

# Big Data Analytics in Tracking COVID-19 Spread

## Utilizing Google Location Data

---

**Yaw Mei Wyin**

Institute of Sustainable Energy, Universiti Tenaga Nasional, Kajang, Malaysia

**Prajindra Sankar Krishnan**

Institute of Sustainable Energy, Universiti Tenaga Nasional, Kajang, Malaysia

**Chen Chai Phing**

Institute of Sustainable Energy, Universiti Tenaga Nasional, Kajang, Malaysia

**Tiong Sieh Kiong**

Institute of Sustainable Energy, Universiti Tenaga Nasional, Kajang, Malaysia

---

**Abstract:** According to mobility data that records mobility traffic using location trackers on mobile phones, the COVID-19 epidemic and the adoption of social distance policies have drastically altered people's visiting patterns. However, rather than the volume of visitors, the transmission is controlled by the frequency and length of concurrent occupation at particular places. Therefore, it is essential to comprehend how people interact in various settings in order to focus legislation, guide contact tracking, and educate prevention initiatives. This study suggests an effective method for reducing the virus's propagation among university students enrolled on-campus by creating a self-developed Google History Location Extractor and Indicator software based on actual data on people's movements. The platform enables academics and policymakers to model the results of human mobility and the epidemic condition under various epidemic control measures and assess the potential for future advancements in the epidemic's spread. It provides tools for identifying prospective contacts, analyzing individual infection risks, and reviewing the success of campus regulations. By more precisely focusing on probable virus carriers during the screening process, the suggested multi-functional platform makes it easier to decide on epidemic control measures, ultimately helping to manage and avoid future outbreaks.

**Keywords:** Contact networks, human mobility simulation, epidemic control policy

## Introduction

Human mobility is a key factor in the analysis of an infectious disease's dissemination, especially for highly contagious diseases like COVID-19 ([Gatto et al., 2020](#); [Kiang et al., 2021](#); [Venkatramanan et al., 2021](#); [Wang et al., 2020](#); [Wu et al., 2020](#)). The most efficient way to decrease the disease's rate of transmission of an outbreak is to identify potential patients early and to implement preventative measures as soon as possible. With the spread of devices with precise localization capabilities and strong wireless networks, it is now possible to track human movement over long periods of time and across a wide area. Many databases that track people's travels are now available to researchers, allowing for more in-depth analysis and epidemic spread modelling at a finer spatial and temporal granularity. Traditional sources of mobility data, where many types of motions can be gathered, include GPS and Google Location History ([Ruktanonchai et al., 2018](#)).

The Google History Location Extractor software platform is suggested in this article. The platform estimates on-campus student contact behaviours through GPS trajectories, simulates on-campus student mobility in response to public policies to assess their efficacy in slowing the epidemic's spreading process, and simulates the spread of the epidemic among actual students' trajectory datasets. The platform allows data mining and exploration on the risks of infection spread, including the identification of potential secondary contacts. It consists of two key functions: a probabilistic model of infectious disease transmission at the individual level; and an individual infection risk exploration.

By simulating how people behave in response to public policies and determining how effectively each policy functions, our platform can help governments make educated judgements. In order to better understand the potential COVID-19 spread among university students, this study suggests the creation of a Google History Location Extractor program as a tool for data analysis. It will provide interactive visualization features and capabilities.

The three functions of our platform—exploring regional and individual infection risk; simulating public policy using trajectory replacement; and modelling probabilistic individual-level infectious disease transmission—are all described in depth in the sections that follow. We also evaluate related prior research in these sections. The next sections of the study deal with mobility pattern modelling, implementation, outcomes, and a discussion of our findings. We summarize the main points and stress the importance of our platform in our conclusion.

## Related Works

Building a network to represent person-to-person interactions can help us understand how diseases spread throughout a society. These networks may be created using a variety of

approaches, such as surveys, statistical methods, census data, and movement data ([Chang et al., 2021](#); [Kumar et al., 2021](#); [Maheshwari & Albert, 2020](#); [Muller et al., 2020](#); [Rechlin et al., 2020](#); [Schlosser et al., 2020](#); [Soures et al., 2020](#); [Yi et al., 2021](#)). As a result of the widespread use of location tracking programs on mobile phones, information about people's locations and the amount of time they spend in particular places is documented in a huge amount of data. New systems, such as Google Community Movement Reports ([Google, 2020](#)), offer tools and aggregated movement data in response to COVID-19. To protect people's privacy, data is gathered and rendered anonymous.

People's behaviour quickly changed when social distance measures were implemented across the country. The effects of these measures on state-by-state mobility trends have been extensively reported in scholarly publications and press pieces ([Dave et al., 2020](#); [Lasry et al., 2020](#); [McMinn & Talbot, 2020](#); [Pan et al., 2020](#); [Weill et al., 2020](#)). These assessments frequently use measures like visit duration, visit frequency, or travel distance to firms to gauge traffic ([Klise et al., 2021](#)). These measurements offer helpful traffic indications, but they do not adequately depict the interactions that happen there ([Klise et al., 2021](#)). It is crucial to take individual clustering and concurrent visits to certain venues into consideration in order to comprehend the potential of disease transmission within a community.

In the early twenty-first century, a number of extensions of epidemiological models were developed after the development of the traditional epidemiological SIR model. The SIS model was created by researchers ([Parshani, Carmi & Havlin, 2010](#)) to simulate epidemic diseases without immunity by eradicating the "recovered" population and leaving the diseased population merely susceptible. The "E" (exposed) abbreviation stands for the incubation period, which is the time when a person is infected but not yet contagious, according to the SEIR model by Prem *et al.* ([2020](#)). Due of its relevance to COVID-19, the SEIR model is employed in this work to simulate the pandemic at the individual level. Other epidemic models can easily be adapted to this simulation model.

Numerous researches examining the connection between human movement and the transmission of infectious illnesses have been published since the COVID-19 epidemic, which is exceedingly contagious, even while the illness is incubating. For instance, to simulate the spread of COVID-19, the study by Kraemer *et al.* ([2020](#)) used real-time human movement data from Baidu Inc. According to the research, isolation and other local control measures are more successful in stopping the spread of illness than travel restrictions that apply just locally. Similar to this, Wang *et al.* ([2020](#)) and Chiang *et al.* ([2022](#)) used Google movement data to examine the connection between COVID-19 spread and human movement. To further forecast the spread of diseases among diverse ethnic and socioeconomic groups, Chang *et al.* ([2021](#)) developed a meta-population SEIR model.

