



Article

Face-to-Face Interactions Estimated Using Mobile Phone Data to Support Contact Tracing Operations

Silvino Pedro Cumbane ^{1,2}, Gyözö Gidófalvi ¹, Osvaldo Fernando Cossa ^{3,4}, Afonso Madivadua Júnior ⁴,
Nuno Sousa ^{3,5} and Frederico Branco ^{6,7,*}

¹ Division of Geoinformatics, Department of Urban Planning and Environment, KTH Royal Institute of Technology, Teknikringen 10A, SE-114 28 Stockholm, Sweden; silvino.cumbane@uem.ac.mz (S.P.C.); gyozo@kth.se (G.G.)

² Division of Geographic Information Sciences, Department of Mathematics and Informatics, Eduardo Mondlane University, Julius Nyerere Street, Maputo 3453, Mozambique

³ Department of Sciences and Technology, Universidade Aberta, 1269-001 Lisbon, Portugal; ofcossa@gmail.com (O.F.C.); nuno.sousa@uab.pt (N.S.)

⁴ Department of Service Quality, Instituto Nacional das Comunicações de Moçambique-INCM, Praça 16 de Junho, Maputo 848, Mozambique; ajunior@incm.gov.mz

⁵ Institute for Systems Engineering and Computers at Coimbra (INESCC), 3030-790 Coimbra, Portugal

⁶ School of Science and Technology, Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal

⁷ Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), 4200-465 Porto, Portugal

* Correspondence: fbranco@utad.pt

Abstract: Understanding people's face-to-face interactions is crucial for effective infectious disease management. Traditional contact tracing, often relying on interviews or smartphone applications, faces limitations such as incomplete recall, low adoption rates, and privacy concerns. This study proposes utilizing anonymized Call Detail Records (CDRs) as a substitute for in-person meetings. We assume that when two individuals engage in a phone call connected to the same cell tower, they are likely to meet shortly thereafter. Testing this assumption, we evaluated two hypotheses. The first hypothesis—that such co-located interactions occur in a workplace setting—achieved 83% agreement, which is considered a strong indication of reliability. The second hypothesis—that calls made during these co-location events are shorter than usual—achieved 86% agreement, suggesting an almost perfect reliability level. These results demonstrate that CDR-based co-location events can serve as a reliable substitute for in-person interactions and thus hold significant potential for enhancing contact tracing and supporting public health efforts.

Keywords: Call Detail Records (CDRs); co-location; face-to-face meetings; contact tracing; Mozambique



Academic Editor: Domenico Ursino

Received: 14 November 2024

Revised: 17 December 2024

Accepted: 21 December 2024

Published: 30 December 2024

Citation: Cumbane, S.P.; Gidófalvi, G.; Cossa, O.F.; Júnior, A.M.; Sousa, N.; Branco, F. Face-to-Face Interactions

Estimated Using Mobile Phone Data to Support Contact Tracing Operations. *Big Data Cogn. Comput.* **2025**, *9*, 4.

<https://doi.org/10.3390/bdcc9010004>

<https://doi.org/10.3390/bdcc9010004>

<https://doi.org/10.3390/bdcc9010004>

Copyright: © 2024 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Infectious diseases are among the top causes of death globally each year. For example, in 1998, only six diseases (AIDS, malaria, tuberculosis, measles, diarrheal diseases, and acute respiratory infections) accounted for nearly 90% of deaths of children and young adults [1]. Recently, in January 2020, the first case of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), or simply COVID-19, was reported, and three years later, it infected and killed over 760 million and 6.9 million people, respectively, as of 10 June 2023 [2]. In many cases, the virus is transmitted through physical interaction between people. To manage the spread of infectious diseases and reduce the societal impact of lockdowns,

contact tracing, case isolation, and quarantine have been recommended as the best non-pharmaceutical measures [3–7]. Contact tracing involves contacting exposed individuals to support and reduce transmission, using technology to maximize responses [8].

Traditionally, after a confirmed or probable case of an infected person is identified, health authorities typically conduct an interview with the individual, often via phone. ECDPC [9] lists the generic steps for the contact tracing process as (a) identifying the contacts of a COVID-19 case; (b) providing contacts with information on self-quarantine, proper hand hygiene, respiratory etiquette, and guidance on what to do if symptoms develop; (c) testing all high-risk exposure contacts, regardless of vaccination status, as soon as possible after identification to facilitate further contact tracing; (d) testing all unvaccinated low-risk exposure contacts; and (e) testing all contacts that become symptomatic. However, this case-by-case approach is very time-consuming and costly. Tracing all contacts can be challenging when individuals have numerous contacts, struggle to remember them accurately, or are unable to provide information on how to reach them.

Therefore, many smartphone applications (apps) have been suggested to support and enhance the traditional contact tracing approach [10]. Based on radio wave sensors, a smartphone can record when two individuals are in close proximity for long enough, creating a high risk of transmission if one of them is infected, such as with the coronavirus [11]. During the coronavirus pandemic, contact tracing apps were mandatory in East Asian countries such as China and especially Taiwan, and their efficacy alongside manual contact tracing methods in identifying new cases have been proven effective [12,13]. The apps use algorithmic analysis of the user's travel history and health status to assign a green code (which demonstrates that they have not been in contact with a confirmed case of COVID-19), yellow code (which indicates that they have a certain probability of having been in contact with a confirmed COVID-19 case), and red code (which shows that they have been in contact with a confirmed case of COVID-19). In South Korea, contact tracing apps like 'Corona 100' have gained popularity and helped public health officials reduce the time required to trace patients' movements by approximately 24 h, thus assisting the public in avoiding potentially infectious areas. However, according to Akinbi et al. [10], there is limited evidence to suggest that the approaches used in East Asian countries would be easily transferable to neoliberal societies such as the USA, UK, France, Germany, and others, which have different political and cultural systems. Therefore, several challenges have been identified in these applications: the set of personal data, including location and health information, raise significant privacy issues. Users may be reluctant to adopt apps that compromise their personal information [14,15]; inadequate security measures can result in data breaches, revealing sensitive information to unauthorized individuals [16,17]; variations in device capabilities and operating systems can affect the accuracy and reliability of contact tracing [18]; and the effectiveness of these apps depends on widespread adoption and consistent usage, which can be influenced by user trust and perceived benefits [19].

To address these challenges, the use of anonymized CDRs has been proposed as a complementary approach. Anonymized CDRs provide aggregated data on user movements and interactions without revealing personal identities, thereby mitigating privacy concerns. This method has been shown to offer valuable insights into population mobility patterns and potential exposure risks. Integrating anonymized CDRs into contact tracing efforts can enhance the effectiveness of public health responses while preserving user privacy and security. Anonymized CDRs provide aggregated data on user movements and interactions without revealing personal identities, thereby mitigating privacy concerns. This method has been shown to offer valuable insights into population mobility patterns and potential exposure risks [20].

This study used anonymized CDRs to estimate face-to-face interactions to support contact tracing operations. This approach does not require users to install any app on their smartphone. It does not compromise users' privacy because the data are anonymized by the operator/regulator, and spatial resolution is at the cell tower level, i.e., it is impossible to extract the exact position of the anonymized user. In addition, it does not depend on user behavior and participation because all the subscribers that use their mobile phones automatically generate CDRs. Moreover, it has no technical constraints because the data capture all subscribers (feature phone and smartphone users and does not require the subscriber to install any app).

To ensure that the results are meaningful and accessible to decision-makers, the concept of Post-Administrative Units (PAUs) was introduced. PAUs provide an aggregated level of geographic boundaries that reflect postal or administrative divisions, allowing stakeholders such as health authorities or urban planners to more readily understand and utilize the data. By mapping cell-tower-based analyses onto PAUs, highly granular, technical information is transformed into a format aligned with recognized territorial demarcations. This approach enables a clearer interpretation of the results, ensuring that the insights derived from the study can effectively inform policy decisions and strategic interventions.

The key contributions of this study are as follows: (1) developing a technique to compare users' usual calling patterns with their calling behavior immediately after co-location events, thereby enabling the identification of significant changes in their interactions, (2) the development of a method to determine the residential and work locations of mobile subscribers through the analysis of CDRs, (3) creating a predictive model capable of inferring likely in-person meetings by examining co-location patterns between individuals, and (4) the proposal of an effective method to trace potential contacts among subscribers during public health emergencies, such as epidemics, providing an ethical and efficient alternative for contact tracing.

The rest of the paper is structured as follows. Section 2 reviews the related work conducted in this area. Section 3 outlines the data used in this study and the proposed method. Section 4 presents the results and discussion. Finally, Section 5 concludes and presents future directions of the research.

