Measuring Levels of Activity in a Changing City

A Study Using Cellphone Data Streams

Leo Ferres [1,3,5], Rossano Schifanella [2,3], Nicola Perra [4,3], Salvatore Vilella [2]
Loreto Bravo [1,5], Daniela Paolotti [3], Giancarlo Ruffo [2], Manuel Sacasa [5]

1. Instituto de Data Science, Faculty of Engineering, UDD, RM, Chile
2. Università degli Studi di Torino, Torino, Italy
3. ISI Foundation, Torino, Italy
4. University of Greenwich, London, UK
5. Telefónica Investigación y Desarrollo, RM, Chile

Notice: this is a preliminary assessment which has not yet been peer-reviewed and will be updated periodically as new data and analysis become available. We consider it a "living document". To the best of our knowledge we have paid due diligence to the science, but errors may remain. If so, we'd appreciate feedback and corrections at lferres@udd.cl.
[v.2020-04-05 19:53:57 -0400]

Main findings

1. After March 16th, when the first few measures were put in place (closing schools, reduced large gatherings), the city of Santiago defaulted to a weekend (Saturday) activity pattern.
2. After March 16, trips between antennas become shorter and more localized.

Introduction

Chile, like many other countries, was closely monitoring the world as COVID19 was spreading, and learned from Europe, China and others in their handling of the virus. Some mitigation measures were already put in place by the Government on late February: on 02/24, some passengers on a cruise ship were quickly put under epidemiological observation; on 02/27, new measures were announced with respect to the strengthening of Chilean hospital infrastructure after a confirmed case in Brazil while, on the same day, the Chilean government announced that the COVID19 test was going to be free for everyone. On 03/02, every passenger arriving in Chile from abroad was to be surveyed and a COVID19 Protocol was enacted. The Ministry of Health **confirmed the first person testing positive for COVID19 on March 3, 2020**.

After the World Health Organization (WHO) declared a COVID19 pandemic, several more radical measures started to be put in place: on 03/12, large gatherings were no more allowed in sports: football matches, in particular, had to be played in empty stadiums, buildings were re-purposed for holding beds. **The authorities confirmed 33 new cases of COVID19 on 03/12**, and quarantined a school in one of the comunas[1] in the Metropolitan Region of Santiago. On 03/14, Caleta Tortel is the first larger administrative region of Chile, the comuna, to declare a quarantine. On 03/15, the president suspended classes and reduced public meetings. On 03/16, the country's Emergency Office declared a "code yellow" state of pandemia, and the president closed the borders, many people at this stage have been instructed to stay home. **On March 18, 2020, with 238 confirmed COVID19 cases, health authorities instructed citizens to respect physical isolation and a state of catastrophe was declared**.

Data

For this work, we used 3.2 billion ($3.2*10^9$) XDR connections to analyze the mobility of Chile after the big events described above, from February 26, and until March 18, 2020. For this report, we have concentrated on the main urban area of the largest region of Chile: The "Region Metropolitana" with half the population of the country, and the most affected area of COVID with 47% of cases as of April 3, with the Araucanía Region in the South a far second with 12%.

There exist several "streams" in cellphone datasets. The most common one in the literature is called **CDR** (short for Call Detail Record) and basically tells us, in anonymized form, who calls who from which antennas and at what time. Antennas have "fixed" latitude and longitude, it is not possible for telcos to count people using more fine-grained data like GPS, which only work with apps like Facebook or Google, maybe aggregated by third parties. There is a second stream called **XDR**, which is the "data" channel, the one that is used to download contents from the web, for example. While CDRs are very sparse in time, XDRs provide a better time resolution. Now, the only information we get with XDR is the timestamp of when the connection happened, and the latitude and longitude of the antenna used to download the content. The XDR dataset looks like the table to the right.The **PHONE_ID** is the anonymized hash of a mobile phone number. Every **CELL_ID** is an identifier for the cell where the connection took place. The field **EVENT_TM** is simply a timestamp in the "YYYMMDDTHHMMSS" format, where "YYYY" is the year with four digits (2020), "MM" is the month with two digits (preceded by 0, for example 02) and "DD" is a two digit number, like 26 above, also preceded by 0 for numbers smaller than 10, if needed. The "HHMMSS" is the zero-filled hours, minutes and seconds, "T" is a separator.