This study was motivated by recent studies on the use of visual analytics in the modelling and control of epidemics. For instance, Guo (2007) suggests a visual analytic tool that combines extraordinarily huge geographical interaction data in order to find patterns that aid decision-making during an epidemic. Using an interactive interface, researchers in Afzal *et al.* (2020) and Afzal *et al.* (2011) compare and assess the effectiveness of various control regimes. Additionally, Dunne *et al.* (2015) illustrate how illness spreads in response to demographic factors, such as population density, allowing users to see both long- and short-term patterns of epidemic spread. In contrast to other studies, this one identifies high-risk students based on the entire campus student trajectory and offers a more detailed epidemic simulation at the individual level.

## Mobility Patterns Modelling

In this study, on-campus students' mobility and network data, including their latitude and longitude coordinates, visited sites, and close proximity (within a 1-metre range) between two mobility data devices over a 14-day period, were analysed. The researchers also used computer modelling to look at movement patterns and determine how likely it was that COVID-19 would spread.

### Direct contact model

In this work, a direct contact paradigm for COVID-19 was utilised from Ghayvat *et al.* (2021). Based on Ghayvat *et al.* (2021), a mobility matrix  $Mob^{mobility}$  was created using data from a worldwide mobility network, with the elements of the matrix being represented as  $M_{p,q}^{mobility}$ . The COVID-19 subject's daily movements are shown by the  $M_{p,q}^{mobility}$ , as well as the number of people who passed the 1-m barrier and drew nearer to the COVID-19 subject. Additionally, it shows the COVID-19 subject's movements during relocation at a specified site  $c$  at a certain time  $t$  and day  $d$ , as well as during the incubation period (maximum  $d$  for COVID-19 is set to 15 days) (Gupta *et al.*, 2020). Equation (1) may be used to compute the direct contact  $DC(t, d)$  (Ghayvat *et al.*, 2021).

$$DC_{(t,d)} = \frac{[\sum_{d=1}^{15} (C19_p^{mobility} \times \sum_{n=1}^N (NS_{n,p}^{mobility}))] \times A_{area}}{D_{C19p,NS_{n,p}}} \quad (1)$$

where

$C19_{p,q}^{mobility}$  = COVID-19 Mobility at site p;

$NS_n$  = A set of healthy people who travelled close to a COVID-19 subject for a predetermined minute at site p;

$n = 1, 2, 3, \dots, N$ ;  $N$  = number of people;

$d$  = day, 1, 2, 3, ..., 15;

$t$ = when the COVID-19 subject and a nearby, healthy person is close to the established threshold limit of 1 m;

$n$ = total number of people who had contact with the COVID-19 subject;

$A_{area}$  = the COVID-19 subject's average mobility and the total number of nearby healthy people depending on the radius  $r_{p,q}$  at site  $p$ ;

$D_{iC19p,NSn,p}$  = distance between a subject with COVID-19 and nearby healthy people.

The above listed criteria are subject to change with time, space, and place. The proposed methodology (Ghayvat *et al.*, 2021) makes use of easily accessible mobility network data to analyze instances of direct interaction with COVID-19 using a probabilistic technique. This method is useful for tracking contact instances in a particular area at a certain time and day. However, by taking into consideration those who have been close to a COVID-19 patient for a certain period of minutes or more, the COVID-19 Pandemic Direct Contact Model can be improved. In order to do this, an Indirect Contact Model from Ghayvat *et al.* (2021) was also integrated, allowing for the identification of those who may have been exposed to COVID-19 through indirect contact.

## Indirect contact model

A probabilistic COVID19 pandemic indirect contact model is used (Ghayvat *et al.*, 2021) to identify people who have indirectly come into touch with COVID-19, in order to analyse the dynamics of indirect COVID-19 transmission. In order to do this, Ghayvat *et al.* (2021) created an indirect mobility matrix,  $iM^{mobility}$ , with  $M_{p,q}^{mobility}$  as its constituent, utilising information from the global mobility network.  $M_{p,q}^{mobility}$  displays the routine behaviours of healthy persons who have spent longer than the predetermined number of minutes around COVID-19 exposed individuals (Dunne *et al.*, 2015). Additionally, it takes into account their relocation moves between the regions  $p$  and  $q$ , as documented during a 15-day period beginning at time  $t$ , day  $d$ , and the  $d_{final\ day}-1$ . Using Equation (2), the indirect infection suspicion  $ID_{(t,d)}$  may be determined (Ghayvat *et al.*, 2021).

$$ID_{(t,d)} = \frac{[\sum_{d=1}^{15}(iC19_p^{mobility} \times \sum_{n=1}^N(iNS_{n,p}^{mobility}))] \times A_{area}}{D_{iC19p,NSn,p}} \quad (2)$$

where

$iC19_{p,q}^{mobility}$  = suspected COVID19 subject Mobility at location  $p$ ;

$iNS_n$  = a list of nearby healthy people that have travelled near a probable COVID-19 subject for predefined duration of minutes at location  $p$ ;

$D_{iC19p,iNSn,p}$  = distance between suspected COVID-19 subject and the neighbouring healthy individuals.

The Haversine formula, which uses the latitude and longitude information of the positions of two persons to get the distance between them in kilometres, is used to calculate the distance between two people,  $(D_{C19_p,NS_{n,p}})$  and  $(D_{iC19_p,iNS_{n,p}})$ , at a certain time. The formula can be expressed as follows:

$$a = \sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat1) \times \cos(lat2) \times \sin^2\left(\frac{\Delta long}{2}\right) \quad (3)$$

$$c = 2 \times \arctan2(\sqrt{a}, \sqrt{1-a}) \quad (4)$$

$$d = R \cdot c \quad (5)$$

Distance in km based on Pythagoras' theorem is given by the equations below:

$$x = (long2 - long1) \times \cos\left(\frac{lat1+lat2}{2}\right) \quad (6)$$

$$y = lat2 - lat1 \quad (7)$$

$$d = R(\sqrt{x \times x + y \times y}) \quad (8)$$

where

$R$ = Earth's radius (6371 km)

$lat1, long1$ = latitude, longitude pair of 1st point

$lat2, long2$ = latitude, longitude pair of 2nd point

$\Delta lat$ = difference between two latitudes

$\Delta long$ = difference between two longitudes

$c$ = Axis interaction calculation

$d$ = distance in kilometres

## Implementation

### Mobility data

Since infectious illnesses are spread by direct contact between people, epidemiologists can anticipate disease outbreaks by understanding the patterns of human movement. In order to respond to and recover from such crisis occurrences, it is essential to comprehend how individuals move through time and geography, which may be done with the use of mobile phone location data. Location history is one of the essential elements that characterizes these mobilities. Different techniques ([Khalel, 2010](#)) can be used to determine a person's location: cell site location method, GPS and Wi-Fi positioning.

To precisely pinpoint the user's location, Google Maps makes use of all three of these technologies. Users can use the Google Takeout service, which offers the data in JSON (JavaScript Object Notation) format to export this information. There are many different types of metadata in the exported JSON file. Some of the most important data points are outlined,

which include heading, activity type, latitudeE7 (latitude multiplied by  $1 \times 10^7$ ), longitudeE7 (longitude multiplied by  $1 \times 10^7$ ), accuracy, timestampMs (timestamp in milliseconds) and altitude.

