University of Glasgow | College of Medical, Veterinary & Life Sciences

Reconstruction of Transmission Trees of FMD Virus During the 2001 UK Epidemic

Mikhail Churakov[1,2], Simon Rogers[1,3], Rowland R. Kao[1,3]

1 Boyd Orr Centre for Population and Ecosystem Health, University of Glasgow, Glasgow, United Kingdom
2 Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom
3 School of Computing Science, University of Glasgow, Glasgow, United Kingdom

Introduction

Reconstruction of transmission trees during an outbreak is not only a demanding and interesting task itself but can also help us to evaluate effectiveness of various control policies, identify mechanisms of transmission and better understand disease dynamics. Different types of data were used previously to determine directions of disease spread: contact tracing, temporal, spatial and genetic data. Cottam et al. [1] proposed a method to combine genetic and temporal data to identify transmission trees of foot-and-mouth disease virus.

In this work we elaborate the approach by Cottam et al. by taking into account information about locations of farms, discuss the assumption about independence of individual transmission links and present a method that can estimate likelihood of a whole tree structure by integrating conditional probabilities of individual links. We apply our approach and compare number of reconstructed transmission trees for a cluster of 15 farms that were infected during the 2001 foot-and-mouth epidemic in the UK.

Background

Cottam et al. [1] used the data of different types sequentially. Firstly, all the possible transmission trees that agreed with genetic data were enumerated. And then they were assessed using the likelihood function based on reporting dates. The likelihood function that farm i infected farm j was deter ined using two additional functions:

- $I_i(t)$ — the probability that farm i was first infected at time t — has a form of beta distribution and depends on the most likely date of infection from lesion dating (determines the mode) and the date it was reported of being infectious (determines the right limit).
- L_k — the probability that the first infected individual on a given farm incubates virus for k days prior to becoming infectious (distribution of incubation periods) — has a form of gamma distribution with mean value of 5 days and 95% of values from 2 to 12 days.

$F_i(t)$ — the probability that farm i is a source of infection at time t can be written as:

$$F_i(t) = \begin{cases} \sum_{\tau=0}^{t} \left(I_i(\tau) \cdot \left(\sum_{k=1}^{t-\tau} L(k) \right) \right) & t \le C_i \\ 0 & t > C_i \end{cases} \quad (1)$$

where C_i is the time at which farm i was culled.

Assuming that the farms couldn't be multiply infected and there are only n possible sources of infection, the likelihood of infection can be written as:

$$\lambda_{ij} = \frac{\sum_{t=0}^{min(C_i,C_j)} \left(F_i(t) \cdot I_j(t) \right)}{\sum_{k=1,k\neq j}^{n} \left(\sum_{t=0}^{min(C_k,C_j)} \left(F_k(t) \cdot I_j(t) \right) \right)} \quad (2)$$

Data of different types are available for 2001 FMD epidemic, these include: spatial data (locations of farms), temporal data (dates of farms being reported infectious and culled), genetic data (several RNA sequences from isolates on infected farms) and additional data about species presented on farms.

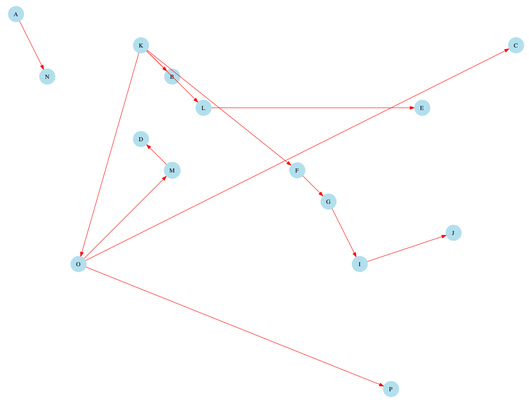


Figure 1: The most likely epidemic tree for 15 observed farms based on genetic and temporal data [1].

