Rabies Hotspot Detection Using Bipartite Network Modelling Approach

DAREN JIAN BING CHIA 1, WOON CHEE KOK 1, NUR ASHEILA ABDUL TAIB 1, BOON HAO HONG 1, KHAIRANI ABD MAJID 2 & JANE LABADIN *1

1Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia; 2Department of Computer Science, Faculty of Defence Science and Technology, Universiti Pertahanan Nasional Malaysia, Kem Sungai Besi, 57000 Kuala Lumpur, Malaysia

*Corresponding authors: ljane@unimas.my

ABSTRACT

Despite entering its fourth year, the rabies outbreak in the East Malaysian state of Sarawak has claimed another nine lives in 2020, culminating with a total of 31 laboratory-confirmed cases of human rabies as of 31st December 2020. One of the outbreak control challenges faced by the authorities within a previously rabies-free area, such as in the case of Sarawak, is the lack of information regarding possible starting sources, notably hotspot locations of the outbreak. Identification of potential high-risk areas for rabies infection is a sine qua non for effective disease interventions and control strategies. Motivated by this and in preparation for future similar incidents, this paper presented a preliminary study on rabies hotspot identification. The modelling approach adopted the bipartite network where the two disjoint sets of nodes are the Location node and Dog (Bite Cases) node. The formulation of the network followed closely the Bipartite Modeling Methodology Framework. Thorough model verification was done in an attempt to show that such problem domain can be modelled using the Bipartite Modeling approach.

Keywords: Bipartite Network Modeling Framework, BRC, Dog, Location, Rabies, Ranking

INTRODUCTION

An outbreak of rabies has killed 29 people to date and affected dozens of localities throughout the East Malaysian state of Sarawak. Various rabies control efforts have been carried out by the local government, which include canine rabies vaccination, human post-exposure prophylaxis and targeted removal of dogs through culling (Taib, Labadin, & Piau, 2019). Despite entering its fourth year, the disease has claimed another nine lives in 2020 alone and dog bite incidents remain relatively high (State Disaster Management Committee, 2020). As Sarawak was historically rabies-free, rabies surveillance and vaccination practice were not conducted prior to the initial outbreak in July 2017 (Navanithakumar et al., 2019). Hence, the government faced some difficulties to control the disease outbreak when it first started out which may be attributable to the lack of information regarding possible hotspots for rabies. Herein we focus on the identification of these possible starting sources or rather, hotspot locations of Sarawak rabies outbreak.

Over the past decades, mathematical modelling of dynamical systems has been a vital tool in analyzing the evolution of disease spread while also offering insights into the most efficient control strategies. According to Siettos and Russo (2013), current mathematical models of dynamical systems generally encompass four different approaches: (1) deterministic models, (2) stochastic models, (3) agent-based models and (4) network models. As such, previous rabies modelling studies mainly investigate the transmission dynamics of rabies using deterministic models (Asamoah, Oduro, Bonyah & Seidu, 2017; Huang, Ruan, Shu & Wu, 2019; Zhang, Jin, Sun & Ruan, 2012), stochastic models (Cao, Feng, Wen, Zu & Gao, 2020; Dürr & Ward, 2015; Hudson, Brookes, & Ward, 2017; Sparkes et al., 2016), agent-based models (Brookes, Dürr, & Ward, 2019) and network models (Laager et al., 2018).
In the case of network epidemiology modelling, individuals or groups of individuals are represented as nodes while the links or rather connections between nodes are known as edges. Disease will transmit from one node to another along the edges. Should the transmission probability along the edge reach a high value, an epidemic is predicted to happen (Craft & Caillaud, 2011). One of the advantages of using this approach is that it could display complex relations of a disease accurately according to real world situations. Another benefit is that this type of modelling generally contains lists of parameters (attributes) associated to nodes and edges. This will allow clear representation of the variations in the between-host or between-edge patterns and relationships. Due to its ability to depict disease transmission dynamics via the nodes and edges while producing accurate results with minimum amount of data supplement, the network-based approach is commonly implemented to model various infectious diseases such as hepatitis B (Chandler, 2017), malaria (Pipatsart et al., 2018), HIV (Zhong, Zhang, & Li, 2018) and dengue (Kok & Labadin, 2019).

