

Infection Risk Score: Identifying the risk of infection propagation based on human contact

Rachit Agarwal
rachitag@iitk.ac.in
IIT Kanpur, India

Abhik Banerjee
abanerjee@swin.edu.au
Swinburne University, Australia

ABSTRACT

A wide range of approaches have been applied to manage the spread of global pandemic events such as COVID-19, which have met with varying degrees of success. Given the large-scale social and economic impact coupled with the increasing time span of the pandemic, it is important to not only manage the spread of the disease but also put extra efforts on measures that expedite resumption of social and economic life. It is therefore important to identify situations that carry high risk, and act early whenever such situations are identified. While a large number of mobile applications have been developed, they are aimed at obtaining information that can be used for contact tracing, but not at estimating the risk of social situations. In this paper, we introduce an infection risk score that provides an estimate of the infection risk arising from human contacts. Using a real-world human contact dataset, we show that the proposed risk score can provide a realistic estimate of the level of risk in the population. We also describe how the proposed infection risk score can be implemented on smartphones. Finally, we identify representative use cases that can leverage the risk score to minimize infection propagation.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; **Mobile computing**; • **Networks** → *Peer-to-peer protocols*.

KEYWORDS

Infection risk score, Contact Tracing, Mobile Computing, Internet of Things, Mobile Health

ACM Reference Format:

Rachit Agarwal and Abhik Banerjee. 2020. Infection Risk Score: Identifying the risk of infection propagation based on human contact. In *1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19 (COVID-19)*, November 3, 2020, Seattle, WA, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3423459.3430754>

1 INTRODUCTION

21st century has already been witness to multiple pandemics in the first two decades, with the biggest being COVID-19 caused by

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
COVID-19, November 3, 2020, Seattle, WA, USA

© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8168-0/20/11...\$15.00
<https://doi.org/10.1145/3423459.3430754>

the SARS-CoV-2 virus. The unprecedented spread of the COVID-19 has led to global efforts by governments to contain the pandemic and to limit the impact of the virus on human society. As with any other infectious disease, the efforts to contain the virus largely focus on (i) minimizing human-to-human contact by enforcing people to maintain a certain distance with others (also known as **social distancing**), (ii) minimizing economic activity (also known as **lockdown**) for a certain period in geographical regions, such as cities, states or even entire countries, and (iii) by performing **contact tracing**, which involves tracking the disease spread by identifying the contacts of the confirmed cases. However, these efforts have been met with varying degrees of success, and the authorities have been trying to use technology as much as possible to elevate their efforts [13].

With the rise of the Internet of Things (IoT) and mobile health [33] (also referred to as *mHealth*), there has been a growth in the number of possibilities related to not only understanding the environment but also detecting diseases early. With regards to the COVID-19 pandemic in particular, governments around the world have looked to leverage the use of smartphone applications for limiting the spread of the disease, given the ubiquity of smartphone usage. While many of these applications focus on providing up-to-date information about the spread of the disease, other applications aim to notify users in real-time when they come in contact with an infected person [11]. These infection tracking applications use a variety of sensors embedded in a smartphone to help detect the transmission in real-time. A common type of sensor used is Bluetooth Low Energy (BLE), which can be used for proximity detection. Multiple applications that leverage BLE for the purpose of monitoring the growth of the COVID-19 pandemic have been introduced in various countries. In India, the Aarogya-Setu Application [20] informs how many infected people are within a certain distance of a person using the application by matching with national database of the infected people. In Australia, the COVIDSafe application [2] provide notifications to the users if their contact is detected with a confirmed infected person. Similar applications have been developed by the governments of many other countries. Further, a collaboration between Apple Inc. and Google has led to the development of an Exposure API that enables developers to build various applications using which application users can know if they came into contact with other infected people [6]. Apart from smartphone applications, other types of technologies are also used to help in the cause of containing pandemic. These include the use of SwipeSense technology to track use of medical equipment and to track whether hospital staff wash their hands regularly¹.

¹<https://www.cncb.com/2020/08/02/hospitals-tracking-covid-19-with-badge-sensors-swipesense-technology.html>

Despite the technological innovations and advancements, the use of applications, such as those described above, and technologies for managing and controlling the spread of the infection is challenging due to multiple reasons. Firstly, a person carrying the disease may not show any symptoms for a long period (e.g. when the infection is in the incubation period). As a result, any close contacts with other people would not be detected as being risky by infection tracking applications, and hence would not help in containing the spread of the infection. Indeed, in the case of COVID-19, a significant fraction of cases have been identified as asymptomatic for the entire duration of infection [8]. These also contribute to community spread of the disease, which can lead to exponential growth in the number of infections. Secondly, once someone is confirmed as infected, he/she is typically isolated and is not allowed to get involved in any social activities until fully recovered. Thirdly, existing methods for managing infection spread are primarily reactive. Counter-measures are often taken after a person is confirmed to be infectious. Subsequently, authorities proceed with counter measures such as lockdown of the specific geographical region. Thus, currently, the scope for detection of the infection spread and its management is limited. Therefore, there is a need for an early estimate of the potential risk in a geographical region to enable authorities to act quickly.

For risk estimation to be effective, it needs to identify people who have greater exposure to the infection, quarantine the exposed people, identify regions with potentially high exposures, and declare a region as hot-spot even before the outbreak happens in that region. Additionally, the risk estimation measure should also be able to identify situations which are likely to lead to transmissions even when there are not any confirmed presence of the known infections. Finally, any such risk estimation needs to be adaptable to a wide range of technology platforms. While a notion of risk score has been introduced as part of the Exposure API by Apple Inc and Google, its main drawbacks is that it only provides a risk measure based on confirmed exposures to infections.

In this paper, we present a risk score that can be used to assess the risk for individuals based on contact events identified using smartphones. A key novelty of the proposed risk score is that it estimates the risk propagation, unlike existing literature that only assess immediate risk. The proposed risk score can be used to assess the level of risk within geographical regions, enabling authorities to act early to contain a potential outbreak. Further, monitoring the risk score can also help individuals take actions.

In particular, the key contributions of this paper are as follows:

- **Infection risk score:** We introduce a risk score that estimates infection propagation by monitoring contact events among individuals. The risk score takes into consideration factors such as the contact proximity, transmission likelihood and vulnerability to a disease.
- **Evaluation using realistic dataset:** We evaluate the infection risk score using a real-world human contact dataset that has previously been used to study infection propagation. Our results show that potentially risky situations are well captured using the infection risk score.