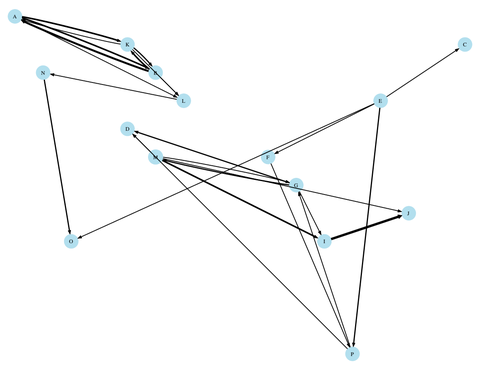
Methods

Adding spatial data

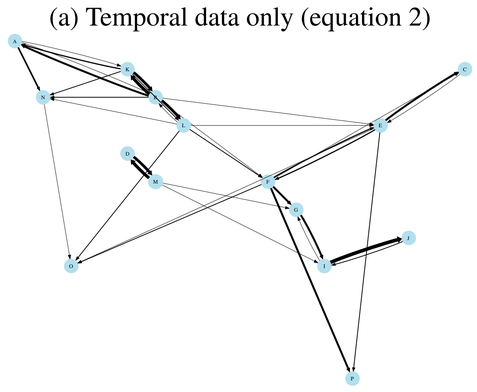
To take into account information that farms are not equally susceptible to infections from other farms we add to the likelihood function (equation 2) a multiplier $K(d_{ij})$ — the distance kernel function that shows at which rate an infectious farm i could have infected a susceptible farm j according to distance between them.

The likelihood function is therefore rewritten to include spatial information:

$$\lambda_{ij} = \frac{\sum_{t=0}^{min(C_i,C_j)} \left(F_i(t) \cdot I_j(t) \right) \cdot K(d_{ij})}{\sum_{k=1,k\neq j}^{n} \left(\sum_{t=0}^{min(C_k,C_j)} \left(F_k(t) \cdot I_j(t) \right) \cdot K(d_{kj}) \right)} \quad (3)$$



(a) Temporal data only (equation 2)



(b) Temporal and spatial data with information about species on farms (equation 3)

Figure 2: The most likely sources of infection for every single observed farm based on different types of data. The thickness of the links corresponds to the value of likelihood function. Closely situated farms (e.g. D and M, A and N) became more likely to infect each other, whereas likelihood of several events (e.g. G and P infecting D) became lower.

Forward simulations

In [1] possible transmission links between farms were considered independently — probability of a farm A infecting a farm B includes all the possible pairs of dates when A and B were infected. This means that we take into account scenarios when a source farm A was infected after a daughter case B. Therefore, by multiplying likelihoods of the edges we can not calculate the likelihood of a tree.

We can obtain likelihoods of trees (rather than likelihoods of certain edges) by forward simulations or integration of conditional probabilities to omit situations described above.

The likelihoods of infection transmissions should be adjusted during the course of epidemic as they depend on the day of epidemic and the current status of every single farm. If farm i was infected at day d_{inf} then function $I_i(t) = \begin{cases} 1 & t = d_{inf} \\ 0 & t \neq d_{inf} \end{cases}$. Which implies that $F_i(t) = \sum_{\tau=0}^{t} \left(I_i(\tau) \cdot \sum_{k=1}^{t-\tau} L(k) \right) = \sum_{k=1}^{t-d_{inf}} L(k)$. This will affect all the likelihoods λ_{ij} that need to be recalculated.

We used spatial SIR epidemic model with infection and removal stages on each step. Pre-calculated likelihood transmission matrix $M[i][j][t]$ was used to determine infection transmission between farms and was updated every time when infection event occurred. Index cases were fixed (A and K farms).

We performed series of forward simulations and obtained number of possible epidemic scenarios — transmission trees with associated infection dates. Epidemiological likelihood of a scenario was calculated as a product of likelihoods of transmission events.

