

# Modelling COVID-19 Hotspot Using Bipartite Network Approach

Boon Hao Hong <sup>1</sup> , Jane Labadin <sup>1</sup> , Wei King Tiong <sup>1</sup> , Terrin Lim <sup>1</sup> ,  
Melvin Hsien Liang Chung <sup>2</sup> 

<sup>1</sup> Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan 94300, Malaysia

<sup>2</sup> Bintulu Divisional Health Office, Bintulu, Sarawak, Malaysia

Corresponding author: Jane Labadin (ljane@unimas.my)

## Abstract

COVID-19 causes a jarring impact on the livelihoods of people in Malaysia and globally. To prevent an outbreak in the community, identifying the likely sources of infection (hotspots) of COVID-19 is important. The goal of this study is to formulate a bipartite network model of COVID-19 transmissions by incorporating patient mobility data to address the assumption on population homogeneity made in the conventional models and focus on indirect transmission. Two types of nodes – human and location – are the main concern in the research scenario. 21 location nodes and 31 human nodes are identified from a patient's pre-processed mobility data. The parameters used in this study for location node and human node quantifications are the ventilation rate of a location and the environmental properties of the location that affect the stability of the virus such as temperature and relative humidity. The summation rule is applied to quantify all nodes in the network and the link weight between the human node and the location node. The ranking of location and human nodes in this network is computed using a web search algorithm. This model is considered verified as the error obtained from the comparison made between the benchmark model and the COVID-19 bipartite network model is small. As a result, the higher ranking of the location is denoted as a hotspot in this study, and for a human node attached to this node will be ranked higher in the human node ranking. Consequently, the hotspot has a higher risk of transmission compared to other locations. These findings are proposed to provide a framework for public health authorities to identify the sources of infection and high-risk groups of people in the COVID-19 cases to control the transmission at the initial stage.

## Keywords

COVID-19; Hotspot; Contact tracing; Bipartite network; Location rank.

**Citation:** Hong, B. H., Labadin, J., Tiong, W. K., Lim, T., & Chung, M. H. L. (2021). Modelling COVID-19 Hotspot Using Bipartite Network Approach. *Acta Informatica Pragensia*, 10(2), 123–137. <https://doi.org/10.18267/j.aip.151>

**Special Issue Editor:** Chiew Kang Leng, Universiti Malaysia Sarawak, Malaysia

**Academic Editor:** Zdenek Smutny, Prague University of Economics and Business, Czech Republic

**Copyright:** © 2021 by the author(s). Licensee Prague University of Economics and Business, Czech Republic.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY 4.0).

# 1 Introduction

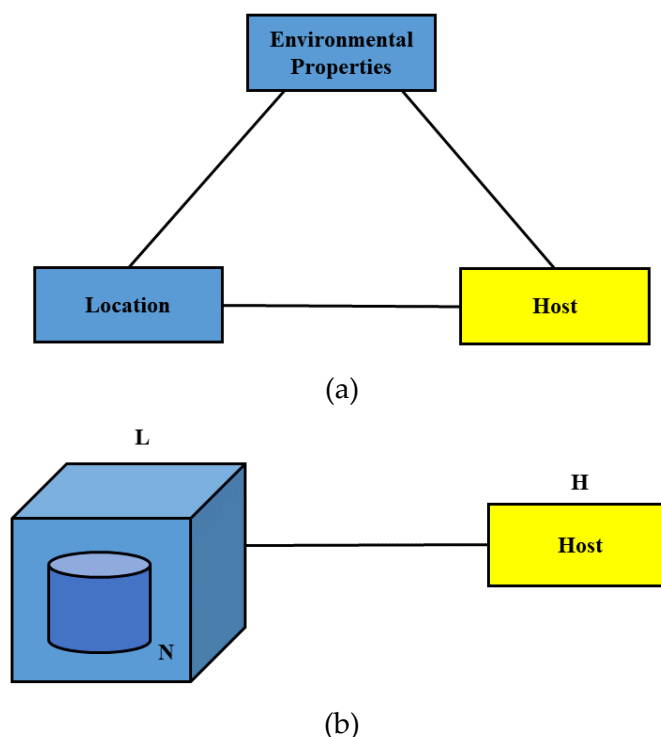
The COVID-19 pandemic that shocked the world in early January 2020 witnessed many countries rushing to characterise the nature of the SARS-COV-2 virus that causes the COVID-19 disease and its highly infectious transmission. Globally, we witnessed unprecedented use of digital technology and innovation in assisting in the detection and prevention of the disease. Major technological companies and numerous governments focused their attention on rapid deployment of digital applications and platforms that utilise digital data such as geolocation, tracker, mobile reception signal and Bluetooth data in efforts to model and predict COVID-19 transmissions (Whitelaw et al., 2020; Gasser et al., 2020). Identifying the sources of infection of COVID-19 is crucial for preventing community transmission and averting formation of new transmission clusters. In order to detect the source of infection, the locations visited by an infectious person as well as the environmental properties of the locations need to be taken into consideration. In this study, the source of infection is referred to as a hotspot, which is the location that has a higher risk of COVID-19 transmission. This definition corresponds to one of the descriptions discussed by Lessler et al. (2017), which is an area that has high efficiency of the disease transmission. A hotspot can be identified by statistical approaches, which are adopted by many studies associated with several factors. For instance, the daily number of new cases and incidence rate in a weekly unit are used to perform the time series forecasting and calculate the Moran index in Brazil for spatial analysis (Gomes et al., 2020). The Moran index is frequently applied to obtain the autocorrelation for spatial data; it has been used to study the relationship between the spatial effect and COVID-19 transmission in China (Huang et al., 2020). The temperature is used to determine the transmission rate in China at a macro level, by integrating a number of incidence and demographic factors (Xie et al., 2020).

Early studies of COVID-19 transmission focussed on the prediction of incidents (Kucharski et al., 2020; Shen et al., 2020; Wu et al., 2020). This is because COVID-19 is a newly emerging disease and the primary public health concern is to flatten the outbreak curve and reduce the burden of medical resources. Many of these studies used conventional approaches such as compartmental models, which assumed the population is well mixed (homogeneous) and thus providing astute simulation to decide on the action that public health authorities need to take. The main issue with the homogeneity assumptions made in the conventional approach is that it does not reflect the real world since people's mobility is influenced by age, social activity, occupation and other factors. To overcome this problem, partial differential equation (PDE) is used to integrate those factors in and combine them with population dynamics (Viguerie et al., 2020; Wang et al., 2020). The spatial factor considered in those studies determined the diffusion of the infection at the macro level, resulting in a computationally complex model.