- **Adaption of risk score using smartphones:** We provide detailed description on how smartphones can be used to implement the infection risk score to track infections.

Finally, we also discuss how the accuracy of infection risk score can be improved by incorporating contextual information, and also present a discussion on potential use cases of the risk score for managing infection spread.

The rest of the paper is organised as follows. Section 2 provides detailed survey of related techniques and existing metrics used to quantify risk and exposure. Section 3 provides details of the proposed risk model. In Section 4 we evaluated the model using real data. Section 5 provides the details on the propose version of the smartphone application. This is followed by perspective uses of the risk score in section 6. We finally conclude in Section 7.

2 RELATED WORK

Recent studies focusing on containing pandemic can be mainly classified into three broad groups: survey based studies, IoT based studies and epidemic model based studies.

In survey based studies, in [19], authors report that factors such as contact with infected person, work overload, medical history of the person, and if the person wore Personal Protective Equipment (PPE) or not play an important role in determining the risk of infection transmission of COVID-19. Similarly, to identify potential exposure, WHO uses risk assessment forms to determine the risk of exposure. Here they ask questions related to if the person wore the PPE as recommended or not [34].

In IoT based studies, there is increased focus on smartphone based infection detection. Many applications and IoT Devices are available that perform contact tracing using proximity checks. A survey of some of these application is present in [11]. We do not survey these applications again and instead present, in brief, new applications and devices that have come-up since the publication of [11]. Recent applications and devices includes EasyBand, a wearable device that vibrates when a marked (infected) Easyband comes in close proximity [27]. Nonetheless, it has issues related to centralized control and communication. In [12], authors used magnetometer based proximity detection, while in [22] authors used multiple sensors to improve the distance estimation accuracy. Such techniques fail in the case when a smartphone lacks certain required sensor. Further, these applications achieve *privacy by architecture* and not *privacy by design*. Many recent application and IoT devices claim to follow privacy guidelines such as those mentioned in [27]. These applications and devices include: (i) Pan European Privacy-Preserving Proximity Tracing (PEPP-Pt)² that uses anonymized ID for communication, (ii) TraceSecure that uses secret sharing technique to identify proximity [4], and (iii) proximity-based privacy-preserving contact tracing (P³CT) that uses ambient signature protocol [21]. Again, while these applications are privacy preserving, they achieve privacy by architecture. In summary, these studies model risk using factors such as distance [4, 12, 20, 22, 27] and duration [6]. These works mainly use either BLE or magnetometer to estimate distance from neighbor. Nonetheless, these works do not quantify the risk, and instead, just provide an estimate of whether a person was in contact with some other person or not.

²<https://www.pepp-pt.org>

On the other hand, from epidemic modeling point of view, there are many studies that quantify risk using different parameters such as: size of cough droplets, rate of cough, volume of particles generated, concentration of pathogens, max distance covered by pathogen in air, pathogen particles lost due to temperature and humidity, time an infected person stayed at a given location, duration of contact with susceptible person, and his pulmonary rate [26]. Most of these factors, until now, cannot be estimated using smartphone. Instead, disease specific average values for these factors can be used as constants while modeling risk score. In [26], authors estimated risk as an aggregation of risk score for both when a person comes in direct contact with other person and when a person gets infected indirectly (a case of *community spreading*).

3 INFECTION RISK SCORE

In this section, we present the *infection risk score* that quantifies risk of catching an infection. Our score considers exposure to pathogen and context in a social network setting. For convenience, *infection risk score* is referred by the term *risk score* in the remainder of the paper.

3.1 Network model: Modeling the population as a temporal network

For any geographical area, we consider the population to be represented by a temporal graph G such that $G(V_t, E_t)$ is a temporal snapshot at time t that is created by individuals in a given area A_a . For the purpose of this paper, we consider that the risk score computation for individuals is done using mobile apps, and hence, each individual is represented using smartphones. Here V_t is the set of smartphones communicating and active at time t and E_t is the set of edges that exists between smartphones in V_t . Let Δt be the time difference between two consecutive temporal snapshots of G . For our model we assume that if two people are in contact, for say 10 epochs, then the edge between them is persistent over $\frac{10}{\Delta t}$ snapshots of the graphs. Each person $i \in V_t$ has a location, l_t , marked by latitude and longitude pair such that $l_t = (la_{i,t}, lo_{i,t})$. Given the interactions, at time t , each person i has a neighborhood, $N_{i,t}$ where each person $j \in N_{i,t}$ has an edge (in E_t) to the person i and is $d_{i,j,t}$ distance apart. Here $d_{i,j,t} < \theta_d$ i.e., i and j are within communication range and at maximum θ_d distance apart.

3.2 Risk score parameters

In this section, we identify the key factors that impact infection propagation.

- (1) **Exposure caused by a neighbor:** Communicable diseases such as COVID-19 generally spread when a person $i \in V_t$ comes in close proximity with a infected person (person j) or touches the surface that infected person has touched [26]. In such a case, the person i is exposed to pathogens from the infected person, which can lead to infection spread. The exposure to a neighboring individual is a key factor determining the likelihood of a transmission event from a neighbor, and we term this as the **neighbor exposure**. For the scope of the current paper, we limit our discussion to the exposure caused when an infected person come in close proximity,

although this may easily be extended to include other modes of propagation.

To determine how the neighbor exposure impacts the spread of infection, we consider that an infected neighbor j exhales $n_{i,j,t} \in \mathbb{R}^+$ pathogens and these pathogens are homogeneously distributed within the permissible θ_d distance. Further, we consider the following assumptions: (i) there is no loss in pathogens, (ii) each time same number of pathogens are exhaled, and (iii) between two consecutive temporal snapshots of the graph (i.e., $G(V_t, E_t)$ and $G(V_{t-\Delta t}, E_{t-\Delta t})$), a person i stays in contact with person j for the Δt time. In such a scenario, the exposure to an infectious disease of the person i at time t with respect to a particular neighbor j is given by $E_{i,j,t} = \Delta t \times n_{i,j,t}$.

In ideal conditions, if a neighbor j is not infected, i.e., he/she does not cough, and wears proper protective gears such as face mask or face shields, $E_{i,j,t} = 0$ because there are no pathogens exhaled by j . In such a case, the whole idea of maintaining social distancing even when people are not infected would fail and susceptible people would be deemed harmless. On the other hand, if some neighbor is infected and coughing badly, $E_{i,j,t} \gg 0$. In this case, other people would ideally limit from meeting the infected person. In such a situation also, barring the infected person, other susceptible people would continue their physical social activities. Let $r_{j,t-\Delta t}$ be the risk score of the neighbor at time $t - \Delta t$. To account for above mentioned aspects and ensure that social distancing is enforced between susceptible people also, we add the previous instance risk score of the neighbor to the exposure caused due to the neighbor, i.e., $E_{i,j,t} = \Delta t \times n_{i,j,t} + r_{j,t-\Delta t}$.

