Assessing the Usefulness of Digital Contact Tracing Using Real-World Contact Data

Chuan Li
Université Paris Cité
Vincent Gauthier
`vincent.gauthier@telecom-sudparis.eu`

Telecom SudParis
Miguel Nunez-del-Prado
World Bank
Hugo Alatrista-Salas
Pôle Universitaire Léonard de Vinci
Hassine Moungla
Université Paris Cité

Article

Keywords:

Posted Date: November 13th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7925876/v1

Additional Declarations: No competing interests reported.

Assessing the Usefulness of Digital Contact Tracing Using Real-World Contact Data

Chuan Li[1,2]**, Vincent Gauthier**[2]**, Miguel Nunez-del-Prado**[3,4]**, Hugo Alatrista-Salas**[5]**, and Hassine Moungla**[1,2]

[1]LIPADE, Université Paris Cité, Paris, 75006, France
[2]SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France
[3]Peruvian University of Applied Sciences, Lima, 15023, Peru
[4]The World Bank, Washington D.C., 20433, U.S.A.
[5]De Vinci Higher Education, De Vinci Research Center, Paris, France
[*]Corresponding author: Vincent Gauthier - vincent.gauthier@telecom-sudparis.eu

ABSTRACT

The significant impact of the global health crisis caused by the novel coronavirus (SARS-CoV-2) on health systems and social behaviors has prompted many countries to adopt digital contact tracing (DCT) technologies as a complement to traditional epidemiological methods. This study evaluates the effectiveness and limitations of Peru's DCT approach using real-world data from 1,660,000 users, including 80,068 confirmed cases of Covid-19. Our study provides original insights into the complex relationship between contact patterns and contagion dynamics based on data collected from Peru's DCT application in 2020. Using temporal network reconstruction via the Link Stream formalism, we thoroughly investigated the temporal interaction networks of DCT users and found that, on average, individuals who contracted the virus had more encounters than those who did not. Our research aims to understand how mobility among DCT users evolved throughout the pandemic and how contact patterns emerged within the collected data. Although limited user engagement hindered precise, individual-level tracing assessments, our findings suggest that the DCT tool has considerable potential to inform public policy on a broader scale. Additionally, the study explores the complex interplay between mobility and socioeconomic indicators, revealing significant disparities in movement patterns among different socioeconomic groups. Finally, we provide the community with a dataset containing our reconstructed large contact networks alongside the given infection status to foster future research on the development of enhanced DCT tools.

Introduction

The COVID-19 pandemic profoundly disrupted social dynamics and exposed critical weaknesses in health systems worldwide, including resource limitations, infrastructure gaps, and institutional fragmentation. To address these challenges and support overstretched healthcare services, many countries adopted digital technologies — such as contact tracing applications and Bluetooth-based proximity detection systems — as complementary tools to traditional epidemiological methods, aiming to monitor infection chains and curb the virus's spread more effectively.

In this context, the mobile application *Perú en tus manos* (Peru in your hands app) was first released on April 1st, 2020 with three main features: (1) digital triage, (2) map with regions considered at risk of contagion, and (3) contact tracing. Unlike most contact tracing applications released in the Europe, *Perú en tus manos* was designed with a centralized infrastructure. This setup enables the government to identify opted-in users, allowing for personalized isolation notifications. Specifically, Peru's DCT application uses the phone's GPS and Bluetooth low energy to gather the device's geographic location and information about potential contacts with other DCT devices in its vicinity. Both sensors gather information when the application is open or running in the background. The location data and information about nearby phones are sent through the telecommunications infrastructure to a central database. This setup is designed to detect and warn individuals who have been in the vicinity of someone with a positive virus diagnosis and, subsequently, to monitor their condition and progress through phone calls. Beyond its primary function as a DCT tool, the application has served as a key component of Peru's broader COVID-19 response strategy. Developed to notify individuals at elevated risk of infection, the application aims to facilitate timely interventions and disrupt transmission chains. This initiative is particularly vital in the context of anticipated surges in infection rates and the country's constrained healthcare infrastructure. In addition to individual-level alerts, the data generated by the application contributes to evidence-based policymaking in both the health and security sectors. In the health domain, it supports the identification of high-risk zones for targeted testing, the allocation of medical resources and personnel, the establishment of isolation centers, and the coordination of sanitation efforts. In the realm of public security, it enables authorities to manage crowd control and detect areas with heightened transmission risk. The DCT application collects anonymized geolocation, Bluetooth contact, and COVID-19 test data to estimate contagion risk. This study, however, was limited to analyzing GPS data, timestamps, and infection status due to access constraints.

The primary contribution of this study is to assess the efficacy of DCT devices in capturing contact patterns during the COVID-19 pandemic, based on data collected by the *Perú en tus manos* DCT application. This dataset encompasses information from 1.66 million users, including COVID-19 infection status for a subset of individuals. Section 2 offers a comprehensive overview of the data utilized and the chronological development of the pandemic in Peru.

Next in Section 3, we conducted an analysis of the temporal dynamics and structural characteristics of interactions between DCT devices. This investigation enabled the identification of regions exhibiting the highest movement intensity, alongside the monitoring of variations in trip length over the observation period. The results indicate a significant correlation between overall mobility patterns in Peru and the implementation of public policies. Notably, a decrease in mobility was observed during an epidemic resurgence, despite the relaxation of mobility restrictions enacted by public authorities. In this section, we used network reconstruction methodologies, using a temporal graph framework known as Link Stream, to investigate the evolution of contact patterns among DCT users throughout the pandemic period. Despite inherent limitations in the dataset, the analysis reveals significant structural differences in contact behavior between infected individuals and those in the control group.

Lastly, in Section 5 we examine the intricate relationship between mobility, socioeconomic measures, and mobility determinant of the risk of infection using SLM spatial lag model. Furthermore, we demonstrate that socioeconomic strata significantly influence individual mobility. Finally, in Section 6 we discusses our key findings and we conclude our study in Section 7.

1 Related Works

Mobile phone location data provides insights into human behavior during disasters, enabling rapid assessment of impacts through the analysis of human mobility patterns in real-time; the estimation of the displaced population extend following disasters, longer-term recovery processes by tracking population movements over time; and inferring the damage to infrastructure by analyzing changes in mobility patterns pre- and post-disaster. These high spatial and temporal resolution assessments provide detailed indicators of human behavior during disasters, significantly more effective than traditional survey methods and more targeted response efforts[1]. Regarding pandemics, location information provides information about human mobility and social interactions, which are essential for understanding and controlling the spread of COVID-19. Similarly, situational awareness was performed by analyzing mobile phone location data to avoid spreading COVID-19, which could allow Cause-Effect Analysis, forecasting trends and outbreaks, and evaluating the impact[2]. On the other hand, Luo *et al.*[3] analyzed how human mobility influenced the spread of COVID-19 in Wuhan, identified high-risk areas, assessed prevention measures, and provided insights for targeted control strategies with a view to emphasizing policies in managing public health emergencies. The radius of gyration has proved useful in analyzing human mobility during the COVID-19 pandemic, providing insight into the associated public health implications. Hernando *et al.*[4] used this indicator to predict mortality patterns in Spain, while Gauvin *et al.*[5] analyzed mobility responses in different provinces in Italy. More recently, Santana *et al.*[6] applied it in the United Kingdom to study changes in mobility behavior, combining it with measures of residual activity. These studies highlight its importance in understanding mobility dynamics and informing pandemic response strategies.

Mobile phone location data also provide a unique lens to study human behavior during disasters, including the socioeconomic changes caused by events such as the COVID-19 pandemic. Some papers have estimated socioeconomic status by analyzing individuals' mobility and social behavior through measures such as social volume and diversity and correlating these with official socioeconomic indicators[7,8]. Similarly, mobile phone location data showed how COVID-19 blocks increased mobile phone use in low-income areas and decreased it in high-income areas due to shifts in internet access[9]. Other research has shown that neighborhoods' socio-economic and geographical characteristics are important in predicting the responsiveness of essential public health and protective measures that can be taken to prevent the harmful spread of COVID-19[10]. To summarize, mobility patterns often change significantly during epidemic disasters, reflecting economic disruption[11], access to resources and shifts in priorities.

Moreover, Contact Tracing (CT) applications have played an important role in the management of epidemics such as COVID-19, using technology to identify, notify and isolate potential cases. Indeed, Muntoni et al.[12] highlight the effectiveness of probabilistic contact tracing (PCT) methods such as belief propagation and simple mean field in managing infectious disease outbreaks. Otherwise, aggregated data from contact tracing applications reveal trends in movement and interaction. A study of data from contact tracing users in Munich shows that contact duration, stopping behavior, and travel characteristics follow distinct statistical distributions that influence contact opportunities and disease transmission[13]. These studies are helping researchers understand how the virus spreads geographically and socially. Other studies are interested in investigating interpersonal contact using data other than that provided by contact tracing apps. For example, using a Gaussian probabilistic algorithm, anonymized geolocation data from mobile devices has been used to detect proximity between devices. In this way, the authors in[14] predict potential outbreaks before traditional contact tracing apps detect them. Similarly, Aleta et al.[15] estimate the probability of contact and Covid 19 transmission using a probabilistic, time-varying contact network based on individuals' stops in different scenarios, such as schools or workplaces.

Digital contact tracing systems proved effective during the COVID-19 pandemic, identifying exposure risks and helping reduce viral transmission when integrated with other public health interventions. However, their usability posed significant challenges, including accessibility for diverse populations, technological literacy requirements, limitations in device compatibility, privacy, data security and trust in institutions[16, 17]. In the same context, Li *et al.*[18] found that personal differences, including privacy concerns and trust in government institutions, significantly influenced technology adoption intentions in the United States. In addition, Oyibo et al.[19] examinated 13 articles to assess user perceptions and acceptance of DCT applications. The authors, therefore, suggest a fishbone diagram depicting key acceptance factors and related recommended design guidelines for adopting these technologies. Other authors have reviewed wireless technologies to be used in contact tracing applications, focusing on wearable devices' technical requirements and adoption challenges[20, 21].

Casaburi et al.[22] aim to understand how the mobility network in Chile responded to repeated COVID-19 intervention periods, considering socio-demographic factors, epidemiological data, and pre-intervention mobility network characteristics. The study utilizes a mobility dataset for Chile provided by Telefónica Chile, which records data-packet exchanges between anonymized users and cell phone towers from March 1, 2020, to April 28, 2021. The researchers applied spatial lag and regression models to investigate adherence to mobility interventions at the municipality level in Chile during the first and second waves of COVID-19. The analysis revealed that wealth, labor structure, COVID-19 incidence, and network metrics characterizing pre-pandemic municipality connectivity played a significant role in shaping mobility changes across the two waves. While some demographic factors like age and gender showed variability in their impact, the underlying structure of the Chilean mobility network remained largely resilient across the intervention periods. The findings suggest that historical mobility data can be valuable for informing future epidemic models and pandemic preparedness.

The study by Quach et al.[23] aimed to systematically review the global literature to comprehensively evaluate the effectiveness of DCT interventions for COVID-19 and identify factors influencing their effectiveness. The authors collected data from 133 peer-reviewed studies evaluating 121 distinct DCT implementations worldwide, supplemented by grey literature to identify a total of 171 real-world DCT interventions. Using a systematic scoping review design, the study synthesized qualitative and quantitative measures of effectiveness—including technical, end-user, population-level, and epidemiological categories—while assessing study quality with the Mixed Methods Appraisal Tool. Results showed that 60% of studies considered DCT effective, particularly when epidemiological impact metrics were evaluated. Key enablers of effectiveness included high uptake, integration with other public health measures, public trust supported by transparency and privacy standards, and strong technical performance. Conversely, barriers included poor technical performance, privacy concerns, lack of awareness, and insufficient integration into health systems. The study concluded that DCT effectiveness largely depends on public trust and its integration within broader pandemic response frameworks, underscoring the need for comprehensive communication strategies and regulatory policies for future pandemic preparedness.

2 Data and COVID-19 Epidemic Context in Peru

In this section, we provide an overview of the key dates marking the evolution of the COVID-19 pandemic in Peru, alongside a description of the datasets in the following sections were collected.

Timeline of the COVID-19 Related Events in Peru

To provide a comprehensive understanding of the pandemic's evolution in Peru, Figure 1 presents the principal milestones of the COVID-19 outbreak in chronological sequence.

On March 6th, 2020, Peru faced the emergence of the coronavirus pandemic with the first confirmed case[24]. On March 10th, classes in public and private schools were suspended until March 30th as a precautionary measure.