2. Related Works

Anonymized CDRs have been used for different purposes, such as human dynamics [21], human mobility patterns [22–25], and disaster management [26–32]. However, few studies have investigated using anonymized mobile CDRs to estimate face-to-face interactions. Jin and Park [33] conducted a survey of 232 college students who owned a cell phone to investigate the relationship between cellphone use, interpersonal motives for using cell phones, face-to-face communication, and loneliness. The authors discovered that the level of face-to-face interaction was positively correlated with the participants' cellphone use and their interpersonal motives for using cell phones. In other words, the more the participants engaged in face-to-face interactions with others, the stronger their motives for using cell phones and the more frequently they used them. Przybylski and Weinstein [34] evaluated the extent to which mobile communication devices shape relationship quality in dyadic settings. The authors found that the presence of mobile communication devices negatively impacts closeness, connection, and the quality of conversations. Another study from Hamann [35] investigated the impact of mobile phones on the ability to communicate. As in [34], the authors found that their mere presence in any social setting affects a person's ability to coherently speak with someone in all aspects of face-to-face communication. Most of the studies presented above focused on investigating the behavior of mobile phone users during face-to-face meetings, not to infer the interactions themselves.

The study that directly used anonymized CDRs to estimate face-to-face interactions was conducted by Calabrese et al. [36]. In this study, the authors utilized anonymized CDRs from over one million customers to explore the relationship between individuals' calls and their physical location. They found that the frequent and recent calls are indications of coordination before face-to-face meetings. However, in this study, the authors only ignored the co-location in living or working areas, while it is known that meetings among people can happen anywhere. In addition, the authors only proved that the call duration between two co-located users is less than the average call duration and did not investigate the duration of calls during the meeting between the co-located users and others [31].

Additionally, anonymized CDRs are used to infer the spatial distribution of displaced populations by analyzing the changes in home cell towers for each anonymized mobile phone subscriber before and after a disaster. The results demonstrate a promising correlation coefficient (70%) between the numbers of arrivals in each neighborhood as determined by CDRs. The authors focused on urban environments, whereas in the present study, it is proposed to analyze data from rural areas, which can provide different results in the local context.

For the present study, it is proposed to (1) investigate the difference between the usual call behavior of users and call behavior right after co-location, (2) estimate and analyze users' home and work locations, (3) explore how co-location events between people working can help infer face-to-face meetings, and (4) demonstrate how the results from this study can assist in supporting contact tracing efforts to curb the spread of an infectious disease.

3. Materials and Methods

3.1. Study Area

Mozambique is located on the southeastern coast of Africa and has a surface area of 801,590 square kilometers. Mozambique borders Zimbabwe, Tanzania, Zambia, Malawi, Eswatini, and South Africa. Its 2500 km long coastline along the Indian Ocean, facing the east of Madagascar. The country is divided into ten provinces and a provincial capital city. It is estimated to have a population of 33 million, of which 62% reside within rural regions. The country is composed of 154 districts, 444 administrative posts, and 1164 localities [37]. However, the data available and used in this study consist of 411 administrative posts. Figure 1 presents the study area.

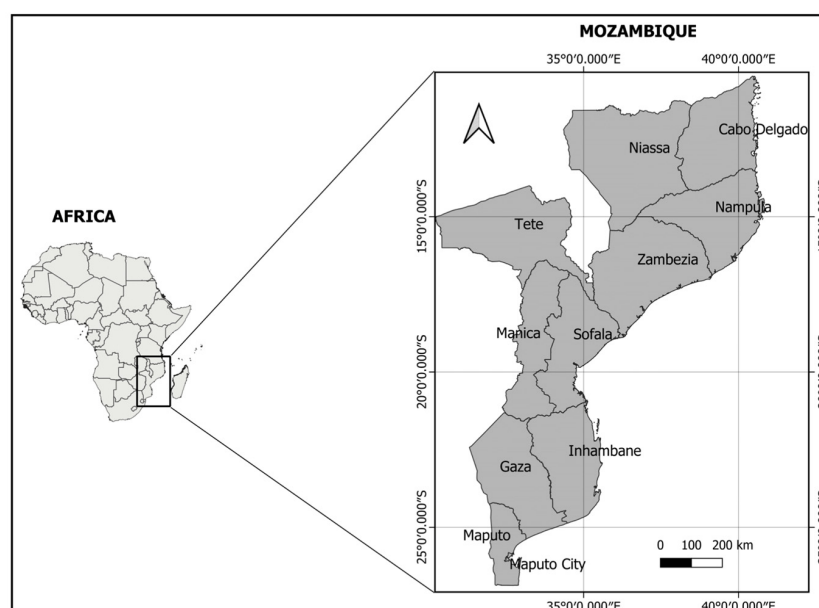


Figure 1. Study area (source: created by the authors).

3.2. Dataset

In this study, anonymized CDRs were used. CDRs are records made by mobile phone operators whenever a subscriber engages in a mobile phone activity such as a call, Short Message Service (SMS), or data use [33]. This study used only call data from one of the biggest mobile operators. Due to contractual agreements and data protection regulations, this dataset cannot be publicly shared or made available to external parties. However, any interested research entity can request access to call traffic data from the Communications Regulator of Mozambique (INCM), provided that it submits a research proposal. Each CDR entry consists of the International Mobile Equipment Identifier (IMEI) of the Caller and Callee, International Mobile Subscriber Identifier (IMSI) of the Caller and Callee, start time of activity, duration of the call, Location Area Code (LAC) of the Caller and Callee, cell-tower Identifier (CELL-ID) of Caller and Callee, activity type (Call, SMS, and Internet), and connection type (2G, 3G, or 4G) [31]. To ensure the protection of subscriber data, the operator anonymized the CDRs by applying an encryption algorithm based on a secret key, making decryption more difficult. Table 1 presents an example of a row of anonymized mobile CDRs.

Table 1. Example of anonymized mobile CDRs [31].

Attribute	Value
IMEI–CALLER (anonymized)	3bd78673f3084c4bcc564580c028b83367c5f8489dc5ba63a68afb3383f0d2ce
IMEI–CALLEE (anonymized)	450325fa618ca1d370b69f54c5c1f485b05a2599273c94187ca3682582653211
IMSI–CALLER (anonymized)	1d68927da0e0681a887a0ac721af3c911aa71d965f1706d83c61bcc82d59e856
IMSI–CALLEE (anonymized)	674f6eb6d350e69326af7ca86335e7d55b7fe89731d2a135cb53eb2ba1b743
START TIME OF ACTIVITY	2019-03-09 21:39:42
DURATION (seconds)	320
LAC-CALLER	5800
CELL-ID-CALLER	694715
LAC-CALLEE	620
CELL-ID-CALLEE	705688
ACTIVITY-TYPE	CALL
CONNECTION-TYPE	3G

As can be seen from Table 1, anonymized CDRs do not include any personally identifiable information, e.g., name, phone number, gender, or age, to preserve user privacy. In addition to the CDRs file, another file contains spatial information (latitude and longitude) and the Cell-ID, which is used to join CDRs with cell tower data.

In this research, 24 days (6 March 2019–29 March 2019) worth of CDRs were used. These data were originally provided by the INCM for the purposes of study [31] and have been reused in the present study, thereby avoiding the need for a formal request for a new dataset from the communications regulator. Figure 2 shows the distribution of daily activities and the number of active subscribers.

From Figure 2, it is clear that there is a smooth proportion between the daily number of activities and the number of subscribers. This is a good indicator of mobile phone use patterns.

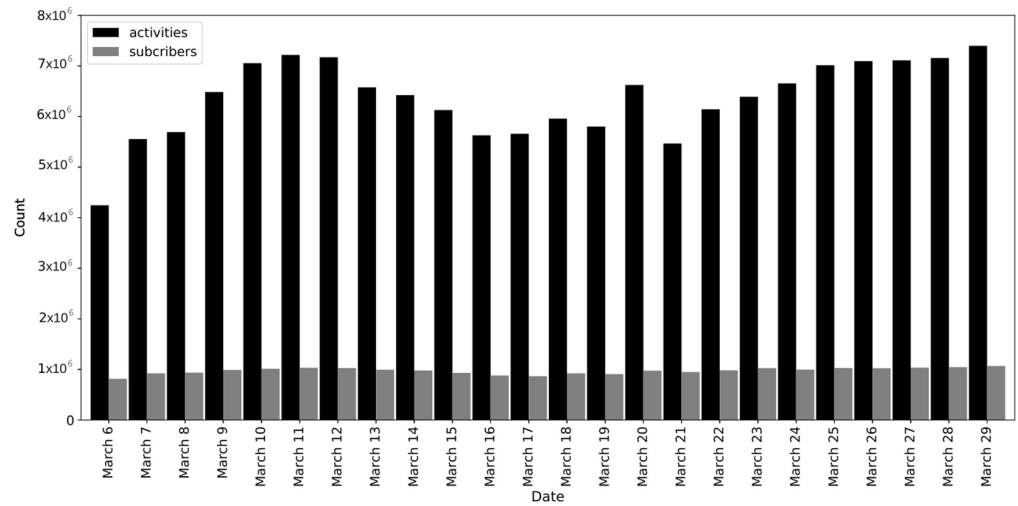


Figure 2. Number of activities and active subscribers per day in the study area (source: created by the authors).

3.3. Methodology

Face-to-face interaction estimation using mobile phone anonymized CDRs consists of multiple steps, some of which were carried out using established techniques from prior research [31]. The proposed method consists of six main steps, namely (1) Voronoi tessellation of the study area, (2) estimation of co-locations, (3) estimation of mobile phone users’ home location and work location, (4) testing the hypotheses, (5) quantify the per cent of agreement of the proposed hypotheses, and (6) assignment of the results at Post-Administrative Units. Figure 3 summarizes the method used in this research.