- (2) **Neighbor weight:** We define the **neighbor weight** as the likelihood that an individual in the vicinity is infectious. Since we aim to estimate the risk even in situations where confirmed infections are not known, the neighbor weight can be estimated based on multiple contextual parameters. For instance, in the case of communicable diseases such as COVID-19, if a neighbor is from a hot-spot area or has a history of the disease then the risk of getting infection from the neighbor is high because the neighbor is coming from a containment zone. Further, impact of diseases like COVID-19 is high on people who have a weak immunity either due to age or have chronic diseases like kidney failure and diabetes. On top, if a person is staying indoor, with a poor ventilation chances of spreading the disease and getting infected increases manyfold [28, 32]. In [28], authors recommend that proper ventilation indoor can reduce infections up-to 60%. Nonetheless, for COVID-19, different countries have different statistics, for example, India having relatively younger population, middle age people are more infected while more older people have died. Let $w_{j,t}$ be the **weight** such that $w_{j,t} \in [0, 1]$ that identifies such contextual information of the neighbor. Summed over all the neighbors of the person i at time t , the total exposure of i from its neighbors $j \in N_{i,t}$ is

thus given by equation 1.

$$E_{i,t} = \sum_{j \in N_{i,t}} w_{j,t} \times (E_{i,j,t} + r_{j,t-\Delta t}) \quad (1)$$

3.3 Risk score formulation

In addition to the neighbor weight and exposure, we define **vulnerability** as the likelihood that an individual exposed to risky situations continues to be at risk. At any time t the risk score of an individual i is the dependent on the risk score at $t - \Delta t$, his vulnerability, and the exposure from the neighbors at t . The total risk, thus, is given by equation 2.

$$r_{i,t} = \frac{v_{i,t} \times r_{i,t-\Delta t} + \sum_{j \in N_{i,t}} w_{j,t} \times (E_{i,j,t} + r_{j,t-\Delta t})}{1 + \sum_{j \in N_{i,t}} w_{j,t}} \quad (2)$$

Here the denominator is the normalization factor. A disease usually has a period between when the person i gets infected from the disease and time when he becomes an active spreader of the disease. For example, for COVID-19, the median incubation period is around 5 to 6 days³. In our scenario, even if a person comes in contact with a person for whom the disease is still in incubation period, the risk exposure is equally high as compared to meeting a person who is an active spreader. Thus, our model does not consider the incubation period.

From the equation (2), the value of $r_{i,t} \in \mathbb{R}^+$. If the person is taken into isolation (i.e., no interaction with neighbors) after getting infected, his risk score will decrease with a factor $v_{i,t}$ and will eventually decay and reach minimum in $t = \lceil \frac{r_{i,t-\Delta t}}{v_{i,t}} \rceil$ time instances. This accounts for the fact that risk to and from such people is minimized when they are in isolation. For simplicity, at $t = 0$ (or the initial condition) for all people we assign them as susceptible and their risk score to $r_{i,0} = 1$. As the actual infection state of a person is unknown, the idea of social distancing mandates to maintain a certain distance even if the person is susceptible. Maintaining social distancing reduces the possibility of getting infected. We assign a non zero value to $r_{i,0}$ to ensure that social distance is maintained and our model captures it. For simplicity, let $r_{i,0} = 1$. As and when a person is officially tagged infected, we assign $r_{i,t} = 2$. Note that, a low value of $r_{i,t}$, is achieved when all the neighbors are susceptible. For a new person joining in, we assume that he is a susceptible person.

Our method only considers ego network of a person for the calculation of the risk score. This enables all the smartphones involved to compute their individual risk scores simultaneously.

4 EVALUATION AND VALIDATION

In this section we provide an evaluation and validation of our risk model using a real-world dataset.

4.1 Dataset

While there are many datasets which have previously been used to study epidemic spread, specially smartphone based datasets that use Call Detail Records (CDRs) and GPS location information, [5], they are (i) not widely used [23], and (ii) mostly generated from a

random population sample which do not reflect true neighborhood size. Instead, we use a dataset of 789 individuals (including students and teachers) obtained on a single day in an American high school that has 158 rooms [25], which has previously been used to study spread of infectious diseases [28]. Here each point of interest (POI) is considered to be a room in the school. The dataset is mainly used to study human contact network for infectious disease transmission. The dataset is collected between 6AM to 4:30PM at an interval of 20 seconds. The granularity of positioning information available is at the level of rooms, and hence, each individual is geo-tagged with the room ID they are in at a particular epoch. We consider that contact events occur between individuals whenever they are in the same room, and all individuals present in a particular room at a given epoch are connected to each other.

The temporal distribution of individuals in the dataset is shown in Fig. 1. Fig. 1(a) shows the heatmap of number of people present in a room at different epochs. The white color represents that nobody was present in a room at the particular epoch. Fig. 1(b) presents total number of people in a school at a given epoch. A sudden increase and a sudden drop in the number of people accounts for the beginning of the school in the morning when people arrive, and the end of the day, when they went back from school. Fig. 1(c) presents the maximum number of rooms occupied by people. Note that at maximum only $\approx 62\%$ rooms are occupied. Fig. 1(d) presents ratio between number of people in the school and rooms occupied at a given time. The maximum average density of people in a room is 9. A sudden increase at the end of the day is because most of the people were present in a single room. Fig. 1(e) presents number of times a given room was occupied during the data collection period. From the figure we infer that (i) some rooms were always empty and nobody went to those rooms, (ii) the entire population is concentrated in only a few rooms and after certain time period there is an exponential decrease in the population indicating the end of classes in the school, (iii) during the day, rooms gradually start to fill up and there is an exponential rise in the population size.

4.2 Dynamics of epidemic spread on the contact network

We evaluate the proposed risk score using both SI (Susceptible-Infected) and SIS (Susceptible-Infected-Susceptible) models. Here, we note that, since the risk score looks at contact events only, we do not evaluate a SEIR (Susceptible-Exposed-Infected-Recovered) model, as identification of "exposed" and "recovered" states are not the focus of the model.