By 15 March, the Peruvian president had closed the borders and declared a 15-day lockdown starting on 16 March, accompanied by a curfew from 8 pm to 5 am. Despite these measures, the first death from the virus was confirmed on 20 March. The lockdown was subsequently extended twice: first on 1 April, and then again on 11 April.

On 3 April 2020, the *Te cuido Perú* app was launched to identify areas with high contagion rates. On 3 May, the government issued a decree outlining a four-step plan to restart the economy, initially targeting restaurants, delivery services, taxis, and the hotel and tourism sectors. The lockdown was extended again on 11 May.

On 24 May, the national lockdown was extended until 30 June and the wearing of masks was made mandatory in a further effort to curb the spread of the virus. On 1 June, the Minister of Economy announced the "second phase" of economic reactivation. Meanwhile, the Ministry of Education authorised the resumption of face-to-face classes in rural areas from 1 July. Selective nationwide social mobilisation measures were implemented on 31 July.

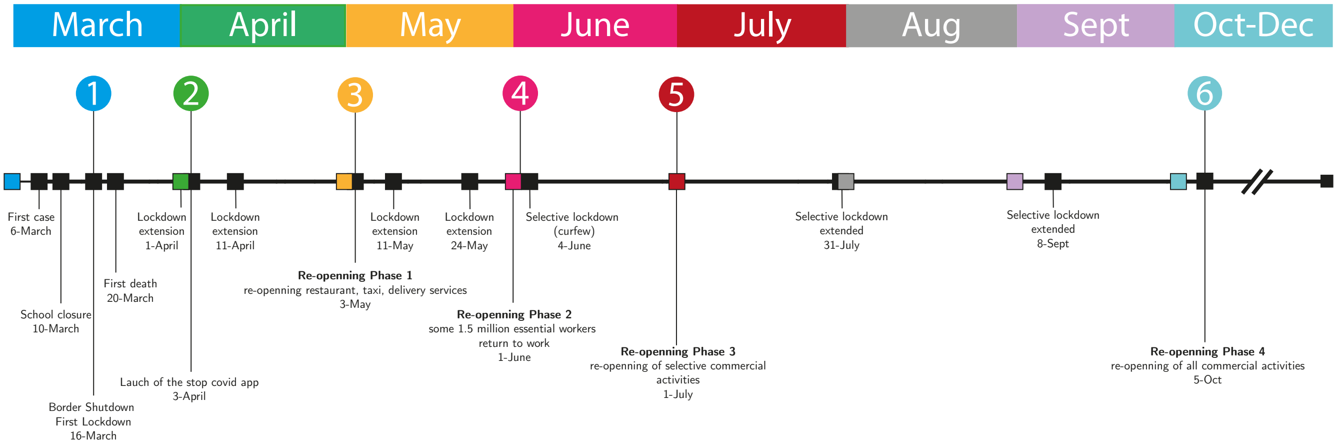

Figure 1. Timeline of the most important events from the beginning of the pandemic of COVID-19 up to September 2020

By late June, although 84% of Peru's economy had resumed production after restrictions eased, population mobility declined as COVID-19 deaths surged to 8,400 (around 500 daily). At that time, Peru ranked sixth worldwide in confirmed cases[25,26]

The "third phase" of the economic reactivation began on 1 July, with the reopening of commercial activities and public transport under certain restrictions. Despite these relaxations, the government extended the lockdown period until 31 July. Strict lockdown measures remained in place in certain provinces, reflecting regional disparities in responses to the pandemic.

At the end of August, the Peruvian government extended the selective lockdown for 90 days, to take effect on 8 September. Social distancing and the mandatory use of masks were expected to continue for the foreseeable future.

In September, vaccine trials began and new economic reactivation initiatives were announced. However, social restrictions, including a ban on holiday gatherings, remained in place and the selective lockdown was extended until 28 February 2021. The final phase of the economic reactivation was finally launched on 5 October.

Despite efforts by the government and wider society, Peru became one of the countries with the highest death tolls from the pandemic worldwide. After conducting an Oaxaca–Blinder decomposition analysis on various variables, the authors in[27] attributed Peru's high infection and fatality rates to factors such as low per capita income, inadequate budget allocation for fighting the pandemic, a high unemployment rate, low healthcare resource density, and a high co-morbidity rate. Inspired by[27,28], we hypothesize that Peru's fragile health system, which includes insufficient health care and a shortage of intensive care units and oxygen supplies, combined with overcrowding and a high prevalence of co-morbidities in the general population, has profoundly impacted the population by fostering fear and reinforcing the need to stay home.

Data collection

In this study, we analyze a subset of data collected through the DCT application *Perú en tus manos*. This dataset comprises anonymized GPS location data, timestamps, and reported infection statuses, all gathered within Peru during the year 2020. Table 1 presents descriptive statistics summarizing the dataset, which includes information from approximately 1.66 million users of the application. Among these, 663,250 were identified as active users, and 80,068 reported a positive SARS-CoV-2 diagnosis. Notably, only 10,131 of these positive cases correspond to registered users who consented to share their location data through the application. The stop locations within the GPS trajectories were extracted using the trajectory segmentation algorithm introduced in[29] and implemented via the Scikit-mobility library[30].

# records with location	844,032,954
# of user who downloaded the app	1,896,228
# user with locations records	1,662,422
# activate user with locations records ($>$ 30 records)	663,250
Collection duration	april - oct 2020
# of positive tested individuals	80,068
# of positive tested individuals that are DCT active users	10,131
# records with stop location	49,478,242
# user with stop locations records	1,410,900

Table 1. Summary statistics of the data collected by the DCT application in Peru in 2020.

3 Preliminary Analysis of Mobility Data from DCT Devices

To emphasize the two primary data components within our dataset, we first examine the temporal dynamics of infection among individuals represented in the data. This analysis is compared with external sources, as shown in Figure 2(A), which presents the normalized daily number of COVID-19 cases based on both the DCT dataset and World Health Organization (WHO) data[31]. The normalization facilitates a direct comparison between the DCT-derived figures and official WHO statistics. A clear upward trend is observed between March and June 2020, corresponding to the initial phase of the pandemic in Peru. The strong agreement between the two data sources, particularly during April and May, highlights the reliability of the DCT dataset in reflecting actual infection trends. Note that the colored labels on the x-axis in the figure correspond to significant events during the pandemic in Peru (see Figure 1).

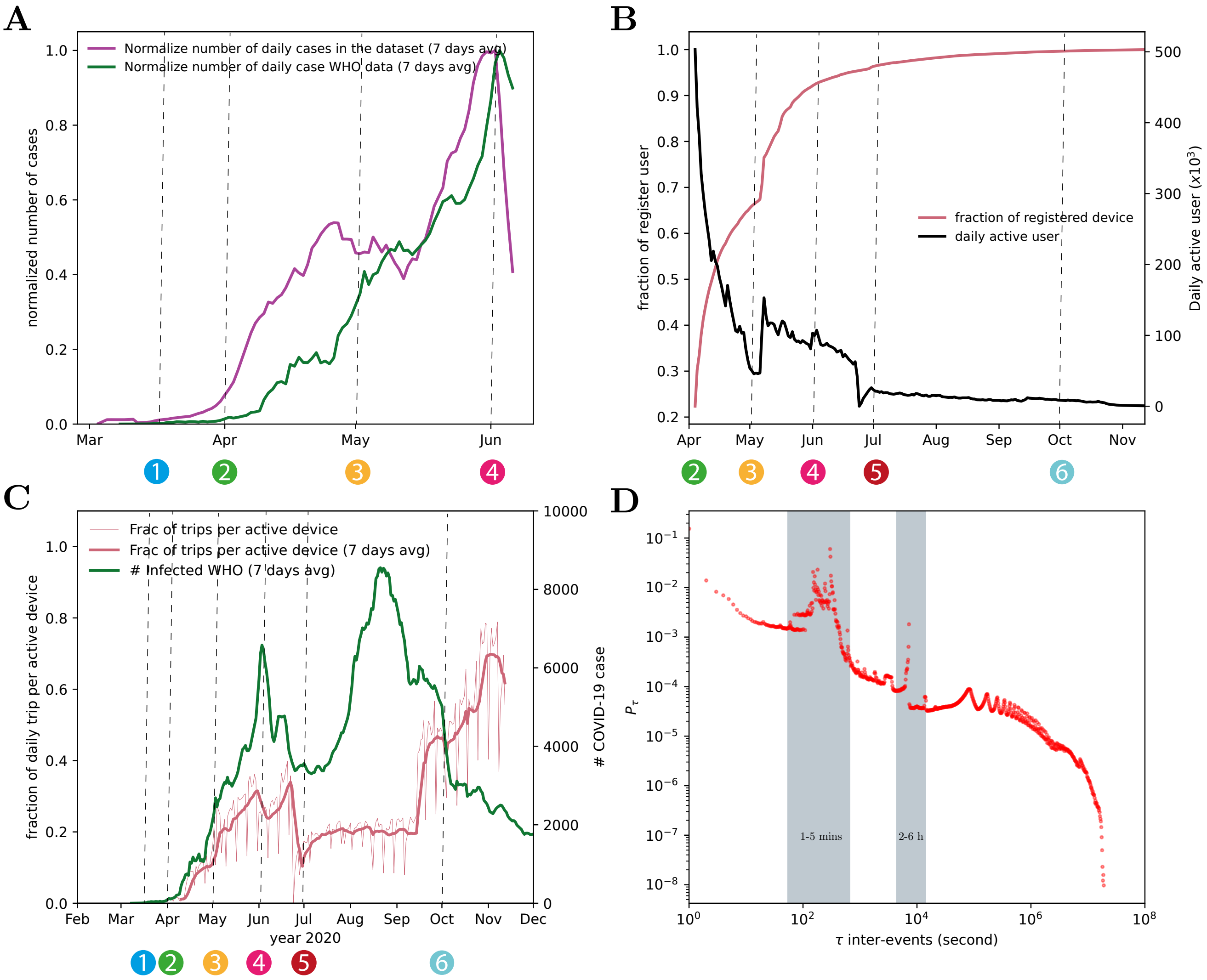

Figure 2. **A**) Evolution of the normalized 7-day average of COVID-19 cases in Peru: WHO data compared with our dataset from April to July 2020. **B**) Fraction of daily active DCT devices from April to December 2020. **C**) Evolution of the number of daily trips per active device in Peru from April to November 2020. **D**) Inter-event of location gathering time in seconds. The coloured circles in Figures **A**, **B** and **C** correspond to the key events shown on the timeline in Figure 1.

Secondly, to evaluate the reliability of our dataset, we analyze the temporal trends in both daily active users and cumulative user registrations throughout the study period. As shown in Figure 2(B), the proportion of registered devices contributing location data (depicted by the red line) reached its maximum in August 2020. Daily user activity is quantified by calculating the number of unique devices that transmitted at least one GPS data point per day. The number of active users demonstrates substantial variability over time. Specifically, daily active user counts declined from approximately 500,000 in April to a minimum in July, followed by a continued downward trend through December, which marks the end of the observation period.

Figure 2(C) below illustrates the number of daily trips per active device (pink line) and the number of reported daily cases of Coronavirus (green line) according to the World Health Organization. The light pink line shows the raw percentage of trips per active device, and the dark pink line shows the percentage smoothed using a seven-day moving average. Temporal trends in mobility data align with mobility policy changes (*i.e.* lockdowns) throughout the year, displaying significant fluctuations. There are noticeable increases in mobility during July and September, times when mobility restrictions were relaxed. In contrast, there is a significant drop in trips during strict lockdown measures (indicated by numbers ① and ② in Figure 2(C)). This data shows a positive correlation with mobility trends, especially evident during the peak infection rates in April and June.

Finally, Figure 2(D) shows the distribution of the time between two consecutive GPS position recordings from the mobile app. The distribution reveals activity patterns displaying remarkable heterogeneity; some applications exhibit high levels of activity, whereas others are significantly less active. However, clear characteristic frequencies are shown in the 1-5 minute and 2-6 hour intervals. The former highlights applications that record the phone as being in motion, while the latter highlights more typical human behavior, such as going to and from work, shopping, and spending time in the mall. The large number of events at the tail end of the distribution mostly consists of users who restart the application after a period of inactivity. Despite the uneven distribution of the sampling interval, most records can be used to analyze user position and trip details.

Despite the significant variability in usage patterns of the DCT application, as illustrated in Figure 2, our findings demonstrate that it is still possible to effectively monitor mobility trends and infection rates. These metrics can serve as valuable indicators for informing the development of epidemic management strategies.