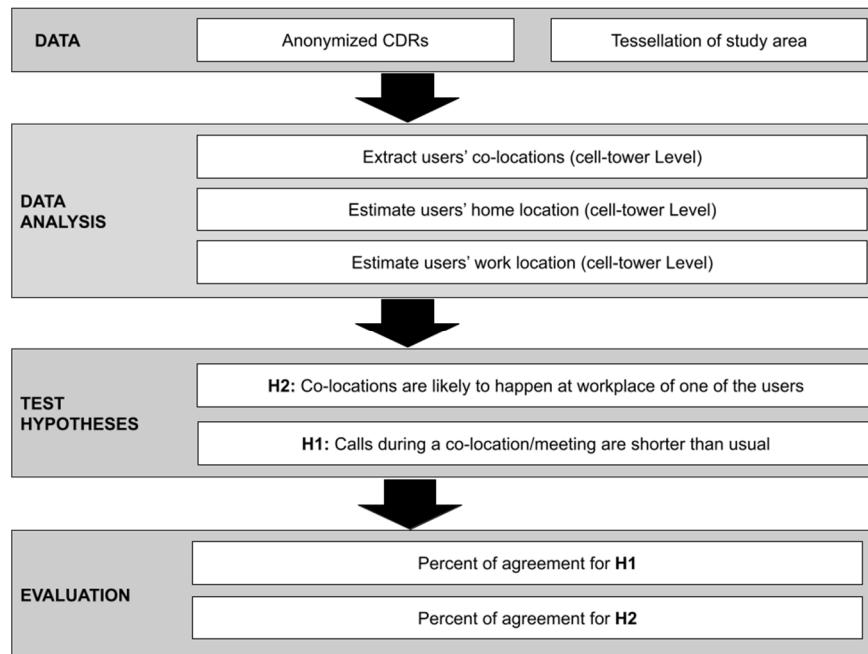


Figure 3. Methodology flow chart (source: created by the authors).

3.3.1. Tessellation of the Study Area

The study area comprises 4918 cell towers, spatially distributed based on population density, meaning that densely populated areas have a higher concentration of cell towers. Hence, each cell tower’s coverage area decreased in highly populated areas. Using cell

tower locations as centroids, Voronoi polygons of the study area were constructed using the principles presented by Boots et al. [38], as shown in Figure 4.

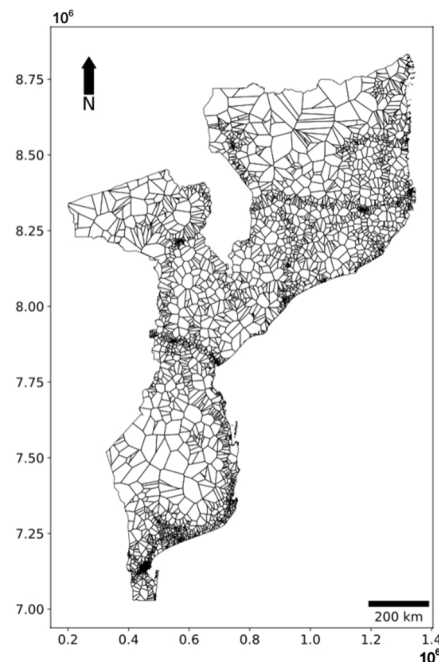


Figure 4. Tessellation of the study area (source: created by the authors).

3.3.2. Co-Locations Extraction

This study defines a co-location event as a call between two subscribers connected to the same cell tower. A co-location event can be seen as a call to coordinate a meeting in a nearby area, referred to as a “coordination knot”, as hypothesized by [39,40]. Of course, two mobile phone users can call each other while sharing the same cell tower without meeting each other. Therefore, an in-depth spatiotemporal investigation of co-location events is needed to find a reasonable proxy for face-to-face meetings.

3.3.3. Home and Work Location Estimation

The underlying principle for estimating the user’s home location (at the cell tower level) is that people typically stay at home during the night. Based on this principle, a cell tower is designated as the user’s home location if it is the most frequently used during nighttime (07:00 p.m.–07:00 a.m.) for the entire analysis period, as proposed by [31]. Similarly, the work location is the most used cell tower during the daytime (07:00 a.m.–07:00 p.m.).

3.3.4. Testing the Hypotheses

To define co-locations as the reasonable proxies of face-to-face meetings, two spatiotemporal-related hypotheses were defined, as follows:

H1: *Co-locations are likely to happen at the workplace of one of the users.*

To this hypothesis, first, co-locations were extracted from the anonymized CDR data. Then, assuming that people usually meet during the daytime, and during this period, people are at work, the co-location place was investigated. The investigation consisted of comparing the co-location place to the work location of the pair. If the co-location place is the same as the work location of one of the users, then the hypothesis is valid. Figure 5 presents the graphical representation of the process to test H1.

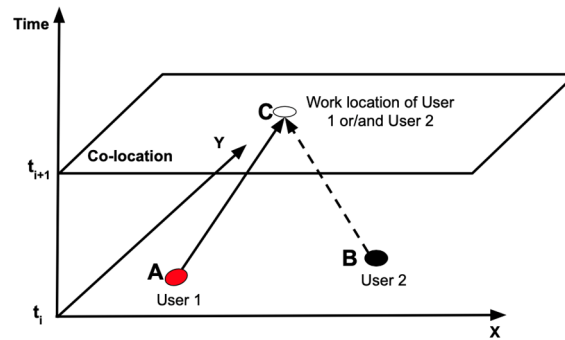


Figure 5. Testing H1 (source: created by the authors).

H2: *Calls while attending a meeting are shorter than usual.*

Co-location events were first extracted from the anonymized CDR data to test this hypothesis. Then, the call duration was investigated right after co-location. Figure 6 shows the graphical representation of the process to test H2.

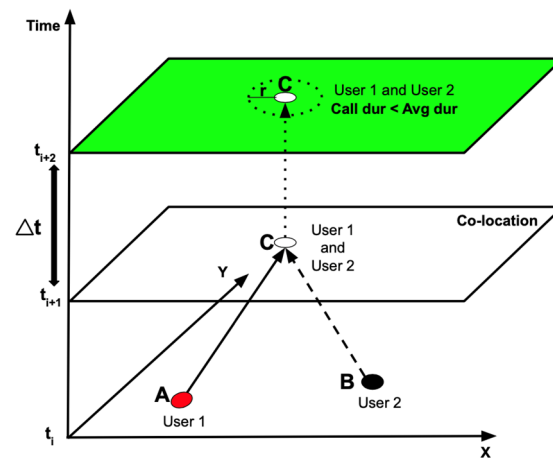


Figure 6. Testing H2 (source: created by the authors).

In Figure 6, at time t_i , User 1 and User 2 are at locations A and B, respectively. At time t_{i+1} , the users co-located at location C and spent Δt , and one of the users received a call while at the exact location. The duration of this call was compared to the average call duration of that user for the entire study period. r is a scalar used to reduce the effect of false displacement, which happens when a user makes or receives a call while their nearest cell tower is fully loaded. The mobile phone operator assigns their mobile activity to another cell tower.

3.3.5. Evaluation

The daily percentage of agreement for each hypothesis was computed using Equation (1) to evaluate the results:

$$P_i = \frac{T_H - F_H}{T_H + F_H} \tag{1}$$

where

- T_H represents the results that agree with the hypothesis;
- F_H represents the results that disagree with the hypothesis and
- i represents the days of the month (March) considered in this study.

The T_H and F_H results were based on the CDR data analysis, which confirms or not what is proposed in each hypothesis.

H1: *Co-locations are likely to happen at the workplace of one of the users.*

To evaluate this hypothesis, co-location place was compared to the work location of the pair. If the co-location place was the same as the work location of one of the users, then the co-location was labeled with true, otherwise, with false. Then, the labeling results for all the co-location activities for a specific day were then aggregated and Equation (1) was applied to compute the daily percentage of agreement for H1.

H2: *Calls while attending a meeting are shorter than usual.*

To evaluate this hypothesis, the call duration of a given user during co-location were compared to their average call duration. If the call duration during the co-location was less than average call duration, then the activity was labeled as true, otherwise, the activity was labeled as false. The labeling results for all the users, for specific day were then aggregated and Equation (1) was applied to compute the daily percentage of agreement for H2.

3.3.6. Assigning the Cell-Tower-Based Results to Post-Administrative Units

All the estimations were at the cell tower level (Thiessen polygons). However, decision-makers are usually interested in seeing the results in administrative units. Post-Administrative Units (PAUs) present the analysis results in this study. There is no match between the Thiessen polygons and PAUs. Therefore, the spatial interpolation proposed by Flowerdew et al. [41] was used for cell-tower-based results to PAUs, as shown in Equation (2):

$$u_i = \frac{a_i * S}{A_t} \quad (2)$$

where

- u_i represents the number of subscribers assigned to a PAU using spatial interpolation method;
- i represents the number of PAUs that share a given cell tower;
- a_i is the intersection area between the coverage area of a cell tower (Voronoi polygon) and PAU;
- A_t represents the total area of the cell tower under consideration and
- S is the total number of subscribers found in the shared cell tower.