Let $S_{i,t}$ be the fraction of people that are susceptible in a region i at time t , $I_{i,t}$ be the fraction of population that is infected in a region i at time t , $N_{i,t} = 1$ be the fraction of total population in the region at time t , β_i be the infection rate in the region i , and γ_i be the recovery rate in the region i . Note that at any given point of time $N_{i,t} = S_{i,t} + I_{i,t}$ because we consider only two states, susceptible and infected. The change in the fraction of susceptible and infected people over time is given by equation (3) [10]. Here the underlying assumptions are that there is a homogeneous mixing of the population and no birth and death happens (the total population is fixed).

³<https://www.mohfw.gov.in/pdf/DGSOrder04of2020.pdf>

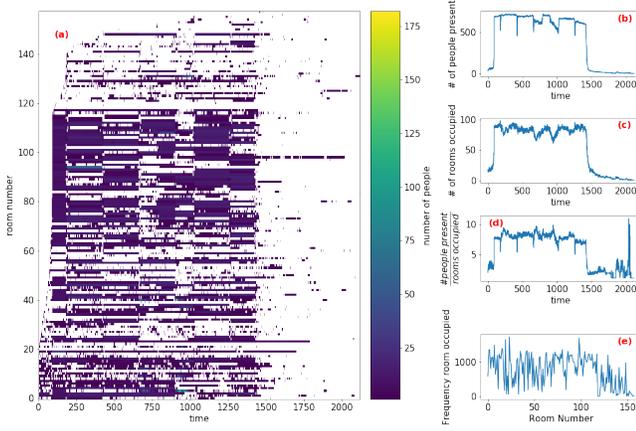


Figure 1: Temporal distribution of people in the rooms. (a) Heatmap showing number of people in each room over time. The white color represents empty room at the particular epoch. (b) Total number of people present in the school over time. (c) Total number of room occupied in the school over time. (d) average density of each room. (e) Number of times a room is occupied.

$$\begin{aligned} \frac{dS_{i,t}}{dt} &= -\frac{\beta S_{i,t} I_{i,t}}{N_{i,t}} + \gamma I_{i,t} \\ \frac{dI_{i,t}}{dt} &= -\frac{dS_{i,t}}{dt} \end{aligned} \quad (3)$$

4.3 Results

Currently, the exact behavior of exposure and vulnerability parameters for pandemics such as COVID-19 is not known. Further, as the dataset is POI based, actual distances are also not available in the dataset. Thus, we assume that the exposure parameter for each person is normally distributed with $\mu = 0.5$ and $\sigma = 0.1$. Further, the vulnerability parameter is also normally distributed with $\mu = 0.5$ and $\sigma = 0.2$.

For our analysis we study following three aspects using different initial condition, infection rate, and recovery rate. First, we identify fraction of people who are identified infected using the SIS and SI epidemic models. This helps us understand the infection spread over time in the population and understand the *dynamics on the contact network*. Second, we measure the ratio between the median risk scores of infected people and susceptible people. A ratio more than one indicates that the risk score of infected people is more, as intended. A higher ratio implies that the risk score can be used to better identify people who are exposed to infection and have high probability to get infected. When there are no infections in a neighborhood, this value tends to 0. Third, we study the fraction of people that are alerted using our model.

To test and study the above-mentioned aspects, as an initial condition, the values for $I_{i,0}$, β_i and γ_i used are $I_{i,0} \in \{0.0, 0.01, 0.5\}$, $\beta_i \in \{0.0, 0.5, 1.0\}$ and $\gamma_i \in \{0.0, 0.75\}$. $I_{i,0} = 0.0$ states that there are no initial infections in the region. $\beta_i = 0.0$ states there are no transmission happening and the disease does not spread via contact. On the other hand, $\beta_i = 1.0$ would state that the disease is highly contagious. Similarly, $\gamma_i = 0.0$ would state that there is no recovery

which is equivalent to SI type epidemic model. $\gamma_i = 0.75$ would mean that the recovery rate is 75% (i.e. similar to the recovery rate of COVID-19 patients in India⁴). The results presented here are averaged over 50 simulations runs and conducted using python.

Figures 2 and 3 present results obtained for the above-mentioned aspects when different values of $I_{i,0}$, β_i and γ_i are used for SIS and SI models respectively. The β_i and γ_i values are assumed to not vary across rooms. Fig. 2 is obtained when $\gamma = 0.75$ while Fig. 3 is obtained when $\gamma = 0.0$. From the Fig. 2, as per SIS model, when there is no infection, dissemination of infection does not occur because subsequent $dI_{i,t}/dt = 0$ (see fig. 2(a)). This lead to ratio of median risk scores (represented as $r_{m,Infected}/r_{m,Susceptible}$) to be 0 as there are no infected people (see Fig. 2(b)). The inset Fig. 2(b') shows the median risk scores of susceptible people ($r_{m,Susceptible}$) and indicates that, even when there are no confirmed infections, crowded situations which carry high risk can be identified as having high risk scores. As our risk model is not dependent on β_i , γ_i , and initial infection, via risk score, we are able to detects potential risky situations (see fig. 2(c), 2(f), and 2(i)) which epidemic models such as SIS model are not able to detect. For cases when $I_{i,0} \geq 0.01$ and $\beta_i \in \{0.0, 0.5, 1.0\}$, we observe that infections either die off (for $\beta_i = 0.0$) or achieve stability (for $\beta_i \in \{0.5, 1.0\}$, see fig. 2(d) and 2(g)). The reason for reduction in infections is the recovery rate, while for stability it is the low number of people present when $epoch > 1500$. The ratio of median risk scores for different β_i and $I_{i,0} \geq 0.01$ is shown in fig. 2(e) and 2(h) where we observe that after few epochs the ratio is < 2 and even reaches < 1 in short duration. This behavior is because (a) the median value of infected identified by SIS model is less than the median value that of susceptible people and (b) the number of infected is less that number of susceptible. From the fig. 2(f) and 2(g) we also see that, irrespective of the epidemic state, most of the people are at high risk.

On the other hand, from Fig. 3, we see that when there is no recovery (i.e., $\gamma_i = 0.0$) and when $I_{i,0} = 0$, the behavior is similar to previous scenario (see fig. 3(a), 3(b), 3(b'), 3(c), 3(f), and 3(i)). Nonetheless, when $I_{i,0} > 0.0$ and $\beta_i \neq 0.0$, the infections eventually reach entire population which is true as there is no recovery (see fig. 3(d) and 3(g)). Further, in this case, due to the above-mentioned reason, ratio of median risk scores is also high (see fig. 3(e) and 3(h)). Ratio equal to 1 is achieved when $\beta_i = 0$.

