

# Towards the eradication of Ebola

Team 41523

February 9, 2015

## **Abstract**

TBD

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Background . . . . .	3
1.2	Transmission . . . . .	4
1.3	Medication . . . . .	5
<b>2</b>	<b>Problem Formulation</b>	<b>5</b>
2.1	Problem restatement . . . . .	5
2.2	Challenges . . . . .	5
<b>3</b>	<b>Disease Modelling</b>	<b>7</b>
3.1	Assumptions . . . . .	7
3.2	Network Building . . . . .	7
3.3	Differential equation model on the network . . . . .	7
3.3.1	Differential Equation model without considering population flow . .	8
3.3.2	Differential Equation model with considering population flow . . .	9
3.3.3	Two step model calibration . . . . .	9
3.4	Cascaded Poisson Process Model . . . . .	11
3.4.1	Proposed model . . . . .	11

# 1 Introduction

## 1.1 Background

The Ebola virus causes an acute, serious illness which is often fatal if untreated.

The current outbreak in west Africa, (first cases notified in March 2014), is the largest and most complex Ebola outbreak since the Ebola virus was first discovered in 1976. There have been more cases and deaths in this outbreak than all others combined.

The most severely affected countries, Guinea, Sierra Leone and Liberia have very weak health systems, lacking human and infrastructural resources, having only recently emerged from long periods of conflict and instability. On August 8, the WHO Director-General declared this outbreak a Public Health Emergency of International Concern.

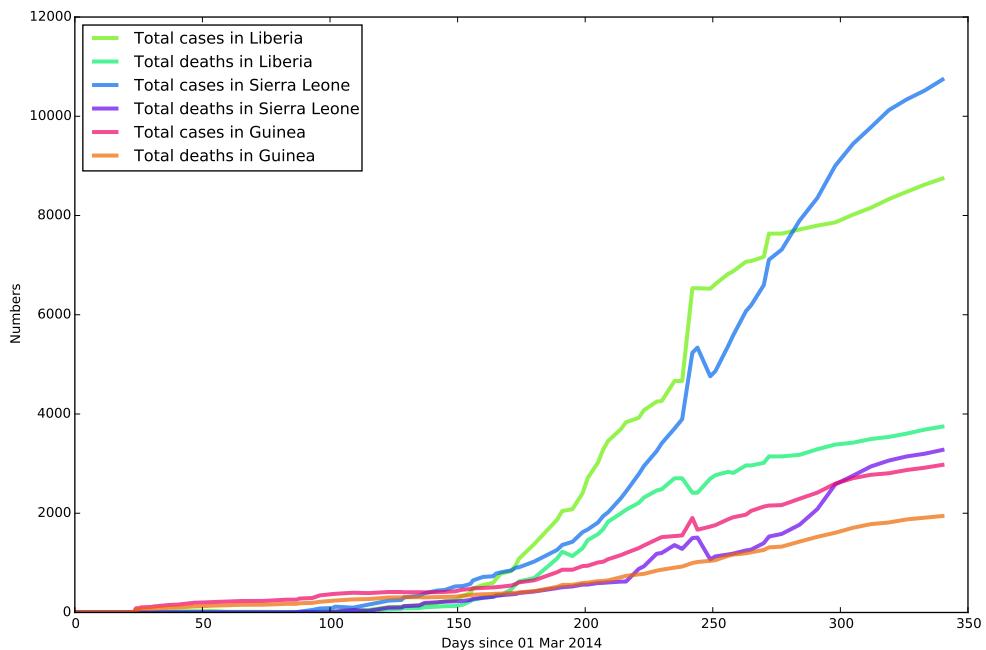


Figure 1: General statistics of Ebola outbreak in 2014

Fig. 1 shows the basic statistics of the Ebola outbreak in 2014. This is the cumulative cases and deaths in the three countries where the outbreak is the severe.<sup>1</sup>

<sup>1</sup>The hump of the curve origin from the inconsistency of the data source, it changed on Nov 12 2014.