The network model is widely used in infectious disease studies for location detection and contact tracing (Block et al., 2020; Kok et al., 2017; Meyers et al., 2006; Weeden & Cornwell, 2020). The network model is developed based on the network theory, which is a branch of graph theory that captures the heterogeneity of a real-world problem. Several studies have used the network model to simulate different scenarios of COVID-19. Block et al. (2020) use the network model to understand the effect of social distancing policy as control measures since the social interaction in different places is addressed by the model. Contact tracing is one of the important control measures adopted worldwide. The efficacy of contact tracing can be simulated using a network model whereby the nodes in the model act as hosts in the real world and the health status of each host will change over time after contact with other nodes in the network (Firth et al., 2020).

Interaction between different sub-populations is also described using the contact network model (Liu et al., 2020; Majumdar & Mehta, 2020; Weeden & Cornwell, 2020). Those models focus on the interaction between hosts in a specific area, but the influence of the location visited by persons is excluded. As previously stated, the host's movement is crucial for identifying the source of infection. To integrate the host's mobility in a network, the bipartite network model (BNM) is proposed for this study. The

relationship of the host, location and environment can be illustrated using an epidemiology triangle (Eze, 2013) as shown in Figure 1(a). The location component and environment properties are highly related and the environmental properties of each location are different from others; thus, the location component can act as a component consisting of environmental properties. Thus, the epidemiology triangle is modified into a basic building block depicted in Figure 1(b).



**Figure 1:** (a) Epidemiology triangle; (b) Basic building block, where *L* is the location, *N* is environmental properties of the location and *H* is the host.

As proposed by Eze et al. (2014), the two-type nodes shown in Figure 1 (b) form a bipartite network when they are quantified according to the problem considered. Eze et al. (2014) illustrated the method successfully in identifying the source of infection for malaria and Kok et al. (2017) applied a similar approach to dengue. Kok et al. (2017) fully integrated the dengue patient mobility data into the network to form the location node. In the location node, several environmental properties such as temperature, humidity and precipitation are used to quantify mosquitoes' life cycle and their survival rate at a particular location. In this study, the droplet produced by the COVID-19 patient carries the virus and is hypothesized to stay in the air and on the surfaces of nearby objects. The environmental properties will thus be quantified to estimate the stability of the virus in a location. As mentioned above, the environmental properties are highly related to the location; thus, the basic building block depicted in Figure 1(b) is the main structure to connect the host (human) and location in this study. Therefore, the BNM is suitable to model the problem concerned with the two discrete entities.

This study aims to predict a hotspot of COVID-19 by formulating a contact network that includes patient movement history two weeks prior to being tested positive. By identifying the high-risk location visited by the patient, the people who visited the location can be isolated beforehand to prevent the disease spreading to more people. Even though there is no case reported of indirect transmission of COVID-19 in Malaysia, there were several cases reported from other countries with the possibility of being infected in public places via fomite or aerosol transmission (Cai et al., 2020; Lauer et al., 2020; Miller et al., 2021). Those findings inspired this study to address indirect transmission in Malaysia by formulating a network to connect the host with the locations visited. These data are called patient mobility data and can be obtained via a mobile tracing application, such as the *MySejahtera* application used in Malaysia. Therefore,

the source of infection may be more accurately determined if there are more mobility data from other patients that can be included in the network. The output of this study is a form of proof-of-concept that the BNM can be employed to identify a hotspot of COVID-19 using a patient's past mobility data. The findings aim to help public health authorities in Malaysia to perform more effective contact tracing and reduce the community transmission cluster.

## 2 Materials and Methods

The methodology adopted in this research is the Bipartite Network-Based Methodology Framework (BNB-RMF) (Liew, 2016). This methodology was applied to model the habitat suitability of the Irrawaddy dolphin and dengue hotspot identification (Liew et al., 2015; Kok et al., 2017). There are a few stages necessary in this methodology: (i) pre-processing of the data used; (ii) formulation of the bipartite graph structure and network; (iii) quantification and ranking of the nodes.

### 2.1 Data pre-processing

The dataset used in this study is a real case contact tracing investigation form provided by the Sarawak Bintulu Division Health Office. Currently, contact tracing of COVID-19 is triggered once a person is tested positive; thus, the dataset provided is based on a case. The investigation form (MoH, 2020) consists of the location visited by the patient with visit time or period. The close contacts mentioned in the investigation form are defined as (i) household member of the COVID-19 patient; (ii) workplace colleague of the COVID-19 patient in close proximity; (iii) social contact with COVID-19 in close proximity; (iv) travelling with the COVID-19 patient in the same transportation. The contact tracing is to trace all the close contacts of the patient in 14 days before the point of diagnosis. All locations listed in the investigation form will be converted into Global Positioning System (GPS) coordinates and any duplicated locations will be removed. Each location will be coded with an Identity Document (ID) and will be referred to as a location node in this study. In order to generate an anonymized dataset, the names of close contacts and patients stated in the investigation form will be replaced with a unique ID to conceal their personal information.

A total of 21 location nodes and 31 human nodes (which include the case patient) are identified from the investigation form and used as the input to formulate the network. 79 data points are formed, where each person having visited a location is defined as a unique data point. Every data point consists of the GPS coordinates of the locations visited, average surface air temperature ( $T$ , in degrees Celsius,  $^{\circ}\text{C}$ ), average relative humidity ( $RH$ , in percentage, %), location type ( $L_{type}$ , in binary, outdoor, 0 or indoor, 1), ventilation rate of the location ( $Q$ , in per hour,  $h^{-1}$ ), the frequency of a person having visited the location ( $Fh$ ), the total duration of stay of one person in the location ( $D$ , in hours,  $h$ ), the frequency of one location visited by a patient ( $Fl$ ) and human status ( $H_s$ , in binary, 1 for patient or 0 for close contact). The environmental data used in this study, which are  $T$  and  $RH$ , are obtained from the World Weather Online platform. The GPS coordinates of each location node and the date and time of visit are used to retrieve the  $T$  and  $RH$  for each location node.

### 2.2 Modelling of bipartite network model

A hotspot is defined as a location that has a higher risk of COVID-19 transmission, and it is the location visited by an infectious person. To identify the hotspot, the parameters of human and location nodes are important so that the link weight can be quantified. A hotspot of COVID-19 in this study is determined by the ranking of the location according to its COVID-19 hotspot ranking value (CHR). Higher ranking indicates a higher risk of transmission. For the human node, the ranking value is termed the COVID-19 infection risk ranking (CRR) and used to predict the potential infectious person. The model formulation process is discussed below.

## 2.2.1 Formulation of bipartite graph structure

A bipartite COVID-19 contact (BCC) graph is formulated by employing the basic building block illustrated in Figure 1(b). The patient and the close contact retrieved from the investigation form are substituted into the host vertex in the figure and defined as the human node (H). The location vertex in the figure is replaced by the location visited by the patient, which is referred to as the location node (L). The edge connecting the human node and the location node in the network is denoted as a link (E). The E represents the interaction between the human and the location and the form when the H visited the L. The weight of the link implies the contact strength between the human and the location node. Higher contact strength indicates a stronger relationship between the human and the given location. The higher contact strength suggests that the location has a higher risk of spreading the virus and the person has a higher possibility to be infected from the location.