As the results presented here show, the growth in risk score values increases when there is increased contact of susceptible individuals with infected individuals. Further, even in situations where infections are unknown, the risk score values are shown to grow when there are more crowded situations and population movement among those. Thus, the risk score values can be used for managing infection spread, even before confirmed infections are identified.

5 RISK SCORE IMPLEMENTATION USING SMARTPHONES

In this section, we demonstrate how the proposed risk score can be implemented as part of a smartphone based infection tracking applications. There are two main components required for estimation

⁴<https://www.financialexpress.com/lifestyle/health/indias-covid-19-recovery-rate-nears-75-case-fatality-rate-one-of-the-lowest-globally-at-1-86/2063108/>

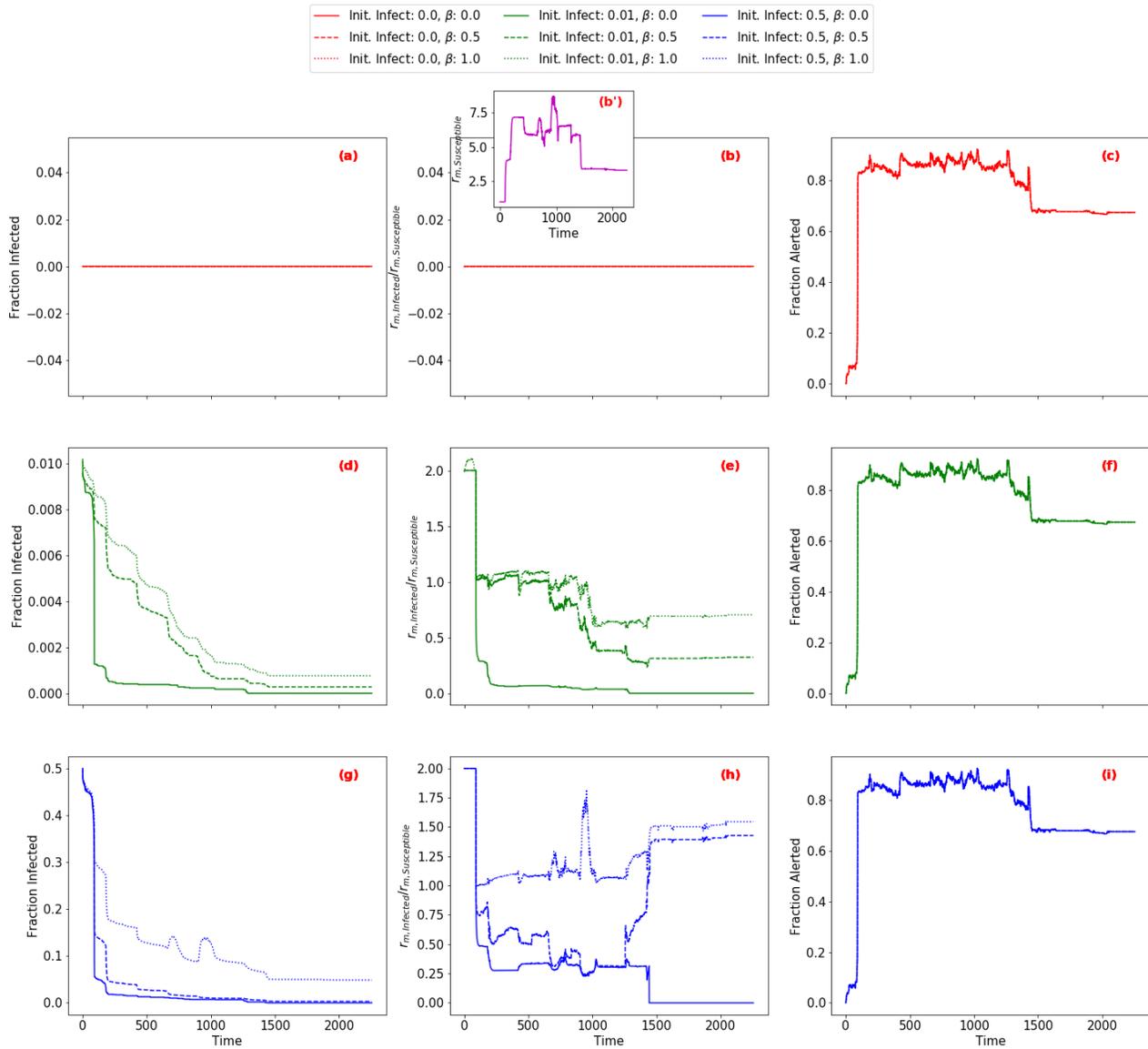


Figure 2: Results for SIS epidemic model, $\beta \in \{0.0, 0.5, 1.0\}$, $\gamma_i = 0.75$ and $I_{i,0} \in \{0.0, 0.01, 0.5\}$

of the risk score on a smartphone - (a) aggregation of risk scores of neighboring smartphones, and (b) computation of the risk score of the smartphone itself. Similar to the infection tracking applications used for COVID-19, we consider that estimation of the exposure is done using BLE. However, unlike existing applications which use centralized data repositories to obtain risk scores of neighboring smartphones (i.e. if they are confirmed to be infected), using our approach, each smartphone can (a) periodically broadcast its own risk score, by embedding this value in the BLE advertising packets, and (b) periodically update its own risk score by aggregating the risk scores of all other smartphones in its neighborhood. Such an approach has the following advantages:

- The risk score computation does not need to depend on a centralized database containing information about infected individuals, which might be outdated.
- The risk score reflect encounters not just with confirmed individuals, but also present environments that are risky from the perspective of infection spread.
- Our approach is better suited for privacy preservation, since no information pertaining to the identity of individuals is stored or communicated.

