Mining Google and Apple mobility data: Temporal Anatomy for COVID-19 Social Distancing

Corentin Cot  
IP2I, CNRS and University of Lyon

Giacomo Cacciapaglia (✉ g.cacciapaglia@ipnl.in2p3.fr)  
IP2I, CNRS and University of Lyon

Francesco Sannino (✉ sannino@cp3.sdu.dk)  
CP3-Origins and DIAS and Università Federico II di Napoli

Research Article

Keywords: Mobility data, Google, Apple, COVID-19

Posted Date: September 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-76030/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. 
Read Full License
Mining Google and Apple mobility data:
Temporal Anatomy for COVID-19 Social Distancing

Corentin Cot,1,2 Giacomo Cacciapaglia,1,2* Francesco Sannino3,4

1Institut de Physique des deux Infinis de Lyon (IP2I), UMR5822, CNRS/IN2P3, F-69622, Villeurbanne, France
2University of Lyon, Université Claude Bernard Lyon 1, F-69001, Lyon, France
3CP3-Origins & the Danish Institute for Advanced Study, Danish IAS, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark
4Dipartimento di Fisica “E. Pancini”, Università di Napoli Federico II & INFN sezione di Napoli, Complesso Universitario di Monte S. Angelo Edificio 6, via Cintia, 80126 Napoli, Italy

*To whom correspondence should be addressed; E-mail: g.cacciapaglia@ipnl.in2p3.fr

September 9, 2020

We employ the Google and Apple mobility data to identify, quantify and classify different degrees of social distancing and characterise their imprint on the first wave of the COVID-19 pandemic in Europe and in the United States. We identify the period of enacted social distancing via Google and Apple data, independently from the political decisions. Interestingly we observe a general decrease in the infection rate occurring two to five weeks after the onset of mobility reduction for the European countries and the American states.

COVID-19 has disrupted our way of living with long lasting impact on our social behaviour
and the world economy. At the same time, differently from earlier pandemics, a very large amount of data has been collected (1–3) thanks, also, to our smartphone dominated society. Smartphones run mobility applications, such as Google and/or Apple Maps, that help humans navigate. The mobility information stemming from these apps has been harvested by Google and Apple, which have subsequently made it publicly available on the following websites: [Google](https://www.google.com) and [Apple](https://www.apple.com).

In this paper we mine these data to quantify and characterise the effects of social distancing measures enacted by various European countries and American states. An early study of mobility effects on the pandemic evolution in China can be found in Ref. (4). The Google mobility data, in Google wordings, *show movement trends by region, across different categories of places*. As categories we will use “Residential” and “Workplace”, which best describe the change in people’s behaviour after the implementation of social distancing measures with respect to a baseline day. The latter is defined, according to Google, as the median value from the 5-week period from the 3rd of January to the 6th of February, 2020, predating the wide spread of the virus in Europe. The data show how visitors to (or time spent in) categorised places changed with respect to the baseline day. For Apple, the available mobility data represent a relative volume of direction requests per country/region, sub-region or city, compared to a baseline volume defined on the 13th of January, 2020. We will be using, from Apple, information about “Driving” and “Walking”, assuming they represent the time spent by people away from home. For the United States (US), only “Driving” data are available.

Another set of data relevant for this work is related to the virus spreading dynamics, which we take from the website [https://ourworldindata.org/](https://ourworldindata.org/). We normalise the data of each country as cases per million inhabitants.

The data relative to the total number of infected cases are effectively parameterised using the High Energy Physics inspired formalism introduced in (5), dubbed *epidemic Renormalisation*.
Group (eRG). The approach has been generalised to take into account the spreading dynamics across different regions of the world in (6) and the evolution of the second wave pandemic across Europe in (7). The advantage of the eRG formalism resides in the limited number of coefficients needed to classify the spreading dynamics for each country. More complicated models have been used in the literature to study the effect of non-pharmaceutical interventions, including mobility, for Europe (8) and the US (9–13), with the latter mostly focusing on local communities.