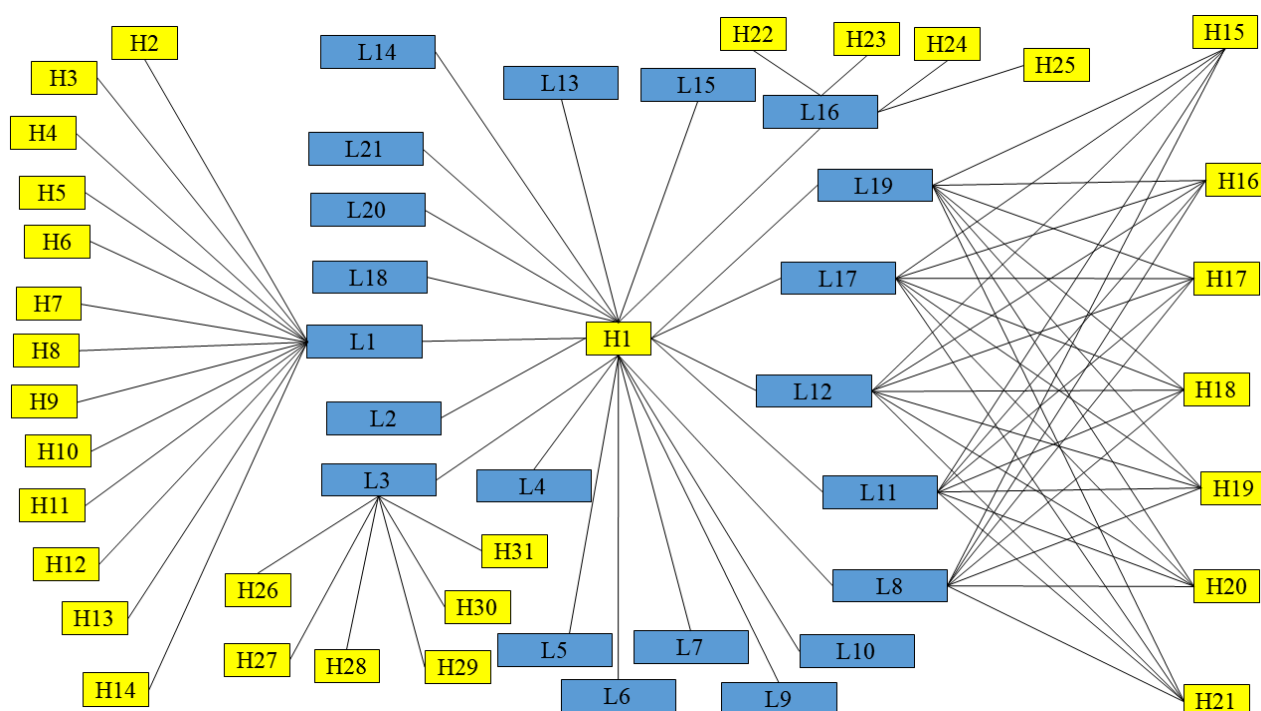
There are 30 close contacts identified from the investigation form; hence, a total of 31 human nodes are listed and labelled as shown in Equation (2). There are 21 locations visited by the patient and labelled accordingly as in Equation (3). The edge between the human node and location nodes are the 79 data points mentioned in Section 2.1 and defined accordingly as in Equation (4). Therefore, the BCC graph is formulated and defined as in Equation (1) and graphically depicted in Figure 2.

$$BCC = BCC(H, L, E) \text{ where} \quad (1)$$

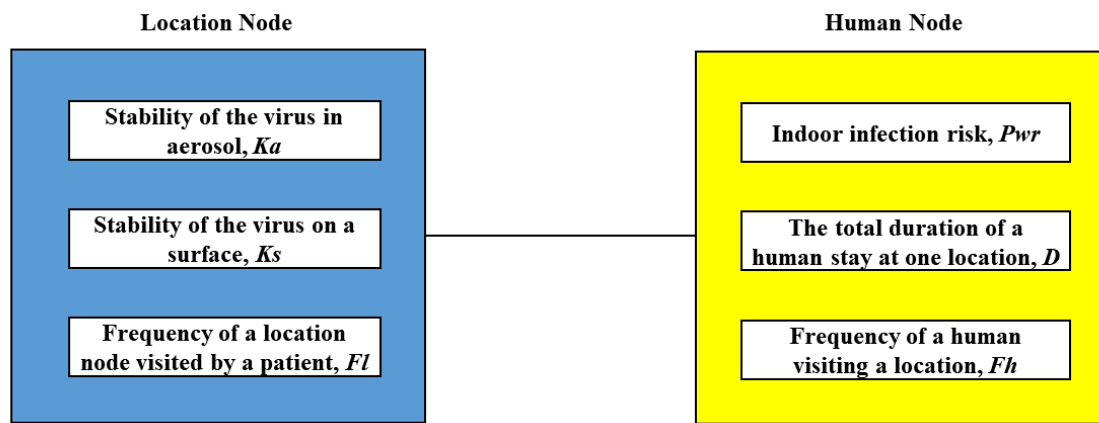
$$H = \{H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, \dots, H31\} \quad (2)$$

$$L = \{L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, \dots, L21\} \quad (3)$$

$$E = \{e1, e2, e3, e4, e5, e6, e7, e8, e9, e10, \dots, e79\} \quad (4)$$



**Figure 2:** Complete bipartite COVID-19 contact graph with human node (H) and location node (L).



**Figure 3:** Overall BCC network with all parameters incorporated into the model.

### 2.2.2 Quantification of parameters for location node

There are three parameters that are chosen to quantify the location node as shown in Figure 3. As mentioned earlier, the aim of the study is to incorporate the indirect transmission of the disease via aerosol. Thus, the first parameter derived to be used in the location node is the stability of the virus in aerosol ( $Ka$ ). Aerosols are generated when people talk, sneeze, and cough; hence, such actions done by an infectious person may result in the aerosols carrying the virus. Aerosol generated by COVID-19 patients can carry an active virus and remain viable for a certain period. Since the decay rate of the virus depends on the decay constant, the decay constant for the virus infectivity is used to quantify the stability of the virus in this study. The stability of the virus is affected by environmental properties such as temperature and humidity (Fears et al., 2020; Santarpia et al., 2020; van Doremalen et al., 2020). Hence, the relationship between the stability of the virus in aerosol and the environment needs to be quantified in this study.