Next, we provide details on the design of the BLE advertising packet as well as how risk score computation can be done individually by a smartphone application. For the purpose of this discussion, we refer to the smartphone performing the risk computation as

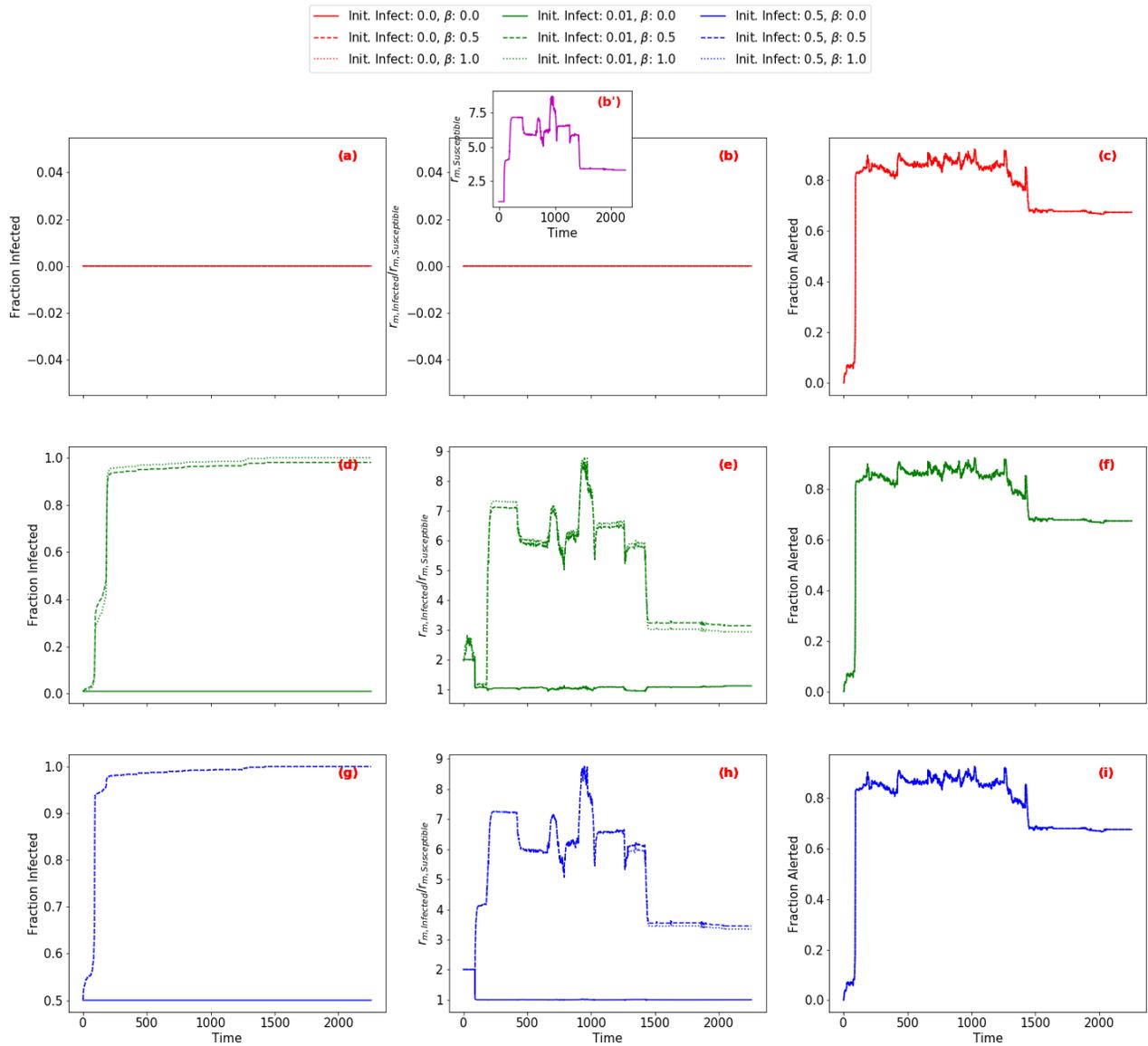


Figure 3: Results for SI epidemic model, $\beta \in \{0.0, 0.5, 1.0\}$, $\gamma_i = 0.0$ and $I_{i,0} \in \{0.0, 0.01, 0.5\}$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31																	
UUID																	r	0	1	.	0	0	w	0	.	5	0																					

Figure 4: BLE Packet Format

the ego node, and all other smartphones in its vicinity as neighbor nodes.

5.1 BLE advertising packet

Existing infection tracking techniques record the BLE Media Access Control (MAC) addresses of nearby smartphones [14] and compare them with a centralized database of infected individuals. Instead,

we discuss how we use the BLE advertising packet to communicate risk score values.

The BLE advertising packet allows including optional payload of up to 31 bytes [17]. We use these available bytes for broadcasting the risk score. Our payload includes:

- (1) A 128-bit unique identifier (**UUID**) which is a fixed value used to identify the service, enabling each smartphone to filter out all nearby beacons broadcasting the risk score.
- (2) A 6 bytes long **Risk score** which includes the risk score value rounded to two decimal places and prefixed by “r”.
- (3) A 5 bytes long **weight of the neighbor** which includes the neighbor weight value rounded to two decimal places and prefixed by “w”.

RSSI	Neighbor weight
> -55 dbm	0.8
> -63 & ≤ -55 dbm	0.5
> -75 & ≤ -63 dbm	0.1
≤ -75 dbm	0.0

Table 1: Neighbor exposure estimation from RSSI values of BLE advertisements received from neighbors

Fig. 4 shows the payload format of the BLE advertising packet.

5.2 Risk score computation

In addition to the risk scores obtained from the BLE advertisements from the neighbors, the weight of the neighbors, and exposure caused by the neighbors are also required for the purpose of risk score computation, along with the vulnerability of the node itself. Note that here “node” means the smartphone.

5.2.1 Neighbor exposure. The exposure from a neighbor is an estimation of the likelihood of a transmission event from a neighbor. For infectious diseases such as COVID-19, the likelihood of transmission increases with close contacts. While BLE signal characteristics, such as received signal strength indication (RSSI) and attenuation, can be used for distance estimation, they are known to be noisy estimators [18]. Hence, for the purpose of estimation of exposure from the neighbor, we use a coarse grained mapping, similar to those used in the Exposure API [7]. Based on the existing studies, Table 1 shows how the exposure values can be mapped from the RSSI values [15]. A higher RSSI value maps to a higher exposure from a neighbor.

5.2.2 Neighbor weight. The neighbor weight is an estimate of the likelihood of a neighboring node to be infectious, which can depend on a range of factors, such as the prevalence of preexisting diseases, age, etc. If such information is available, the derived neighbor weight is included in the BLE advertising packet. However, while such information may not always be available at an individual level, approximate measures are often available at a population level, which can be used as fixed values for all smartphones in a geographic region. For instance, neighbor weight may be derived from the basic reproduction number (R_0) [9] value for a particular epidemic for a given geographical region.