Evolution of the Trips Length During the Pandemic

Trip length plays a crucial role in human mobility in general[32]. Here, we use this metric to characterise how travel restrictions have been applied during the different phases of the pandemic. One of the most used metrics to capture the length of the trip is the radius of gyration[33] (*c.f.* Appendix C) which measures the maximum length of regular travels an individual makes from his household (*c.f.* Appendix B, and was used to estimate new cases of COVID-19[4].

Figure 3 shows the radius of gyration behavior during the pandemic period of eight months from April to November 2020 by analyzing the daily average radius of gyration across the country. More in detail, Figure 3(A) depicts the Probability Density Function (PDF) of the radius of gyration in logarithmic scale showing that trips less than 1 *Km.* are more likely to occur than longer ones. The heavy-tailed distribution is characteristic of human mobility patterns, where the majority have confined travel ranges, and a few traverse extensive distances, which has been described in the literature[33,34]. However, Figure 3(A) shows sharp transition between different geographical scale from local (0-10 *km.*), regional (11-99 *km.*), and national (100+ *Km.*) movements, which provide a indication that the mobility was limited across these boundary which suggest that the lockdown measure had an limiting effect on the overall travel distance.

Figure 3(B) illustrates three different phases of mobility related to COVID waves in relation with the mean and percentiles (*pct*). The first phase of the economic reactivation, from May to June, shows a low average radius of gyration across the country due to the strict lockdown measures imposed by the government (see green circle ③). The second and third re-opening phases (circles ④ and ⑤), from July to September and after September respectively, show an increase in the average radius of gyration compared to the previous phases. Finally, Figure 3(C) depicts the spatial distribution of the average radius of gyration in Lima city. We have taken the Peru capital city, Lima as an example because, it has a one third of the population of the country. Regarding the spatial distribution, the radius of gyration increases as we move further away from urban centers evincing a strong difference of behavior between the areas located in the city center compared to those located in the suburban area. For instance, the radius of gyration is less than 0.1 km in Lima center, increasing to over 5 km on the outskirts of the city. Consequently, people living in the city center have high accessibility to facilities needed to sustain their daily life (hospital, shopping malls), which reduces the need to travel long distances. On the contrary, people living the suburban area have to travel further to access the same set of facilities. Additionally, people living in the suburban area have strong tendency to travel to the city center for work, despite the closure measure, which explains the higher radius of gyration of the suburban area. Our observations are supported by the analysis developed in the following subsections, where we demonstrate that the areas with the highest degree of connectivity are located in the urban areas.

4 Spatiotemporal Modeling of DCT Device Interactions Using Bipartite Stream Graphs

Temporal Evolution of Point of Interests (POIs) Connectivity

In order to identify the locations where the maximum number of interactions occur with optimal efficiency, it is essential to consider both the duration of these interactions and the variability of users' activity over time. To conduct a more thorough investigation of the mobility of DCT devices, we employ a bipartite stream graph model[35]. This model enables us to capture the temporal and structural nature of the interactions between DCT devices. For instance, Figure 4 represents a bipartite stream graph $\mathscr{B}$ where the DCT devices v_d are connected to a Point of Interest (POI), representing and hexagon or a giver area, v_p during a given time interval $[t_s, t_e[$. A point of interest is defined as an H3[36] location at a given resolution (in the present work,

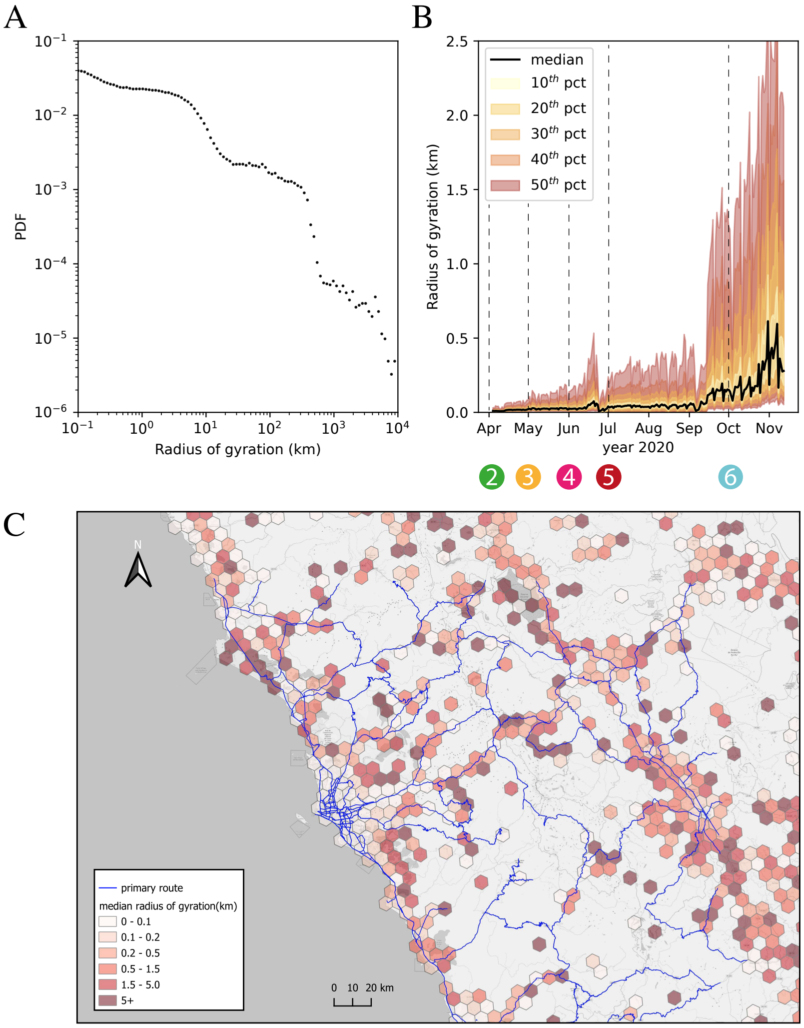

Figure 3. Radius of gyration. A) distribution of the radius of gyration. B) Temporal evolution of the daily mean radius of gyration across the country. C) Spatial distribution of the mean radius of gyration around the capital city of Lima. The coloured circles in Figure **B** correspond to the key events shown on the timeline in Figure 1.

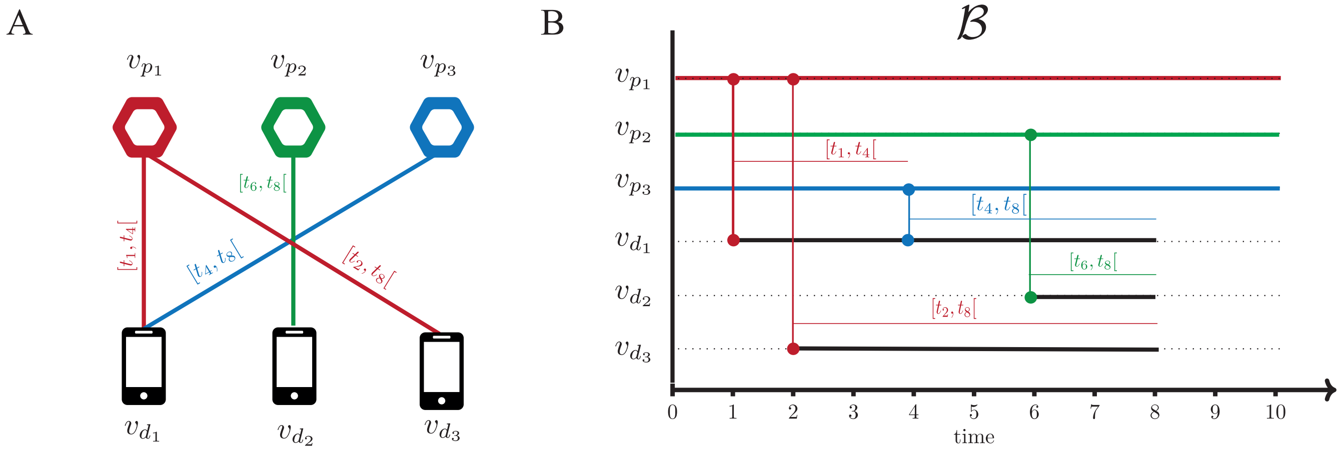

Figure 4. A) Bipartite relation in time between mobile phone device and POIs B) $\mathscr{B}$ Bipartite Stream graph representation of the mobile device with the DCT app and POIs interactions, which explicitly models the temporal dynamics of the system. Temporal edges (vertical solid line) indicate active interactions between nodes within specific time intervals with a POI for instance (e.g. the red vertical line show the interaction between the POI v_{p_1} and device v_{d_1} during interval $[t_1, t_4[$). Where the solid horizontal line represents the activity of a device or a POI (e.g. the device v_{d_1} was active during the interval $[t_1, t_8[$).

the hexagon is set at resolution level 8). Together, figures 4(A) and 4(B) provide a complementary perspective: the bipartite graph in Figure 4(A) conveys the overall structural relationships, while the stream graph in Figure 4(B) emphasizes the temporal evolution of these interactions.

More formally, we create a bipartite stream graph $\mathscr{B} = (T, V_D, V_P, W, E)$, where T is a set of time instant that define the duration of the data collected in our dataset $T = [01/04/2020, 01/01/2021]$. The vertices $v_d \in V_D$ is the finite set of user equipped with a contact tracing device and the vertices $v_p \in V_P$ the finite set of POI. W is a set of temporal nodes $W \subseteq T \times V$ that model the period of activity of a node (device in our dataset are not always active), for instance in the Figure 4, $W = [t_0, t_{10}] \times \{v_{p_1}, v_{p_2}, v_{p_3}\} \cup [t_1, t_8] \times \{v_{d_1}\} \cup [t_6, t_8] \times \{v_{d_2}\} \cup [t_2, t_8] \times \{v_{d_3}\}$. Finally E set of edges representing a interactions between a devices and a POI over specific time intervals, for instance in the Figure 4, $E = [t_1, t_4[\times \{(v_{p_1}, v_{d_1})\} \cup [t_2, t_8[\times \{(v_{p_1}, v_{d_3})\} \cup [t_4, t_8[\times \{(v_{p_3}, v_{d_1})\} \cup [t_6, t_8[\times \{(v_{p_2}, v_{d_2})\}$. In our model, a device can only belong to one POI at a time.

Given our stream graph $\mathscr{B}$, we define in Eq. 1 the POIs mean daily degree $d^i(v_p)$ as number of devices $v_d \in V_D$ that are in the POI v_p during a given day i weighted by their stay duration. We also define $E^i = \{(u, v, T_{uv}) \in E | T_{uv} \in T_i\}$ is the set of all edges that occurred during the interval T_i (day i), and the neighbor of a node v during the day i is define as the following $\mathscr{N}^i(v) = \{(T_{uv}, u), (u, v, T_{uv}) \in E^i\}$, them we compute the POIs v_p daily degree $d^i(v_p)$ as follows:

$$d^i(v_p) = \sum_{(T_{u_d v_p},\, u_d) \in \mathscr{N}^i(v_p)} \frac{|T_{u_d\, v_p}|}{|T_i|} \tag{1}$$

It is now straightforward to compute the average temporal degree and the daily average temporal degree of the bipartite stream graph as $\bar{d^i} = \frac{1}{|V_p|} \sum_{v \in V_p} d^i(v)$. In order to accurately represent the evolution of the average temporal degree of the stream graph over time in the presence of a large variance of active devices in our dataset, we normalize the POIs daily average degree $\bar{d^i}$ according to the number of devices active during day i, as follows:

$$\tilde{d} = \frac{1}{|I|} \sum_{i \in I} \frac{A^{max}}{|A^i|} \bar{d^i} \tag{2}$$

Where A^i is the set of active users during the day i, and $A^{max} = \max_{i \in I} |A^i|$. The daily mean temporal degree allows us to characterize how the intensity of spatial distributions of gatherings of people evolves over time (and thus the potential for virus transmission). In Figure 5(A), we present the daily mean temporal degree of POIs, which confirms the general trend that we previously described: the degree of POIs was low during the early days of the pandemic due to lockdown measures that limited the formation of large gatherings. Once the peak of the pandemic was over (after the summer), the average POI degree increased as people started to travel again. To elaborate further, as illustrated in Figure 5(A), the variation in the daily average temporal degree of POIs predominantly affects the tail of the distribution, as demonstrated in Figure 5(B). Figure 5(A) shows the distribution of the average temporal degree for different days in the dataset. For days with a low average temporal degree, the impact is mostly on the POIs located in the tail of the distribution. This suggests that, for the most part, only the highest degree POIs experience a reduction in degree during the lockdown period. This finding provides compelling evidence that individuals tend to avoid highly populated areas during lockdowns or in the early stages of a virus spreading. Figure 12 depicts the POI average temporal degree distribution ($\tilde{d}$) compute across all our dataset. It closely follows a truncated power law, indicated by the red dashed line, with a powerlaw exponent $\alpha = 1.60$. This pattern reveals a heavy-tailed distribution where most POIs have low average temporal degrees, but a small number exhibit significantly higher values. The truncation, visible as a sharp drop at the highest degrees, suggests an upper limit to POI connectivity. Overall, the figure highlights the strong heterogeneity in POI activity, with a few highly connected locations playing a disproportionate role in the temporal network.