4. Results and Discussion

First, the users' home and work locations at the cell tower level were estimated and assigned to PAUs. Figure 7 shows the distribution of users working and living, and the percentage increase in users in each PAU during daytime.

Figure 7a,b clearly shows that the number of subscribers living in each PAU does not differ much from the number of people working in the same area.

This might be related to two situations: (a) people work in their residential areas; (b) there is a replacement of people in each PAU, i.e., the number of outflow for working is equal to the number of inflow for working. Figure 7c shows the detailed increase in people during daytime in each PAU. This result suggests a slight increase in users in each PAU (up to 30%).

After extracting co-locations, an evaluation was conducted to extract only those co-locations at the work location. Then, the cell tower distribution of the co-locations at the cell tower level was assigned to PAUs, as shown in Figures 8 and 9.

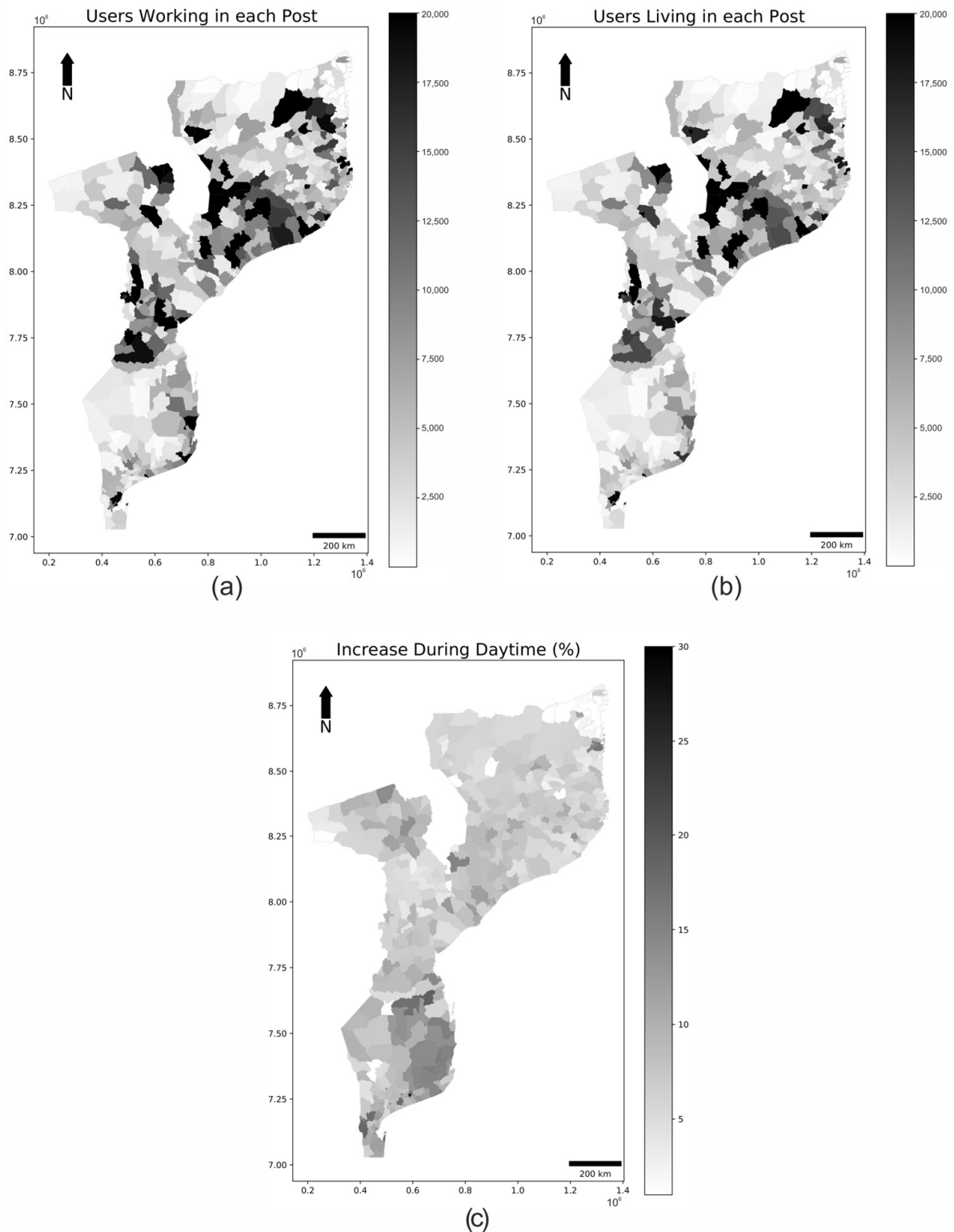


Figure 7. Total number of users (a) working and (b) living, and (c) the percent of increase during daytime in each PAU (source: created by the authors).

Figure 8 shows that the co-location events have a similar daily distribution for each cell tower from 6 March to 15 March (the day that cyclone Idai struck Beira city in central Mozambique). From 16 March to 17 March, all the districts along the corridor of the cyclone from east to west were affected and did not register mobile phone activities. Therefore, there was a significant reduction in co-locations along this corridor.



Figure 8. The co-location was distributed at the work location from 6 March to 17 March 2019 (source: created by the authors).

Figure 9 suggests that the affected regions' cell towers failed to register the activities for more than 4 days. However, from 22 March, there was a clear tendency for recovery of the cell tower; hence, the number of co-locations increased significantly up to the last day (29 March), when over 90% of the cell towers were operational.

The daily percentage of agreement was computed to evaluate the first hypothesis. Figure 10 shows the daily percentage of agreement.

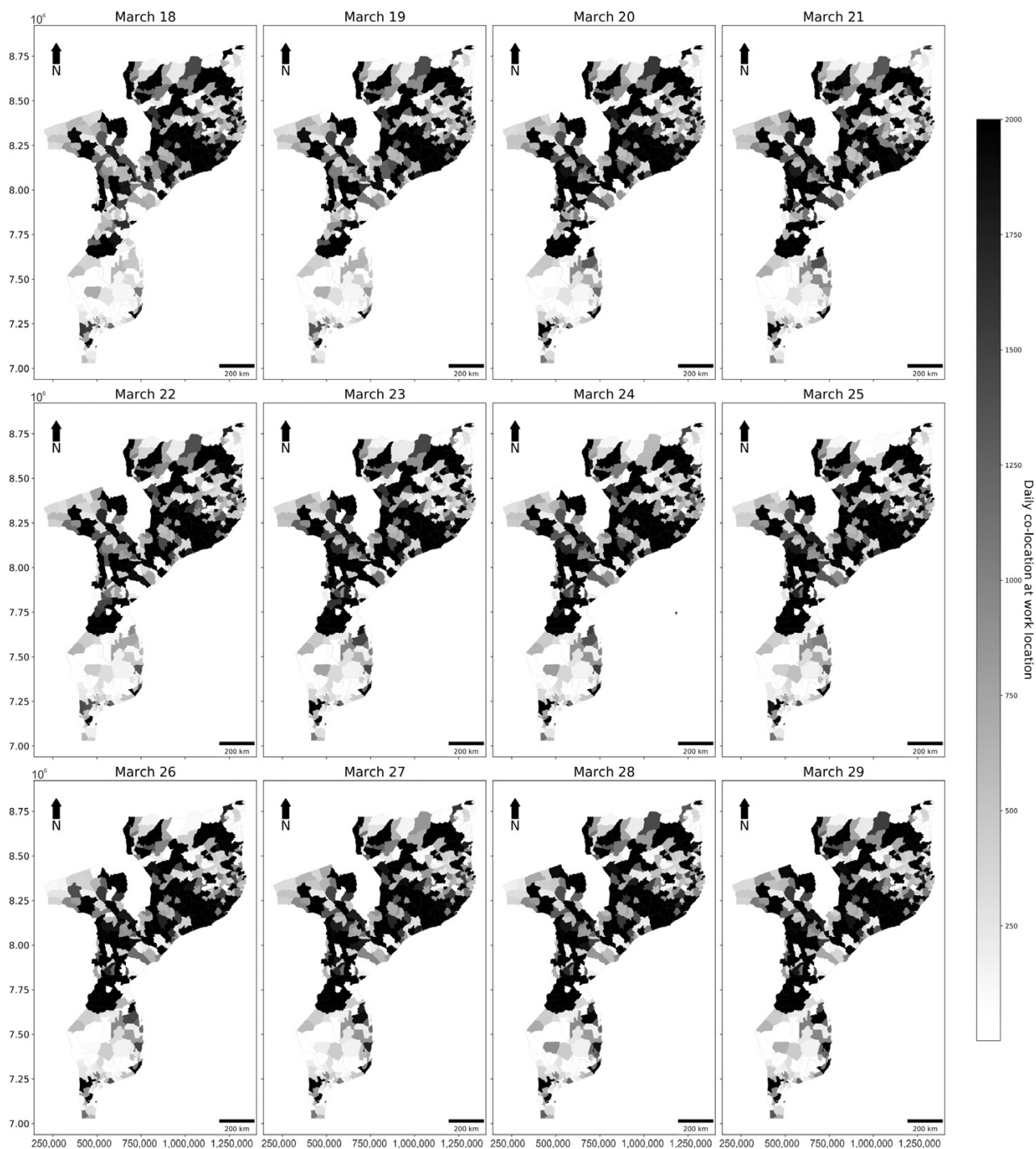


Figure 9. Co-location will be distributed at the work location from 18 March to 29 March 2019 (Source: Created by the Authors).