To address the effect of temperature and humidity on the infectivity of the virus, Dabisch et al. (2021) conducted an experiment to analyse the virus infectivity under several conditions. The temperature condition was adjusted from 10°C to 40°C and the condition of relative humidity was tested from 20% to 70%. They then performed a stepwise regression analysis to obtain a regression model. This regression model is utilized to obtain the equation of stability of the virus in aerosol ( $Ka$ ) with temperature ( $T$ ) and relative humidity ( $RH$ ) for this study as shown in Equation (5):

$$Ka = 0.16030 + 0.041018 \left( \frac{T - 20.615}{10.585} \right) + 0.02176 \left( \frac{RH - 45.235}{28.665} \right) - 0.15457 \left( \frac{T - 20.615}{10.585} \right). \quad (5)$$

Note that a higher decay constant indicates a rapid decrease in the infectivity of the virus in aerosol in a location. The stability of the virus on a surface is considered in this study and used as the second of the derived parameters in the location node. Several studies claim that the virus can stay on various surfaces and remain infective from an hour to several days (Guo et al., 2020; Santarpia et al., 2020; van Doremalen et al., 2020). The virus generated by the patient via droplets can fall on the surface of an object and can then be touched by people. The infectivity of the virus on the surface will be influenced by several factors such as temperature, relative humidity and surface type.

In order to obtain the relationship for infectivity of the virus on a surface, an experiment is conducted by Biryukov et al. (2020) where the temperature is controlled from 24°C to 35°C and the relative humidity is adjusted from 20 to 80%; the surfaces chosen are stainless steel and acrylonitrile butadiene styrene (ABS) plastic, which are materials most seen in public places. A linear regression analysis is performed to fit the



experimental data, and the half-life of the virus is the quantity of interest in the experiment. Since the decay constant will act as the unit of measurement for virus stability, the half-life equation obtained from that experiment will be converted into a decay constant format as shown in Equation (6):

$$K_s = \frac{\ln 2}{32.426272 - 0.622108T - 0.153707RH} \quad (6)$$

The third parameter is  $Fl$ , which is defined as the frequency of a location node visited by a patient. The number of visits made by a close contact is not taken into account because only the patient is assumed to carry and spread the virus. Thus, the  $Fl$  is quantified to integrate the impact of the patient's location visiting time. To quantify the  $Fl$ , a link matrix is formed where  $L_i$  is the location node  $i$  and  $H_j$  is the human node with human type; ( $H_s$ ) is 1. Since there is only one patient in this network,  $j = 1$  with  $Fh_{1:i}$  representing the number of times the patient visited the location  $i$ , where  $i \in \{1, 2, 3, \dots, 21\}$ . Therefore, the  $Fl$  is mathematically defined as in Equations (7)-(8):

$$Fl_i = \sum Link_{Matrix_{BCC\ network}(L_i H_1)} \times Fh_{1:i} \text{ where} \quad (7)$$

$$Link_{Matrix_{BCC\ network}(L_i H_1)} = \begin{cases} 1, & \text{if } L_i \text{ is linked to } H_1 \\ 0, & \text{if } L_i \text{ is not linked to } H_1 \end{cases} \quad (8)$$

### 2.2.3 Quantification of parameters for human node

There are also three parameters chosen to quantify the human node as illustrated in Figure 3. The first parameter concerned is the indoor infection risk ( $P_{wr}$ ), affecting the close contact only. The indoor infection risk  $P_{wr}$  is quantified using the Wells-Riley equation, which is based on the concept of 'quantum of infection' proposed by Wells (1955) and formulated by Riley et al. (1978). Measles was the first infectious disease to apply the Wells-Riley equation for risk assessment. The 'quantum of infection' is a hypothetical infectious dose unit exhaled by an infectious person. The quantum generation rate,  $q$ , found in the Wells-Riley equation, cannot be directly measured but it can be estimated, as suggested by Dai & Zhao (2020), based on the epidemiology data such as the number of incidents and the basic reproduction number,  $R_0$ . The equation to quantify the  $q$  is defined by Dai & Zhao (2020) and shown in Equation (9):

$$q = A + B_1 R_0 + B_2 R_0^2 \quad (9)$$

Where the  $A$ ,  $B_1$ , and  $B_2$  are constant numbers which are -30.27958, -44.81536, and 19.67934 respectively. The  $R_0$  for Malaysia is 3.08, as reported by Gill et al., (2020); therefore, the  $q$  is found to be  $18.38 \text{ h}^{-1}$ . The pulmonary ventilation rate of a susceptible  $p$  is estimated as  $0.3 \text{ m}^3$  per hour based on Duan (2013). The ventilation rate of a location ( $Q$ ) found in the Wells-Riley equation, is quantified based on the location type in accordance to the value recommended by the American Society of Heating, Refrigerating and Air-Conditioning Engineers guideline (ASHRAE, 2013).  $I$  indicates the number of infectious persons in the location, and  $t$  refers to the total duration of stay ( $D$ ) by the close contact. Putting all the information described earlier together, the quantification of  $P_{wr}$  is therefore defined as Equation (10):

$$P_{wr_{H_j:L_i}} = \begin{cases} 0, & \text{if } H_{sH_j} = 1, \\ \left[ 1 - \exp\left(\frac{-Iqpt}{Q}\right) \right] \times L_{type} \times Link_{Matrix_{BCC\ network}(H_j L_i)}, & \text{otherwise.} \end{cases} \quad (10)$$

The parameter ( $D$ ) mentioned earlier is the second parameter for the human node. It is defined as the total duration of a person's stay at one location and is measured in hours. This parameter is calculated as the cumulative duration of an infectious person having visited a particular location. The final parameter

required for the human node is the frequency of a person visiting a location. This is simply defined as the number of times a person visited a location  $Fh_{j:i}$ , shown mathematically in Equation (11), where  $i \in \{1, 2, 3, \dots, 21\}$  and  $j \in \{1, 2, 3, \dots, 31\}$