Without further ado, following (5, 6), we introduce $\alpha(t)$ below

$$\alpha(t) = \ln (I(t)) ,$$

where $I(t)$ is the total number of infected cases per million inhabitants in a given region and $\ln$ indicates its natural logarithm. The function $\alpha(t)$ turns out to be well described by the following logistic function:

$$\alpha(t) = \frac{ae^{\gamma t}}{b + e^{\gamma t}} .$$

Here, $a$ represents the logarithm of the final number of infected cases per million inhabitants, $b$ denotes the temporal shift from the start of the pandemic and $\gamma$ measures the flatness of the curve of the number of new infected cases. Here, and in the following, we will measure the time $t$ in weeks, so that $\gamma$ is measured in inverse weeks. It was argued in (5, 6) that, aside from the trivial temporal shift provided by $b$ and for the first wave of the pandemics, two numbers are sufficient to characterise the evolution of the number of infected cases per each region, i.e. $a$ and $\gamma$. This fact helps studying the correlation between mobility data and the virus spreading dynamics for each region. By going beyond the previous parameterisation, we will discover a finer temporal structure directly related to the effects of the imposed lockdown and social distancing measures in the different regions.

In this work we focus on a selection of European countries and all of the US states. In
Europe, we considered countries with more than 3 million inhabitants and for which the data were available. Note that we will only consider the period during which the first wave of the COVID-19 was raging in Europe and in the US.

Using Google and Apple data, we provide a rationale to identify the timing of the social distancing measure actualisation in each region. European countries and US states adopted different degrees of social distancing measures during the first wave of the COVID-19 pandemic. Moreover the severity of the measures changed during the spreading of the epidemic within each region of the world. This is why we defined the beginning of the impact of social distancing measures in terms of the reduction in the mobility of individuals, rather than on political decisions.

We mine Google’s Residential and Workplace mobility data since they show movement trends across different places compared to a reference period before the implementation of any measure. The Residential and Workplace data are best suited to quantify when and to what extent people reduced their mobility and increased social isolation. Similarly, for Apple, we choose Driving and Walking data for Europe and Driving for the US states, expressing them in terms of a percentage reduction. Note that the Apple data refer to variations in the number of searches done on the Maps app, more details to be found on the website [Apple](#).

We define an immobility indicator \( M \), as described in the section Methods in the supplementary material, in terms of an average mobility reduction in the chosen categories. The average is taken over six weeks after the beginning of social distancing. This indicator allows to sort the European countries and the American states based on the hardness of social distancing. We also define regions with the highest rate of mobility reduction (low mobility, LM) and regions with the least reduction (high mobility, HM), for Europe and the US separately. The results are shown in Fig. [1], where we indicate the LM regions in cyan and the HM regions in red, with a colour gradient representing different shades of mobility being proportional to the value of the
Figure 1: The COVID-19 Mobility Map for Europe and the US. The two maps represent respectively the European and US states with different shades of mobility from the highest (HM) in bright red to the lowest (LM) in cyan. At the bottom of the figure there are three tadpole-like plots showing correlations between the four mobility reduction categories: Residential and Workplace from Google, Driving and Walking from Apple. The head of the tadpoles correspond to the average over 6 weeks after social distancing begins, while the tail indicates a 8 week average. The colour code in the three plots reflects the maps one. The maps are drawn with Wolfram Mathematica.

indicator. For Europe, the countries with the smallest mobility grossly correspond to those that imposed a lockdown, while the highest mobility country is Sweden, where no measures were imposed. Nevertheless, even for Sweden the mobility data show a significant variation that allows us to define the beginning of social distancing despite the political decisions. Similarly for
the US states, the lowest LM corresponds to states in the North-East, California and Hawaii, which imposed lockdown measures. We also noticed that the beginning of the measures, as defined by the mobility data in the US, corresponds to the dates when the schools were closed in each state (1).

To validate our conclusions, at the bottom of Fig. 1 we show the correlations between Google and Apple mobility data for Europe (left and central plot) and the US (right plot). Each region is represented by a tadpole-like symbol, with the head corresponding to the 6-week average and the tail to the 8-week average. We label each country and state by using the same colour code as in the maps. The plots show a clear correlation between the percentage change in each category. We also checked that the same correlation persists when comparing Google to Apple categories.