Figure 10 shows that the daily percentage of agreement was greater than 80% for the whole study period. This result suggests that the first hypothesis states that meetings between users usually happen at their workplace has almost perfect agreement.

After evaluating the first hypothesis, the second hypothesis was tested, which states that calls during meetings/co-locations are shorter than usual. The test consisted of analyzing the duration of calls during the co-location. Since some users have very sparse CDRs, we focused on those activities that occurred most within an hour after co-location. We assumed that users meet within a few minutes after the co-location call, and the meeting duration is at least an hour. Figure 11 presents the daily average call duration (which

represents the typical call duration for each subscriber) and the call duration distribution right after co-location.

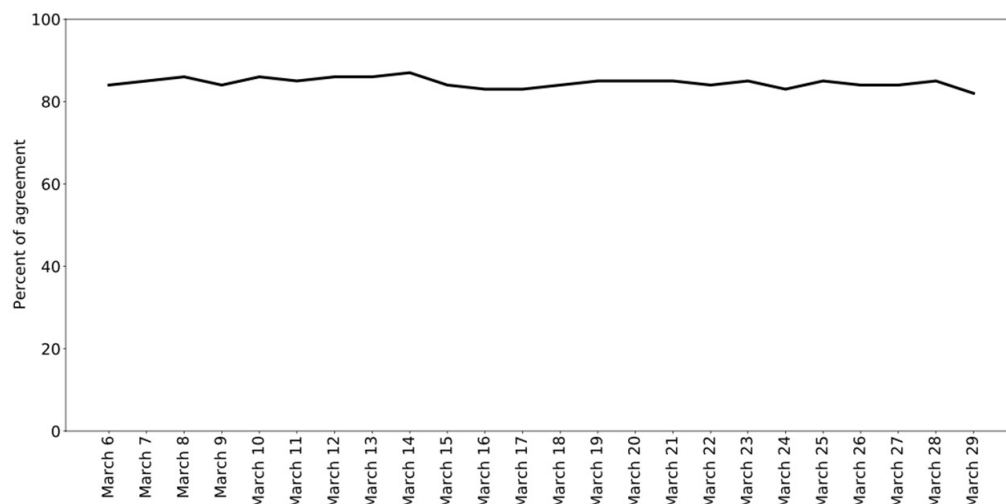


Figure 10. Percent of agreement for H1 (source: created by the authors).

From Figure 11, it is clear that overall, there are similar distributions between the average duration of calls and the duration of calls right after co-location. It is also clear that the distribution of the average duration of calls is greater than the distribution of the duration of calls right after co-location for the whole study period for calls with a duration less than or equal to 300 s (5 min). However, there is an opposite trend for calls with a duration greater than 5 min, i.e., the average call duration distribution is lesser than the call duration distribution right after co-location.

A similar process to the one carried out for the first hypothesis was executed to evaluate the second hypothesis, i.e., the percentage of agreement was computed by comparing the co-locations that satisfied the hypothesis and those that failed. Figure 12 presents the daily percentage of agreement for hypothesis 2.

Figure 12 clearly shows that similar to the first hypothesis, the second hypothesis has a daily percentage of agreement greater than 80% for the whole study period. This suggests that this hypothesis has almost perfect agreement.

Assuming that the co-locations represent the proxy of face-to-face meetings if a given user is diagnosed with an infection (e.g., COVID-19), their contacts for the last couple of days can be represented as in Figure 13.

Figure 13 illustrates how the co-location data can be applied to contact tracing during an infectious disease outbreak. The figure uses a timeline that shows the daily number of meetings a hypothetical infected user has with other users leading up to the day they were confirmed positive. Day “T” represents the day before the user was confirmed as positive, and “n” represents the virus incubation period, which is 14 days for COVID-19 [42,43]. The figure shows that the infected user had varying numbers of daily meetings with different users throughout the incubation period. User 13 and User 14, who had contact with the infected user on Day “T–n”, are highlighted as potential sources of infection. It is noted that User 13 may still be infected, while User 14 may have already recovered. The figure demonstrates that identifying and isolating all individuals who had contact with the infected user during the incubation period can be an effective strategy to contain the spread of the virus [44]. It is important to note that this is a hypothetical example using a small sample of users to demonstrate the principle. In a real-world scenario, contact tracing would involve a significantly larger network of individuals.

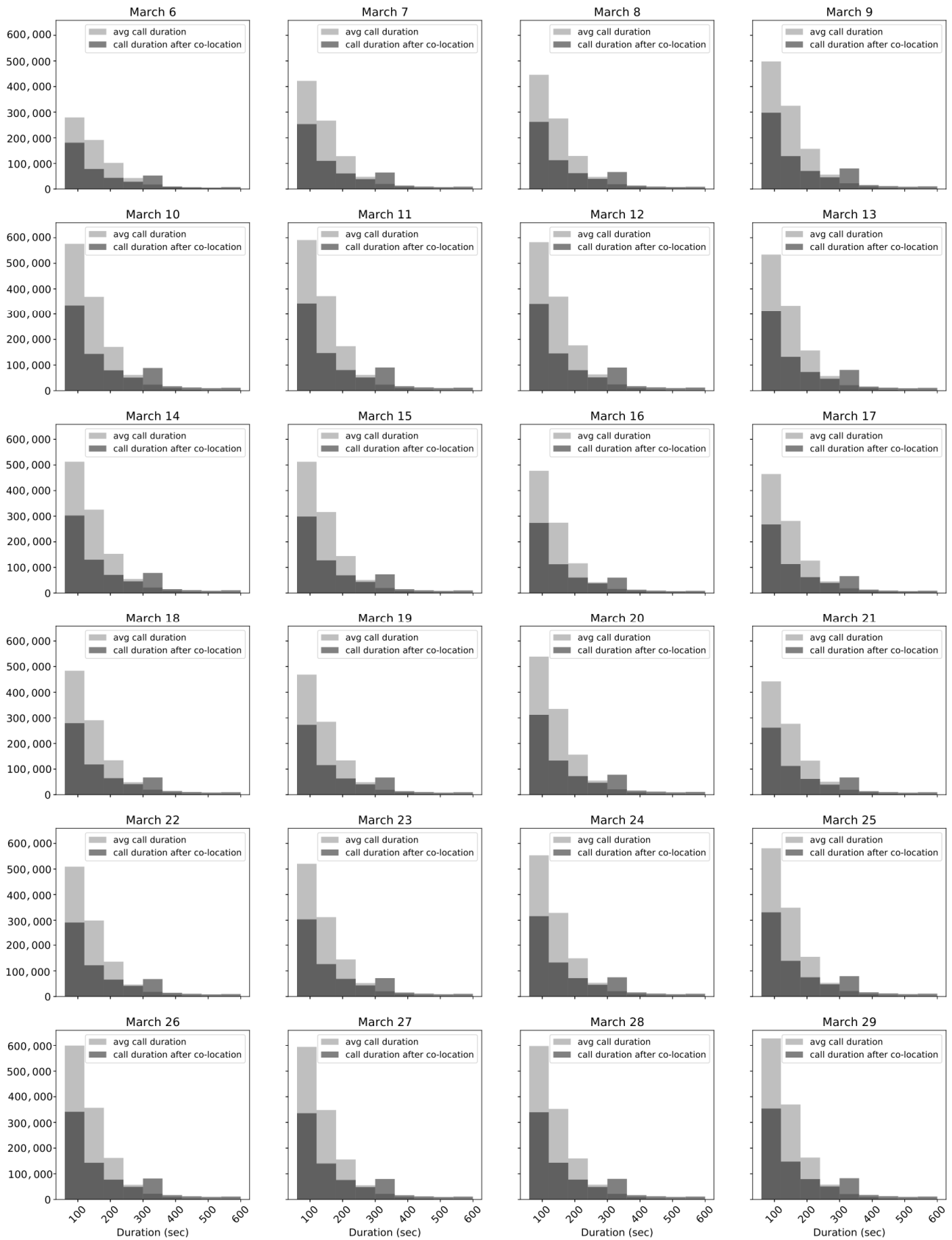


Figure 11. Distribution of the average duration of calls and the duration of the call right after co-location (source: created by the authors).