---

[1] A minor administrative region of Chile, roughly similar to a "county" in the United States, see here. We keep the Spanish name because there is in effect no direct translation.

PHONE_ID	CELL_ID	EVENT_TM
fa50fc29fd235b9d61abbd11a06975f3214f6506d3e0f37e648317df2358524c	1269024	20200226T073059
fa50fc29fd235b9d61abbd11a06975f3214f6506d3e0f37e648317df2358524c	1260577	20200226T070056
f748ad1fc4ba44409147104a2f2ab6ad6a4f2e3203b1c9999fb2a29dea2c18a0f	1071903	20200226T073600
3a882aefb350b663b560d908cac4a99f3240e2e3e637d5445faad84c36052ae0	1589281	20200226T070219
ea02c57369bb2b4705bc28695c331d736524a8531567cb79019dfaf79d60e653	159855904	20200226T072003

Figure 1: A sample of the dataset

Antennas do not really follow a "coverage" pattern, but rather a "demand" pattern: there are more antennas where there are more connections. In the case of SCL, the antenna-placement, at least for the subset of our dataset, looks like the figure below, where the black dots are cells and the color shows the density of cells in that region. Usually, it is the case that more demand happens in areas that are commercial, rather than residential, so darker colors in the map means more connections/demand/activity.
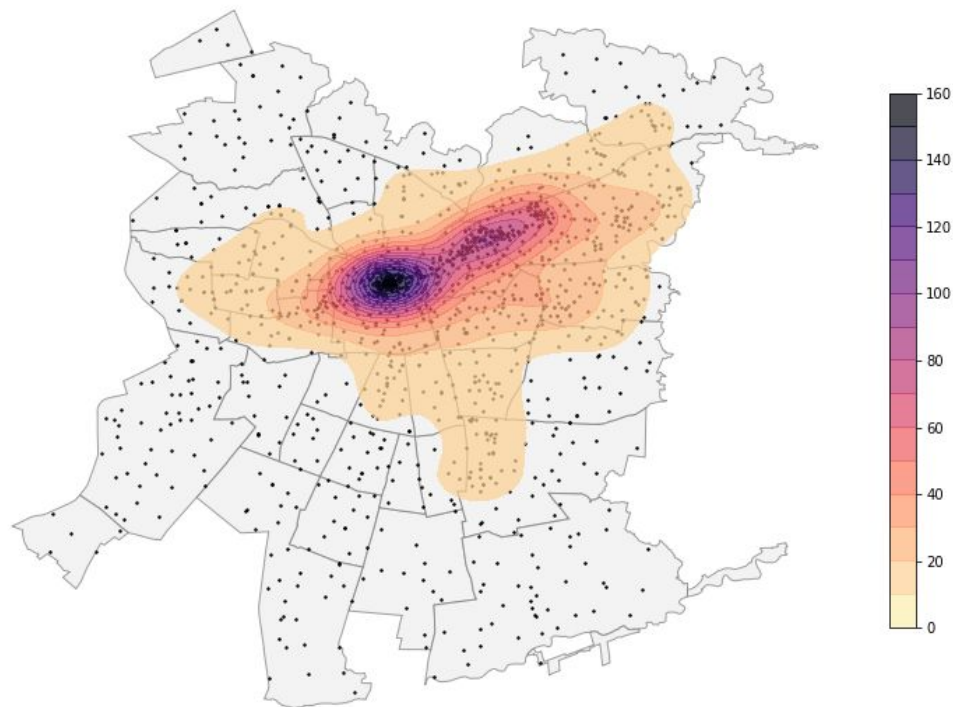