We now analyse possible correlations between mobility data and the parameters of the logistic function \( \alpha(t) \) such as the infection rate \( \gamma \) and the log of the total number of infected cases \( a \). To our surprise we find that \( \gamma \) is uncorrelated to the degree of mobility reduction. This implies that mobility changes have little impact on the velocity of diffusion of the disease. Of course, mobility data only capture one aspect of the social distancing, thus they do not offer a complete picture of the situation in various regions. This surprising finding can be interpreted in various ways. On the one hand, the result may imply that the main factor behind a reduction of \( \gamma \) could lie in the behaviour of individuals in social occasions (mask wearing, proximity, greeting habits, to mention a few); on the other hand, it is quite possible that the value of \( \gamma \) does not represent the effect of the social distancing measures, as it derives from a global fit over a wide timescale. In other words, the fit values include both the measure and the pre-measure periods.

To push further the analysis, we now explore whether social distancing measures (as defined via the Apple/Google mobility data) lead to distinct temporal patterns in the European countries under study and the American states. In the eRG approach, \( \gamma \) is the natural parameter to use for
this task. We assume that, after the measures are enacted, there are two distinct temporal regions describing the time dependence of the number of infected cases. These two regions, B and C in the illustrative plot in Fig. 2, are naturally described by two different gammas. We confirm that such an analysis is possible via a MonteCarlo analysis. We then move to the actual data and discover that two distinct temporal regions with their own gammas do emerge for several regions. In Fig. 2 we show the outcome of the fit to the data in terms of the time interval $\Delta t$ between the beginning of social distancing and its effects measured when the infection rate $\gamma$ changes. We discover that most countries display a similar $\Delta t$. By fitting the distributions in Fig. 2 to a gaussian, we find that to the two sigma level we have $\Delta t = 2.7 \pm 1.7$ weeks for Europe and $\Delta t = 3.3 \pm 1.6$ weeks for the US. The high compatibility of the two ranges shows the emergence of a universal time scale for social distancing to be effective.

Another important result is the general and strong reduction of the infection rate measured within and after $\Delta t$ both for Europe and the US, as shown in the left panels of Fig. 2 and summarised by the red histograms of Fig. 2d.

**Discussion**

We analysed the mobility data released by Google and Apple to quantify the effects of social distancing on the COVID-19 spreading dynamics in Europe and in the US. We were able to classify different shades of social distancing measures for the first wave of the pandemic. After identifying the countries according to their level of immobility, we observed a strong decrease in the infection rate occurring two to five weeks after the onset of mobility reduction for the countries studied here. Another important result is the discovery of a universal time scale after which social distancing starts showing its impact.

We have also provided an actual measure of the impact of social distancing for each region, showing that the effect amounts to a reduction by $20\% - 40\%$ in the infection rate for most
Figure 2: **Temporal anatomy of COVID-19 social distancing effects.** In panel a), we show $\Delta t$ and the percentage variation $\Delta \gamma$ in the infection rate for the European countries considered in this study. In panel b), we show the same for all the US states. In panels c) and d) we display the same results in the form of histograms, for Europe and the US separately, highlighting that $\Delta t$ clusters around similar values. In panel e), we illustrate the subdivision of the first wave epidemic curve in three temporal regions: A before social distancing as defined via mobility data occurs, B until an effect is observed in the epidemic curve as a change in $\gamma$, and C covering the later times. $\Delta t$ equals to the duration of the period B.

countries in Europe and 30% – 70% in the US. This is, to the best of our knowledge, a first global and direct measure of the impact of social distancing. Interestingly, even countries that did not impose political measures, like Sweden, show a reduction of the infection rate similar to the ones experiencing a lockdown, suggesting that a certain degree of social restrain occurred
regardless of the political decisions. Our results are compatible with early analysis of local social distancing measures taken in China (4), where mobility data inter-cities from Baidu was used within a compartmental model.