## 1.2 Transmission

It is thought that fruit bats of the Pteropodidae family are natural Ebola virus hosts. Ebola is introduced into the human population through close contact with the blood, secretions, organs or other bodily fluids of infected animals such as chimpanzees, gorillas, fruit bats, monkeys, forest antelope and porcupines found ill or dead or in the rainforest.

Ebola then spreads through human-to-human transmission via direct contact (through broken skin or mucous membranes) with the blood, secretions, organs or other bodily fluids of infected people, and with surfaces and materials (e.g. bedding, clothing) contaminated with these fluids.

It is reported that [1], 72% of infections occurred in households or in the general community , 17.5% in hospitals, and 10.4% at funerals. This observation is important for our modelling.

Human-to-human transmission gives rise to the wide spread such disease. It has spread between countries starting in Guinea then spreading across land borders to Sierra Leone and Liberia, by air (1 traveller only) to Nigeria, and by land (1 traveller) to Senegal.

Figure. 2 shows the location of Ebola infected areas. We also got the geological location and other information such as population for the purpose of model calibration.

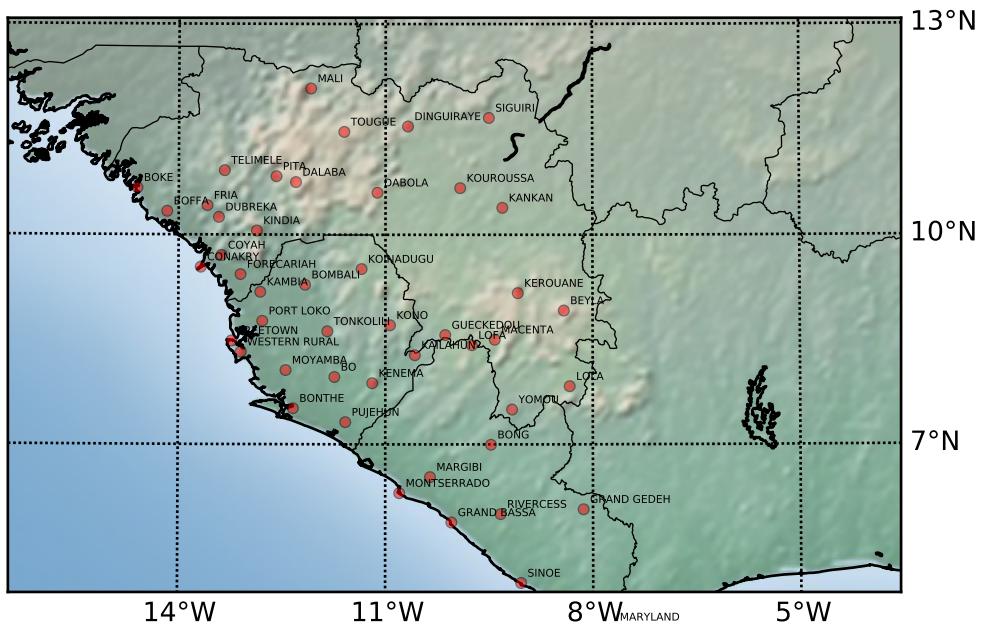


Figure 2: Location of Ebola infected cities

From observing the outbreak data in Fig. 3, we found there exists a correlation between infected areas in time. This observation validated that Ebola spread between countries on the land.

### 1.3 Medication

Supportive care-rehydration with oral or intravenous fluids- and treatment of specific symptoms, improves survival. There is as yet no proven treatment available for EVD. However, a range of potential treatments including blood products, immune therapies and drug therapies are currently being evaluated. No licensed vaccines are available yet, but 2 potential vaccines are undergoing human safety testing.

In the modelling, we assume medicine and vaccines are found for treating Ebola.