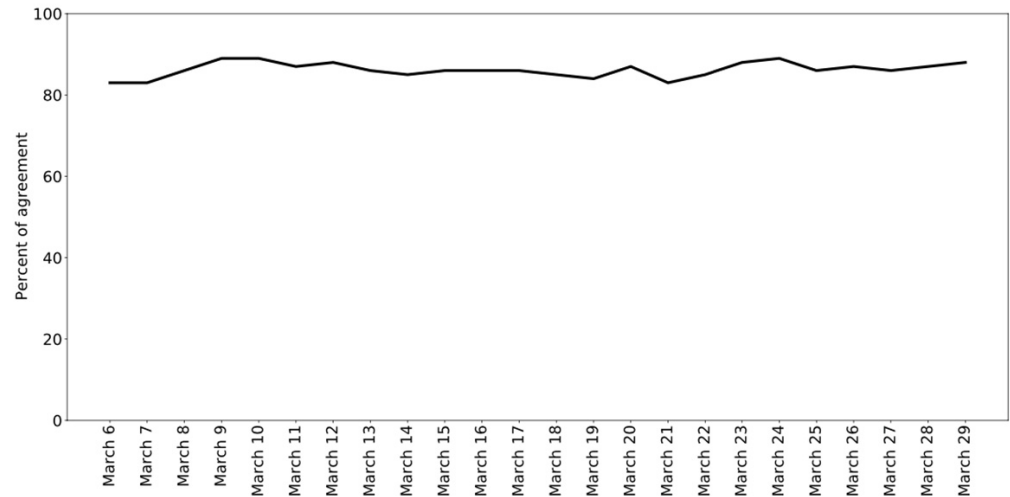


Figure 12. Percent of agreement for H2 (source: created by the authors).

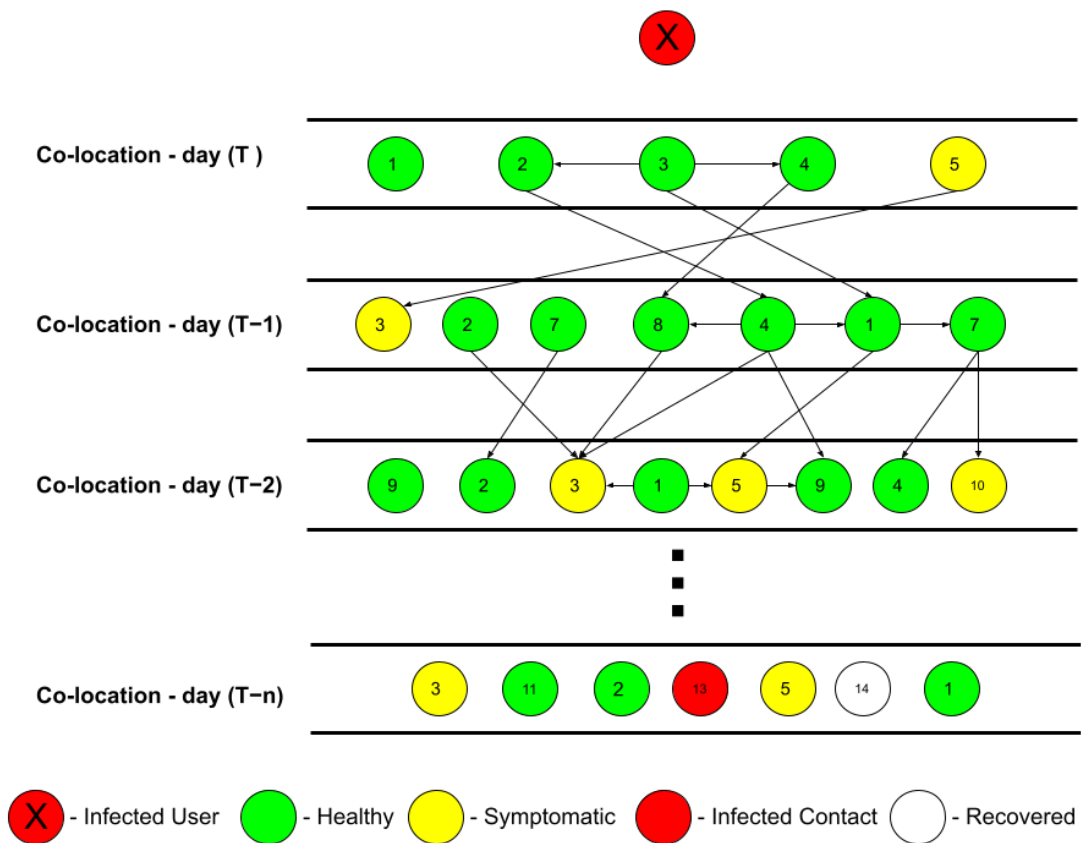


Figure 13. Contact tracing for an infected user (source: created by the authors).

To measure the dynamics of face-to-face meetings in the study area, the co-location events were aggregated by subscription, and the results are shown in Figure 14.

Figure 14 shows the daily distribution of the total co-location per user. These results suggest that over 50% of users co-locate once, and around 40% collocate more than once and less than six times daily. Assuming that co-location is a proxy of face-to-face meetings, there is a considerable number of contacts between people, which could accelerate spread in the case of a surge of an infectious disease. To gain a deeper understanding of the distribution of co-location events, the daily statistics are computed and presented in Table 2.

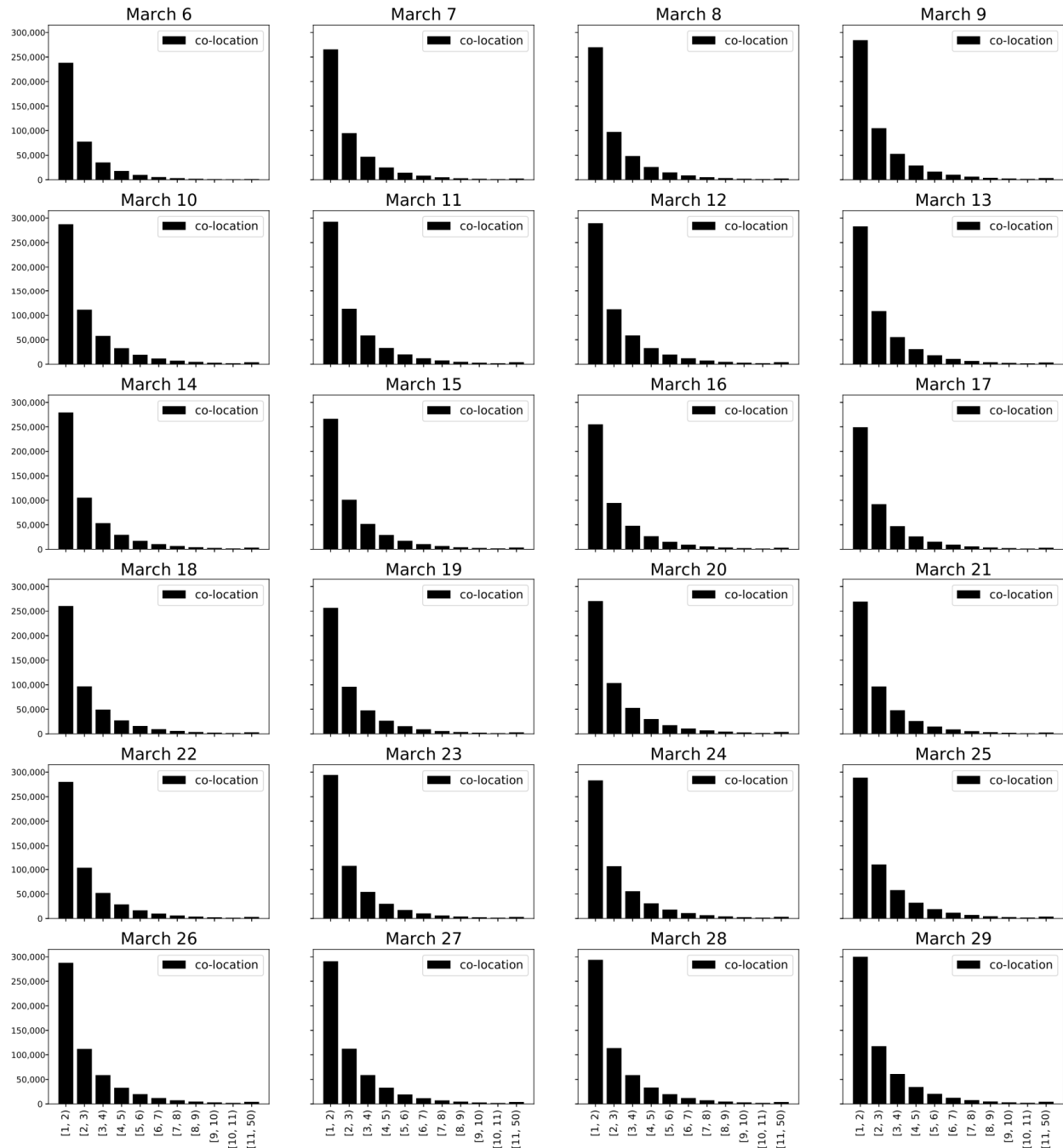


Figure 14. Distribution of co-location events (source: created by the authors).