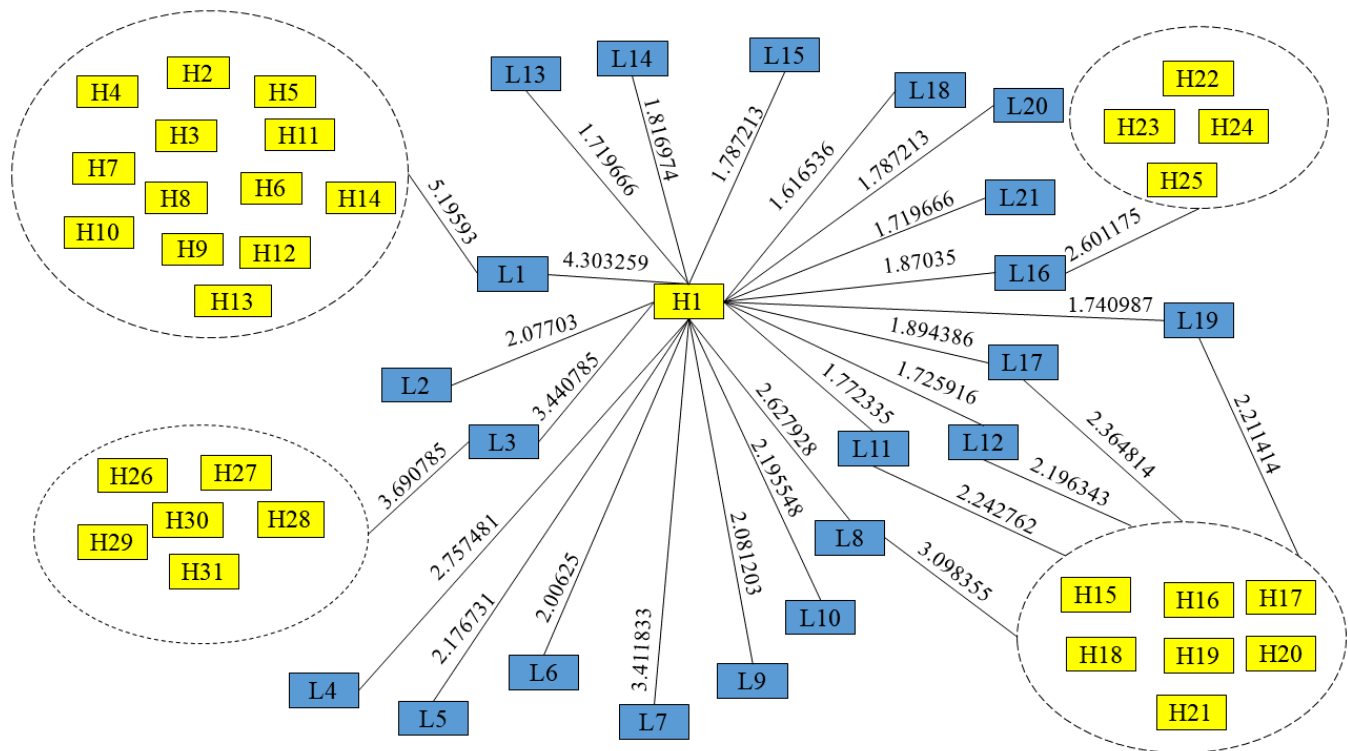
$$Fh_{j:i} = \begin{cases} n, & \text{if } H_j \text{ visited } L_i \text{ } n \text{ times where } n \in \mathbb{Z}^+ \\ 0 & \text{if } H_j \text{ not visit } L_i \end{cases} \quad (11)$$

## 2.2.4 Quantification of link weight

In order to convert the BCC graph depicted in Figure 2 into a network, the weight of the edge needs to be quantified. The link weight is defined as the contact strength between the human and the location node, it is termed the COVID-19 contact strength (CCS). A higher value of the CCS indicates a stronger intensity between the human node and the location node in the network. There are four rules introduced by Liew et al. (2015): (i) summation, (ii) multiplication, (iii) summation of products, and (iv) multiplication of sums. The summation rule is used in dengue hotspot identification (Kok et al., 2017). Since the aspect of this study is the same as the dengue hotspot study, which is disease hotspot identification, the summation rule is applied for the link weight quantification at this stage. All of the parameter values are normalized before the link weight is quantified. The link weight quantification is formulated in Equation (12), where  $i \in \{1, 2, 3, \dots, 21\}$  and  $j \in \{1, 2, 3, \dots, 31\}$ :

$$\begin{aligned} CCS_{i:j} &= \left( \sum Location_{NodeParameters_i} \right) + \left( \sum Human_{NodeParameters_{j:i}} \right) \\ &= (Ka_i + Ks_i + Fl_i) + (Pwr_{H_j:L_i} + D_{j:i} + Fh_{j:i}). \end{aligned} \quad (12)$$

The complete BCC network is illustrated in Figure 4 with the link weight value shown on all edges. To provide better visualization for the network, the human nodes from H2 to H31 are clustered according to the location contacted with H1. The link weight value will be substituted to each data point on the link matrix as an input to generate the ranking of the human node and the location node.



**Figure 4:** Bipartite COVID-19 contact network with clustered human nodes.



## 2.2.5 Implementation of ranking algorithm

To rank the human nodes and the location nodes, the Hypertext-Induced Topic Search (HITS) algorithm is used in this study. The HITS is a web page searching algorithm with a query-dependent algorithm, which is more dynamic than other types of web search algorithms. It not only considers the link popularity or node degree, but the link weights of the network are also taken into account. Thus, this is one of the reasons for choosing HITS as the ranking algorithm in this study. Another reason is that the HITS involves the authority and hub where the hub matrix is pointing to the authority matrix, which corresponds to the location node and human node in this study. The HITS has been used to rank the hotspots of mosquito-borne diseases: malaria (Eze, 2013) and dengue (Kok, 2017) and suitability of dolphin habitats (Liew, 2016). All the mentioned studies have used the BNM to identify the hotspots and habitat suitability, proving the HITS to be suitable for bipartite network analysis.

The CCS matrix computed by Equation (12) is used to generate a hub matrix and an authority matrix as the inputs for the HITS to determine the corresponding eigenvalues and eigenvectors for each location node and human node. The R language is chosen to implement the HITS algorithm by using the power method to identify the principal eigenvalues and eigenvectors. The eigenvector of the location node is defined as CHR while the eigenvector of human nodes is defined as CRR. These CHR and CRR values range from 0 to 1 and thus, the ranking is based on these values evaluated using Equation (13):

$$CHR = \frac{Eigenvector_i}{Max(Eigenvector_i, Eigenvector_{i+1}, Eigenvector_{i+2}, \dots, Eigenvector_n)} \quad (13)$$

The CRR values are computed using the same Equation (13), where the *Eigenvector* is the principal eigenvector generated from the HITS algorithm and  $n$  is the number of nodes, which is 21 for CHR and 31 for CRR.

## 3 Results and Discussion

The BCC network is considered fully formulated when all the parameters defined earlier are quantified. As mentioned, there are three parameters for the human node and three for the location node. In terms of node degree, the highest one is location node L1 with a total of 14 links attached to human nodes H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13 and H14. This is followed by L8, L11, L12, L17 and L19, sharing the same number of node degrees, which is 7.

The ranking results of the location nodes are listed in Table 1. As expected, L1 is placed first in the ranking results since it has the highest CHR value and its highest node degree is 14. However, L3 has only 7 node degrees but it was ranked higher than L11, L12, L17 and L19. According to the literature that used the same methodology, the node degree of location is not a principal factor that affects the location node ranking (Kok et al., 2017; Ying et al., 2015). In fact, the CCS of L3 with the linked human nodes are higher than the CSS values of L11, L12, L17 and L19. This statement implies that the other parameters of the human nodes and location nodes signify the ranking. Consistent with the literature, this study shows that the statement is true based on the ranking shown in Table 1. Thus, not only does the node degree of location play a role in identifying the hotspot, but also other factors need to be taken into consideration. With respect to the aim of this study, as anticipated, location L1 is found to be a hotspot since it is the workplace of the infected person H1. Nevertheless, locations L8 and L3 should also be focussed on because both of them are ranked after L1.