Finally, Figure 5(C) shows the spatial distribution of the average temporal degree of the POIs around the capital city of Lima, where we can see that the POIs with the highest degree are located in the city center and the surrounding areas. To account the degree due to visiting devices in Figure 5(D), we explore the temporal degree of the POIs when we filter the interaction between the devices and the POIs where the device is not in his home cell during the day i. We define E^i_{fil} the set of filtered temporal edges such as:

$$E^i_{fil} = \{(u,v,T_{uv}) \in E | T_{uv} \in T_i \wedge Home(u) \neq v\} \tag{3}$$

Here the home cell of a device $Home(u)$ is defined as the cell where the device spends most of its time during the night (*c.f.* Section B). We compute $\tilde{d}^{filter}$ as defined above, but using only a subset of the interaction $E^i_{fil} \subset E^i$ (device to POIs) where a device is not present in its home cell during the day i. If we consider only the interaction where the device is not in its home cell, the spatial distribution of the average temporal degree of the POIs around the capital Lima remains the same, suggesting that the city center is the main attractor of the mobility of DCT devices.

To verify this observation, the centripetality (Γ) and anisotropy (Λ) metrics[37] are computed (*c.f.* Appendix D) to study the directionality of urban mobility. Higher anisotropy reflects a disproportionate distribution of traffic flows across different directions, with a concentration of movement along specific routes. In contrast, high centripetality indicates that the majority of traffic flows are directed towards the city center. To identify the city center, we computed the Maximum Centripetality Algorithm (MCA) over the mobility pattern (*c.f.* Figure 14). The result highlights that the city center region derived from the MCA algorithm corresponds to POIs that have the highest temporal degree of connectivity (*c.f.* figures 5(C) and 5(D). This result suggests that the city center serves as the primary attractor of urban mobility, with the POIs that have the highest connectivity are consistently clustered in near city center. This highlights the role of these areas as the central hub of Lima's mobility network.

Temporal Contact network during the pandemic

In this section, we focus on the analysis of the temporal contact between DCT devices. To better understand the contact pattern between mobile users carrying a DCT device, we projected the previously defined bipartite stream graph $\mathscr{B} = (T,V,W,E)$ (*c.f.* Figure 6(A)) onto a stream graph $\mathscr{S}$ highlighting the contact pattern between DCT devices (*c.f.* Figure 6(B)). More formally, the stream graph $\mathscr{S} = (T,V_d,W_d,E_d)$, where each $v \in V_d$ is a user equipped with a DCT device. W_d is a set of temporal nodes $W_d \subseteq W \cap (T \times V_d)$ which define the set of the time instant a node is active. For example in Figure 6(B), $W_d = ([t_1,t_4] \cup [t_5,t_8]) \times \{v_{d_1}\} \cup [t_7,t_9] \times \{v_{d_2}\} \cup ([t_2,t_8] \cup [t_8,t_9]) \times \{v_{d_3}\} \cup [t_6,t_9] \times \{v_{d_4}\}$ and define T_u as the the set of time intervals in which the node u is active (i.e. the set of time intervals in which the contact tracing application is active). E_d is the set of links $E_d = \cup_{(t,v) \in W_d} \{(t,uw)\ s.t.\ (t,v,u) \in E$ and $(t,v,w) \in E\}$. In other words, in $\mathscr{S}$ two nodes (DCT devices) are linked together at a given time instant if they have a neighbor (a POI) common in V_p at this time. For example, in Figure 6(C), and $E_d = ([t_2,t_4[) \times \{v_{d_1}v_{d_3}\} \cup [t_6,t_8[\times \{v_{d_1}v_{d_4}\} \cup [t_8,t_9[\times \{v_{d_2}v_{d_3}\}$ and we denote by T_{uv} the set of times intervals where the connection uv is present (the set of times that the DCT device u is in contact with v).

Figure 6 summarizes the transformation steps needed to generate the temporal contact networks based on the original bipartite graph between devices ($v_{d_1}, v_{d_2}, v_{d_3}, v_{d_4}$) and points of interest (POIs, $v_{p_1}, v_{p_2}, v_{p_3}$). Each edge between a device and a POI is assigned a time interval, denoting the period during which the interaction occurred. Figure 6(B,C) shows the temporal contact network between mobile devices, approximated by projecting the bipartite graph onto the device layer. This projection establishes edges between devices that were at the same point of interest (POI) during overlapping time intervals, as illustrated below. For instance, the presence of an edge between vertices v_{d_1} and v_{d_3} during the overlapping interval $[t_2,t_4]$ indicates their shared presence at v_{p_1} at that particular moment. A similar observation can be made between v_{d_3} and v_{d_4}, who share an edge during $[t_6,t_9]$ due to their shared interaction with v_{p_3}.

The projected network is an approximation of the real temporal contact network between mobile devices present in the DCT. The use of the projected network to model contacts gave rise to two strong hypotheses. Firstly, devices connected to the

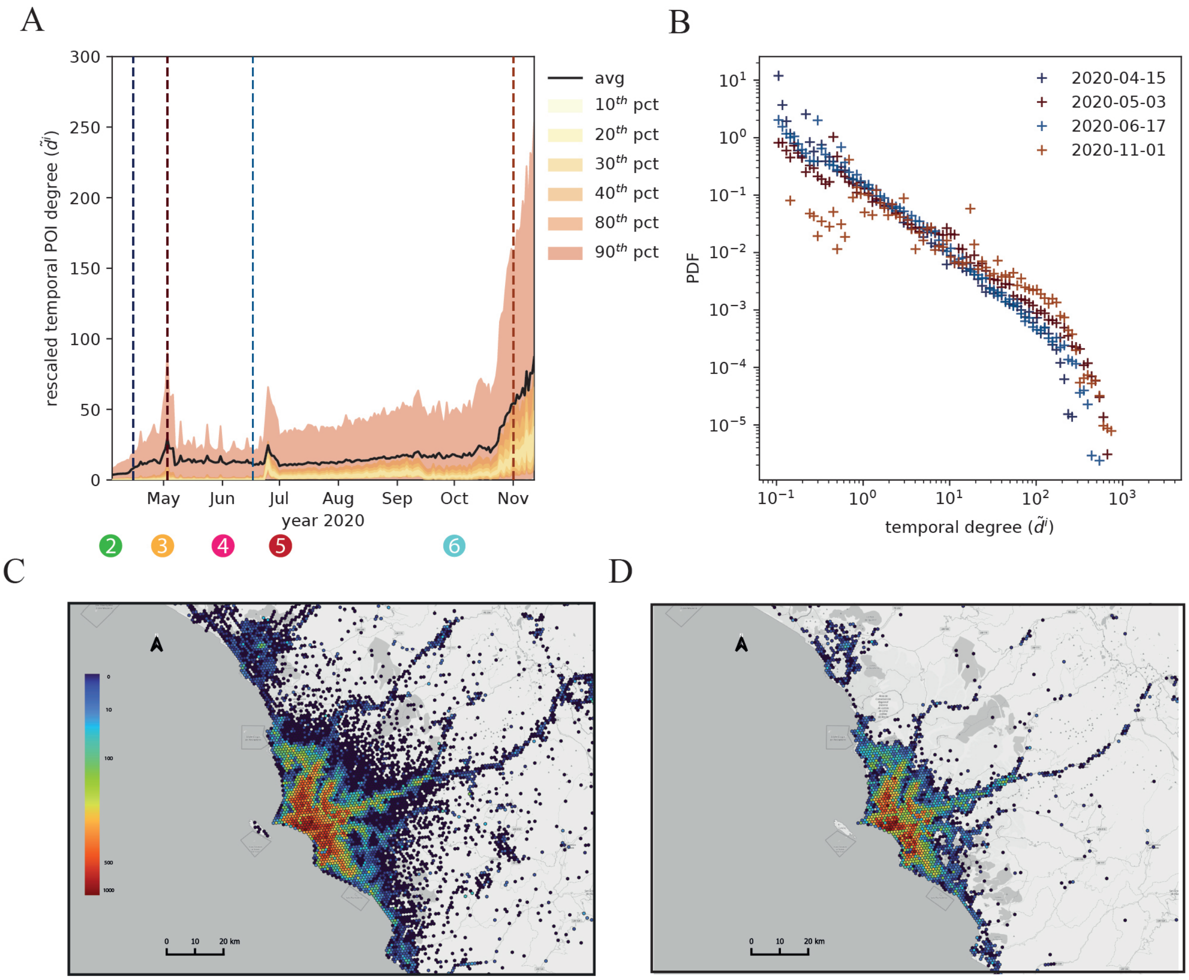

Figure 5. **A**) Daily mean temporal degree of the POIs. **B**) Distribution of the POIs mean temporal degree during the year 2020. **C**) Spatial distribution of the mean temporal degree of the POIs around the capital city of Lima. **D**) Spatial distribution of the filtered mean temporal degree of the POIs around the capital city of Lima. The coloured circles in Figure **A** correspond to the key events shown on the timeline in Figure 1.

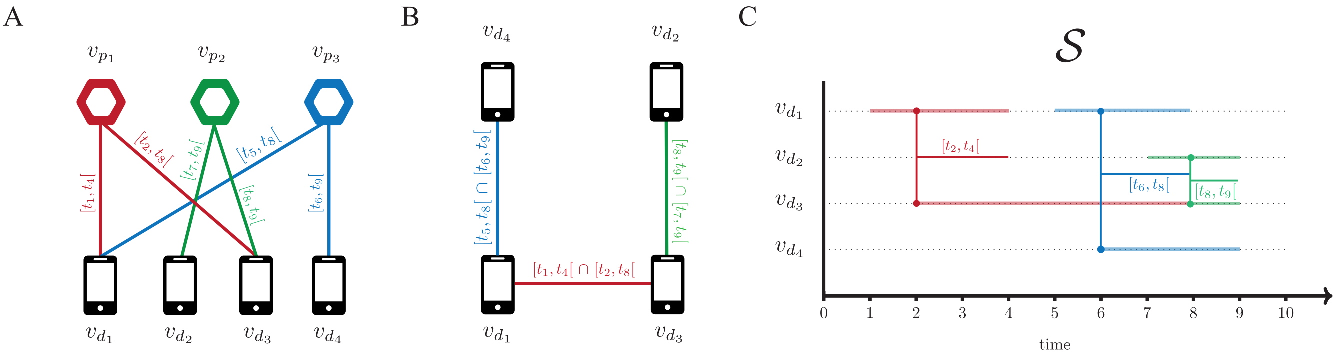

Figure 6. A) $\mathscr{B}$ Bipartite stream graph of the device POIs interactions, B) $\mathscr{S}$ is The projected temporal network of bipartite stream graph $\mathscr{B}$ C) The stream graph representation of $\mathscr{S}$

same POIs during the same time period are considered to be in contact with each other. Secondly, the device is considered to be remaining at a given POI between two records of its position. It is evident that the resulting contact networks will be significantly influenced by the spatial granularity of the POIs (larger POIs will be more connected to the projected network). To circumvent these potential pitfalls, we conducted a series of tests utilizing different POI's size for our analysis (H3 resolution level 8, 9, and 10). It is anticipated that the design of the contact network generated will be influenced by these two hypotheses. However, it is expected that the resulting network will provide a reasonable approximation of the true contact pattern of DCT users. With the stream graph $\mathscr{S}$ generated, we can now easily compute the temporal degree of a node $v \in V_p$ as follow:

$$d(v) = \sum_{u \in \mathscr{N}(v)} \frac{|T_{uv}|}{|W|} \tag{4}$$

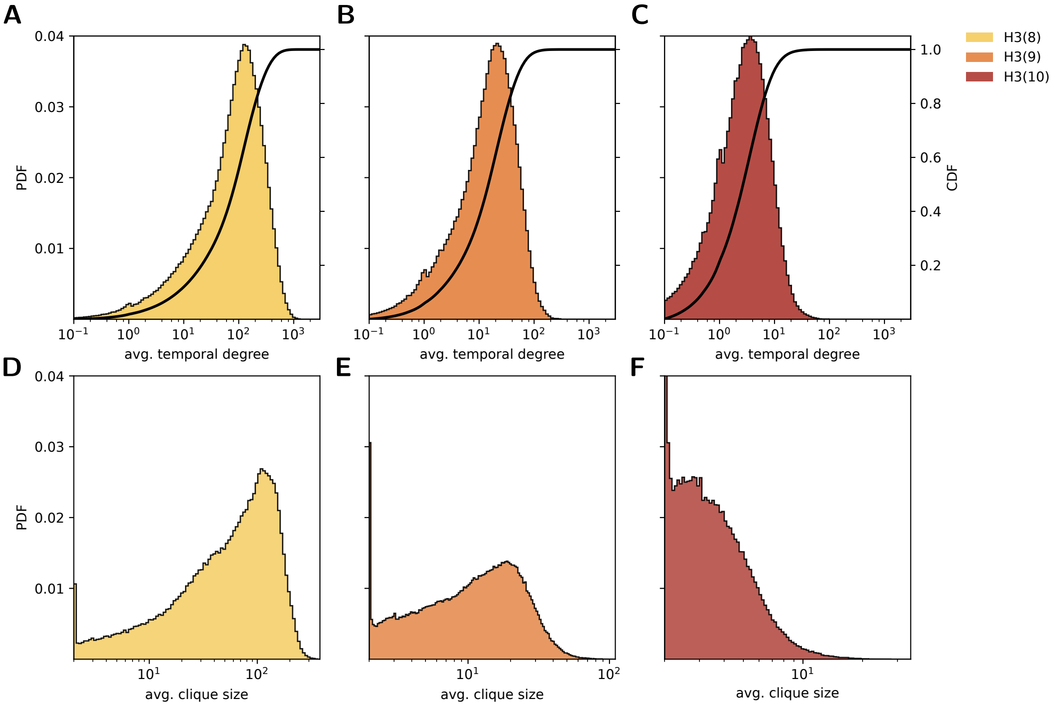

Figure 7. Distribution Temporal degree of node in the contact network A) H3 resolution level 8, B) H3 resolution level 9. C) resolution level 10. Distribution of the clique size D) H3 resolution level 8, E) H3 resolution level 9, F) H3 resolution level 10

In Figure 7 (A, B, C), we show the distribution of the temporal degree of mobile devices in DCT in the stream graph $\mathscr{S}$ for different H3 resolution levels (8, 9, and 10, respectively). The results demonstrate that the degree distribution remains consistent across various spatial granularity, despite being shifted as expected. As anticipated, the coarser the spatial granularity, the higher the device connectivity, which can be easily explained by the fact that a larger POI area will increase the possible contact duration between two devices, which in turn increases the temporal degree of nodes in contact.

To confirm the effectiveness of our method for calculating the average temporal degree, we performed a comparison with similar approach to the work described in[38], which examine the temporal changes in clique size and clique duration (*c.f.* Figure 13. To detect temporal cliques, we segmented the stream graph $\mathscr{S}$ into intervals of ten minutes each and identified all maximum cliques of size k (a fully connected subgraph with k nodes). The probability distribution of a DCT device that belongs to a temporal clique of size k is shown in Figure 13(A). The same methodology was applied to compute $P(\tau = k)$, the distribution of the duration of the clique of size k for different levels of aggregation (H3 levels 8, 9, 10, respectively) (*c.f.* Figure 13(B)). This led to a conclusion that is consistent with the findings on the temporal degree distribution: the effect of the level of aggregation does not alter the fundamental trend of the distribution of both the clique duration and size distributions. Lastly, in Figure 7 (D, E, F) The results reveal a degree of similarity in the distributions of average clique size and average temporal node degree. While both distributions represent a comparable underlying mechanism, they also capture subtle, distinct behaviours, thereby reinforcing the robustness of our findings. Furthermore, our method of computing temporal degree has a significant advantage in that it eliminates the need for data segmentation and allows for the full utilization of contact durations throughout the analysis.

In order to assess the correlation between infection susceptibility and the number of contacts, two experiments were conducted. The DCT users were divided into two distinct groups: The initial group comprised users who had contracted the virus (DCT users who had a positive PCR test in our dataset) at some point during the study period, while the second group constituted the control group. In Figure 8 (A, B) illustrate the group size distribution of both populations, as well as the cumulative distribution of the group size distribution. As illustrated in both figures, the infected population demonstrated a higher propensity to cluster in groups ranging from 1 to 30 individuals, in comparison to the control group. Nevertheless, members of the infected group are less inclined to attend large gatherings (hot-spots).

In the second experience, we computed the average temporal degree of an infected person seven days before their infection status is declared with a PCR test. We then compare this with a control population. We repeat this process for two different months (May and June) and for different spatial aggregation levels (*c.f.* Figure 8 (C,D,E,F,G,H,I,J,K)). The results obtained showed a consistent trend: infected populations were more likely to have an average temporal degree a larger than the control group.

Finally, Figure 9 (A, B) shows the average clique size and average temporal degree of the infected user and the user in the control group computed on all our dataset. Both metrics are consistently higher at all spatial levels of aggregation. Based on this observation, we hypothesized that the infected population have on average a different behavioural patterns, they are more likely to be contact with larger number of people. Moreover we show here that both the group size and the average temporal degree even computed based on large spatial proximity seems to be relevant indicator of person contact pattern. As illustrated in Figure 9 (C), our analysis reveals that individuals with greater temporal connectivity—measured by their average temporal degree—consistently face a higher likelihood of infection, regardless of the spatial scale considered. The relationship between temporal degree and infection risk follows a log-linear pattern. To assess infection probabilities, we categorized individuals into groups based on five-percentile ranges of their average temporal degree. While degree centrality is widely recognized as a key indicator for pinpointing influential nodes in disease transmission, real-world contact data is inherently imperfect, plagued by noise and gaps. Although some studies have assessed how well network measures hold up under such conditions, achieving a complete and accurate network map remains impractical in most contexts, except for highly controlled systems like animal trade networks[39–41].

These findings highlight the potential of alternative methods—specifically, those that infer dynamic contact patterns from spatiotemporal distance—as a valuable supplement to conventional approaches relying on Bluetooth-based proximity networks. The complete set of anonymized spatio-temporal contact and infection data used in this analysis is publicly available and described in detail in Open Dataset Description and Availability section, enabling full reproducibility and further research on contact behaviour modelling.

5 Mobility Drivers and Socioeconomic Disparities in COVID-19 Transmission Dynamics

In this section, we analyze the influence of mobility and social strata on infection dynamics across metropolitan Lima by leveraging a comprehensive Origin–Destination (OD) dataset (*c.f.* Table 1) and WHO reported infected population (*c.f.* Section 2) at H3 at resolution 6.

Integrating Mobility and Spatial Dependence to Understand COVID-19 Transmission

Understanding the mobility factors impacting infection dynamics is essential for accurately modelling epidemic dynamics. To analyse infection dynamics, we apply Spatial-Lag Smoothing to mobility metrics, which often show temporal gaps and localized inconsistencies ("spatial noise") that can distort mobility patterns[42]. Using key mobility metrics such as *within flow*, *in flow*, *out flow*, *average contacts*, *radius of gyration*, and *presence* derived from the Origin–Destination (OD) flows matrix (*c.f.* Table 2) in the standardized H3 grid. Then, a neighborhood graph that maximizes neighboring features while preserving the spatial structure of hexagonal tiling is built. For each POI, we compute weighted averages of mobility metrics from neighboring cells to define the spatial weight matrix W. This smoothing reduces abrupt anomalies. Furthermore, the socioeconomic status (*social strata*) of each POI is inferred from Census-based data[43] (*c.f.* Appendix F). To account for spatial dependence in infection dynamics, the analysis is extended with the Spatial Lag Models in Equation 5.

$$\mathbf{y} = \lambda \tilde{\mathbf{W}} \mathbf{y} + \sum_{j=1}^{p} \beta_j \mathbf{X}_j + \varepsilon \qquad (5)$$

where $\mathbf{y}$ is the vector normalizing each entry of the infection count in a POI; λ quantifies the degree of spatial dependence, capturing the influence of infections in spatially connected POIs; $\tilde{\mathbf{W}}$ is the normalized spatial weight matrix. Additionally, in the second part of the equation, $\sum_{j=1}^{p} \beta_j \mathbf{X}_j$ represents the aggregated contribution of p independent mobility variables, where each β_j is the coefficient quantifying the effect of the mobility metric $\mathbf{X}_j$ matrix where each row is the POI and each column is

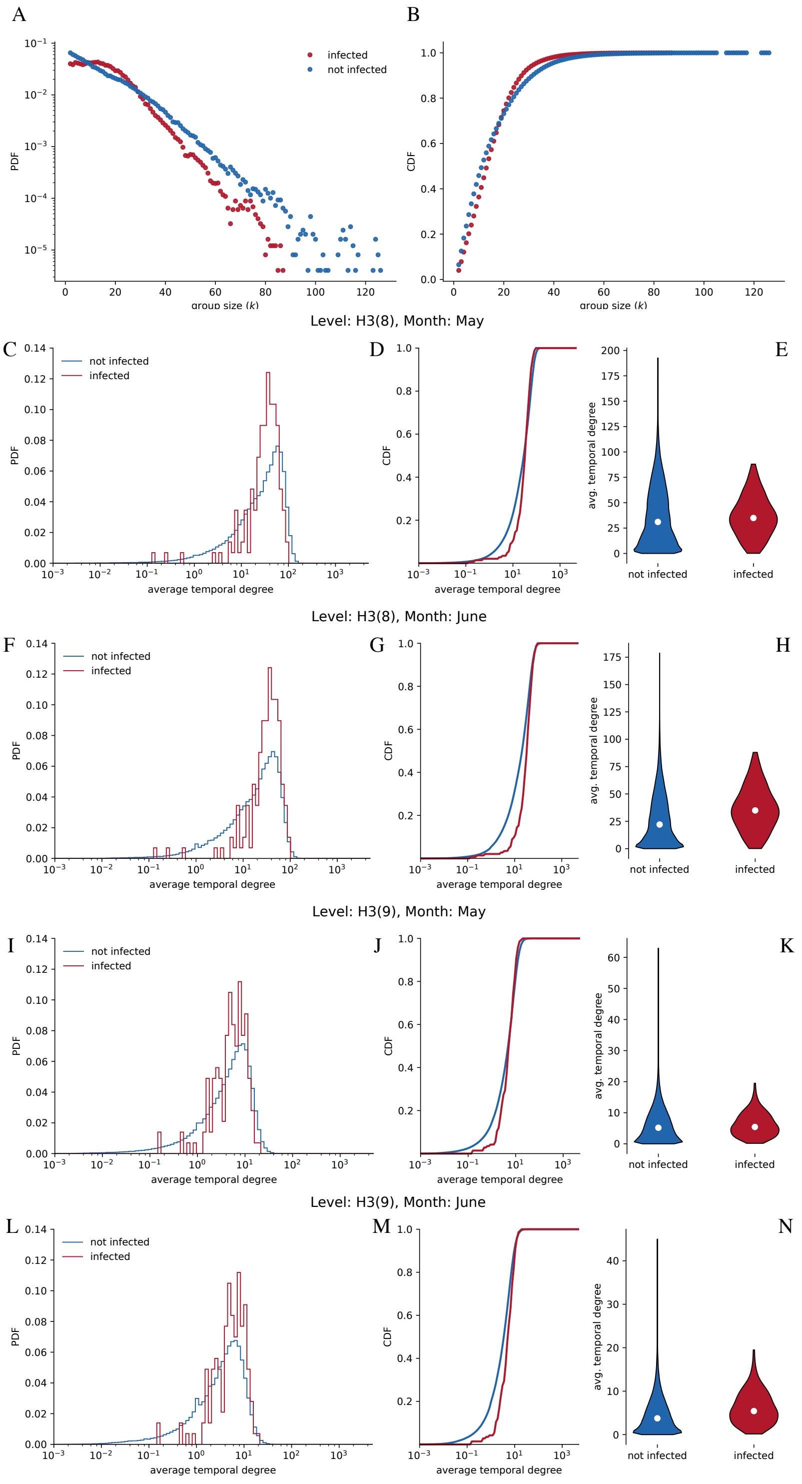

Figure 8. A) Probability Density Function (PDF) of group size for both infected and control group population, B) Cumulative Distribution Function (CDF) of group size for both infected and control group population, C, F, I, L) Probability density function of the average temporal degree of DCT devices where the contact pattern are computed according different spatial level 8 and 9 for the month of May and June, D, G, J, M) Cumulative distribution function of the average temporal degree of DCT devices where the contact pattern are computed according to spatial level 8 and 9 for for the month of May and June, E, H, K, N) Mean value of the temporal degree for both populations.