Along with place searches and route finding, Google Maps offers a variety of features. Through GPS location data, Google Timeline, a feature of Google Maps, continuously monitors user activity, including the locations visited and the mode of transportation utilized to get there. This unprocessed information gives the user's position at a certain time and date.

The location history information from the Android smart phones of 50 volunteer university students was examined in this study. Due to the pilot study's focus on the possible spread of COVID-19 among university students, the location data was gathered for 14 days, and only students who were staying on campus were included. The Google History Location Extractor and Indicator was used to examine the location records to spot possible COVID-19 spread. In order to check for potential COVID-19 dissemination scenarios, the study involved charting the positions of the students on a map with comprehensive data of the date, time, area, and distance between people. The Venn diagram used to examine the location data of two students is shown in Figure 1.

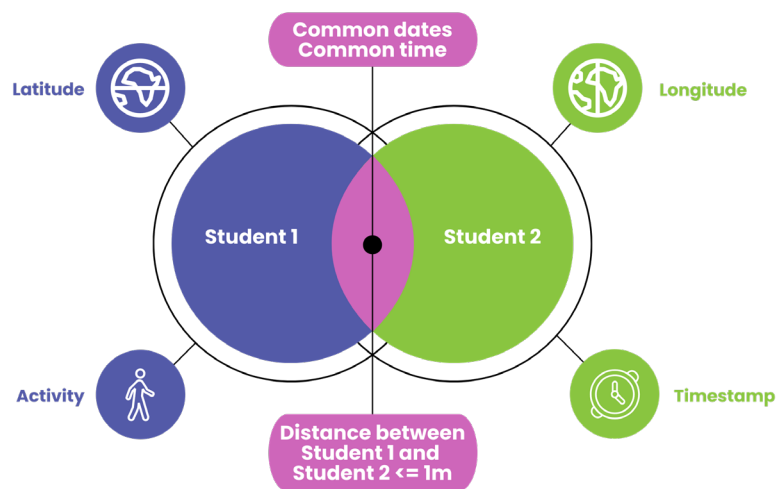


Figure 1. Venn diagram of location analysis for two-students

## Software implementation

Using software to extract and visualize the downloaded data, this study examined the location history information of 50 students' JSON files. As shown in Figure 2, the program, dubbed "Google History Location Extractor", has a number of functions, such as data extraction, graphical user interface presentation, data storing in CSV or KML format, and mobility visualization on Google Maps. Additionally, for more in-depth study, the data may be exported to a Microsoft SQL Server database format. Figure 3 shows the flowchart for the entire



software program, which consists of data collection, import, filtering, distance computation, and visualization.

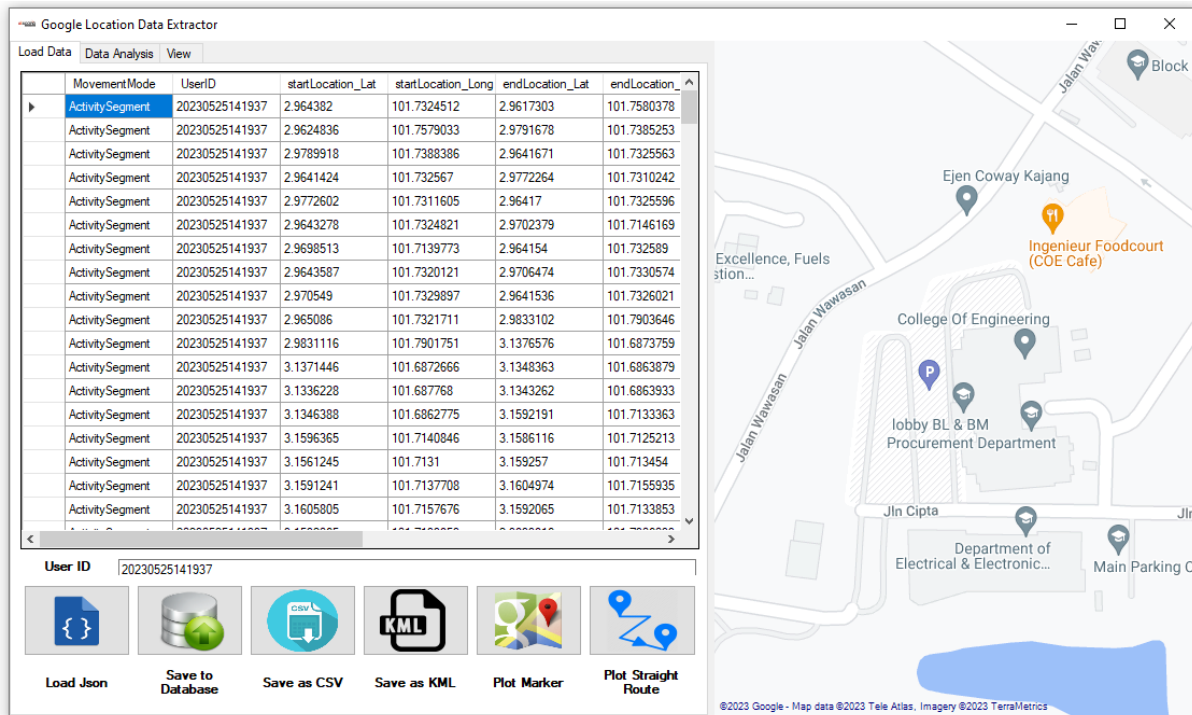


Figure 2. User interface of Google Historical Location Extractor software

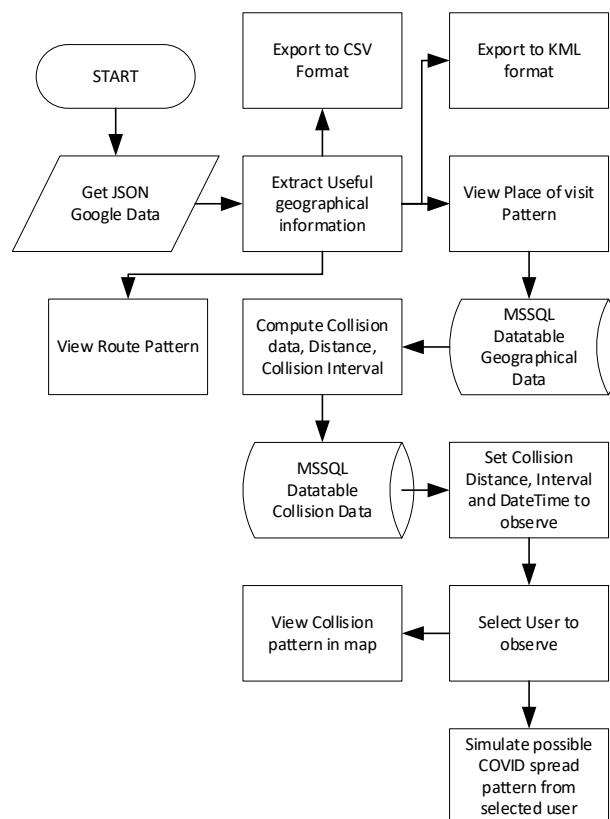


Figure 3. Entire system architecture

The process of this study started with 50 students submitting their Android smartphone location history during a 2-week period, which served as the source of the data for this