**Table 1:** Location ranking with CHR of BCC Network.

Ranking	Ranked location	Principal Eigenvector	CHR
1	L1	0.975249	1
2	L8	0.094208	0.09659884
3	L3	0.072006	0.07383377
4	L17	0.070077	0.07185522
5	L11	0.066062	0.06773823
6	L19	0.06503	0.06668079
7	L12	0.064535	0.06617241
8	L7	0.055977	0.05739758
9	L4	0.045241	0.04638934
10	L10	0.036022	0.03693591
11	L5	0.035713	0.03661934
12	L9	0.034146	0.03501226
13	L2	0.034077	0.03494206
14	L16	0.03305	0.03388918
15	L6	0.032916	0.03375132
16	L14	0.029811	0.03056712
17	L15	0.029322	0.03006645
18	L20	0.029322	0.03006645
19	L13	0.028214	0.02893009
20	L21	0.028214	0.02893009
21	L18	0.026522	0.02719513

The ranking of the human node tabulated in Table 2 can be used as a potential indicator to predict the risk of infection for close contact and identify the ‘super spreader’ if there are more than one patient in the network. Based on the data obtained and hence the results, since there is only one patient (H1) in the network, then, as expected, the human node with the highest CRR is H1, implying that H1 is the person spreading the virus in this network. On the other hand, the close contacts, H2 to H14, hold the same CRR values and are clustered into the same rank. They are the close contacts of H1 via location node L1, and since L1 was ranked first among the location nodes, this implies that this cluster of close contacts has a higher risk than other human nodes in this network.

**Table 2:** Human ranking with CRR of BCC Network.

Ranking	Ranked Human	Principal Eigenvector	CRR
1	H1	0.319135187	1
2	H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14	0.260510833	0.924232
3	H15, H16, H17, H18, H19, H20, H21	0.045822581	0.806119
4	H26, H27, H28, H29, H30, H31	0.013662691	0.702686
5	H22, H23, H24, H25	0.004419703	0.643522

To ensure that the implementation of the HITS algorithm is correct, a benchmark verification is performed using UCINET to generate a benchmark model. UCINET is a network analysis software that can support many network analysis types, for instance centrality measurement, elementary network analysis and subgroup identification. One of the built-in functions of UCINET is network centrality analysis, where the input data is a hub matrix and an authority matrix to obtain the eigenvector of the network and is termed  $CHR_B$  for the location node and  $CRR_B$  for the human node. Since the hub matrix and authority matrix are also used in this study to generate the ranking, it is suitable to choose UCINET as the benchmark model. The root mean square error (RMSE) between  $CHR_B$  and  $CHR$  as well as between  $CRR_B$  and  $CRR$  will be evaluated and used as an indicator to verify the model. The model is verified if the RMSE value is less than 0.05. The RMSE found when comparing between  $CRR_B$  and  $CRR$  is 0.000224 and the RMSE obtained when comparing between  $CHR_B$  and  $CHR$  is 0.001419. Since the RMSE are found to be less than 0.05 for both instances, the implementation process is verified.

At the point of writing, there are no new cases reported from the close contacts (human node) of this formulated network. A possible explanation for this might be that the patient in this study, which is H1, has probably not developed the symptoms yet. According to Cevik et al. (2020), 5-6 days are required to develop the symptoms once exposed to the virus, and the virus transmission capacity of COVID-19 is highest in the first week after symptoms have developed. The diagnosis results obtained by viral detection do not correspond to the infectivity of the patient (Cevik et al., 2020). Thus, the viral load of the patient (H1) in this network may not be sufficient enough to infect other people before being diagnosed as positive.

Another possible explanation for this is that the patient's exact onset date may be much earlier than the tracing done as reported in the investigation form. The reverse transcription polymerase chain reaction (RT-PCR) is a common method for detecting the pathogen of COVID-19 in the human body. A swab sample from the upper respiratory tract, which is between the nose and the mouth, is taken for the RT-PCR test. The RT-PCR can capture the virus in the sample even 17 days after the symptoms appear (Cevik et al., 2021). This shows that the patient's point of exposure is crucial for estimating the actual onset date. The dataset used in this study did not state the point of exposure of H1, and this is difficult to estimate because H1 is the only patient in this network. Hence, there is a possibility that some close contacts are left untraced if the actual onset date of H1 is much earlier.

To create a full picture of this study, the transmission ability of asymptomatic patients is suggested to be included in the model. An asymptomatic patient is a patient who has not developed any symptoms of COVID-19 before and after diagnosis, mostly detected through close contact tracing. Several studies have claimed that the transmission ability of asymptomatic patients is lower than that of symptomatic and pre-symptomatic patients (Buitrago-Garcia et al., 2020; Qiu et al., 2021). A pre-symptomatic patient is a patient who has no symptoms at the point of diagnosis but develops them after that, while a symptomatic patient is a patient who has developed symptoms before the point of diagnosis, and the symptoms can be further classified into three levels: (i) mild-moderate illness, (ii) severe illness, and (iii) critical illness. The parameters used in the model are only concentrated on the effect of virus transmission under environmental conditions of the location at the current stage. The formulation lacks the consideration of the kinetic dynamics of the virus in the human body. Even though the transmission potential of the asymptomatic patient is low, it should not be neglected. It is proposed that this could be included in the human node quantification stage, thus ensuring that the network obtained will be richer as it adds more human and location nodes to the network.

## 4 Conclusion

The main goal of this study is to identify hotspots of COVID-19 transmission by formulating a bipartite network model based on patient mobility data. There are 30 close contacts and 1 patient with 21 locations visited by the patient in the network. The stability of the virus on surfaces and in aerosols are the derived parameters of the location node in the network as it determines the effect of temperature and relative humidity of the location on the virus. The indoor infection risk is another derived parameter of the human node to evaluate the risk of infection by considering the ventilation rate of a particular location. The BCC network was formulated by quantifying the link weight between each pair of the bipartite nodes and was verified using UCINET. Even though there is only one patient (H1) in this network model, the findings show that BNM can identify the hotspots, which for this study were found to be L1, followed by L8 and L3. It is evident from the results that in identifying disease hotspots, it is insufficient to rely only on the locations with a higher node degree; other factors must also be taken into consideration such as location size, location activity type, and density of people in the location. The limitation of this study is the data considered as it happened to be imbalanced: only one patient appears in the network and there are no subsequent cases linked with the patient, H1.