Figure 2: Antenna density, darker color means more antennas. Demand (more antennas) is correlated to places where the floating population arrives to work, rather than where they reside.
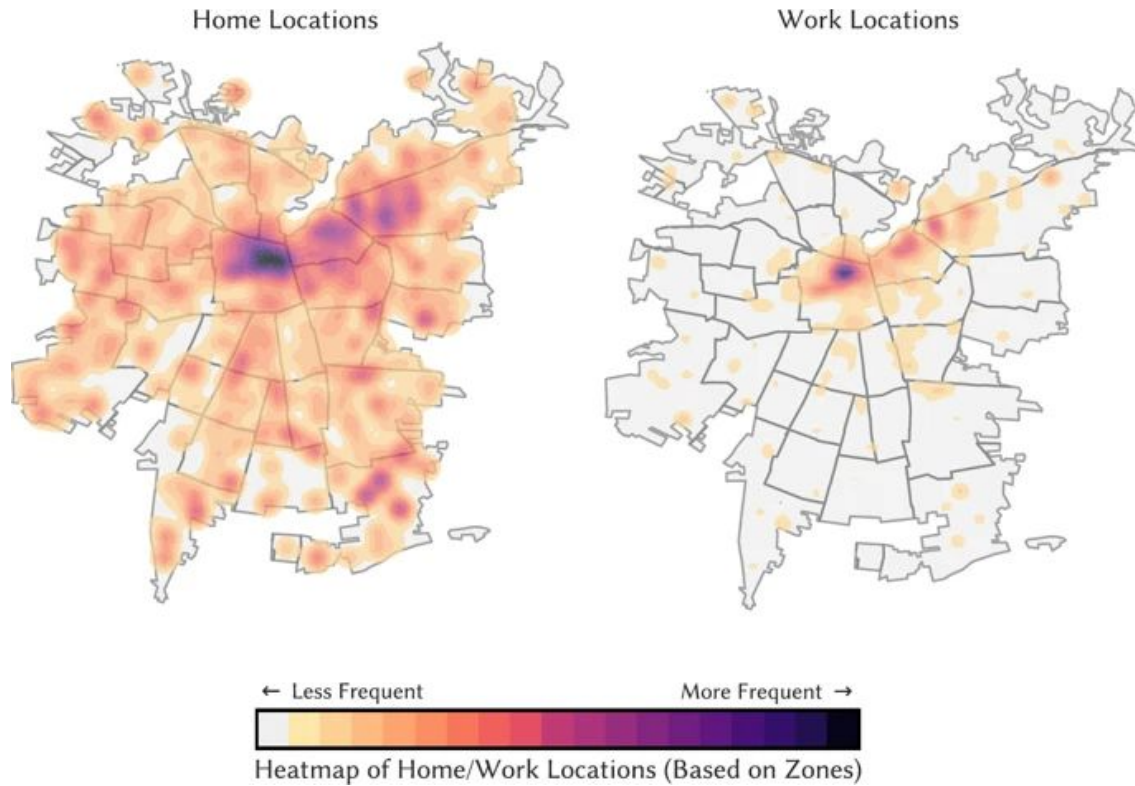
Figure 3: Heatmaps of home and work locations based on XDR analysis (see here).

Figure 3 shows home/work location derived from similar XDR data for another study in Santiago. Here we see how work is located where antennas also have higher density, while home location is distributed more sparsely (still with some concentration downtown and in the comunas of the north east). So, if the emergency measures on the 16th were indeed successful, we would expect an increase in the "outer" comunas (the more residential ones, where antennas are sparser), and a decrease in "work" comunas, where antennas are denser.