research. The location history data was obtained in JSON format and converted to table format using the import technique in VB.net. Data was filtered according to accuracy. The common location sites were identified using the distance parameter, keeping only spots that are closer than one metre. Filtering speeds up the analytical process by removing irrelevant data. Lastly, a map is used to represent the findings, because it is an efficient way to convey the specifics and conclusions of the investigation. A visualization of the closeness between two students within a given distance is made using Google's Direction API.

Identifying hotspots through the use of mobility tracking data serves as a crucial tool in monitoring and mitigating the spread of contagious diseases among volunteers. While it is important to implement comprehensive infection monitoring strategies, hotspot identification plays a pivotal role in proactive surveillance and prevention.

Hotspot identification provides valuable insights into areas where the risk of disease transmission is heightened. By analyzing the density of mobility tracking data points, we can identify regions with a higher concentration of volunteers and potential interactions. This information serves as an early warning system, enabling researchers and health authorities to allocate resources effectively and implement targeted interventions to prevent future infections.

The identification of hotspots not only facilitates prompt responses but also aids in understanding the dynamics of disease transmission. By studying the characteristics of these hotspots, such as population density, demographic information, and environmental factors, we can gain valuable insights into the underlying mechanisms of disease spread. This knowledge can inform future prevention and control strategies, contributing to more effective infection monitoring and mitigation efforts.

Furthermore, hotspot identification provides a foundation for implementing complementary infection monitoring strategies. By focusing resources on these high-risk areas, researchers can prioritize regular testing protocols and enhance contact tracing efforts. This targeted approach ensures that potential infections are detected and contained at an early stage, minimizing the risk of further transmission within the volunteer group.

It is important to emphasize that hotspot identification should not be viewed as a standalone measure, but rather as an integral part of a comprehensive infection monitoring framework. By integrating hotspot identification with other strategies, such as regular testing, contact tracing, and preventive measures, researchers can establish a robust system that actively monitors and mitigates future infections. Hence, the identification of hotspots plays a crucial role in monitoring the spread of contagious diseases among volunteers. By leveraging mobility tracking data, we can proactively identify areas of heightened risk, enabling targeted

interventions and resource allocation. Hotspot identification should be integrated with comprehensive infection monitoring strategies to enhance surveillance capabilities and facilitate effective prevention and control efforts.

## Results

In this section, the SEIR model's deterministic compartment model for the spread of an infectious disease, which serves as the foundation for the investigation of infection risk and is further explained in the subsections that follow, will be used to develop the individual student detection method.

### SEIR model

Depending on where a population is in the disease's course, the SEIR model (Kantor, 2021) splits them into four different categories. The susceptible population is the initial category and is at danger of contracting the illness. Given that SARS-CoV-2 is a novel virus, it is presumed that everyone who has never had the disease is vulnerable. The exposed population, which has been in contact with the virus but is not yet infectious, is the second category. The infected population, which has caught the virus and may spread it to others, makes up the third category. The population that has recovered from the illness and is no longer prone to the disease makes up the last group.

The following equations show how the compartment model works.

$$\frac{ds}{dt} = -\beta si \quad (9)$$

$$\frac{de}{dt} = \beta si - \alpha e \quad (10)$$

$$\frac{di}{dt} = \alpha e - \gamma i \quad (11)$$

$$\frac{dr}{dt} = \gamma i \quad (12)$$

where  $s + e + i + r = 1$  is an invariant.

The rate processes are modelled as follows:

$\beta$  is the model parameter, with units of 1/day.

$s$  is the size of the susceptible population.

$i$  is the size of the infective population.

$\alpha e$  is the rate at which the exposed population becomes infective, where  $e$  is the size of the exposed population. The average period of time in the exposed state is the incubation period of the disease, and equal to  $\frac{1}{\alpha}$ .

$\gamma i$  is the rate at which the infected population recovers and becomes resistant to further infection ( $r$  is the size of the recovered, resistant population).  $i$  is the size of the infective population. The average period of infectious state is  $\frac{1}{\gamma}$ .