Due to this, the model validation is not included in this paper. Another issue that was not addressed in this study was the transmission ability of asymptomatic patients, which is hypothesized to affect the results because the attack rate of those patients is distinct from that of symptomatic patients. Therefore, further study is needed to analyse the difference between several sets of networks by using different sets of data. In addition, the influence of each parameter needs to be investigated by performing significance and sensitivity analyses. It is hoped that the outcome of this study will be able to assist public health authorities in increasing the efficacy of contact tracing and identifying the source of infection for every patient so that the outbreak of COVID-19 in Malaysia will be contained.

## Additional Information and Declarations

**Acknowledgement:** We thank the Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak for direct and indirect assistance for the research. Our heartfelt thanks also go to Bintulu division Health Department for collecting and providing the research data.

**Funding:** This work was funded by Ministry of Higher Education, GL/F08/KPT/02/2020.

**Conflict of Interests:** The authors declare no conflict of interest.

**Author Contributions:** J.L., B.H.H.: Conceptualization, Formal analysis. B.H.H.: Data curation, Methodology, Software, Resources, Visualization, Writing – original draft. J.L.: Supervision, Funding acquisition, Project administration. M.H.L.C.: Investigation. J.L., W.K.T., T.L.: Validation. J.L., B.H.H., W.K.T., T.L., M.H.L.C.: Writing – review & editing.

**Data Availability:** The data that support the findings of this study are available from the corresponding author.

## References

- ASHRAE. (2013). *ANSI-ASHRAE Standard 62.1-2013 Ventilation for Acceptable Indoor Air Quality*. American Society of Heating, Refrigerating and Air-Conditioning Engineers. <http://www.myiaire.com/product-docs/ultraDRY/ASHRAE62.1.pdf>
- Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behaviour*, 4, 588–596. <https://doi.org/10.1038/s41562-020-0898-6>
- Buitrago-Garcia, D., Egli-Gany, D., Counotte, M. J., Hossmann, S., Imeri, H., Ipekci, A. M., Salanti, G., & Low, N. (2020). Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLOS Medicine*, 17(9), e1003346. <https://doi.org/10.1371/journal.pmed.1003346>
- Cai, J., Sun, W., Huang, J., Gamber, M., Wu, J., & He, G. (2020). Indirect virus transmission in cluster of COVID-19 cases, Wenzhou, China, 2020. *Emerging infectious diseases*, 26(6), 1343. <https://doi.org/10.3201/eid2606.200412>
- Cevik, M., Kuppalli, K., Kindrachuk, J., & Peiris, M. (2020). Virology, transmission, and pathogenesis of SARS-CoV-2. *BMJ*, 371, m3862. <https://doi.org/10.1136/bmj.m3862>
- Cevik, M., Tate, M., Lloyd, O., Maraolo, A. E., Schafers, J., & Ho, A. (2021). SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe*, 2(1), e13–e22. [https://doi.org/10.1016/s2666-5247\(20\)30172-5](https://doi.org/10.1016/s2666-5247(20)30172-5)
- Dabisch, P., Schuit, M., Herzog, A., Beck, K., Wood, S., Krause, M., Miller, D., Weaver, W., Freeburger, D., Hooper, I., Green, B., Williams, G., Holland, B., Bohannon, J., Wahl, V., Yoltz, J., Hevey, M., & Ratnesar-Shumate, S. (2021). The influence of temperature, humidity, and simulated sunlight on the infectivity of SARS-CoV-2 in aerosols. *Aerosol Science and Technology*, 55(2), 142–153. <https://doi.org/10.1080/02786826.2020.1829536>
- Dai, H., & Zhao, B. (2020). Association of the infection probability of COVID-19 with ventilation rates in confined spaces. *Building Simulation*, 13(6), 1321–1327. <https://doi.org/10.1007/s12273-020-0703-5>
- Duan, X. (2013). *Exposure factors handbook of Chinese population*. China Environmental Science Press.
- Eze, M. O. (2013). *Web Algorithm Search Engine Based Network Modeling of Malaria Transmission*. Doctoral dissertation. Universiti Malaysia Sarawak.
- Fears, A. C., Klimstra, W. B., Duprex, P., Hartman, A., Weaver, S. C., Plante, K. S., Mirchandani, D., Plante, J. A., Aguilar, P. V., Fernández, D., Nalca, A., Totura, A., Dyer, D., Kearney, B., Lackemeyer, M., Bohannon, J. K., Johnson, R., Garry, R. F., Reed, D. S., & Roy, C. J. (2020). Persistence of Severe Acute Respiratory Syndrome Coronavirus 2 in Aerosol Suspensions. *Emerging Infectious Diseases*, 26(9), 2168–2171. <https://doi.org/10.3201/eid2609.201806>
- Firth, J. A., Hellewell, J., Klepac, P., Kissler, S., Kucharski, A. J., & Spurgin, L. G. (2020). Using a real-world network to model localized COVID-19 control strategies. *Nature Medicine*, 26(10), 1616–1622. <https://doi.org/10.1038/s41591-020-1036-8>
- Gasser, U., Ienca, M., Scheibner, J., Sleight, J., & Vayena, E. (2020). Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. *The Lancet Digital Health*, 2(8), e425–e434. [https://doi.org/10.1016/s2589-7500\(20\)30137-0](https://doi.org/10.1016/s2589-7500(20)30137-0)
- Gill, B. S., Jayaraj, V. J., Singh, S., Mohd Ghazali, S., Cheong, Y. L., Md Iderus, N. H., Sundram, B. M., Aris, T. B., Mohd Ibrahim, H., Hong, B. H., & Labadin, J. (2020). Modelling the Effectiveness of Epidemic Control Measures in Preventing the Transmission of COVID-19 in Malaysia. *International Journal of Environmental Research and Public Health*, 17(15), 5509. <https://doi.org/10.3390/ijerph17155509>