Results

On March 16, the government imposed "social distancing" and, in particular, closed schools, universities and reduced large gatherings of people. Our first analysis looked into whether there were any effects of this measure, or whether people would simply keep living their daily lives.
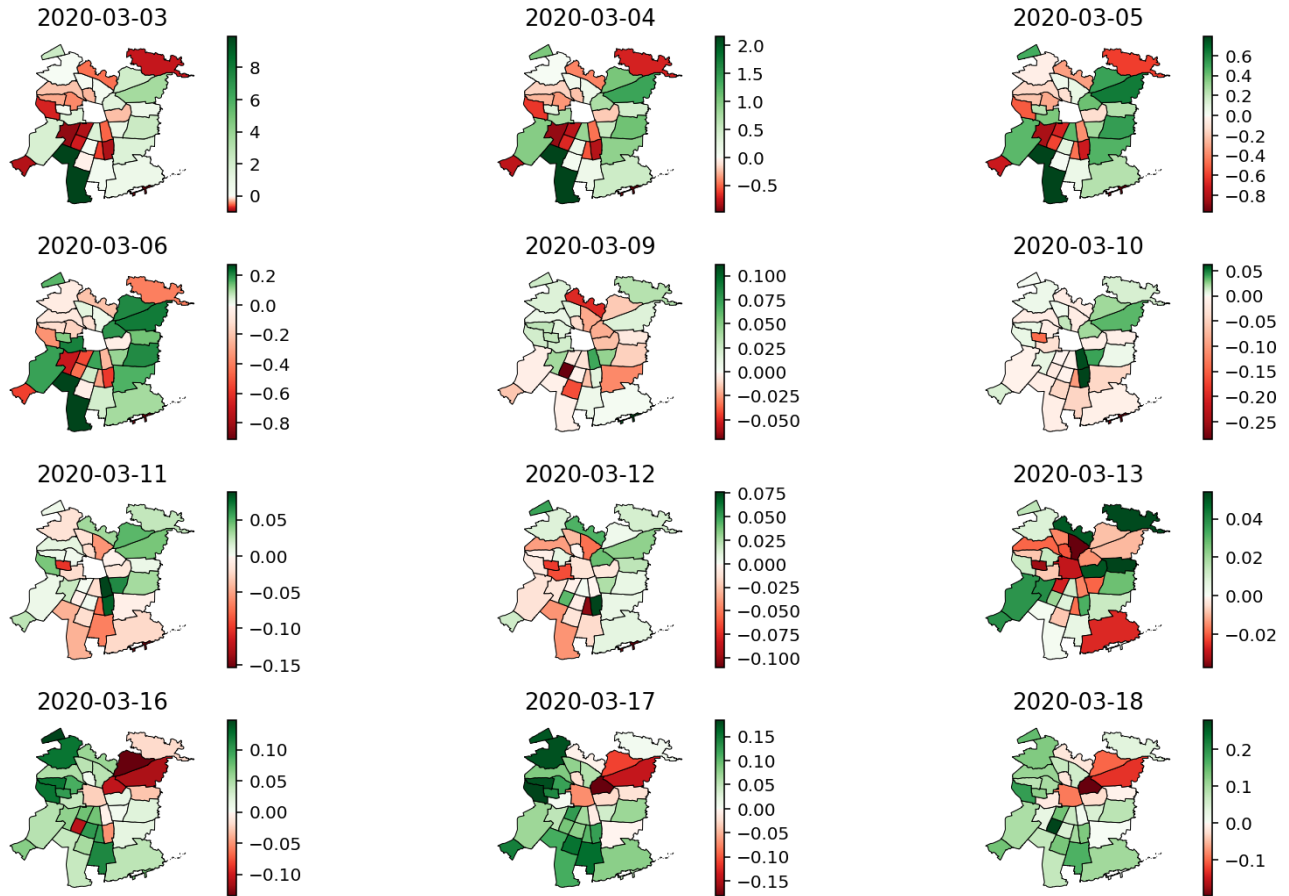
Figure 4: Delta of activity by comuna using a rolling window of 4 days.

The figure above shows the change in the patterns of connections in the greater Santiago area. All regions of Chile were calculated, but this report will only show Santiago. Each image in the figure shows one day, from March 3, to March 18, the last day in our dataset: two days after the closure of universities, schools, etc. The color red means "decreasing", while green means "increasing". The increase vs. decrease is calculated using a 5-day rolling window: that is, for example, we calculate the number of active users in each comuna on March 10, and then we compare it to the average number of active users in the 4 days before, in this case March 4-9 (without weekends, because connection patterns during weekends are quite different, more about this later). What we see in the maps is the *delta* of connections between that day and the average of the rolling window to emphasize relative changes on the fraction of the traffic in that comuna over the total daily number of active users.