## 2 Problem Formulation

### 2.1 Problem restatement

The world medical association has announced that their new medication could stop Ebola and cure patients whose disease is not advanced.

The ultimate goal is to bring Ebola under control. To achieve this goal, a sensible model has to be built, taking not only the spread of the disease, the quantity of the medicine needed, possible feasible delivery systems, locations of delivery, speed of manufacturing of the vaccine or drug into consideration. Then use this model to optimize the eradication of Ebola.

It is natural to break this problem into two parts, the first part focuses on the Ebola itself. For this part, we need to build a model to describe how Ebola spreads and progresses.

The second part is to build an optimization model to optimize the eradication of Ebola based on the intrinsic properties of Ebola.

### 2.2 Challenges

Previous research focuses on the statistical level of disease modelling, thus detailed plan is not easily made from that. So it is challenging to build a fine grind model to assist detailed area-level healthcare plan-making.

The second challenge lies in the spread of disease. Many a time only one traveller from one area to another will cause infection and disease spread in that new area. Traditional statistical models views the population in as a whole, thus they are not able to track this

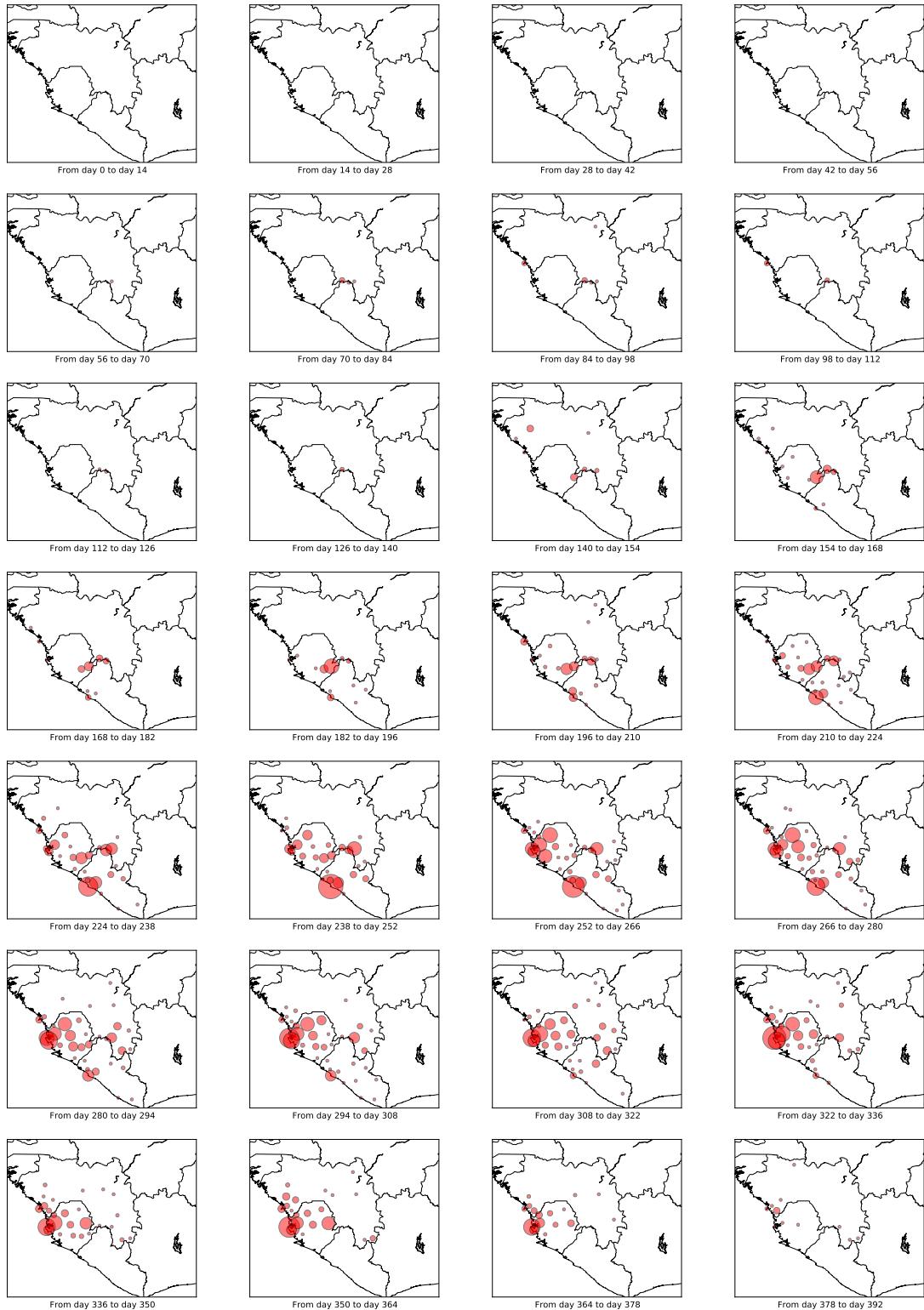


Figure 3: Spread of Ebola disease

phenomenon. The challenge is to view different areas as a network and consider the effect of population migration.

