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 $CR(t)$ 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 $MRC(t)$. 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 $CR(t)$, $M(t)$, 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 $CR(t)$, and