- Gomes, D. S., Andrade, L. A., Ribeiro, C. J. N., Peixoto, M. V. S., Lima, S. V. M. A., Duque, A. M., Cirilo, T. M., Góes, M. A. O., Lima, A. G. C. F., Santos, M. B., Araújo, K. C. G. M., & Santos, A. D. (2020). Risk clusters of COVID-19 transmission in northeastern Brazil: prospective space-time modelling. *Epidemiology and Infection*, 148. <https://doi.org/10.1017/s0950268820001843>
- Guo, Z.-D., Wang, Z.-Y., Zhang, S.-F., Li, X., Li, L., Li, C., Cui, Y., Fu, R.-B., Dong, Y.-Z., Chi, X.-Y., Zhang, M.-Y., Liu, K., Cao, C., Liu, B., Zhang, K., Gao, Y.-W., Lu, B., & Chen, W. (2020). Early Release - Aerosol and Surface Distribution of Severe Acute Respiratory Syndrome Coronavirus 2 in Hospital Wards, Wuhan, China, 2020. *Emerging Infectious Diseases*, 26(7), 1583. <https://doi.org/10.3201/eid2607.200885>
- Huang, R., Liu, M., & Ding, Y. (2020). Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. *The Journal of Infection in Developing Countries*, 14(3), 246–253. <https://doi.org/10.3855/jidc.12585>
- Kok, W. C. (2017). *A Computational Approach to Predict the Spread of Dengue*. Master thesis. Universiti Malaysia Sarawak.
- Kok, W. C., Labadin, J., & Perera, D. (2017). Modeling Dengue Hotspot with Bipartite Network Approach. In Alfred R., Iida H., Ag. Ibrahim A., Lim Y. (eds) *Computational Science and Technology, ICCST 2017* (pp. 220–229). Springer. [https://doi.org/10.1007/978-981-10-8276-4\\_21](https://doi.org/10.1007/978-981-10-8276-4_21)
- Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Sun, F., Jit, M., Munday, J. D., Davies, N., Gimma, A., Zandvoort, K. van, Gibbs, H., Hellewell, J., Jarvis, C. I., Clifford, S., Quilty, B. J., Bosse, N. I., & Abbott, S. (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *The Lancet Infectious Diseases*, 20(5), 553–558. [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9). <https://doi.org/10.7326/M20-0504>
- Lessler, J., Azman, A. S., McKay, H. S., & Moore, S. M. (2017). What is a Hotspot Anyway? *The American Journal of Tropical Medicine and Hygiene*, 96(6), 1270–1273. <https://doi.org/10.4269/ajtmh.16-0427>
- Liew, C. Y. (2016). *Bipartite-Network-Based Modeling of Habitat Suitability*. Doctoral dissertation. Universiti Malaysia Sarawak.
- Meyers, L. A., Newman, M. E. J., & Pourbohloul, B. (2006). Predicting epidemics on directed contact networks. *Journal of Theoretical Biology*, 240(3), 400–418. <https://doi.org/10.1016/j.jtbi.2005.10.004>
- Miller, S. L., Nazaroff, W. W., Jimenez, J. L., Boerstra, A., Buonanno, G., Dancer, S. J., Kurnitski, J., Marr, L. C., Morawska, L., & Noakes, C. (2021). Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event. *Indoor Air*, 31(2), 314–323. <https://doi.org/10.1111/ina.12751>
- MoH. (2020). Annex 12: Management of Closed Contacts of Confirmed Case. Retrieved from <http://covid-19.moh.gov.my/garis-panduan/garis-panduan-kkm>
- Qiu, X., Nergiz, A. I., Maraolo, A. E., Bogoch, I. I., Low, N., & Cevik, M. (2021). Defining the role of asymptomatic and pre-symptomatic SARS-CoV-2 transmission – a living systematic review. *Clinical Microbiology and Infection*, 27(4), 511–519. <https://doi.org/10.1016/j.cmi.2021.01.011>
- Riley, E., Murphy, G., & Riley, R. (1978). Airborne spread of measles in a suburban elementary school. *American Journal of Epidemiology*, 107(5), 421–432. <https://doi.org/10.1093/oxfordjournals.aje.a112560>
- Santarpia, J. L., Rivera, D. N., Herrera, V. L., Morwitzer, M. J., Creager, H. M., Santarpia, G. W., Crown, K. K., Brett-Major, D. M., Schnaubelt, E. R., Broadhurst, M. J., Lawler, J. V., Reid, S. P., & Lowe, J. J. (2020). Aerosol and surface contamination of SARS-CoV-2 observed in quarantine and isolation care. *Scientific Reports*, 10(1), 12732. <https://doi.org/10.1038/s41598-020-69286-3>
- Shen, M., Peng, Z., Xiao, Y., & Zhang, L. (2020). Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. *bioRxiv*, 2020.2001.2023.916726. <https://doi.org/10.1101/2020.01.23.916726>
- van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., Tamin, A., Harcourt, J. L., Thornburg, N. J., Gerber, S. I., Lloyd-Smith, J. O., de Wit, E., & Munster, V. J. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New England Journal of Medicine*, 382(16), 1564–1567. <https://doi.org/10.1056/nejmc2004973>
- Viguerie, A., Veneziani, A., Lorenzo, G., Baroli, D., Aretz-Nellesen, N., Patton, A., Yankeelov, T. E., Reali, A., Hughes, T. J. R., & Auricchio, F. (2020). Diffusion–reaction compartmental models formulated in a continuum mechanics framework: application to COVID-19, mathematical analysis, and numerical study. *Computational Mechanics*, 66(5), 1131–1152. <https://doi.org/10.1007/s00466-020-01888-0>
- Wang, Z., Zhang, X., Teichert, G. H., Carrasco-Teja, M., & Garikipati, K. (2020). System inference for the spatio-temporal evolution of infectious diseases: Michigan in the time of COVID-19. *Computational Mechanics*, 66(5), 1153–1176. <https://doi.org/10.1007/s00466-020-01894-2>
- Weeden, K., & Cornwell, B. (2020). The Small-World Network of College Classes: Implications for Epidemic Spread on a University Campus. *Sociological Science*, 7, 222–241. <https://doi.org/10.15195/v7.a9>
- Wells, W. F. (1955). *Airborne Contagion and Air Hygiene: An Ecological Study of Droplet Infections*. Harvard University Press.
- Whitelaw, S., Mamas, M. A., Topol, E., & Spall, H. G. C. V. (2020). Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health*, 2(8), e435–e440. [https://doi.org/10.1016/S2589-7500\(20\)30142-4](https://doi.org/10.1016/S2589-7500(20)30142-4)
- 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)



**Xie, Z., Qin, Y., Li, Y., Shen, W., Zheng, Z., & Liu, S.** (2020). Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors. *Science of the Total Environment*, 744, 140929. <https://doi.org/10.1016/j.scitotenv.2020.140929>

**Ying, L. C., Labadin, J., Chai, W. Y., Tuen, A. A., & Peter, C.** (2015). Applying Bipartite Network Approach to Scarce Data: Modeling Habitat Suitability of a Marine Mammal Species. *Procedia Computer Science*, 60, 266–275. <https://doi.org/10.1016/j.procs.2015.08.126>

---

**Editorial record:** The article has been peer-reviewed. First submission received on 15 April 2021. Revisions received on 17 May 2021 and 6 June 2021. Accepted for publication on 6 June 2021. The editors coordinating the peer-review of this manuscript were Chiew Kang Leng  and Zdenek Smutny . The editor in charge of approving this manuscript for publication was Zdenek Smutny.

**Special Issue:** Inspiring Technologies for Digital Inclusivity. Selected papers from the 12<sup>th</sup> International Conference on Information Technology in Asia (CITA'21).

---

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951

---