In spite of a perusal of the works above-mentioned, modelling studies on rabies in the context of Sarawak remain scarce. For instance, Taib et al. (2019) developed a deterministic, compartmental model to study the spread of rabies in Sarawak and evaluated the effectiveness of different rabies control strategies (Taib et al., 2019). The work however did not focus on detecting rabies hotspots in Sarawak. In this paper, we attempt to identify potential high-risk areas of rabies infection in Sarawak by utilizing the Bipartite Network-Based Modeling Framework (BNB-MF) approach that follows heavily from Liew (2016) and Kok (2018) studies. Additionally, identification of rabies hotspots using bipartite network modelling approach, to the best of our knowledge, has yet to be explored.

MATERIALS & METHODS

This research applies a modified version of Bipartite-Network-Based Modeling (BNM) approach originally used to model habitat suitability of Irrawaddy dolphin (Liew, 2016) and malaria hotspots (Kok, 2018). The two types of nodes considered in our bipartite contact network model are location node and dog node. Liew’s study (Liew, 2016) is chosen as our main reference due to the animal attributes involved. To be more specific, dogs have higher similarity to dolphins (Liew, 2016) (in terms of mobility) as compared to mosquitoes (Kok, 2018). The modified BNM approach used in this study is known as BNB-MF which involves three stages as follows:

i. formulation of bipartite graph
ii. development of bipartite rabies contact (BRC) network model
iii. verification of the bipartite rabies contact (BRC) model

Data collection

Information on dog bite cases, location address of exposure, and human rabies positive cases between 1st April 2017 to 31st July 2017 were extracted via online sources such as the official Sarawak Disaster Information website and online news. Records stored on the Sarawak Disaster Information website are actively being updated and maintained daily. This will ensure accuracy of data gathered for this study. A total amount of 17 epidemiological weeks of dog bite cases was collected during the study period with only six cases of dog bites clinically confirmed to be rabies-positive.

Data pre-processing

As location addresses collected were raw data without GPS coordinates provided, the location data were digitized prior to parameters quantification. The distance between different location nodes was also calculated to avoid possible data redundancy appearing in our BRC network. Based on Figure 1, the first step is to manually convert each new location address into GPS coordinates (longitude and latitude), by utilising Google Maps (https://www.google.com/maps). For instance, one of the recorded location addresses read as “Kampung Krait, Jalan Batang Kedup, 94700 Serian, Sarawak”. By using the Google Maps website, the value of longitude and latitude of a specific location can be retrieved. As for the location mentioned above, the longitude and latitude are discovered to be 110.660221 and
1.049387, respectively. The second step was to declare each different location identification and record them into the database. The output obtained from the first step was further processed by using a location declaration generator. The generator code was created using R programming and the output created is a distance matrix which records the distance between different location nodes.

**Implementation of ranking algorithm**
Ranking of the location nodes will be required in order to determine the possible rabies hotspots in Sarawak. As this study only uses two types of nodes (bipartite), Hypertext-Indexed Topic Search (HITS) search algorithm is applied towards the BRC network. In addition, HITS algorithm is justified over Page-Rank search algorithm due to its capability to link both the hub and authority matrices. In this step, the Rabies Contact Strength (RCS) values, which correspond to the link weight values in the BRC network, are used as the input for the HITS algorithm. The output from this algorithm is a pair of principal eigenvectors, which is taken as the location ranking of the BRC network. We labelled this location ranking as the Rabies Hotspot Ranking (RHR), which can also serve as a measurement of dog density.

**Model verification**
Two methods of model verification are applied: namely (1) benchmark verification and (2) analytical verification. Benchmark verification is conducted using UCINET 6 in order to compare the results obtained through the BRC network model with that of another benchmark system (Kok & Labadin, 2019). The root-mean-square-error (RMSE) is calculated by comparing the ranking values from the selected benchmark system and that from the BRC network model. When the RMSE does not exceed 0.05, the BRC network model is verified. As location is the primary concern here, UNICET will focus on generating the benchmark ranking values of the location nodes (\(RHR_b\)). The location ranking of the BRC network (\(RHR_{BRC\ network}\)) and the \(RHR_b\) are normalized in the range 0 to 1 before proceeding with the calculation of RMSE. The RMSE values for the location nodes (\(RMSE_{Loc}\)) are computed using the formula below, where \(N_{Loc}\) = number of location nodes.

\[
RMSE_{Loc} = \sqrt{\frac{1}{N_{Loc}} \sum_{i=1}^{N_{Loc}} (RHB_{bi} - RHB_{BRC\ networki})^2}. 
\]

In analytical verification, the degree of closeness between ranking values for each location and dog nodes will be measured. Spearman’s Rank Correlation Coefficient (SRCC) is used to verify a network model with small data size which is appropriate in this model. SRCC values usually range from -1 to +1, whereby +1 indicates a very strong positive association between the ranks, whereas -1 signifies strong negative association of the ranks and 0 represents non-existent association between the ranks. Table 1 illustrates a range of values used to interpret SRCC. In this study,
it is hypothesized that there exists a high correlation with a coefficient of more than 0.8 between the RHR of each location node and the hub matrix of edges linked to it. According to Liew (2016), a minimum of 0.7 SRCC between RHR and hub matrix of the location nodes should be acquired to verify the models created in this study.