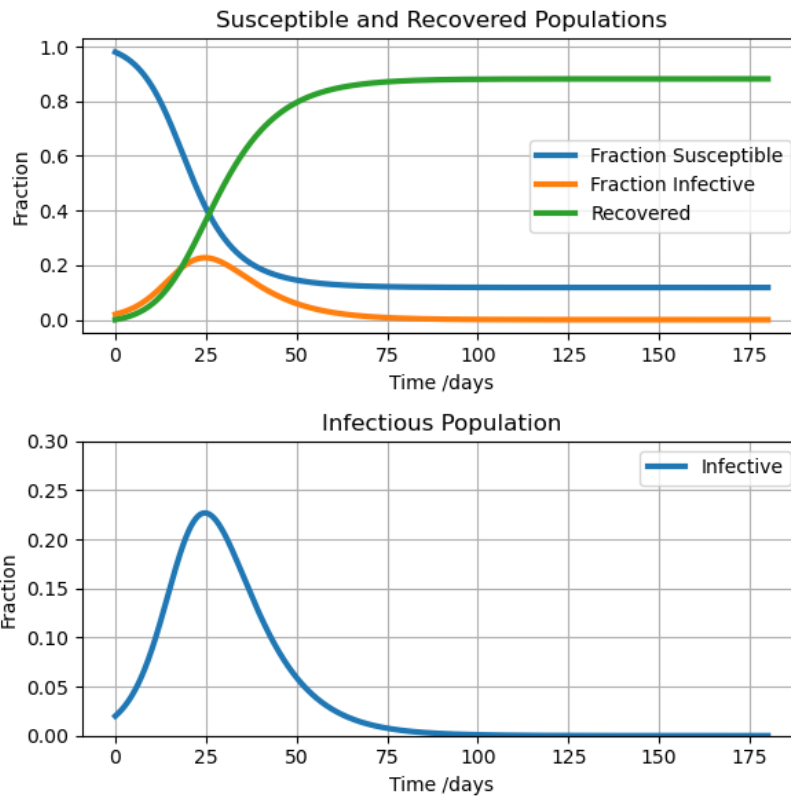


Figure 4. Simulation of the SIR model

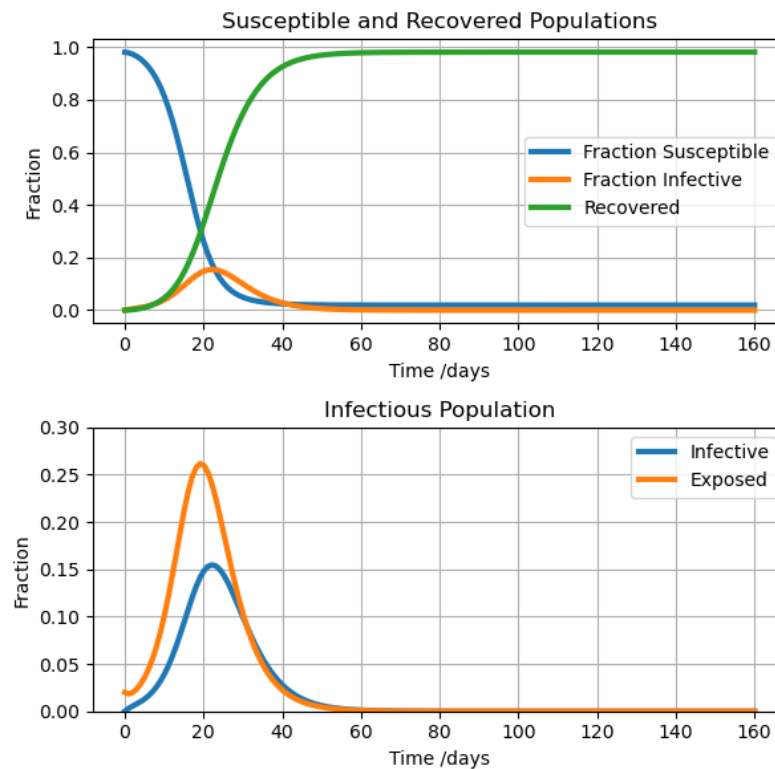
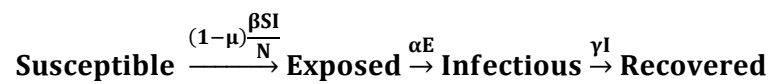


Figure 5. Simulation of the SEIR model

Figure 5 depicts the introduction of an exposed population compartment, which appears to delay the spread of the illness but does not appear to reduce the number of afflicted students, in contrast to Figure 4, which is depicted without the exposed population compartment. The number of students is set at 50.

The purpose of campus regulations is to reduce the spread of the virus by social isolation techniques that work to stop infected people from infecting vulnerable others. A control parameter called  $\mu$  is included to account for the efficacy of these measures in modelling: a value of 0 indicates no control; and a value of 1 indicates complete isolation of infectious persons. The goal of the model is to look at how social isolation policies may affect how an epidemic turns out.

The following diagram shows how the compartment model works (Kantor, 2021):



The rate processes are modelled as follows:

The number  $(1 - \mu) \frac{\beta SI}{N}$  represents the frequency with which an infected population transmits a disease to a vulnerable population  $\mu$ , where  $\mu = 0$  denotes the lack of effective interventions and  $\mu = 1$  denotes the total cessation of disease transmission. It describes the efficacy of public health measures in preventing the spread of illness.

This produces a system of four equations (Kantor, 2021) after substitution:

$$\frac{ds}{dt} = -(1 - \mu)\beta si \tag{13}$$

$$\frac{de}{dt} = (1 - \mu)\beta si - \alpha e \tag{14}$$

$$\frac{di}{dt} = \alpha e - \gamma i \tag{15}$$

$$\frac{dr}{dt} = \gamma i \tag{16}$$

where  $s + e + i + r = 1$  is an invariant.

Figure 6 shows how campus mobility restriction policies have slowed the spread of epidemics, decreased the percentage of infected students, reduced the demand on medical resources, decreased the number of students who eventually contract the disease, and saved lives for illnesses with non-zero mortality.

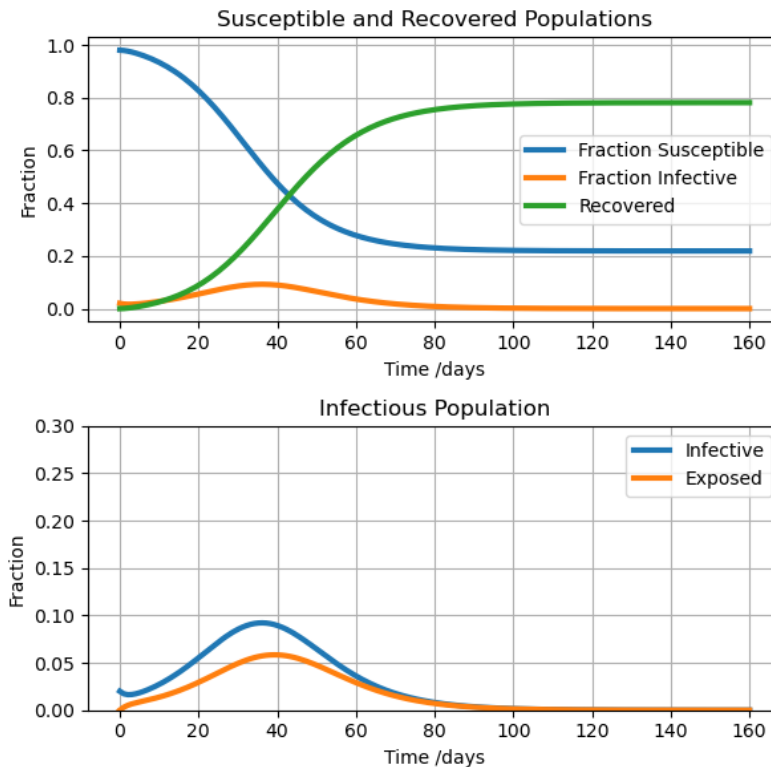


Figure 6. Simulation of the SEIR model with Control Parameter,  $\mu$

## Infection risk exploration

The purpose of this project was to use data analysis to create a graphical user interface for identifying prospective encounters amongst students on campus based on their travel patterns. This was accomplished using a tracing program that classifies contacts into various danger levels according on the order of their encounters.

Using Google Mobility Data and Google Direction API, which show different combinations of latitude and longitude on a single map, Figure 7 shows the location and routes followed by two students. The whereabouts of two different students are mapped over the same days, showing that they often visited the same location. Common locations for many students are shown in Figure 8.

The proximity analysis parameters are shown in the Data Analysis of Figure 9, and the outcomes are shown in the resulting table at the bottom of the Figure. Date, time, latitude, longitude, visited place, and levels of communication are among the information in the table. In the table, Level 1 denotes direct touch; Level 2 and Level 3 denote indirect contact from Level 1 and Level 2, respectively. With the help of the settings, users may enter Start Date, Start Time, Interval (minute), Collision Distance (metre), Collision Interval (seconds), and Collision User (the user who will be chosen for proximity analysis).