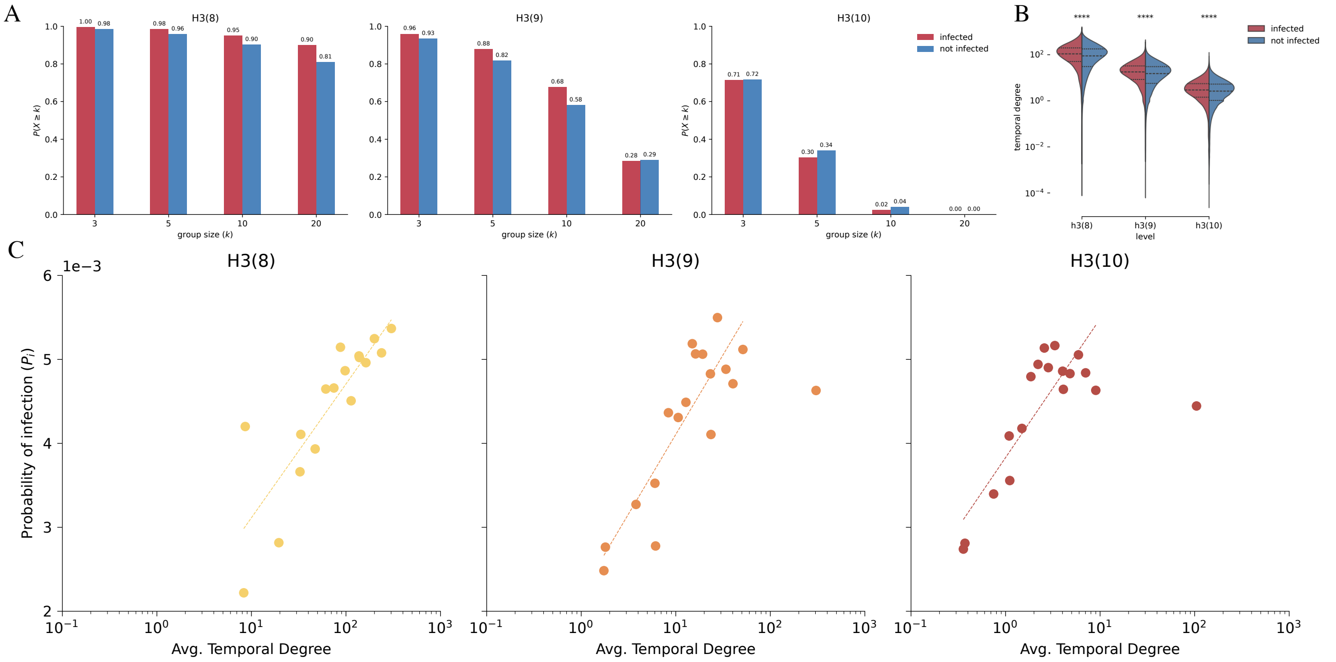

Figure 9. A) The probability distribution of a DCT device that belongs to a temporal clique of size greater or equal than k for H3 level 8, 9 and 10 computed during the period cover by our dataset, B) Average temporal degree for H3 level 8, 9 and 10, C) Average temporal degree binned every 5 percentile with the experimental probability of getting infected for different H3 level (8,9,10)

a variable characterizing the POI (*c.f.* Table 2) on infection dynamics, while ε is the error term. To address heteroscedasticity and potential outliers, the model is estimated using *Iteratively Reweighted Least Squares* (IRLS)[44]. Our primary objective is to understand the underlying mechanisms driving infection rates rather than to optimize prediction accuracy. To capture both local spatial effects and long-range mobility-driven transmission, we further extend the model by integrating ring-based contiguity and OD-based adjacencies in Equation 6.

$$y(t) = \lambda \left(\alpha \tilde{\mathbf{W}}_{od} + (1-\alpha)\tilde{\mathbf{W}}_{sp}\right) y(t) + \sum_{j=1}^{p} \beta_j \mathbf{X}_j(t) + \varepsilon, \quad 0 \leq \alpha \leq 1. \tag{6}$$

In Equation 6, the parameter α determines the relative contribution of mobility flows versus local spatial adjacency. $\tilde{\mathbf{W}}_{od}$ is the OD-based weight matrix capturing significant travel flows between non-contiguous POIs; while $\tilde{\mathbf{W}}_{sp}$ is the spatial contiguity-based weight matrix. This dual-adjacency framework enables the model to capture both short-range infection spillovers and longer-range transmission dynamics driven by high-volume travel.

Figures 10 (A) and 10 (B) display the regression coefficients for the aggregated and weekly models, respectively. The aggregated model (*c.f.* Figure 10 (A)) achieves high accuracy with a coefficient of determination $R^2 = 0.95$, while the weekly model (Figure 10 (B)) has a lower but robust $R^2 = 0.77$, reflecting greater complexity at finer temporal scales. Both models show strong positive effects of inflow and within-region flow on transmission. Outflow is more influential in the aggregated model ($\beta = 0.749$) than in the weekly model ($\beta = -0.180$), suggesting that outward movement may stabilize infections over longer periods. Socioeconomic factors also play a role: the social index (1 = poorest to 5 = wealthiest) negatively correlates with infection rates, especially in the aggregated model ($\beta = -0.118$), indicating higher risk in lower-income areas. Presence at home strongly reduces infections ($\beta = -1.545$) in the aggregated model, likely reflecting reduced mobility. Spatial dependence is strong in both models ($\lambda = 0.94$ aggregated, $\lambda = 0.77$ weekly), confirming that infections in neighboring regions influence each other and highlighting the importance of spatial modeling. Comparing Figure 10 (C) and Figure 10 (D), we observe that the aggregated models have a better predictive accuracy.

These findings highlight that integrating spatial and mobility data significantly enhances epidemiological forecasting. The aggregated model offers strong predictive power, while the weekly model captures changing trends. Together, they show that considering both spatial proximity and human movement yields more accurate, actionable insights for public health, especially

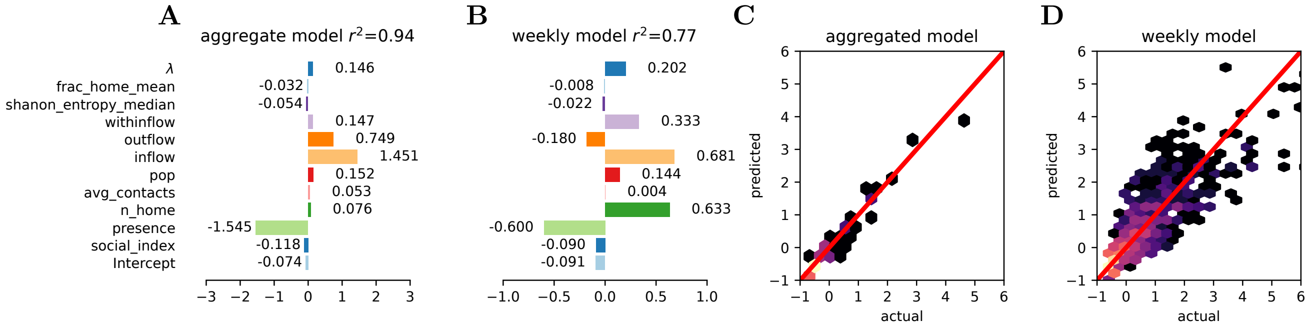

Figure 10. Performance of spatial regression models incorporating both spatial adjacency and mobility features. A, B) Regression coefficients of key predictors for the aggregated and weekly models, respectively, including OD flow components and social indicators. C, D) Predicted vs. actual infection counts for the weekly and aggregated models. Color intensity reflects data point density. The red diagonal line represents the ideal fit. The aggregated model (right) achieves higher predictive accuracy, while the weekly model captures finer-grained temporal dynamics.

for identifying vulnerable regions and targeting interventions. Incorporating OD flow data and spatial neighbor weights provides a fuller understanding of infection dynamics.

Socio-Economic Disparities in Mobility Patterns and COVID-19 Transmission Dynamics

In the present subsection, we analyze the role of socio-economic factors in shaping disparities in mobility and infection transmission during the COVID-19 pandemic.

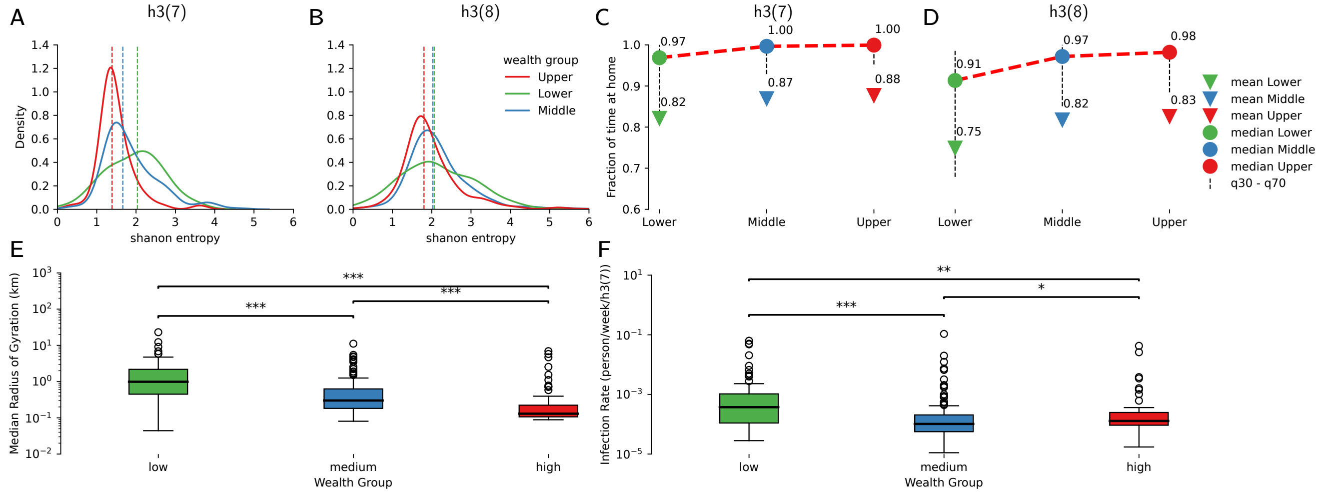

Figure 11. Socio-economic disparities in individual-level entropy metrics during the COVID-19 pandemic, measured via Shannon entropy of movement patterns at H3 resolutions 7 (A, C, E) and 8 (B, D, F)

To this end, we investigate differences in individual movement patterns across wealth groups using entropy-based mobility metrics. First, the Shannon entropy was calculated for each user in our dataset, based on the observed frequencies of visits to distinct POIs defined at H3 spatial resolution levels 7 and 8. Subsequently, each user was assigned to a social stratum derived from their inferred home cell Lower (green), Middle (blue), and Upper (red). Based on the Shannon entropy value, we plotted the kernel density distribution in Figure 11(A) and Figure 11(B), respectively. Contrary to initial expectations, the distribution for the Upper wealth group is markedly sharper and more concentrated, with a peak at a lower entropy value compared to the Middle and Lower groups. This indicates that wealthier individuals tended to visit a smaller number of distinct locations, thereby exhibiting more spatially constrained movement patterns and lower entropy. This trend is further substantiated in Figure 11(C,D), which present the *stay-at-home rate* (SHR), a behavioral metric derived from individual-level mobility data. The SHR is defined as the ratio of stops occurring in a user's home H3 cell (HP) to the total number of stops (TP) as show in Equation 7.

$$\mathrm{SHR} = \frac{\mathrm{HP}}{\mathrm{TP}} \tag{7}$$

This measure captures the proportion of time users spent at their inferred home location, offering insight into compliance with or capacity for home-based isolation[45]. In Figure 11(C,D), the mean values (triangles) and medians (circles) of the SHR are reported for each wealth group, accompanied by vertical lines representing the inter-quartile range (30th–70th percentile) and annotations of the fraction of users falling within this central distribution band. Consistently across H3 resolutions 7 and 8, individuals in the lower wealth group exhibit the lowest average and median SHR, while those in the upper wealth group demonstrate the highest. These results align with the hypothesis that wealthier individuals were more capable of sustained self-isolation, likely due to factors such as remote work feasibility, access to private transport, or secondary residences. In contrast, lower-income users maintained greater physical mobility, often driven by necessity, such as continued commuting for in-person employment or accessing essential services. Interestingly, the Lower wealth group also displays the broadest interquartile spread in SHR, underscoring a higher degree of behavioral heterogeneity within this population. This is also represented in Figure 11(A,B), with `Shannon entropy median` showing a negative effect ($\beta = -0.054$) and weekly negative effect ($\beta = -0.022$).