Using smartphone based open-source mobility data, we showed that it is possible to provide a temporal anatomy of social distancing. We discovered the emergence of a characteristic time scale related to when social distancing effects have a measurable impact. This timing can also be used to quantify the impact of social distancing by determining the variation in infection rate per country. Finding similar reduction, however, does not imply that the countries have a similar number of infected cases per million inhabitants. It simply means that there has been a change in social behaviour. The result of this study, based on the simple eRG approach, lays the basis for an effective tool for the authorities to evaluate the timing and impact of the imposition of social distancing measures, in particular related to movement restrictions.

References

1. G. A. Wellenius, et al., Impacts of state-level policies on social distancing in the united states using aggregated mobility data during the covid-19 pandemic (2020).

2. M. N. Lurie, J. Silva, R. R. Yorlets, J. Tao, P. A. Chan, The Journal of Infectious Diseases (2020). Jiaa491.

3. A. S. Islind, M. Óskarsdóttir, H. Steingrímsdóttir, Changes in mobility patterns in europe during the covid-19 pandemic: Novel insights using open source data (2020).

4. S. Lai, et al., Nature (2020).

5. M. Della Morte, D. Orlando, F. Sannino, Front. in Phys. 8, 144 (2020).

6. G. Cacciapaglia, F. Sannino, in press on Nature Sci. Rep. (2020).
7. G. Cacciapaglia, C. Cot, F. Sannino, *in press on Nature Sci. Rep* (2020).

8. S. Flaxman, *et al., Nature* (2020).

9. A. Kabiri, A. Darzi, W. Zhou, Q. Sun, L. Zhang, How different age groups responded to the covid-19 pandemic in terms of mobility behaviors: a case study of the united states (2020).

10. F. Nielsen, G. Marti, S. Ray, S. Pyne, Clustering patterns connecting covid-19 dynamics and human mobility using optimal transport (2020).

11. I. E. Fellows, R. B. Slayton, A. J. Hakim, The covid-19 pandemic, community mobility and the effectiveness of non-pharmaceutical interventions: The united states of america, february to may 2020 (2020).

12. B. Hong, B. Bonczak, A. Gupta, L. Thorpe, C. E. Kontokosta, Exposure density and neighborhood disparities in covid-19 infection risk: Using large-scale geolocation data to understand burdens on vulnerable communities (2020).

13. F. Vanni, D. Lambert, L. Palatella, Epidemic response to physical distancing policies and their impact on the outbreak risk (2020).

**Acknowledgements**

G.C. and C.C. acknowledge partial support from the Labex-LIO (Lyon Institute of Origins) under grant ANR-10-LABX-66 (Agence Nationale pour la Recherche), and FRAMA (FR3127, Fédération de Recherche “André Marie Ampère”).

**Author contribution**

This work has been designed and performed conjointly and equally by the authors. G.C., C.C. and F.S. have equally contributed to the writing of the article.
Competing interests

The authors declare no competing interests.
The COVID-19 Mobility Map for Europe and the US. The two maps represent respectively the European and US states with different shades of mobility from the highest (HM) in bright red to the lowest (LM) in cyan. At the bottom of the figure there are three tadpole-like plots showing correlations between the four mobility reduction categories: Residential and Workplace from Google, Driving and Walking from Apple. The head of the tadpoles correspond to the average over 6 weeks after social distancing begins, while the tail indicates a 8 week average. The colour code in the three plots reflects the maps one. The maps are drawn with Wolfram Mathematica.
Figure 2

Temporal anatomy of COVID-19 social distancing effects. In panel a), we show $\Delta t$ and the percentage variation $\Delta \gamma$ in the infection rate for the European countries considered in this study. In panel b), we show the same for all the US states. In panels c) and d) we display the same results in the form of histograms, for Europe and the US separately, highlighting that $\Delta t$ clusters around similar values. In panel e), we illustrate the subdivision of the first wave epidemic curve in three temporal regions: A before social distancing as defined via mobility data occurs, B until an effect is observed in the epidemic curve as a change in $\gamma$, and C covering the later times. $\Delta t$ equals to the duration of the period B.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Mobilitysupplementary.pdf