As we analyze the last three images in the Figure (March 16-18), we see a decrease in activity (active users) in the regions where demand was higher before the 16th, presumably "more commercial" areas of Santiago (Santiago, Providencia, Vitacura, Las Condes, Huechuraba and Lo Barnechea), while the levels of difference seem to increase (green goes .1, .15, .2), meaning % change in the number of active users, in comunas where the connections are usually fewer.

That way, people seem to be adopting "weekend" patterns, going back to more residential areas. Figure X below shows the difference before and after the intervention on the 16th. By Tuesday (one day after the "suggestion" to stay at home), Santiago stopped activity to the levels of a Saturday, approaching but not quite getting to, the activity levels on a Sunday. The pattern is clear: 1 day to Saturday levels, longer for Sunday.
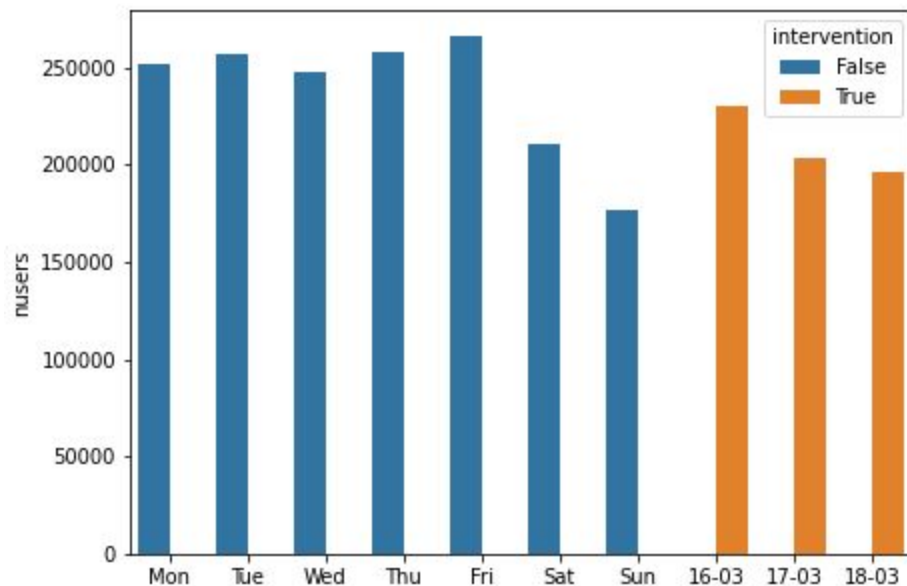

Figure 5: Comparison between the three days after the intervention and the mean unique users per day of the week