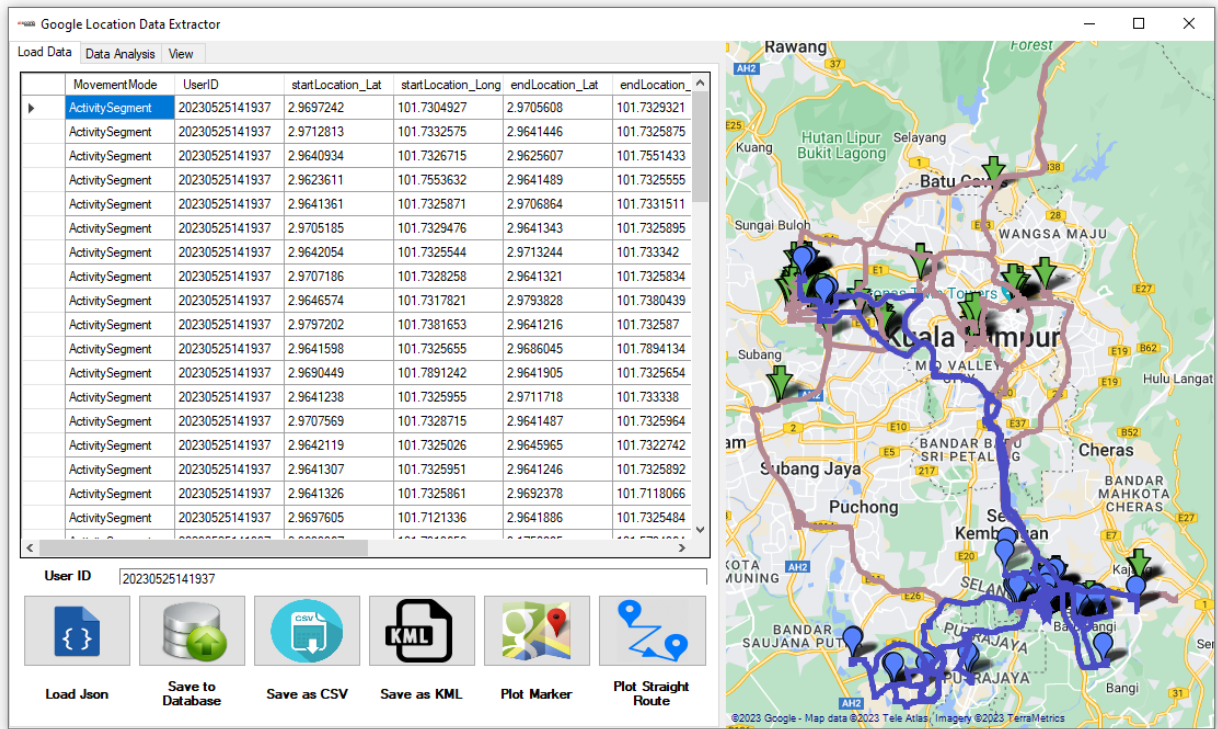


Figure 7. Location history of students displayed on the same map

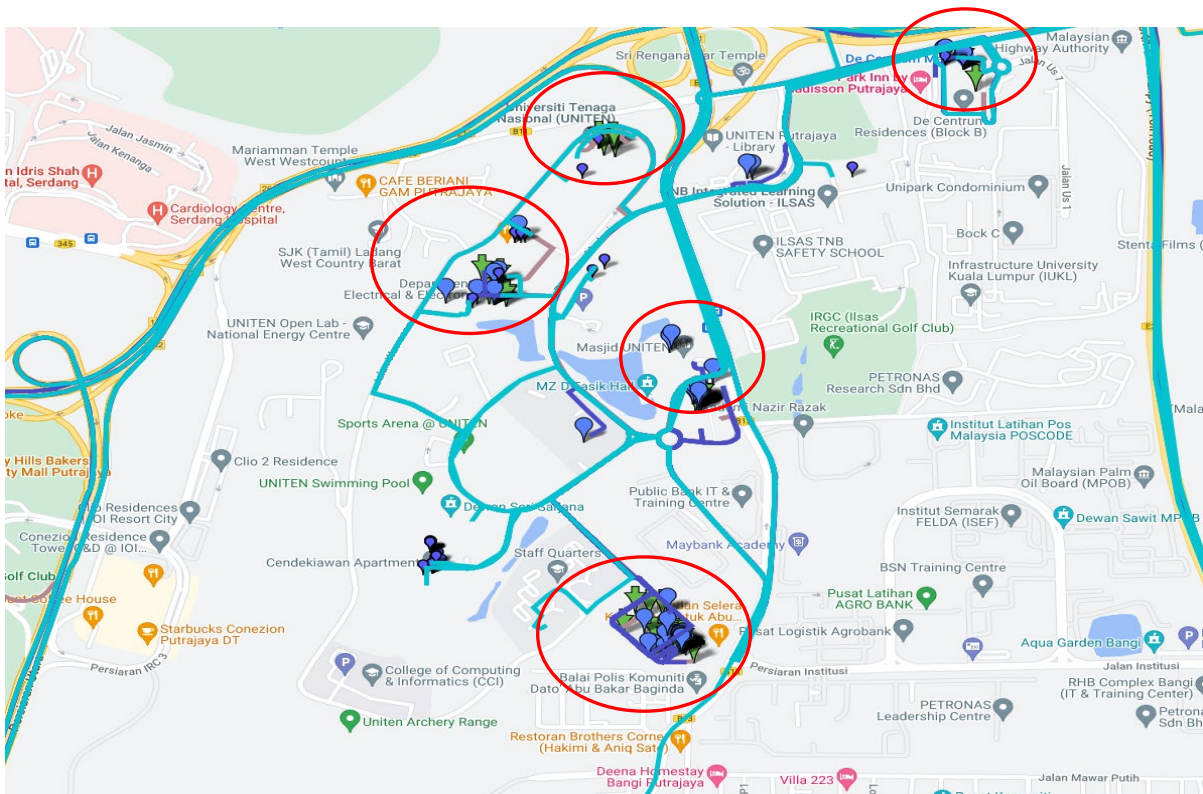


Figure 8. Common student locations on campus

Multiple vulnerable people (level 1 and 2) surround an infected person (chosen under the Collision User option) in Figure 9. Based on their level of contact with the infected user, these vulnerable people are divided into three tiers. Due to their close contact with the infected user, primary contacts (level 1) are at the highest risk of infection. Primary contacts have brought



secondary and tertiary contacts (levels 2 and 3, respectively) into indirect touch with the infected user.