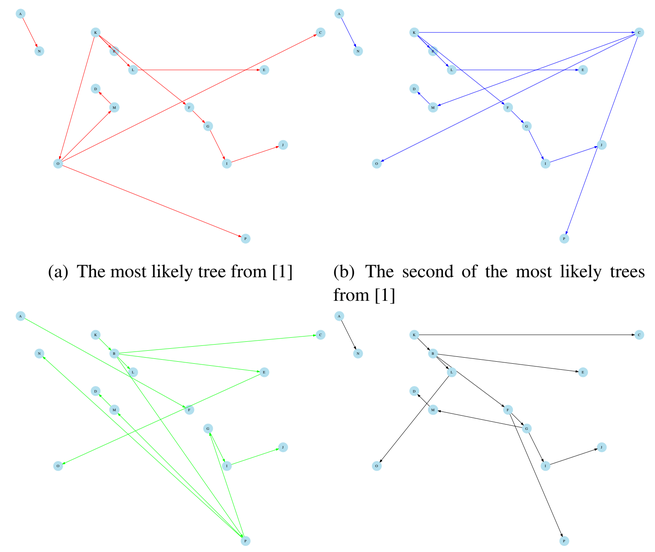
Sampling epidemic scenarios

Ideally, we would like to generate every possible epidemic scenario, but enumeration of all the tree structures (and, on top of it, all the possible timings of events) even for 15 farms would take enormous amount of time.

Using the described methods, we can calculate epidemiological likelihood of a certain epidemic scenario. To obtain the likelihoods of tree structures we need to sum up all the relevant scenarios. But the number of different scenarios for a fixed tree structure is enormous. Therefore, we decided to use Metropolis-Hastings algorithm to sample from a distribution of the likelihoods associated with scenarios for a fixed tree structure.

Results

We performed series of 10000 iterations of the Metropolis-Hastings algorithm for several tree structures:



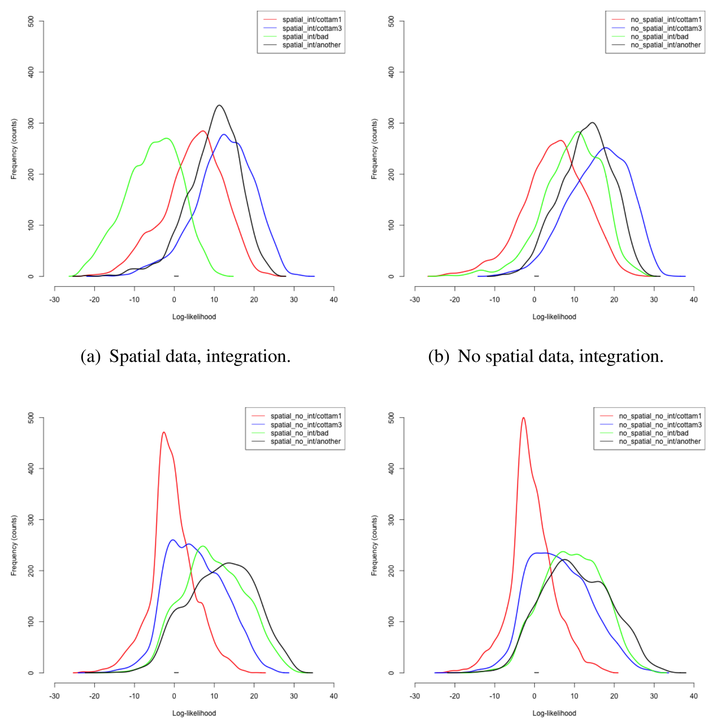
(a) The most likely tree from [1]

(b) The second of the most likely trees from [1]

(c) The tree obtained during the forward simulations (bad)

(d) The tree obtained during the forward simulations (another)

Figure 3: The transmission trees used in analysis.



(a) Spatial data, integration.

(b) No spatial data, integration.

(c) Spatial data, no integration.

(d) No spatial data, no integration.

Figure 4: Results of Metropolis-Hastings algorithm for different epidemiological models.

Discussion

Compared to previous work, we considered the full conditional probability to generate the likelihoods of simulated transmission trees. This requires different methods of comparing tree structures, as we showed significant variation of likelihoods (see Fig. 4) depending on actual timing of transmission events.

We also showed that adding spatial data to the likelihood function changes the predicted transmission pathways of disease (see Fig. 4).

Because our approach is simulation-based, we cannot be guaranteed of having identified the most likely epidemiological scenario, nevertheless our comparison does show that considerable caution must be maintained in attributing causality to any single epidemiological link.

References

[1] Eleanor M Cottam, Gaël Thébaud, Jemma Wadsworth, John Gloster, Leonard Mansley, David J Paton, Donald P King, and Daniel T Haydon. Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proceedings. Biological sciences / The Royal Society*, 275(1637):887–895, April 2008.