Still, the real differences are difficult to see given the different ranges of the maps. In Figure 6 below, we disaggregate by comunas. What is important here is not that the comunas changed, but that we now have a way to measure by how much and where.
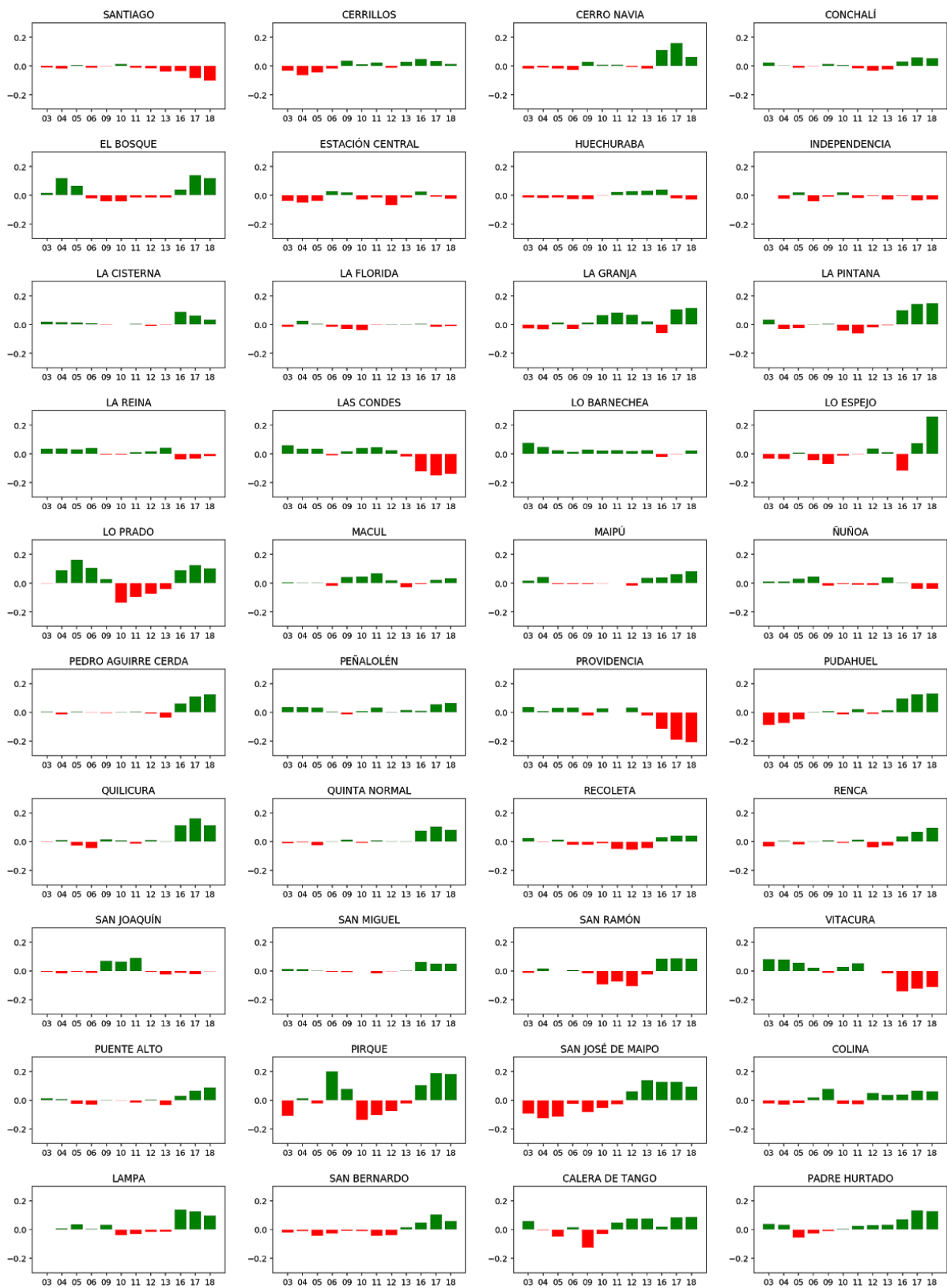
Figure 6: Disaggregated information by comuna, per day. Red means "decrease", green means "increase".

The data also allow for much more refined (micro) analyses aimed at capturing the impact of the governmental measures towards social distancing. It is important to notice how these might be used as input for realistic epidemic models and thus to provide realistic forecasts about the unfolding of the virus.

We build the mobility networks describing the movements of individuals each day from the 26th of February to the 18th of March. Nodes are antennas. Directed and weighted links between them describe the number of people that we see "transition" (i.e. move) from one antenna to the other.

As a first step, we compare the mobility networks describing the same time-window of three days (Monday, Tuesday and Wednesday) before (9-11 of March) and after (16-18 of March) the implementation of social distancing measures. Overall we observe a reduction of about 15% of movements between antennas. Interestingly, this reduction is very similar to what is observed comparing the 9-11 of March with the weekend before or after and in line with observations in other countries, see for example the case of Italy. Also, the average number of antennas connected by the movement of people between them is reduced by about 7%.