5.2.3 Vulnerability. The vulnerability of the ego node is an estimate of how quickly an individual can recover when exposed to infection, and as with the neighbour weight, this depends on a range of factors such as preexisting conditions, age, etc, as well as the nature of the disease itself [3]. When available, such information is incorporated in the computation of the risk score by the ego node.

Currently, the only data shared between the smartphones are the neighbor weights and risk score values. These values are computed on individual smartphones and shared with the neighbors. As no other parameter is shared other than computed values of neighbor weights and risk score and no other information about the neighbor is shared, we enable privacy by design. As a proof of

concept implementation, we can also provide an alpha version of a smartphone application upon request for the readers to test.

6 FUTURE DIRECTIONS

While the risk score proposed here focuses on parameters that can be readily measured using smartphones, it can be developed further, both in terms of increasing its accuracy towards risk estimation, as well as applying it to individual use cases, which we highlight in this section.

6.1 Increasing accuracy of risk score

The accuracy of the risk score proposed in this paper can also be increased by incorporating additional contextual information, where available. Some examples of this are:

- **Indoor and Outdoor location detection:** The likelihood of infection spread has been known to be higher in indoor environments compared to outdoor [28]. This can be incorporated into the risk score by first, automatically detecting the indoor/outdoor context [1], and secondly, by incorporating it into the risk score itself.
- **Identification of exposure context:** As outlined previously in section 5.2, by identification of the infection context in real-time, the risk score computation can be made more accurate. This can include detection of respiratory symptoms to better estimate the exposure [16, 29].

In addition to the points above, a general challenge with all infection tracking applications is that they do not cater to the entire population, since people may not always have access to smartphones and other IoT devices.

6.2 Use cases

The proposed risk score is applicable towards monitoring and managing the spread of infection for population groups, such as over a geographical region, as well as for individuals.

- (1) **Risk score at different spatial scale:** The proposed risk score, in addition to computing score of an individual can be used to compute the risk score at any spatial scale (i.e., a country, a city, a building, a house, a room). For instance, considering A_a be the area for which risk score has to be computed, such as a district, and let L_a be the group of people in that region at time t . The risk score of region A_a at time t is defined as equation 4.

$$r_t^{A_a} = \frac{\sum_{i \in L_a} r_{i,t}}{|L_a|} \quad (4)$$

Here, $|L_a|$ represents the number of people in A_a . Consider a region, R to be comprised of many A s, the total population at time t , N^t is thus $\sum_{i \in R} |L_i|$. Some examples include:

- (a) **Monitoring of geographical regions by government authorities:** As evidenced by the COVID-19 pandemic, the infection spread often starts from small geographical regions, which can grow exponentially if early actions are not taken. Our proposed risk score can be used to obtain an early estimate of the likelihood of infection transmissions within a geographical region. Subsequently, preventative actions, such as increased testing, can be taken, without

even resorting to lockdowns that have economic and social impacts.

- (b) **Monitoring of individual buildings:** An important aspect of managing the spread of infections is to reduce the likelihood of spread in controlled environments, such as office buildings, hotels, hospitals, etc. In such scenarios, risk score can be used to monitor behavior of individuals within such a region, and take quick actions even before any infection is confirmed. Some examples of such use cases are:
- (i) **Hotels:** Guest movements and interactions among guests at hotels can have significant consequences to the infection spread in a pandemic, as has been seen in the case of COVID-19 [30]. The risk score can be used to act quickly by enforcing close monitoring of the individuals who are found to be in risky situations.
 - (ii) **Hospitals:** In order to handle increasing case loads during a pandemic, hospitals typically have dedicated wards. In such cases, it is critical to minimize the likelihood of transmission from such dedicated wards to other wards in the hospital [24], which can be done through monitoring of the risk scores of patients, doctors and other hospital staff.
 - (iii) **Office buildings:** Managing the recovery from a pandemic is equally important to managing its spread, and the risk score can be used as a part of the plans used for businesses and office buildings [31].

Similar scenarios may be envisioned for other closed environments, such as residential buildings, supermarkets, shopping malls, airports, etc.

- (2) **Individual monitoring:** Risk score can also be used to provide real-time alerts to individuals to take action. For instance, it can be used to provide prompts to wear mask if one is detected to move from a less risky region to more risky one. Further, risk score can be used to provide personalized alerts for individuals. For instance, vulnerable people (i.e. who are likely to be affected more due to pre-existing conditions), can be alerted early by using a lower alert threshold.

7 CONCLUSION

In this paper, we introduced a risk score that estimates infection propagation by leveraging the neighborhood of an individuals at a given time. On top, our risk score also takes into consideration factors transmission likelihood and vulnerability to a disease. Our results show that our risk score is able to capture potential risky situations. To further leverage our risk score we demonstrate how our risk score can be implemented in a contact tracing applications and as a proof of concept make it available upon request. Nonetheless, as future directions, we provide use cases and potential parameters that can be included in the risk score to make it more robust.

ACKNOWLEDGEMENTS

The authors contributed equally in the research.

REFERENCES

- [1] R. Agarwal, S. Chopra, V. Christophides, N. Georgantas, and V. Issarny. 2019. Detecting Mobile Crowdsensing Context in the Wild. In *20th IEEE International*