In order to prioritize quarantine and contact screening, particularly when resources are few, a precise classification of contacts may be highly helpful, given the exponential development in the number of connections with each level. The screening of primary contacts must be prioritized, and a full screening must be carried out as quickly as feasible. Additionally, because secondary and tertiary connections also pose a danger, screening for them should come right after screening for main contacts.

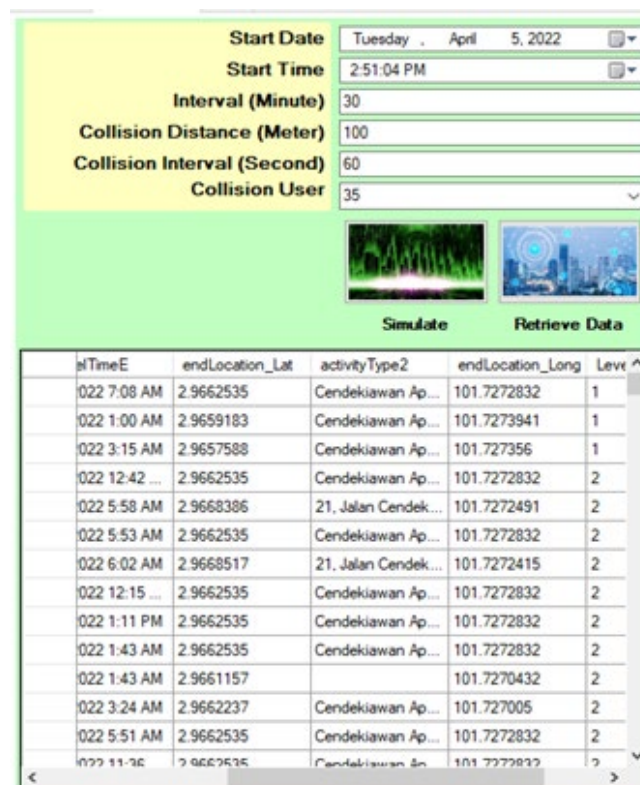


Figure 9. Graphical user interface settings for proximity analysis

## Discussion

The purpose of this study was to learn more about the intricate relationships between students and how those relationships can affect how quickly a virus spreads. We identified probable transmission sites and the linkages between hotspots by examining the mobility networks of university students. A thorough description of the transmission network was produced as a consequence of the research, which made use of the data that was readily accessible regarding residency and involvement in on- and off-campus activities. The created visualization tool showed cross-paths that the system had found, indicating linkages that suggested recent touch. In order to manage the outbreak more effectively and stop further transmission within the university community, visualization and information from these links and networks could



support the implementation of targeted mitigation activities, like isolating cases and quarantining contacts.

There are at least three restrictions on this work. First, certain data were missing or unknown, due to inadequate mobility data, and these data were not included in the network analysis. Second, meetings outside could not accurately represent transmission histories. There is a need for adherence to advised COVID-19 mitigation techniques, such as reducing social gathering sizes, social distance, wearing masks, improving hand cleanliness, and enhancing testing, to avoid the fast spread of COVID-19 in on- and off-campus university settings. The third drawback of this paper is the lack of data gathered from uninfected people or on adherence to mitigation techniques, such social isolation, mask usage, and hand cleanliness. On university and college campuses, promoting virtual activities, such as those connected to sizable student meetings, may assist to reduce the danger of transmission. Collaboration between university administrators, student organizations, and health authorities is essential to ensure that COVID-19 mitigation measures are followed.

## Conclusion

This work proposes a data analysis approach that makes use of a GUI (graphical user interface) to anticipate social distance violations among students on university and college campuses and to detect common movement patterns. Important characteristics, like longitude, latitude, distance, and the time and date of contact within a specified range, are included in the GUI that is used to display analyzed information on a map. A student's past locations may be tracked and compared to those of others, making it possible to identify contacts who could be targeted for screening or quarantine of high-risk persons.

The GUI platform also incorporates region mining, which enables quick calculation of possible dangers and early detection of suspected diseases. These features give users, like the Centers for Disease Control and Prevention, an intuitive user interface to model potential outcomes and arrive at judgements.

The GUI has a number of benefits, such as the capacity to analyze massive quantities of student data concurrently, the ability to determine whether a student has visited a high-risk or polluted region, and its independence from extra hardware for analysis. However, it should be noted that using location services might reduce battery life of handsets and necessitate continual Internet access.

Overall, this data analysis technique has the potential to offer insightful information for monitoring and suppressing the spread of COVID-19 on university and college campuses when

paired with the GUI. The viability and efficacy of utilizing this technology as a component of an all-encompassing mitigation plan require more study.

## Acknowledgements

This work is supported by IDRC Grant No. 109586 – 001 for Artificial Intelligence Framework for Threat Assessment and Containment for COVID-19 and Future Epidemics while Mitigating the Socioeconomic Impact to Women, Children and Underprivileged Groups.

A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia on 26–28 July 2023.

## References

- Afzal, S., Ghani, S., Jenkins-Smith, H. C., Ebert, D. S., Hadwiger, M., & Hoteit, I. (2020). A Visual Analytics Based Decision Making Environment for COVID-19 Modeling and Visualization. 2020 IEEE Visualization Conference (VIS), Salt Lake City, UT, USA, pp. 86–90. <https://doi.org/10.1109/VIS47514.2020.00024>
- Afzal, S., Maciejewski, R., & Ebert, D. S. (2011). Visual analytics decision support environment for epidemic modeling and response evaluation. 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, RI, USA, pp. 191–200. <https://doi.org/10.1109/VAST.2011.6102457>
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589, 82–87. <https://doi.org/10.1038/s41586-020-2923-3>
- Chiang, W. H., Liu, X., & Mohler, G. (2022). Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *International Journal of Forecasting*, 38(2), 505–520. <https://doi.org/10.1016/j.ijforecast.2021.07.001>
- Dave, D., McNichols, D., & Sabia, J. J. (2021). The Contagion Externality of a Superspreading Event: The Sturgis Motorcycle Rally and COVID-19. *Southern Economic Journal*, 87(3), 769–807. <https://doi.org/10.1002/soej.12475>
- Dunne, C., Muller, M. J., Perra, N., & Martino, M. (2015). VoroGraph: Visualization Tools for Epidemic Analysis. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 255–258. <https://doi.org/10.1145/2702613.2725459>
- Guo, D. (2007). Visual analytics of spatial interaction patterns for pandemic decision support. *International Journal of Geographical Information Science*, 21(8), 859–877, <https://doi.org/10.1080/13658810701349037>
- Khalel, A. M. H. (2010). Position Location Techniques in Wireless Communication Systems (Dissertation). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:bth-4796>
- Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., and Andrea Rinaldo. (2020). Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency

- containment measures. *Proceedings of the National Academy of Sciences, USA*, 117(19), 10484–10491. <https://doi.org/10.1073/pnas.2004978117>
- Ghayvat, H., Awais, M., Gope, P., Pandya, S., & Majumdar, S. (2021). Recognizing Suspect and Predicting the Spread of Contagion Based on Mobile Phone Location Data (COUNTERACT): A system of identifying COVID-19 infectious and hazardous sites, detecting disease outbreaks based on the internet of things, edge computing, and artificial intelligence. *Sustainable Cities and Society*, 69, 102798. <https://doi.org/10.1016/j.scs.2021.102798>
- Google. (2020). Google COVID-19 Community Mobility Reports, 2020. Available from <https://www.google.com/covid19/mobility>
- Gupta, R., Bedi, M., Goyal, P., Wadhwa, S., & Verma, V. (2020). Analysis of COVID-19 tracking tool in India: Case study of *Aarogya Setu* mobile application. *Digital Government: Research and Practice*, 1(4), 28. <https://doi.org/10.1145/3416088>
- Kantor, J. (2021). 3.9 modeling and control of a campus outbreak of Coronavirus Covid-19. Retrieved March 13, 2023, from <https://jckantor.github.io/CBE30338/03.09-COVID-19.html>
- Kiang, M. V., Santillana, M., Chen, J. T., Onnela, J.-P., Krieger, N., Engø-Monsen, K., Ekapirat, N., Areechokchai, D., Prempee, P., Maude, R. J., & Buckee, C. O. (2021). Incorporating human mobility data improves forecasts of Dengue fever in Thailand. *Scientific Reports*, 11, 923. <https://doi.org/10.1038/s41598-020-79438-0>
- Klise, K., Beyeler, W., Finley, P., & Makvandi, M. (2021). Analysis of mobility data to build contact networks for COVID-19. *PLoS One*, 16(4), e0249726. <https://doi.org/10.1371/journal.pone.0249726>
- Kraemer, M. U. G., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., Open COVID-19 Data Working Group, du Plessis, L., Faria, N. R., Li, R., Hanage, W. P., Brownstein, J. S., Layan, M., Vespignani, A., Tian, H., Dye, C., Pybus, O. G., & Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490), 493–497. <https://doi.org/10.1126/science.abb4218>
- Kumar, N., Oke, J. & Nahmias-Biran, Bh. (2021). Activity-based epidemic propagation and contact network scaling in auto-dependent metropolitan areas. *Scientific Reports*, 11, 22665. <https://doi.org/10.1038/s41598-021-01522-w>
- Lasry, A., Kidder, D., Hast, M., Poovey, J., Sunshine, G., Winglee, K., *et al.* (2020). Timing of community mitigation and changes in reported COVID-19 and community mobility—four US metropolitan areas, February 26–April 1, 2020. *Morbidity and Mortality Weekly Report*, 69(15), 451–457. <https://doi.org/10.15585/mmwr.mm6915e2>
- Maheshwari, P., & Albert, R. (2020). Network model and analysis of the spread of Covid-19 with social distancing. *Applied Network Science*, 5, 100. <https://doi.org/10.1007/s41109-020-00344-5>
- McMinn, S., & Talbot, R. (2020). Mobile Phone Data Show More Americans Are Leaving Their Homes, Despite Orders. NPR, The Coronavirus Crisis. Available from <https://www.npr.org/2020/05/01/849161820/mobile-phone-data-showmore-americans-are-leaving-their-homes-despite-orders>

- Muller, S. A., Balmer, M., Neumann, A., & Nagel, K. (2020). Mobility traces and spreading of COVID-19. medRxiv preprint. <https://doi.org/10.1101/2020.03.27.20045302>
- Pan, Y., Darzi, A., Kabiri, A., Zhao, G., Luo, W., Xiong, C., Zhang, L. (2020). Quantifying human mobility behaviour changes during the COVID-19 outbreak in the United States. *Scientific Reports*, 10(1):1–9. <https://doi.org/10.1038/s41598-020-77751-2>
- Parshani, R., Carmi, S., & Havlin, S. (2010). Epidemic threshold for the susceptible-infectious-susceptible model on random networks. *Physical Review Letters*, 104(25), 258701. <https://doi.org/10.1103/PhysRevLett.104.258701>
- Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. E., Davies, N., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Jit, M., & Klepac, P. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *Lancet Public Health*, 5(5), E261–E270. [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6)
- Rechtin, M., Feldman, V., Klare, S., Riddle, N., & Sharma, R. (2020). Modeling and Simulation of COVID-19 Pandemic for Cincinnati Tri-State Area. arXiv preprint: 200606021. <https://doi.org/10.48550/arXiv.2006.06021>
- Ruktanonchai, N. W., Ruktanonchai, C. W., Floyd, J. R. & Tatem, A. J. (2018). Using Google Location History data to quantify fine-scale human mobility. *International Journal of Health Geographics*, 17, 28. <https://doi.org/10.1186/s12942-018-0150-z>
- Schlosser, F., Maier, B. F., Jack, O., Hinrichs, D., Zachariae, A., & Brockmann, D. (2020). COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences (USA)*, 117(52), 32883–32890. <https://doi.org/10.1073/pnas.2012326117>
- Soures, N., Chambers, D., Carmichael, Z., Daram, A., Shah, D. P., Clark, K., Potter, L., & Kudithipudi, D. (2020). SIRNet: Understanding social distancing measures with hybrid neural network model for COVID-19 infectious spread. arXiv preprint, 200410376. <https://doi.org/10.48550/arXiv.2004.10376>
- Venkatramanan, S., Sadilek, A., Fadikar, A., Barrett, C. L., Biggerstaff, M., Chen, J., Dotiwala, X., Eastham, P., Gipson, B., Higdon, D., Kucuktunc, O., Lieber, A., Lewis, B. L., Reynolds, Z., Vullikanti, A. K., Wang, L., & Marathe, M. (2021). Forecasting influenza activity using machine-learned mobility map. *Nature Communications*, 12, 726. <https://doi.org/10.1038/s41467-021-21018-5>
- Wang, Y., Wang, Y., Chen, Y., & Qin, Q. (2020). Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *Journal of Medical Virology*, 92(6), 568–576. <https://doi.org/10.1002/jmv.25748>
- Weill, J. A., Stigler, M., Deschenes, O., & Springborn, M. R. (2020). Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proceedings of the National Academy of Sciences*, 117(33), 19658–19660. <https://doi.org/10.1073/pnas.2009412117>
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan,

China: a modelling study. *The Lancet*, 395(10225), 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)

Yi, H., Ng, S. T., Farwin, A., Pei Ting Low, A., Chang, C. M., & Lim, J. (2021). Health equity considerations in COVID-19: geospatial network analysis of the COVID-19 outbreak in the migrant population in Singapore. *Journal of Travel Medicine*, 28(2), taaa159. <https://doi.org/10.1093/jtm/taaa159>