Table 2 suggests that 6 March and 29 March registered the most minor and highest co-location events during the study period, respectively. The number of co-location events increased initially and stabilized on 10 March. From 14 March, the number of co-location events reduced and increased until 23 March, stabilizing until the end of the study period. This last reduction in co-location events can be associated with cyclone Idai, which hit Beira city (in Central Mozambique) on 14 March 2019, causing massive flooding, leaving entire communities submerged under 10 m of water, and damaged infrastructure and roads, with about 3 million people were affected [31]. Along its path, cyclone Idai destroyed many infrastructures, including electricity and telecommunications towers. Therefore, four-to-five days after the cyclone hit Beira city (in the central part of Mozambique), many cell towers were out of order and failed to register mobile phone activities.

Table 2. Daily statistics of co-location events (source: created by the authors).

Day	Count	Minimum	Maximum	Mean	Standard Deviation
6 March	392,965	1	43	1.850	1.542
7 March	468,776	1	30	2.003	1.720
8 March	479,930	1	33	2.020	1.743
9 March	515,802	1	41	2.081	1.837
10 March	541,751	1	35	2.163	1.915
11 March	551,659	1	37	2.168	1.916
12 March	547,261	1	39	2.174	1.930
13 March	525,967	1	31	2.120	1.854
14 March	513,656	1	36	2.104	1.845
15 March	494,928	1	31	2.136	1.891
16 March	465,234	1	47	2.093	1.852
17 March	455,831	1	33	2.104	1.861
18 March	476,565	1	36	2.099	1.846
19 March	468,592	1	36	2.091	1.838
20 March	505,751	1	34	2.157	1.926
21 March	479,126	1	30	2.031	1.764
22 March	508,910	1	33	2.079	1.821
23 March	533,269	1	34	2.070	1.814
24 March	526,407	1	31	2.125	1.859
25 March	542,437	1	38	2.151	1.896
26 March	545,085	1	34	2.177	1.931
27 March	546,915	1	34	2.151	1.893
28 March	552,719	1	33	2.151	1.897
29 March	568,771	1	36	2.173	1.921

In Table 2, the maximum number of user co-location events varies from 30 to 47. Suppose we assume that co-location is a proxy of face-to-face meetings. In that case, users with this number of co-locations with different pairs can be seen, for example, an IT employee in a building who needs to communicate with their colleagues to support them in solving some issues. In real life, except in meetings, it is not common for someone to meet more than 30 people in one day. Usually, a person attends, on average, two meetings, as seen in Table 2 and proved by Yu [20].

5. Conclusions and Future Work

This article proposed using anonymized CDRs to estimate co-locations as a proxy of face-to-face meetings. This approach involves extracting the co-location events as a call between two users connected to the same cell tower. Two hypotheses define the co-location as a proxy of face-to-face meetings. The first hypothesis (H1) states that user meetings usually happen at their workplaces, and the second hypothesis (H2) states that calls during meetings are shorter than usual. To test H1, the work location for each user was computed as the most used cell tower during the study period. Then, co-locations at workplaces and other places for each user were computed. The results were then aggregated at the cell tower level, and the daily percentage of agreement was calculated, which was over 80%, suggesting that there is an almost perfect agreement. H2 was tested by computing the daily average call duration for each user and comparing it with the duration of a call right after the co-location. The results were then aggregated at the cell tower level.

Similarly, as in H1, the daily percentage of agreement was calculated, which was over 80%, suggesting that there is an almost perfect agreement. These results proved that co-location can be seen as a coordination call to meet nearby and can hence be used as a proxy for face-to-face meetings. While this study concluded that co-locations can be

used as a proxy of face-to-face meetings only by using anonymized CDRs, it would be interesting to conduct a survey on the daily meetings for a sample of users and compare it with the results of the proposed approach. In addition to this, this study only used call data. Still, it would be interesting to see the contribution of SMS data and data usage (connection to the internet) since people often communicate through SMS or by using apps like WhatsApp, Facebook, Instagram, Twitter, etc. Additionally, it would be interesting to conduct an analysis of the data during weekends and holidays. This would provide further insight into activity patterns. By comparing data from workdays to weekends and holidays, we could identify distinct behavioral patterns, such as increased movement in recreational areas during leisure times and possibly reduced activity in work-related zones.

Integrating mobile phone data and other digital tools into public health strategies provides a powerful means for health authorities to respond swiftly and effectively to respiratory pandemics, enhancing their ability to protect public health and save lives.

A key contribution of the present study is the validation of two hypotheses to use co-location events as proxies for face-to-face interactions. The first hypothesis assumes that meetings between individuals often occur at their workplace. The study tested this by identifying work locations based on the most frequently used cell towers during daytime hours. The results showed over 80% agreement, indicating that co-location at workplace cell towers could reliably represent face-to-face interactions. The second hypothesis posits that calls made during these co-location events tend to be shorter than usual. The analysis of call durations confirmed this with a similarly high-agreement level, reinforcing the validity of using co-location events as indicators of face-to-face meetings.

Another significant contribution is the method developed to estimate users' home and work locations from CDR data. The study assumes that the cell towers most frequently used during night hours indicate home locations, while those used during daytime hours reflect work locations. This spatial understanding enhances the ability to contextualize co-location events and differentiate between meetings occurring at work and other locations.

The study also extended its analysis to rural areas, an important expansion since previous research has predominantly focused on urban settings. By including rural regions, the study demonstrated the broader applicability of its methods and provided insights into the unique challenges and dynamics in these areas. For example, the findings suggested that the distribution of cell towers and the mobility patterns of individuals in rural settings might influence the nature of co-location events.

Additionally, the research explored the resilience of its methods in the context of a real-world event Cyclone Idai in Mozambique. The cyclone disrupted telecommunications infrastructure, allowing the researchers to evaluate how such disruptions affect mobile phone activity and co-location events. This aspect of the study provided an opportunity to assess the robustness and adaptability of the proposed approach under extreme conditions.

The study's contributions also have significant implications for public health. The ability to use anonymized CDRs for contact tracing supports more efficient identification of potential transmission chains during infectious disease outbreaks.

Finally, the study offers recommendations for future research. It suggests integrating additional types of mobile data, such as SMS and internet usage, to refine the understanding of communication patterns further. Moreover, it calls for surveys to validate the inferred interactions with actual user behaviors, strengthening the empirical foundation of its findings. These forward-looking suggestions ensure that the research paves the way for continued innovation in the field.

Like any other type of study, no matter how careful one is, the possibility of threats to its validity cannot be ruled out, and it is the researcher's responsibility to identify them and define control actions to mitigate them. The use of mobile phone data raises significant

privacy issues. Ensuring that data are anonymized and secure is crucial, but there is always a risk of re-identification, especially when data are combined from multiple sources. The absence of a method that allows for the hypotheses to be confirmed with users (in person) of the numbers under study constitutes a threat to the study's validity.

Author Contributions: Conceptualisation, S.P.C., G.G., O.F.C. and A.M.J.; methodology, S.P.C. and G.G.; investigation: S.P.C., G.G., O.F.C. and A.M.J.; writing—original draft, S.P.C., G.G., O.F.C. and A.M.J.; writing—review and editing, N.S. and F.B.; supervision, S.P.C.; funding acquisition, F.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to institutional indications.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO. Who report on infectious diseases. *Wkly. Epidemiol. Rec. = Relev. Épidémiologique Hebd.* **1999**, *74*, 279.
2. WHO. WHO Coronavirus (COVID-19) Dashboard. 2023. Available online: <https://covid19.who.int/> (accessed on 10 June 2023).
3. Aleta, A.; Martin-Corral, D.; Pastore y Piontti, A.; Ajelli, M.; Litvinova, M.; Chinazzi, M.; Dean, N.E.; Halloran, M.E.; Longini, I.M., Jr.; Merler, S.; et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **2020**, *4*, 964–971. [[CrossRef](#)] [[PubMed](#)]
4. Girum, T.; Lentiro, K.; Geremew, M.; Migora, B.; Shewamare, S. Global strategies and effectiveness for COVID-19 prevention through contact tracing, screening, quarantine, and isolation: A systematic review. *Trop. Med. Health* **2020**, *48*, 91. [[CrossRef](#)]
5. Braithwaite, I.; Callender, T.; Bullock, M.; Aldridge, R.W. Automated and partly automated contact tracing: A systematic review to inform the control of COVID-19. *Lancet Digit. Health* **2020**, *2*, e607–e621. [[CrossRef](#)]
6. Quilty, B.J.; Clifford, S.; Hellewell, J.; Russell, T.W.; Kucharski, A.J.; Flasche, S.; Edmunds, W.J.; Atkins, K.E.; Foss, A.M.; Waterlow, N.R.; et al. Quarantine and testing strategies in contact tracing for SARS-CoV-2: A modelling study. *Lancet Public Health* **2021**, *6*, e175–e183. [[CrossRef](#)]
7. Juneau, C.-E.; Briand, A.-S.; Collazzo, P.; Siebert, U.; Pueyo, T. Effective contact tracing for COVID-19: A systematic review. *Glob. Epidemiol.* **2023**, *5*, 100103. [[CrossRef](#)]
8. Wren, H. 6 Contact Tracing Best Practices—And How Technology Can Help. 2023. Available online: <https://www.zendesk.com/blog/contact-tracing/> (accessed on 13 June 2023).
9. European Centre for Disease Prevention and Control. Contact Tracing in the European Union: Public Health Management of Persons, Including Healthcare Workers, Who Have Had Contact with COVID-19 Cases—Fourth Update. 2021. Available online: <https://www.ecdc.europa.eu/en/covid-19-contact-tracing-public-health-management> (accessed on 13 June 2023).
10. Akinbi, A.; Forshaw, M.; Blinkhorn, V. Contact tracing apps for the COVID-19 20 pandemic: A systematic literature review of challenges and future directions for neo-liberal societies. *Health Inf. Sci. Syst.* **2021**, *9*, 18. [[CrossRef](#)]
11. EDPS. Contact Tracing with Mobile Applications. 2020. Available online: https://edps.europa.eu/sites/edp/files/publication/20-05-08_techdispatch-tracing_en.pdf (accessed on 13 June 2023).
12. Cheng, H.-Y.; Jian, S.-W.; Liu, D.-P.; Ng, T.-C.; Huang, W.-T.; Lin, H.-H. Contact tracing assessment of COVID-19 transmission dynamics in taiwan and risk at different exposure periods before and after symptom onset. *JAMA Intern. Med.* **2020**, *180*, 1156–1163. [[CrossRef](#)]
13. Wang, C.J.; Ng, C.Y.; Brook, R.H. Response to COVID-19 in taiwan: Big data analytics, new technology, and proactive testing. *JAMA* **2020**, *323*, 1341–1342. [[CrossRef](#)]
14. Sam Cannicott. The Ethics of Contact Tracing Apps: International Perspectives. 2020. Available online: <https://cdei.blog.gov.uk/2020/05/12/the-ethics-of-contact-tracing-apps-international-perspectives/> (accessed on 13 June 2023).
15. Maccari, L.; Cagno, V. Do we need a contact tracing app? *Comput. Commun.* **2021**, *166*, 9–18. [[CrossRef](#)]
16. Dar, A.B.; Lone, A.H.; Zahoor, S.; Khan, A.A.; Naaz, R. Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions. *Comput. Sci. Rev.* **2020**, *38*, 100307. [[CrossRef](#)] [[PubMed](#)]
17. Antonioli, D.; Tippenhauer, N.O.; Rasmussen, K. Bias: Bluetooth impersonation attacks. In Proceedings of the 2020 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 18–21 May 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 549–562.