Figure 11(E,F) further depict the relationship between socio-economic status and individual mobility dynamics during the pandemic. The boxplots display the distribution of the *median radius of gyration* (log-transformed) across low, middle, and high wealth groups. A pronounced inverse gradient is observed, wherein average mobility range declines with increasing wealth: individuals in the low-income group exhibit a mean radius of 1.21 km, compared to 0.47 km and 0.26 km for the middle- and high-income groups, respectively. Non-parametric statistical testing using the Kruskal–Wallis test, followed by Dunn–Šidák post-hoc corrections, confirms that all pairwise differences are statistically significant $p < 0.001$ for low vs. middle, middle vs. high and low vs. high. These findings suggest that individuals in lower wealth brackets exhibit broader spatial movement, likely driven by occupational demands requiring physical presence and limited proximity to essential services. In contrast, higher-income individuals demonstrate markedly more localized and constrained mobility patterns, potentially facilitated by the flexibility of remote work, access to private transportation, and the possibility of relocating to less densely populated or better-resourced environments. This stratification in mobility behavior parallels disparities in infection risk. The low-income group displays the highest average *weekly infection rate* (3.17×10^{-3}), substantially exceeding those of the middle (1.40×10^{-3}) and high-income (1.36×10^{-3}) cohorts. This dual burden—greater physical mobility and elevated infection risk—highlights the disproportionate vulnerability of economically disadvantaged populations during periods of mobility-restricting public health interventions.

These patterns reflect more than individual behavioral choices; they are emblematic of structural inequalities that shape lived realities. Contrary to the prevailing assumption that higher socio-economic status enables more varied and extensive mobility, our results consistently demonstrate the opposite: wealthier individuals exhibited significantly lower *Shannon entropy*, higher *stay-at-home rates*, and smaller *median radius of gyration*, indicative of more spatially homogeneous and constrained behavior. Crucially, these reduced mobility patterns correspond with lower infection risk. The ability to limit movement—and thus exposure—was not equally distributed, but instead stratified along socio-economic lines. For individuals in lower wealth brackets, mobility was not a discretionary activity but a necessity, reflecting the demands of in-person labor, fragmented access to services, and limited infrastructural support. As such, the observed elevation in infection rates among these groups cannot be fully understood without accounting for the mobility patterns imposed by socio-economic constraints. Our findings underscore the importance of addressing structural determinants in both mobility behavior and public health vulnerability during pandemics. These results are corroborated with the data from official sources in Figure 15, where districts with less unemployment have the smallest death rate, while districts with more unemployment, implying people leaving their house to search for a job, have more fatalities.

6 Discussion

In this section, we analyze the scope of this research through the lens of DCT's real-world applicability. Firstly, with regard to the use of the DCT application and its adoption during the pandemic in Peru, the main observation is that the DCT had a low penetration rate of 5.7%, compared to other countries where adoption rates reached 80% in Israel and between 10% and 30% in Europe. In the Latin American context, adoption rates vary significantly. In Mexico and Argentina, they are less than 1%, while in Colombia and Uruguay, they stand at 16.8% and 19%, respectively[46,47]. In order to guarantee the requisite sensitivity and predictive accuracy in the results, it is essential to achieve a critical mass of adopters[48,49]. For instance, due to the low adoption rate of the DCT application among the population, we found that there were insufficient cases of direct contact between infected individuals and susceptible individuals who subsequently became infected to be considered significant. In the case of Peru, adoption surged at the onset of the pandemic but subsequently declined, which had a limited effect on the effectiveness of the

DCT in terms of analysis and infection prediction capabilities. The low adoption rate and declining user engagement over time have limited the usefulness of the DCT application at the micro level for establishing contact patterns over time. The main reasons for people choosing not to use the DCT tools are usability issues, concerns over data privacy[48], limited smartphone ownership[49], and even the fear that the DCT application can be used to detect curfew violations and impose sanctions[50].

Although the adoption rate in the present effort has been limited compared to similar studies using Call Detail Records[9, 22], we were still able to derive contact patterns on a large scale and the factors influencing mobility. It is also important to highlight that a majority of users relied on the DCT's hotspot infection maps to avoid areas with elevated numbers of cases. Furthermore, the app has been used to conduct self-assessments of symptoms, thereby helping to determine the necessity for testing. Furthermore, in Figure 2 (A), the purple curve representing the normalized number of daily cases inferred from the DCT. This curve demonstrates a faster and sharper increase in cases compared to the green curve, which reflects official WHO-reported cases. This suggests that the DCT dataset was more sensitive and responsive in detecting early trends, possibly due to higher user engagement and real-time data capture. However, around June, the purple curve drops sharply, likely due to a significant decline in active users, which compromises the representativeness and reliability of the data. In contrast, the WHO curve continues to rise steadily, reflecting broader population-level trends that are less affected by individual user behavior. This discrepancy underscores the trade-off between the immediacy of user-generated data and the stability of official reporting. Although DCTs can provide early signals, their utility is heavily dependent on sustained user participation.

Secondly, despite the lack of user engagement for the application, we show that the DCT data provides a comprehensive overview of mobility patterns in Peru during the Covid-19 pandemic. The data collected helps identify areas where people are still moving around. This information could allow policymakers to understand how effectively public policy compliance is being enforced and what additional measures may be needed to ensure people's safety. By comparing data from different periods of lockdown or relaxed restrictions, it is possible to observe shifts in travel habits and identify areas where new policies need to be put into place or existing ones may need adjusting. This insight has proven invaluable for decision-makers during this unprecedented crisis as they work towards keeping individuals safe while still allowing them access to essential services such as grocery stores and pharmacies.

For example, Figure 5(A) shows the daily mean temporal degree of POIs. This confirms the general trend we previously described, whereby the degree of POIs was low during the early days of the pandemic due to lockdown measures, when people were avoiding large gatherings. Once the peak of the pandemic had passed, the average POI degree increased as people started to go shopping and to the market again. The same conclusions apply to Figure 3(B), where we depict the radius of gyration distribution, showing that the average trip length changes according to the pandemic situation. This provides compelling evidence that individuals tend to avoid highly populated areas and reduce their mobility during lockdowns or in the early stages of a virus spreading.

Thirdly, our study highlights the differences in contact pattern between infected and uninfected individuals. Notably, it shows that infected individuals are more likely to have a higher average temporal degree, indicating more frequent or intense interactions that may contribute to transmission dynamics (*c.f.* Figure 8). This demonstrates that useful contact patterns can be derived even with limited spatial granularity. This could facilitate the large-scale extraction of contact information, such as mobile phone metadata. Indeed, adopting a less stringent interpretation of wireless proximity could enable us to achieve better population coverage and develop a system for gathering contact information that is more privacy-preserving than the Bluetooth low energy beaconing system currently used by contact tracing applications.

Fourthly, our empirical approach addresses gaps identified by Quach et al.[15] around contextual and granular effectiveness assessments, especially in low- and middle-income settings where peer-reviewed evidence remains sparse. Moreover, our analysis highlights practical challenges such as low adoption and privacy concerns with direct data, offering novel insights for improving future digital public health interventions. Thus, while the Quach et al. review provides valuable overarching lessons and evidence synthesis, our study represents a significant improvement in offering direct, actionable, and nuanced analysis of DCT effectiveness in a real-world, resource-constrained setting.

Finally, we found evidence in our dataset that the socio-economic level of a population affected its infection rate. We found that people in more precarious situations were more likely to take longer trips and less likely to stay at home. This increases their probability of becoming infected and spreading the virus, thereby making them more vulnerable. Therefore, these indicators could inform public interventions aimed at providing financial assistance, food or job opportunities to help reduce the mobility of the most vulnerable populations.

7 Conclusion

In this paper, we sought to contribute to the ongoing discourse on DCT by offering an evaluation of its effectiveness. To this end, we analysed a subset of the*Perú en tus manos* dataset, which comprises GPS locations, timestamps and self-reported infection status from 1.66 million users. Despite the low adoption rate (approximately 5.7%) and limited individual-level tracing, we found that mobility trends during the pandemic were strongly correlated with public health interventions, suggesting that DCT users complied with the public health policy implemented in Peru. Using a graph reconstruction technique, we analysed the contact patterns between DCT users and demonstrated that people who contracted the virus tended to have more contacts than those in the control group.

Despite its limitations, the data collected from DCT applications has been extremely important for policymakers and public health officials in assessing the effectiveness of social distancing strategies over time. In the broader context of the disaster risk management (DRM) cycle, particularly during the response phase, this information can be used to estimate the potential number of infections. This estimation is critical for budget planning, enabling the relevant authorities to allocate sufficient resources for diagnostic testing, vaccine procurement, medical personnel, and intensive care unit (ICU) capacity.

It is evident that inadequate user engagement is a matter of concern with regard to the efficacy of DCT. Without wishing to generalise, it is believed that analogous issues have been encountered in other countries. The primary factors identified as contributing to suboptimal user engagement include privacy concerns and inadequate app design. In order to prepare for future events such as the SARS-CoV-2 pandemic, it is necessary to consider alternative solutions that could provide insight into overall contact and mobility patterns if users are reluctant to engage with mobile applications that record these patterns, even if privacy concerns are technically addressed. The key finding of this study is that valuable information can still be extracted, even in the presence of larger spatial aggregation. This indicates the possibility of creating methodologies that collect information at a more coarse level, but with a wider population sample. One such methodology could involve collecting data on people's mobility and contacts using mobile phone metadata, provided that such data collection is made safe in terms of privacy and is acceptable to the population.

Declaration

Ethics Approval and Consent to Participate

The present section of the document addresses the critical concerns surrounding data protection and user privacy in the context of the DCT application build under the funding of the National Council of Science and Technology CONCYTEC "Special Projects: Response to COVID-19", through project No 70744: "COVID: COntrol de VIrus Dinamico". Given the nature of the application, which collects sensitive information such as personal identifiers, location data, and infection history, ensuring user privacy is paramount. To mitigate these concerns, the DCT tool implements several measures:

(1) Upon registration, users' personal information—including names and ID numbers—is transformed into an anonymized identifier using a SHA-256 hash function. This process aims to protect individual identities while still allowing for effective contact tracing. (2) The application was designed to retain data only to fulfill its public health objectives. This includes location data and contact history, which are crucial for identifying potential exposure risks. (3) The participation in the DCT program is voluntary, requiring explicit consent from users to collect and process their data. Users are informed about what data will be collected and how it will be used, empowering them to make informed decisions. Despite these protective measures, challenges remain regarding public trust and the potential for misuse of collected data. Privacy concerns have been cited as significant barriers to user adoption of DCT technologies. As such, ongoing efforts to educate the public about data security practices and the benefits of participation are essential to enhance user engagement and ensure the effectiveness of the DCT application in managing public health crises.

Contributions

All the authors contributed extensively to the manuscript. C.L., V.G., and M.N. contributed to the model development. C.L., V.G., M.N., and H.A. designed the experiments and wrote the main manuscript. H.M. revised and suggested the paper and helped with the formatting review and editing of the paper. All authors have read and agreed to the publication of the manuscript.

Competing interests

The authors declare no competing interests.

Funding Declaration

No funding.