Due to the nature of mobility these average numbers provide only a very general overview. In order to get a better understanding, we analyse the distributions of the variation of key quantities. In Figure 7 we provide a summary. Panel A) shows the variation of the weight, w, (i.e. number of trips between two antennas) between the two periods. The plot shows that the large majority of links do not register a big change. However, few of them show either a large reduction (on the right of zero) or increase (on the left of zero) of traffic. Panel B) shows a similar plot, done however for the in and out strength, s. These two quantities capture the total number of people antennas see arrive (in strength) or depart (out strength). As in panel A), the large majority of antennas report a small variation (note the peak around zero) and few of them register either a large reduction or increase in total incoming or outgoing traffic. We notice however a small asymmetry in the plot that indicates how a reduction is more likely than an increase. In panel C) we report the same type of plot done however considering the number of antennas reached by (in degree) or that can be reached from (out degree) a particular antenna. Here we see how the peak of the distribution is shifted from zero to larger values, meaning that, across the board, social distancing measures have decreased the "exploration" of individuals and thus the number of connected antennas by means of their mobility. Finally, in panel D) we report the distribution of the Jaccard coefficient of each antenna. This quantity goes from zero to one and compares the similarity of two sets. In particular, we compare the set of antennas that can be reached from (neighbors) and reached by (predecessors) before and after the implementation of social distancing. Values close to zero describe a radical difference between the two sets. Values close to one describe instead two very similar sets. The plot shows that intermediate to high values (between 0.4 and 0.6) are the most common. Interestingly we note a peak also for small values which highlights how some variations in the mobility patterns are visible. Overall the picture emerging from all these plots is the following. The implementation of

social distancing has induced some changes in the mobility of individuals. However, such changes are not radical. In fact, most of the key quantities do not register big variations. Nevertheless, we do see quite large variations in some locations.
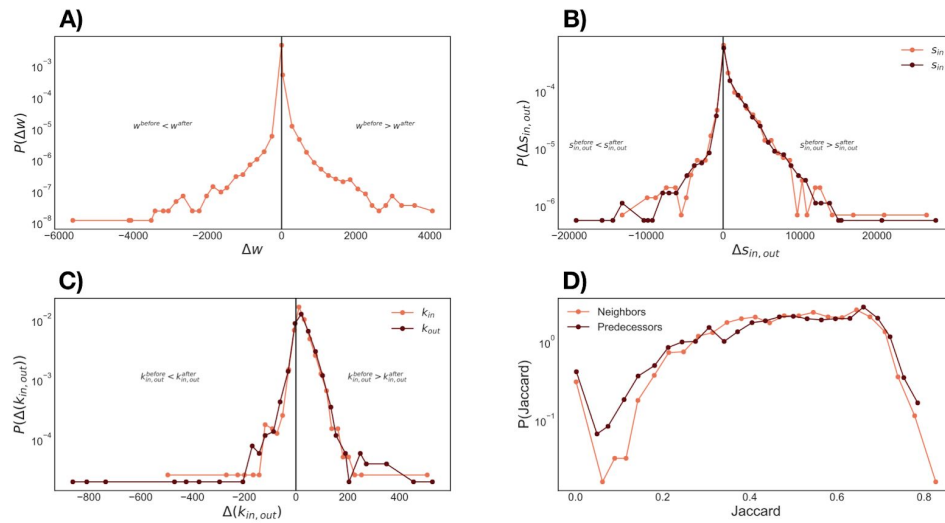