Parameter significance analysis
Parameter significance analysis is conducted to identify the critical parameters and to determine the relative importance of each individual parameter considered in the BRC network. The analysis involved two types; (1) “leave-one-out” analysis for each individual parameter and (2) “multiple-out” analysis for a combination of more than one parameter. For “leave-one-out” analysis, one parameter from the BRC network model is chosen to be excluded at a time, during computation of RHR and other subsequent processes of BNB-MF. As for “leave-multiple-out” analysis, two or more parameters will be excluded from the BRC model and its subsequent processes. This analysis also requires the calculation of SRCC value.

Formulation of bipartite network model
In order to develop our BRC network model, we first need to define the bipartite graph. In this study, we identified two crucial components in rabies disease transmission – location and host properties. The traditional epidemiological triangle has been modified (Figure 2) to serve as the foundation of our bipartite graph structure. The host component (H) represents an animal or human potentially prone to the rabies disease. Next, the environment component (E) specifies exterior causes that can impact the probability of disease infection or transmission. Since this research is dedicated to locating the hotspots of rabies disease, the location component (L) is essential and cannot be removed from the modified disease triangle. Hence, E is derived into a L. It is also assumed that only stray dogs transmit rabies virus to humans and that all strays are susceptible to rabies.

The bipartite graph structure is made up of a set of location nodes (X), a set of dog nodes (D) and a set of relation edges denoted by R. The edge represents a link created when a dog in set D visits a location in set L. Hence, we can define the bipartite graph to be \( G = (D, X, R) \) which consists of two distinct node sets D and X, whereas R represents the edge. Therefore, the bipartite graph G presents the basic building block (Figure 3) for the formulation of BRC network.

![Figure 2. Basic building block of bipartite graph.](image1)

![Figure 3. Basic building block of BRC network.](image2)

According to graph theory, a bipartite network is a weighted bipartite graph in which the affinity between two nodes is measured based on their link weights (Rayfield, Fortin, & Fall, 2011). To formulate an example of our bipartite rabies contact (BRC) network as in Figure 4, we take the previously defined bipartite graph \( G = (D, X, R) \) whereby D and X represent sets of dog nodes and location nodes, respectively, whereas R represents the link edges. Then, let us also assume that set D, the set of dogs (utilizing dog bite cases as proxy) includes these four elements, \( D = \{ D_1, D_2, D_3, D_4 \} \); whereas set X, the set of locations has five elements by which \( X = \{ X_1, X_2, X_3, X_4, X_5 \} \).
whereas set $R$, the set of links in between elements $X$ and $D$, consists of eight elements $R = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8\}$. As a summary, the network can be expressed as \{D1X2, D1X3, D2X3, D3X1, D3X4, D4X3, D4X4, D4X5\}.

To build a network, the computation of link weights corresponding to each link edge is required. We labelled these link weights as the RCS values. The RCS value is defined as the summation of parameter values for both the location and dog nodes. That is:

$$RCS_{i:j} = \sum_{i} \text{location node parameters}_i + \sum_{j} \text{dog node parameters}_j.$$ 

Therefore, suitable attributes or parameters are required to accurately quantify these contact strength values (or link weights) in between the nodes. As seen in Table 2, we have identified a total of five parameters for the location nodes as well as dog nodes set.

![Figure 4. An example of bipartite rabies contact (BRC) network](image)

**Table 2. Parameters for location node and dog node.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_d$</td>
<td>Dog population density</td>
</tr>
<tr>
<td>$R_0$</td>
<td>Reproduction number</td>
</tr>
<tr>
<td>$F_l$</td>
<td>Number of times a location is visited by a dog</td>
</tr>
<tr>
<td>$\beta_c$</td>
<td>Dog biting rate</td>
</tr>
<tr>
<td>$F_d$</td>
<td>Number of times a dog visited a location</td>
</tr>
</tbody>
</table>

For the location node parameters, the dog population density parameter ($P_d$) for a given location node can be written as:

$$P_d = \frac{\text{number of dogs in a given area, } N_d}{\text{area of the location node (km}^2\text{)}}$$

