spacecraft telemetry anomaly detection based on parametric causality and double-criteria drift streaming peaks over threshold. *Appl Sci*. 12(4):1803.
47 Wilson N, Kvalsvig A, Barnard LT, Baker MG. 2020. Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. *Emerging Infect Dis*. 26(6):1339.
48 Fritz M. 2022. Wave after wave: determining the temporal lag in COVID-19 infections and deaths using spatial panel data from Germany. *J Spat Econ*. 3(1):1–30.
49 Crawford FW, *et al.* 2022. Impact of close interpersonal contact on COVID-19 incidence: evidence from 1 year of mobile device data. *Sci Adv*. 8(1):5499.
50 Gostic KM, *et al.* 2020. Practical considerations for measuring the effective reproductive number, R_t. *PLoS Comput Biol*. 16(12):e1008409.
51 Lundberg SM, Lee S-I. 2017. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. 4768–4777. https://doi.org/10.5555/3295222.3295230
52 Koher A, Jørgensen F, Petersen MB, Lehmann S. 2022. Monitoring public behavior during a pandemic using surveys: proof-of-concept via epidemic modelling. arXiv, arXiv:2210.01472, preprint: not peer reviewed.
53 Wardle J, Bhatia S, Kraemer MUG, Nouvellet P, Cori A. 2023. Gaps in mobility data and implications for modelling epidemic spread: a scoping review and simulation study. *Epidemics*. 42:100666. https://doi.org/10.1016/j.epidem.2023.100666
54 Gallotti R, Maniscalco D, Barthelemy M, De Domenico M. 2022. The distorting lens of human mobility data. arXiv, arXiv:2211.10308, preprint: not peer reviewed.
55 Bergeri I, *et al.* 2022. Global SARS-CoV-2 seroprevalence from January 2020 to April 2022: a systematic review and meta-analysis of standardized population-based studies. *PLoS Med*. 19(11):e1004107.
56 Bonaccorsi G, *et al.* 2020. Economic and social consequences of human mobility restrictions under COVID-19. *Proc Natl Acad Sci USA*. 117(27):15530–15535.
57 Galeazzi A, *et al.* 2021. Human mobility in response to COVID-19 in France, Italy and UK. *Sci Rep*. 11(1):1–10.
58 Mazzoli M, Valdano E, Colizza V. 2021. Projecting the COVID-19 epidemic risk in France for the summer 2021. *J Travel Med*. 28(7):taab129.
59 Smolyak A, Bonaccorsi G, Flori A, Pammolli F, Havlin S. 2021. Effects of mobility restrictions during COVID19 in Italy. *Sci Rep*. 11(1):1–15.
60 Shepherd HER, Atherden FS, Chan HMT, Loveridge A, Tatem AJ. 2021. Domestic and international mobility trends in the united kingdom during the COVID-19 pandemic: an analysis of facebook data. *Int J Health Geogr*. 20(1):1–13.
61 Grantz KH, *et al.* 2020. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat Commun*. 11(1):1–8.
62 Ienca M, Vayena E. 2020. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat Med*. 26(4):463–464.
63 Guidotti E, Ardia D. 2020. COVID-19 data hub. *J Open Source Softw*. 5(51):2376.
64 Scott DW. 2015. *Multivariate density estimation: theory, practice, and visualization*. New York (NY): John Wiley & Sons.
65 Perilla JR, Woolf TB. 2012. Towards the prediction of order parameters from molecular dynamics simulations in proteins. *J Chem Phys*. 136(16):04B619.
66 Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 16:321–357.
67 Lundberg SM, Erion GG, Lee S-I. 2018. Consistent individualized feature attribution for tree ensembles. arXiv, arXiv:1802.03888, preprint: not peer reviewed.