- Conference on Mobile Data Management* (Hong Kong). IEEE, 170–175.
- [2] Australian Government Department of Health. 2020. COVIDSafe App. <https://www.health.gov.au/resources/apps-and-tools/covidsafe-app> (Accessed 06/09/2020).
- [3] BBC. 2020. Coronavirus: How long does it take to recover? <https://www.bbc.com/news/health-52301633> (Accessed on 09/22/2020).
- [4] J. Bell, D. Butler, C. Hicks, and J. Crowcroft. 2020. TraceSecure: Towards Privacy Preserving Contact Tracing. *arXiv* (April 2020), 1–22. arXiv:2004.04059 [cs.CR]
- [5] V. Blondel, A. Decuyper, and G. Krings. 2015. A survey of results on mobile phone datasets analysis. *EPJ Data Science* 4, 1 (Aug. 2015), 1–55.
- [6] J. Clover. 2020. Apple’s Exposure Notification System: Everything You Need to Know. <https://www.macrumors.com/guide/exposure-notification/> (Accessed 13/07/2020).
- [7] Google. 2020. Define meaningful exposures: Google API for Exposure Notifications. <https://developers.google.com/android/exposure-notifications/meaningful-exposures> (Accessed on 09/19/2020).
- [8] J. He, Y. Guo, R. Mao, and J. Zhang. 2020. Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. *Journal of Medical Virology* Early Access (2020), 1–11.
- [9] J. M. Heffernan, R. J. Smith, and L. M. Wahl. 2005. Perspectives on the basic reproductive ratio. *Journal of The Royal Society Interface* 2, 4 (Sept. 2005), 281–293. <https://doi.org/10.1098/rsif.2005.0042>
- [10] H. Hethcote. 2000. The Mathematics of Infectious Diseases. *SIAM Rev.* 42, 4 (2000), 599–653.
- [11] M. Islam, I. Islam, K. Munim, and A. Islam. 2020. A Review on the Mobile Applications Developed for COVID-19: An Exploratory Analysis. *IEEE Access* 8 (Aug. 2020), 145601–145610.
- [12] S. Jeong, S. Kuk, and H. Kim. 2019. A Smartphone Magnetometer-Based Diagnostic Test for Automatic Contact Tracing in Infectious Disease Epidemics. *IEEE Access* 7 (Jan. 2019), 20734–20747.
- [13] O. Khazan. 2020. The Most American COVID-19 Failure Yet. <https://www.theatlantic.com/politics/archive/2020/08/contact-tracing-hr-6666-working-us/615637/> (Accessed 12/09/2020).
- [14] P. Kindt, T. Chakraborty, and S. Chakraborty. 2020. How Reliable is Smartphone-based Electronic Contact Tracing for COVID-19? *arXiv* (May 2020), 1–13.
- [15] D. Leith and F. Stephen. 2020. Coronavirus Contact Tracing: Evaluating the Potential Of Using Bluetooth Received Signal Strength for Proximity Detection. *arXiv* (May 2020), 1–11. arXiv:2006.06822 [eess.SP]
- [16] D. Liaquat, R. Wu, T. Son, A. Gershon, H. Alshaer, E. de Lara, and F. Rudzicz. 2018. Towards Ambulatory Cough Monitoring using Smartwatches. In *C41 Health Services Research in Pulmonary Disease: Thematic Poster Session*. American Thoracic Society, A4929–A4929.
- [17] Joakim Lindh. 2016. Bluetooth Low Energy Beacons (Rev. A). <https://www.ti.com/lit/an/swra475a/swra475a.pdf>. (Accessed on 09/24/2020).
- [18] A. Mackey, P. Spachos, L. Song, and K. Plataniotis. 2020. Improving BLE beacon proximity estimation accuracy through bayesian filtering. *IEEE Internet of Things Journal* 7, 4 (April 2020), 3160–3169.
- [19] M. Mhang, M. Dzobo, I. Chitungo, and T. Dzinamarira. 2020. COVID-19 Risk Factors Among Health Workers: A Rapid Review. *Safety and Health at Work* 11, 3 (Sept. 2020), 262–265.
- [20] National Informatics Center, India. 2020. Aarogya Setu Mobile App. <https://www.mygov.in/aarogya-setu-app/> (Accessed 06/07/2020).
- [21] P. Ng, P. Spachos, S. Gregori, and K. Plataniotis. 2020. Epidemic Exposure Notification with Smartwatch: A Proximity-Based Privacy-Preserving Approach. *arxiv* (July 2020), 1–11. arXiv:2007.04399 [cs.CR]
- [22] K. Nguyen, Z. Luo, and C. Watkins. 2020. Epidemic contact tracing with smartphone sensors. *Journal of Location Based Services* 14, 2 (Sept. 2020), 92–128.
- [23] N. Oliver, B. Lepri, H. Sterly, et al. 2020. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances* 6, 23 (April 2020), 1–6.
- [24] T. Oraby, M. Tyshenko, H. Balkhy, et al. 2020. Analysis of the Healthcare MERS-CoV Outbreak in King Abdulaziz Medical Center, Riyadh, Saudi Arabia, June-August 2015 Using a SEIR Ward Transmission Model. *International journal of environmental research and public health* 17, 8 (2020), 2936.
- [25] M. Salathé, M. Kazandjieva, J. Lee, P. Levis, M. Feldman, and J. Jones. 2010. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences* 107, 51 (Dec. 2010), 22020–22025.
- [26] M. Shahzamal, R. Jurdak, R. Arablouei, M. Kim, K. Thilakarathna, and B. Mans. 2017. Airborne Disease Propagation on Large Scale Social Contact Networks. In *Proceedings of the 2nd International Workshop on Social Sensing* (Pittsburgh, PA, USA). ACM, 35–40.
- [27] M. Shukla, M. Rajan, S. Lodha, G. Shroff, and R. Raskar. 2020. Privacy Guidelines for Contact Tracing Applications. *arXiv* (April 2020), 1–10. arXiv:2004.13328 [cs.LG]
- [28] T. Smieszek, G. Lazzari, and M. Salathé. 2019. Assessing the Dynamics and Control of Droplet- and Aerosol-Transmitted Influenza Using an Indoor Positioning System. *Scientific Reports* 9, 1 (Feb. 2019), 1–10.

- [29] X. Sun, Z. Lu, W. Hu, and G. Cao. 2015. SymDetector: Detecting sound-related respiratory symptoms using smartphones. In *Proc. of ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan). ACM, 97–108.
- [30] J. Taylor. 2020. Hotel quarantine linked to 99% of Victoria’s Covid cases, inquiry told. <https://www.theguardian.com/australia-news/2020/aug/18/hotel-quarantine-linked-to-99-of-victorias-covid-cases-inquiry-told> (Accessed on 09/23/2020).
- [31] Business Victoria. 2020. COVIDSafe Plan. <https://www.business.vic.gov.au/coronavirus-covid-19/covid-safe-business/covid-safe-plan> (Accessed on 09/23/2020).
- [32] V. Vuorinen, M. Aarnio, M. Alava, and Others. 2020. Modelling aerosol transport and virus exposure with numerical simulations in relation to SARS-CoV-2 transmission by inhalation indoors. *Safety Science (Early Access)* 130 (Oct. 2020), 1–23.
- [33] C. Wood, M. Thomas, J. Budd, T. Mashamba-Thompson, K. Herbst, D. Pillay, R. Peeling, A. Johnson, R. McKendry, and M. Stevens. 2019. Taking connected mobile-health diagnostics of infectious diseases to the field. *Nature* 566, 7745 (Feb. 2019), 467–474.
- [34] World Health Organization. 2020. *Health workers exposure risk assessment and management in the context of COVID-19 virus: interim guidance, 4 March 2020*. Technical Report. World Health Organization. 1–8 pages. <https://apps.who.int/iris/handle/10665/331340>