## 3 Disease Modelling

### 3.1 Assumptions

TBD

### 3.2 Network Building

The Ebola infected countries are divided into areas based on its political divisions. We denote the areas as nodes,  $N_i$ . And there exists links between every two nodes. We assign a weight  $\sigma_{i,j}$  on each edge to describe the ‘proximity’ of two areas.

Because we mainly consider the population flow, so we adopted the gravity model[2][3] in economics to estimate the population flow between to areas.

$$\sigma_{i,j} = \frac{Po_i^{\beta_1} Po_j^{\beta_2}}{d_{i,j}^{\beta_3}}, i \neq j$$

We denote the population of node  $i$  as  $Po_i$  and the geological distance between node  $i$  and node  $j$ . For simplification, we set:

$$\beta_1 = 1, \beta_2 = 1, \beta_3 = 2$$

And to calculate the distance between to nodes, we simply used the euclidean distance between two nodes.

We fetched the data of 55 infected areas. The link and the strength of links is visualized in 4, and the area of circles stands for the population.

### 3.3 Differential equation model on the network

We proposed an assisted model based on the network structure and an assisted traditional SIR model.

Other assumptions have to be made. We classify the people into one of the following states:

- S: Suspected, not infected but possible for infections
- I: Infected, but not go to hospital

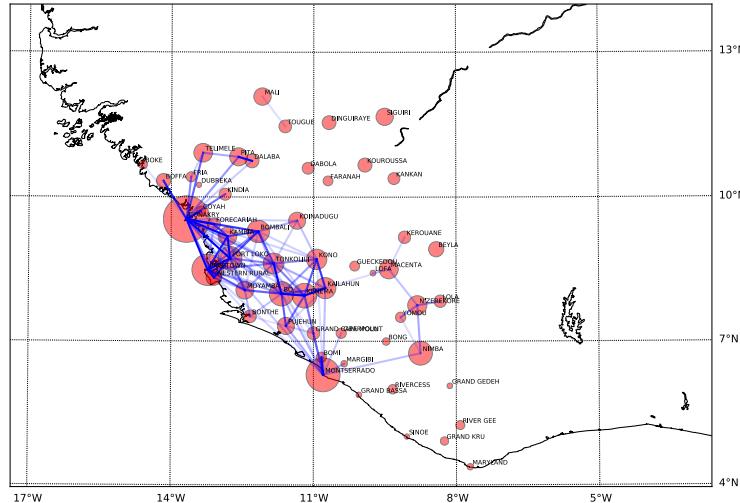


Figure 4: Underlying network of population flow

- H: Infected, and in hospital
- R: Removed, recovered or passed away

For simplification, we assume the ratio of death or recover is constant and can be observed from statistical data.