Since we assumed that stray dogs can travel up to 10 km$^2$, the area of each location node is set to be 10 km$^2$ therefore we have:

$$P_d = \frac{N_d}{10}$$

Reproduction number ($R_0$) among the stray dog population measures how factors such as rate of stray dog control and contact rate of rabid dogs among the dog population affects the birth rate of puppies. To align with the assumptions in our study, $R_0$ from (Asamoah et al., 2017) is modified to be:

$$R_0 = \frac{\gamma}{\gamma + \rho'}$$
where $\gamma$ gives the dog rabies incubation period and $\rho$ represents dog vaccination rate. The parameter $F_i$ is created to recognize the effect of dogs visiting a location and is defined as the frequency of a specific location node being visited by any dog node of the BRC network. Also, $F_i$ can be written in terms of a link matrix of the BRC network, $M(X_i, D_j)$, where

$$ F_i = \sum_{j=1}^{m} [M(X_i, D_j) \times F_{d_{ij}}], $$

and

$$ M(X_i, D_j) = \begin{cases} 1, & \text{if } D_j \text{ is visited by } X_i \\ 0, & \text{if } D_j \text{ is not visited by } X_i \end{cases}, $$

such that $X_i$ = location node of BRC network and $D_j$ = dog node of BRC network, whereas $i \in (1, 2, ..., n)$ and $j \in (1, 2, ..., n)$.

For the dog node parameters, we define the dog biting rate to be:

$$ \beta_c = \beta S_H I_D, $$

where $\beta$ = dog-to-human biting rate, $S_H$ = susceptible humans, and $I_D$ = infected dogs. The dog-to-human biting rate can be estimated from:

$$ \beta = \frac{\text{no. of confirmed human cases}}{\text{total no. of human in all location nodes}}. $$

The parameter $F_d$ measures the number of times that a dog visited a location and can be represented by:

$$ F_{d_{ij}} = \begin{cases} n, & \text{if } D_j \text{ is visited by } X_i \text{ where } n \in Z^+ \\ 0, & \text{if } D_j \text{ is not visited by } X_i \end{cases}, $$

Hence, after defining all parameters, the RCS values can be expressed as:

$$ RCS_{ij} = \sum \text{location node parameters}_i + \sum \text{dog node parameters}_j, $$

$$ RCS_{ij} = \sum [P_{d_{ij}} + R_{0_{ij}} + F_{ij}] + \sum \beta_c + F_{d_{ij}}. $$

The higher the RCS values, the stronger the strength between the two nodes; hence signifying the existence of more attachment between the location node and the specified dog node.
RESULTS

Between 1st April 2017 to 31st July 2017, we have determined 11 location nodes and 10 dog nodes. As shown in Figure 5, the RCS values are calculated based on the quantified concomitant parameters. Next, by using the RCS values as input, results of the RHR are computed via the HITS algorithm. The algorithm ranked the location nodes based on the RHR values which are tabulated in Table 3. In terms of node degree, X3, X4, and X6 had the highest degree of 2 as two dog nodes are attached to X3, X4, and X6 each (Figure 5). Higher-degree nodes are generally inclined to have stronger edges (Pavlopoulos et al., 2018). Based on Table 3 however, the X4 location node has the highest ranking value, hence it is worth noting that the node degree would not indicate the ranking of the location node directly. Therefore, X4 can be considered as a hotspot of rabies based on the BRC network model. This also indicates that X4 could be a possible location with the highest dog density.

Benchmark verification was performed by comparing the results obtained with the benchmark system produced by UCINET 6. The RMSE value of 0.02753 (corrected to four significant figures) which is less than the threshold value of 0.05 is obtained from the ranking of location nodes. This verifies our BRC network model. Additionally, analytical verification is performed and findings showed that RHR and hub matrix of the location nodes are correlated as proven by the SRCC value of 0.9818 obtained. The value is very close to 1, putting it at a position far above the threshold value of 0.7. This finding supports the claim made in benchmark verification, which further approves the BRC network model created in this study.