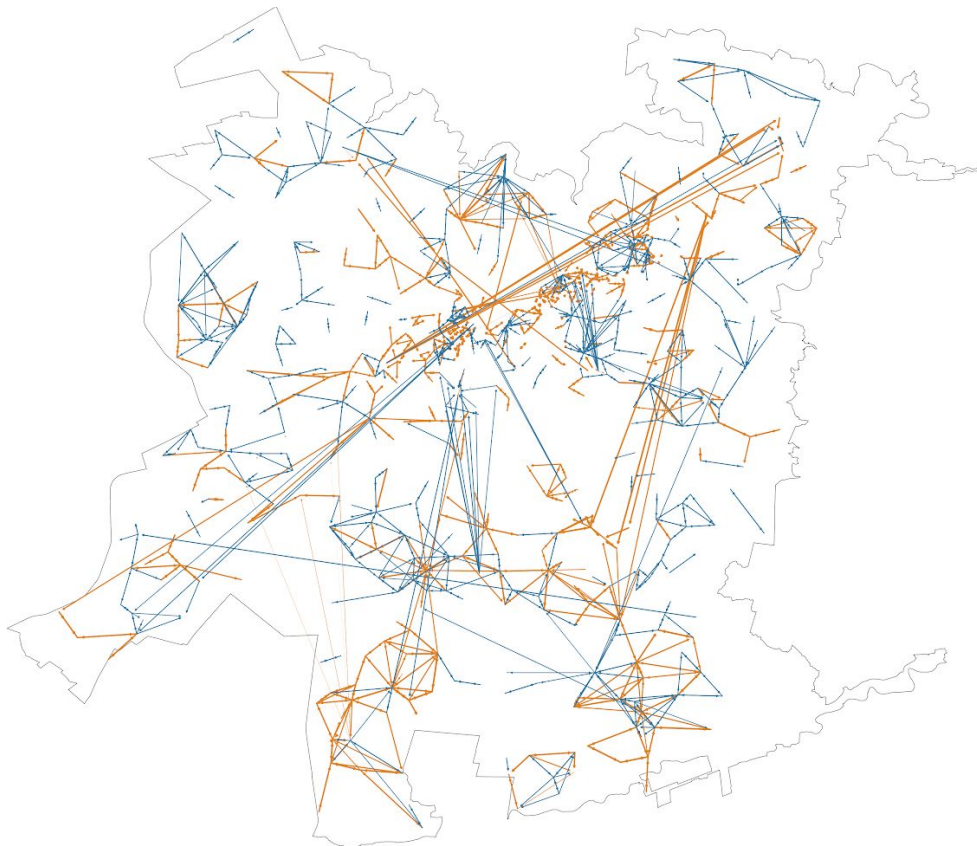**Figure 7**: Distribution of the variations of key quantities of the network of mobility.

Figure 8: Top 1000 links that registered a large reduction (orange) or a large increase (blue) after the implementation of social distancing measures.

To deepen the analysis and go beyond comparisons of statistical distributions, next we turn our attention to the networks themselves. In Figure 8 we show in orange the top 1000 mobility links that registered the largest reduction in traffic after the implementation of the social distancing measures. In blue instead we show the top 1000 links that have registered an increase after the interventions. It is interesting to notice how many of the orange links are concentrated in the commercial region. On the other hand, the blue links are, for the most part, connecting regions outside.

13-15/03 (weekend)   **16-18/03**

Figure 9: Links that recorded the most significant increase in traffic (respect to the patterns observed in two Monday-Wednesday windows) on the weekend before (left) or in the period after (right) the implementation of social distancing. For visualization purposes, we plot only the top 1000 links in both cases.

As mentioned above, the total reduction in movements between antennas observed after the start of social distancing measures is similar to the weekends. In order to explore the extent of such similarity, we did the following. We considered the mobility networks from 2-5th and from 9-11th (Monday-Wednesday) and estimated the median as well as confidence intervals of the traffic observed in each link on those six days. We then extracted the "significant" links both in the weekend (13-15/03) and after the implementation of the measures (16-18/03) as those that reported an average weight above the 95% confidence intervals estimated in the two Monday to Wednesday windows.

In Figure 9 we report the top 1000 of such links. A visual inspection of the two networks, show a good level of similarity. In fact, about 40% of the top links on the right (after the implementation of social distancing) are also present in the top links of the weekend before (on the left). It is important to stress that such links are "heavier" than those observed between the Mondays,

Tuesdays, and Wednesday before. This observation highlights how the implementation of social distancing did not halt mobility, but shifted it towards weekend patterns and traffic levels.

Preserving privacy

These experiments were carried out to be used during decision making and generally be helpful in these critical times when information is everything, while also adhering to very strict protocols to protect people's privacy. No individual or identifiable information is made available at any point. All information in these studies are aggregated at the level of either comunas or antennas, making it virtually impossible to de-identify.

For privacy reasons, the data **will not be made publicly available**. We are open to collaboration with academia, like we have done here, and like others have done already (see [here](#) and [here](#)).

Acknowledgements

This study would not have been possible without Telefónica and its country subsidiaries. They came on board during the early stages of the project just to help out. Telefonica's speed of operation, their commitment to social good, and their vision of data-driven decision making have been an inspiration to the academic team, and other teams worldwide.