18. Chidambaram, S.; Erridge, S.; Kinross, J.; Purkayastha, S. Observational study of uk mobile health apps for COVID-19. *Lancet Digit. Health* **2020**, *2*, e388–e390. [[CrossRef](#)] [[PubMed](#)]
19. Huang, Z.; Guo, H.; Lim, H.Y.; Chow, A. Awareness, acceptance, and adoption of the national digital contact tracing tool post COVID-19 lockdown among visitors to a public hospital in singapore. *Clin. Microbiol. Infect.* **2021**, *27*, 1046–1048. [[CrossRef](#)]
20. Yu, S.H. PrivyTRAC—Privacy and Security Preserving Contact Tracing System. In *Security and Privacy in Communication Networks, Proceedings of the 16th EAI International Conference, SecureComm 2020, Washington, DC, USA, 21–23 October 2020*; Proceedings, Part I 16; Springer International Publishing: Cham, Switzerland, 2020; pp. 515–525.
21. Reades, J.; Calabrese, F.; Sevtsuk, A.; Ratti, C. Cellular census: Explorations in urban data collection. *IEEE Pervasive Comput.* **2007**, *6*, 30–38. [[CrossRef](#)]
22. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779–782. [[CrossRef](#)]
23. Wang, P.; González, M.C.; Hidalgo, C.A.; Barabási, A.-L. Understanding the spreading patterns of mobile phone viruses. *Science* **2009**, *324*, 1071–1076. [[CrossRef](#)]
24. Yuan, Y.; Raubal, M. Extracting dynamic urban mobility patterns from mobile phone data. In *Geographic Information Science, Proceedings of the 7th International Conference, GIScience 2012, Columbus, OH, USA, 18–21 September 2012*; Proceedings 7; Springer: Berlin/Heidelberg, Germany, 2012; pp. 354–367.
25. Batran, M.; Mejia, M.G.; Kanasugi, H.; Sekimoto, Y.; Shibasaki, R. Inferencing human spatiotemporal mobility in greater maputo via mobile phone big data mining. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 259. [[CrossRef](#)]
26. Wilson, R.; zu Erbach-Schoenberg, E.; Albert, M.; Power, D.; Tudge, S.; Gonzalez, M.; Guthrie, S.; Chamberlain, H.; Brooks, C.; Hughes, C.; et al. Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake. *PLoS Curr.* **2016**, *8*. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4779046> (accessed on 20 January 2024). [[CrossRef](#)]
27. Gething, P.W.; Tatem, A.J. Can mobile phone data improve emergency response to natural disasters? *PLoS Med.* **2011**, *8*, e1001085. [[CrossRef](#)]
28. Bengtsson, L.; Lu, X.; Thorson, A.; Garfield, R.; Von Schreeb, J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in haiti. *PLoS Med.* **2011**, *8*, e1001083. [[CrossRef](#)]
29. Lu, X.; Bengtsson, L.; Holme, P. Predictability of population displacement after the 2010 haiti earthquake. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11576–11581. [[CrossRef](#)] [[PubMed](#)]
30. Cumbane, S.; Gidofalvi, G.; Shindyapin, A. Mobile phone data as alternative source of information for affected disaster areas detection after a severe cyclone. In *AGU Fall Meeting Abstracts; 2021*; Volume 2021, p. NH14C–04. Available online: <https://ui.adsabs.harvard.edu/abs/2021AGUFMNH14C..04C/abstract> (accessed on 7 July 2023).
31. Cumbane, S.P.; Gidófalvi, G. Spatial distribution of displaced population estimated using mobile phone data to support disaster response activities. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 421. [[CrossRef](#)]
32. 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.* **2022**, *94*, 101777. [[CrossRef](#)]
33. Jin, B.; Park, N. In-person contact begets calling and texting: Interpersonal motives for cell phone use, face-to-face interaction, and loneliness. *Cyberpsychol. Behav. Soc. Netw.* **2010**, *13*, 611–618. [[CrossRef](#)]
34. Przybylski, A.K.; Weinstein, N. Can you connect with me now? how the presence of mobile communication technology influences face-to-face conversation quality. *J. Soc. Pers. Relatsh.* **2013**, *30*, 237–246. [[CrossRef](#)]
35. Hamann, J. The Effects of Cellphones on Face-to-Face Communication. 2018. Available online: <https://medium.com/@jhamann/the-effects-of-cellphones-on-face-to-face-communication-6de063929b0f> (accessed on 14 June 2023).
36. Calabrese, F.; Smoreda, Z.; Blondel, V.D.; Ratti, C. Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PLoS ONE* **2011**, *6*, e20814. [[CrossRef](#)]
37. ISSUU. Quem é Quem: Governo, Vice-Ministros e Secretários de Estado (2020–2014). 2021. Available online: https://issuu.com/bureauinfopub/docs/quem___quem_30_outubro_2020_digital/23 (accessed on 7 July 2023).
38. Boots, B.; Okabe, A.; Sugihara, K. Spatial tessellations. *Geogr. Inf. Syst.* **1999**, *1*, 503–526.
39. Lambiotte, R.; Blondel, V.D.; De Kerchove, C.; Huens, E.; Prieur, C.; Smoreda, Z.; Van Dooren, P. Geographical dispersal of mobile communication networks. *Phys. A Stat. Mech. Its Appl.* **2008**, *387*, 5317–5325. [[CrossRef](#)]
40. Diminescu, D.; Licoppe, C.; Smoreda, Z.; Ziemlicki, C. Tailing untethered mobile users: Studying urban mobilities and communication practices. In *The Reconstruction of Space and Time*; Routledge: London, UK, 2017; pp. 17–37.
41. Flowerdew, R.; Green, M.; Kehris, E. Using areal interpolation methods in geographic information systems. *Pap. Reg. Sci.* **1991**, *70*, 303–315. [[CrossRef](#)]
42. Dhoubib, W.; Maatoug, J.; Ayouni, I.; Zammit, N.; Ghammem, R.; Fredj, S.B.; Ghannem, H. The incubation period during the pandemic of COVID-19: A systematic review and meta-analysis. *Syst. Rev.* **2021**, *10*, 101. [[CrossRef](#)]

43. Baud, D.; Qi, X.; Nielsen-Saines, K.; Musso, D.; Pomar, L.; Favre, G. Real estimates of mortality following COVID-19 infection. *Lancet Infect. Dis.* **2020**, *20*, 773. [[CrossRef](#)] [[PubMed](#)]
44. Wells, C.R.; Townsend, J.P.; Pandey, A.; Moghadas, S.M.; Krieger, G.; Singer, B.; Mc-Donald, R.H.; Fitzpatrick, M.C.; Galvani, A.P. Optimal COVID-19 quarantine and testing strategies. *Nat. Commun.* **2021**, *12*, 356. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.