Open Dataset Description and Availability

To support transparent and reproducible research in the field of computational epidemiology and mobility analytics, we release the full *anonymized contact and infection dataset* utilized in this study. The dataset—publicly available via Zenodo under DOI: `https://doi.org/10.5281/zenodo.15690187` (CC-BY 4.0)—is structured in two complementary components:

- **Spatio-temporal contact events.**

  - *H3-10 resolution:* 215,124,846 records (`par000000_contact_h3_10.zip`... `par000003_contact_h3_10.zip`)
  - *H3-9 resolution:* 820,508,676 records (`par000000_contact_h3_9.zip`... `par000017_contact_h3_9.zip`)
  - *Schema* (common to both resolutions): `o_s` [timestamp] — contact start time (UTC); `contact_duration_sec` [long] — duration in seconds; `src_device_id`, `dst_device_id` [int] — hashed device identifiers. Additional fields `date` and `year_month` are included in the H3-9 dataset for temporal grouping.

- **Infection records (lookup tables).**
  - *H3-10 resolution:* 7,415 records (`covid_infected_date_anonymized_h3_10.zip`)
  - *H3-9 resolution:* 8,026 records (`covid_infected_date_anonymized_h3_9.zip`)
  - *Schema:* `anonymous_id` [int] — hashed user identifier; `date` [string, format: `yyyy-MM-dd`] — self-reported infection onset date.

The dataset spatially covers the entirety of the Peruvian territory using the H3 geospatial indexing system. This allows for scalable spatial analysis at two granularities: an average hexagonal edge length of approximately 1.2 km at resolution 9, and 0.5 km at resolution 10.

Data privacy and ethical compliance. To ensure privacy protection, all personal identifiers have been irreversibly anonymized using the SHA-256 hashing algorithm. No raw GPS traces, personal metadata, or free-text entries are included. Infection records are limited to hashed identifiers and dates of symptom declaration, consistent with ethical guidelines for responsible data sharing.

Research utility. This dataset enables a wide range of analytical applications, including but not limited to: the reconstruction of temporal contact networks, estimation of exposure risk and secondary transmission rates, simulation-based validation of epidemic models, and spatial-temporal hotspot detection. We encourage researchers to cite the Zenodo DOI in all derivative work and to comply with the CC-BY 4.0 licensing conditions.

References

1. Yabe, T., Jones, N. K., Rao, P. S. C., Gonzalez, M. C. & Ukkusuri, S. V. Mobile phone location data for disasters: A review from natural hazards and epidemics. *Comput. Environ. Urban Syst.* **94**, 101777 (2022).
2. Oliver, N. *et al.* Mobile phone data for informing public health actions across the covid-19 pandemic life cycle. *Sci. advances* **6**, eabc0764 (2020).
3. Luo, M. *et al.* Population mobility and the transmission risk of the covid-19 in wuhan, china. *ISPRS Int. J. Geo-Information* **10**, 395 (2021).
4. Hernando, A., Mateo, D., Bayer, J. & Barrios, I. Radius of gyration as predictor of covid-19 deaths trend with three-weeks offset. *medRxiv* 2021–01 (2021).
5. Gauvin, L. *et al.* Socio-economic determinants of mobility responses during the first wave of covid-19 in italy: from provinces to neighbourhoods. *J. The Royal Soc. Interface* **18**, 20210092 (2021).
6. Santana, C. *et al.* Covid-19 is linked to changes in the time–space dimension of human mobility. *Nat. Hum. Behav.* **7**, 1729–1739 (2023).
7. Pappalardo, L. *et al.* An analytical framework to nowcast well-being using mobile phone data. *Int. J. Data Sci. Anal.* **2**, 75–92 (2016).
8. Šćepanović, S., Mishkovski, I., Hui, P., Nurminen, J. K. & Ylä-Jääski, A. Mobile phone call data as a regional socio-economic proxy indicator. *PloS one* **10**, e0124160 (2015).
9. Zanella, A. F., Rubrichi, S., Smoreda, Z. & Fiore, M. Modeling and understanding the impact of covid-19 containment policies on mobile service consumption in french cities. *EPJ Data Sci.* **13**, 68 (2024).
10. Boz, H. A. *et al.* Investigating neighborhood adaptability using mobility networks: a case study of the covid-19 pandemic. *Humanit. Soc. Sci. Commun.* **11**, DOI: 10.1057/s41599-024-02881-1 (2024).
11. Alatrista-Salas, H., Gauthier, V., Nunez-del Prado, M. & Becker, M. Impact of natural disasters on consumer behavior: Case of the 2017 el niño phenomenon in peru. *PLOS ONE* **16**, e0244409, DOI: 10.1371/journal.pone.0244409 (2021).

12. Muntoni, A. P., Mazza, F., Braunstein, A., Catania, G. & Dall'Asta, L. Effectiveness of probabilistic contact tracing in epidemic containment: The role of superspreaders and transmission path reconstruction. *PNAS Nexus* **3**, pgae377, DOI: 10.1093/pnasnexus/pgae377 (2024). https://academic.oup.com/pnasnexus/article-pdf/3/9/pgae377/59151486/pgae377.pdf.

13. Tonetto, L., Adikari, M., Mohan, N., Ding, A. Y. & Ott, J. Contact duration: Intricacies of human mobility. *Online Soc. Networks Media* **28**, 100196 (2022).

14. Crawford, F. W. *et al.* Impact of close interpersonal contact on covid-19 incidence: Evidence from 1 year of mobile device data. *Sci. Adv.* **8**, eabi5499, DOI: 10.1126/sciadv.abi5499 (2022). https://www.science.org/doi/pdf/10.1126/sciadv.abi5499.

15. Aleta, A. *et al.* Quantifying the importance and location of sars-cov-2 transmission events in large metropolitan areas. *Proc. Natl. Acad. Sci.* **119**, e2112182119 (2022).

16. Abeler, J., Bäcker, M., Buermeyer, U., Zillessen, H. *et al.* Covid-19 contact tracing and data protection can go together. *JMIR mHealth uHealth* **8**, e19359 (2020).

17. Buckee, C. O. *et al.* Aggregated mobility data could help fight covid-19. *Science* **368**, 145–146 (2020).

18. Li, T. *et al.* What makes people install a covid-19 contact-tracing app? understanding the influence of app design and individual difference on contact-tracing app adoption intention. *Pervasive Mob. Comput.* **75**, 101439 (2021).

19. Oyibo, K., Sahu, K. S., Oetomo, A. & Morita, P. P. Factors influencing the adoption of contact tracing applications: systematic review and recommendations. *Front. Digit. Heal.* **4**, 862466 (2022).

20. Shubina, V., Ometov, A. & Lohan, E. S. Technical perspectives of contact-tracing applications on wearables for covid-19 control. In *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 229–235 (IEEE, 2020).

21. Ng, P. C., Spachos, P., Gregori, S. & Plataniotis, K. Epidemic exposure notification with smartwatch: A proximity-based privacy-preserving approach. *arXiv preprint arXiv:2007.04399* (2020).

22. Casaburi, P. *et al.* Resilience of mobility network to dynamic population response across covid-19 interventions: Evidences from chile. *PLOS Comput. Biol.* **21**, e1012802 (2025).

23. Quach, H.-L., Walsh, E. I., Hoang, T. N. A., Terrett, R. N. L. & Vogt, F. Effectiveness of digital contact tracing interventions for covid-19: A systematic scoping review. *Public Heal.* **242**, 146–156 (2025).

24. Online resource (2025).

25. Ministerio de Salud del Perú. Covid 19 en el perú – ministerio de salud (2025).

26. Johns Hopkins University, Center for Systems Science and Engineering (CSSE). Covid-19 map (2025).

27. Cajachagua-Torres, K. N. *et al.* Covid-19 and drivers of excess death rate in peru: A longitudinal ecological study. *Heliyon* **8**, e11948, DOI: https://doi.org/10.1016/j.heliyon.2022.e11948 (2022).

28. Schwalb, A. & Seas, C. The covid-19 pandemic in peru: what went wrong? *The Am. journal tropical medicine hygiene* **104**, 1176 (2021).

29. Zheng, Y. Trajectory data mining: an overview. *ACM Transactions on Intell. Syst. Technol. (TIST)* **6**, 1–41 (2015).

30. Pappalardo, L., Simini, F., Barlacchi, G. & Pellungrini, R. scikit-mobility: A python library for the analysis, generation, and risk assessment of mobility data. *J. Stat. Softw.* **103**, 1–38, DOI: 10.18637/jss.v103.i04 (2022).

31. World Health Organization. Who | peru (2025). Accessed: 2025-08-25.

32. Gambs, S., Killijian, M.-O. & del Prado Cortez, M. N. Show me how you move and i will tell you who you are. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*, 34–41 (2010).

33. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *nature* **453**, 779–782 (2008).

34. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. A tale of many cities: universal patterns in human urban mobility. *PloS one* **7**, e37027 (2012).

35. Latapy, M., Viard, T. & Magnien, C. Stream graphs and link streams for the modeling of interactions over time. *Soc. Netw. Analysis Min.* **8**, DOI: 10.1007/s13278-018-0537-7 (2018).

36. Uber Technologies, Inc. H3: A hexagonal hierarchical geospatial indexing system (2025). Documentación oficial del sistema H3.

37. Zhao, P., Wang, H., Liu, Q., Yan, X.-Y. & Li, J. Unravelling the spatial directionality of urban mobility. *Nat. Commun.* **15**, DOI: 10.1038/s41467-024-48909-7 (2024).

38. Iacopini, I., Karsai, M. & Barrat, A. The temporal dynamics of group interactions in higher-order social networks. *Nat. Commun.* **15**, DOI: 10.1038/s41467-024-50918-5 (2024).

39. Bucur, D. & Holme, P. Beyond ranking nodes: Predicting epidemic outbreak sizes by network centralities. *PLOS Comput. Biol.* **16**, e1008052, DOI: 10.1371/journal.pcbi.1008052 (2020).

40. Smieszek, T. & Salathé, M. A low-cost method to assess the epidemiological importance of individuals in controlling infectious disease outbreaks. *BMC medicine* **11**, 35 (2013).

41. Génois, M. & Barrat, A. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Sci.* **7**, 1–18 (2018).

42. Anselin, L. & Smirnov, O. Efficient algorithms for constructing proper higher order spatial lag operators. *J. regional science* **36**, 67–89 (1996).

43. Instituto Nacional de Estadistica e Informática INEI. Planos estratificados de lima metropolitana a nivel de manzanas 2020. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1744/libro.pdf (2020). Accessed: date-of-access.

44. O'Leary, D. P. Robust regression computation using iteratively reweighted least squares. *SIAM J. on Matrix Analysis Appl.* **11**, 466–480 (1990).

45. Lucchini, L. *et al.* Socioeconomic disparities in mobility behavior during the covid-19 pandemic in developing countries. *EPJ Data Sci.* **14**, 25, DOI: 10.1140/epjds/s13688-025-00532-2 (2025).

46. Ordóñez, S. & Ross, G. Tracing latin america's contact tracing. https://www.wilsoncenter.org/blog-post/tracing-latin-americas-contact-tracing (2021). Accessed: March 12, 2025.

47. Comparitech Research Team. Contact-tracing app adoption by country: How many people are willing to compromise on their privacy? https://www.comparitech.com/blog/vpn-privacy/contact-tracing-app-adoption-by-country/ (2021). Accessed: March 12, 2025.

48. Nabeel, A., Al-Sabah, S. K. & Ashrafian, H. Digital contact tracing applications against covid-19: a systematic review. *Med. Princ. Pract.* **31**, 424–432 (2022).

49. Kleinman, R. A. & Merkel, C. Digital contact tracing for covid-19. *Cmaj* **192**, E653–E656 (2020).

50. Hiperderecho. Así funciona la aplicación móvil perú en tus manos: técnica y legalmente (2020).

51. Csáji, B. C. *et al.* Exploring the mobility of mobile phone users. *Phys. A: statistical mechanics its applications* **392**, 1459–1473 (2013).

52. Phithakkitnukoon, S., Smoreda, Z. & Olivier, P. Socio-geography of human mobility: A study using longitudinal mobile phone data. *PloS one* **7**, e39253 (2012).

53. Centro Nacional de Epidemiologia, prevención y Control de Enfermedades – MINSA. Fallecidos por covid 19. https://www.datosabiertos.gob.pe/dataset/fallecidos-por-covid-19-ministerio-de-salud-minsa (2020).

54. Instituto Nacional de Estadística e Informática INEI. Encuesta Nacional de Hogares (ENAHO). https://www.datosabiertos.gob.pe/dataset/encuesta-nacional-de-hogares-enaho-2020-instituto-nacional-de-estad%C3%ADstica\protect\discretionary{\char\hyphenchar\font}{}{}e\protect\discretionary{\char\hyphenchar\font}{}{}inform%C3%A1tica\protect\discretionary{\char\hyphenchar\font}{}{}inei (2020).