The critical parameters are determined by using two types of analysis which are “leave-one-out” and “multiple-out”. The SRCC values are recalculated after one or more parameters excluded from the network, then a comparison between the computed values with the original ranking is done. The SRCC value showed a sudden drop below the threshold value of 0.7 when parameters $P_d$, $R_0$ and $\beta c$ are excluded. This suggests that $P_d$, $R_0$ and $\beta c$ may be significant parameters. The decrease in SRCC value is relatively more when $R_0$, $F_t$ and $F_d$ are excluded from calculation. It is observed that SRCC values for both location and dog node decreased to 0.6970 and 0.6768 respectively. This indicates that $F_t$ and $F_d$ are essential parameters, due to $F_t$ and $F_d$ being the two main parameters that links the location and dog node. Thus, $F_t$ and $F_d$ must be kept in the model. Moreover, the exclusion of both $P_d$

Table 3. Location ranking of BRC network model.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Ranked Location</th>
<th>RHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X4</td>
<td>1.000000e+00</td>
</tr>
<tr>
<td>2</td>
<td>X6</td>
<td>5.833356e-01</td>
</tr>
<tr>
<td>3</td>
<td>X9</td>
<td>5.096758e-01</td>
</tr>
<tr>
<td>4</td>
<td>X10</td>
<td>4.807880e-01</td>
</tr>
<tr>
<td>5</td>
<td>X5</td>
<td>1.192827e-01</td>
</tr>
<tr>
<td>6</td>
<td>X3</td>
<td>1.426934e-09</td>
</tr>
<tr>
<td>7</td>
<td>X2</td>
<td>4.723061e-15</td>
</tr>
<tr>
<td>8</td>
<td>X8</td>
<td>1.265446e-16</td>
</tr>
<tr>
<td>9</td>
<td>X7</td>
<td>7.596613e-17</td>
</tr>
<tr>
<td>10</td>
<td>X1</td>
<td>1.425781e-17</td>
</tr>
<tr>
<td>11</td>
<td>X11</td>
<td>6.638513e-18</td>
</tr>
</tbody>
</table>
and $R_0$ suggests a low correlation to the original location and dog node ranking. When $P_d$ and $R_0$ are excluded from the calculation, the SRCC value still maintains a value above the threshold value. This possibly indicates that these two parameters are not significant parameters.

**DISCUSSIONS & CONCLUSION**

This study sought to construct a network model that detects hotspot areas of a rabies epidemic outbreak. To achieve this objective, a network model is proposed to describe the contact between the dog and the location that the dog visited. A modified epidemiological triangle is initially formulated in the first stage of BNB-MF in order to develop the bipartite graph structure. Two main components are identified to be the crucial components in rabies disease transmission – location and host properties.

In stage 2 of BNB-MF, a data pre-processing algorithm was developed and deployed to measure GPS coordinate of locations. The data pre-processing algorithm is also used to generate a distance matrix, which can be used to calculate distance between location nodes. Then, parameters for location and dog nodes are quantified, as well as the RCS using a summation rule. Hence, the BRC network consisting of 11 location nodes and 10 dog nodes is formed. Next, to rank each location and dog nodes, an adapted HITS algorithm is applied towards the BRC network to generate RHR value. Based on the results of our BRC network, it can be concluded that X4 is possibly the hotspot for rabies disease transmission. For stage 3 of BNB-MF, UCINET 6 is utilized to verify the BRC model formulated. A parameter significance analysis is conducted to measure the relative importance of each parameter quantified. Results of the analysis support the existence of a link between the parameters of the BRC network model.

Furthermore, due to the fact that this study is the first to ever utilize the bipartite network modelling method in rabies hotspot detection, no direct reference or immediate comparison of results can be done. This work only considers verification and analysis of results to validate the reliability of the model formulated. Also, this study only considers the stray dog population despite the fact that several human rabies cases in Sarawak have reported domestic dogs as the biting animal. As such, $F_l$ and $F_d$ are significant parameters in this study albeit difficult to obtain in real world. One of the possible methods to retrieve the data for parameter $F_d$ is by conducting interviews for the residents in the location. If the location is at micro-level such as a shop lot or a small community, then it is fairly easy to determine the parameter $F_d$ of a dog to the location. Consequently, the parameter $F_l$ can be calculated based on the quantification equation of $F_l$.

Besides, some other potential parameters that could potentially affect the transmission dynamics of rabies are yet to be tested, not to mention testing these parameters will require an additional set of data specifically recording attributes such as the environmental properties of the research areas (e.g., temperature, rainfall of a location node) or rabies disease state transition period. Hence, a potential future work might be to explore several other parameters and further understand the association link between all parameters involved in rabies hotspots detection study. Other than that, studies for locations other than Sarawak should be conducted so as to test and verify this current modelling approach.

**ACKNOWLEDGEMENTS**

The authors thank Universiti Malaysia Sarawak for all support in completing this research with the grant number F08/SpFRGS/1601/2017. Our heartfelt thanks also go to the authorities involved in maintaining Sarawak Disaster Information website, for supplementing us with research data.
REFERENCES