### 3.3.1 Differential Equation model without considering population flow

$N$  denotes for the whole population. Because there are three main causes of infection, *household transmission, clinical transmission, and funeral transmission*.

So the rate is proportional to three terms, corresponding to the three causes.

$$\frac{dS}{dt} = -\frac{\beta_1 SI}{N} - \frac{\beta_3 SH}{N} - \frac{\beta_2 S}{N} \frac{dD}{dt}$$

$$\frac{dI}{dt} = \frac{\beta_1 SI}{N} + \frac{\beta_3 SH}{N} + \frac{\beta_2 S}{N} \frac{dD}{dt} - kI$$

$$\frac{dH}{dt} = kI - \gamma H$$

$$\frac{dR}{dt} = \gamma H$$

Because the ratio of death or recover is constant and can be observed from statistical data, so

$$D = \eta R.$$

### 3.3.2 Differential Equation model with considering population flow

The underlying network is the main dynamic of disease sprawling.

For every node  $V_i \in V$

$$\begin{aligned}\frac{dS_i}{dt} &= -\frac{\beta_1 S_i I_i}{N_i} - \frac{\beta_3 S_i H_i}{N_i} - \frac{\beta_2 S_i}{N_i} \frac{dD_i}{dt} + \sum_{j \in V, j \neq i} \theta(\sigma_{i,j} S_j - \sigma_{j,i} S_i) \\ \frac{dI_i}{dt} &= \frac{\beta_1 S_i I_i}{N_i} + \frac{\beta_3 S_i H_i}{N_i} + \frac{\beta_2 S_i}{N_i} \frac{dD_i}{dt} - k I_i + \sum_{j \in V, j \neq i} \theta(\sigma_{i,j} I_j - \sigma_{j,i} I_i) \\ \frac{dH_i}{dt} &= k I_i - \gamma H_i \\ \frac{dR_i}{dt} &= \gamma H_i\end{aligned}$$

Because the ratio of death or recover is constant and can be observed from statistical data, so

$$D_i = \eta R_i.$$

### 3.3.3 Two step model calibration

The first step of calibration is to determine the parameters for the overall data.

The parameters are  $\beta_1, \beta_2, \beta_3, k, \gamma$ .

Write the equations in matrix form.

$$\vec{x} = \begin{pmatrix} S \\ I \\ H \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & -\beta_1 & -\beta_3 - \beta_2 \eta \gamma \\ 0 & \beta_1 - k & \beta_3 + \beta_\eta \gamma \\ 0 & k & -\gamma \end{pmatrix}$$

Considering the number infected is small comparing to the population. So

$$S_i \approx N_i, \frac{dS_i}{dt} \approx 0$$

$$\dot{\vec{x}} = A \vec{x}$$

The roots of equation  $|pI - A| = 0$  are:

$$p_1 = 0$$

$$p_{2,3} = \frac{-k + \beta_1 - \gamma + \sqrt{(\beta_1 - \gamma - k)^2 - 4(\beta_1 - \gamma - mk + k\gamma)}}{2}$$

The general solution is

$$\psi = A \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + B \begin{pmatrix} -m \frac{p_2+k}{p_2} \\ m \\ p_2 - \beta_1 + k \end{pmatrix} + C \begin{pmatrix} -m \frac{p_3+k}{p_3} \\ m \\ p_3 - \beta_1 + k \end{pmatrix}$$

From statistical data we set  $\beta_1 : \beta_2 : \beta_3 = 2 : 2 : 1$

So

$$\psi = B \frac{7}{6} \beta \left( 1 + \frac{1}{6.3(\beta - 0.163 + \sqrt{\beta^2 + 0.193\beta})} \right) e^{(\beta - 0.163 + \sqrt{\beta^2 + 0.193\beta})t}$$

We fit the data to the first 2/3 data and get  $\beta = ?$  TBD

The second step of calibration considered the population flow.

TBD

$$\theta = 1.4 \times 10^{-9}$$

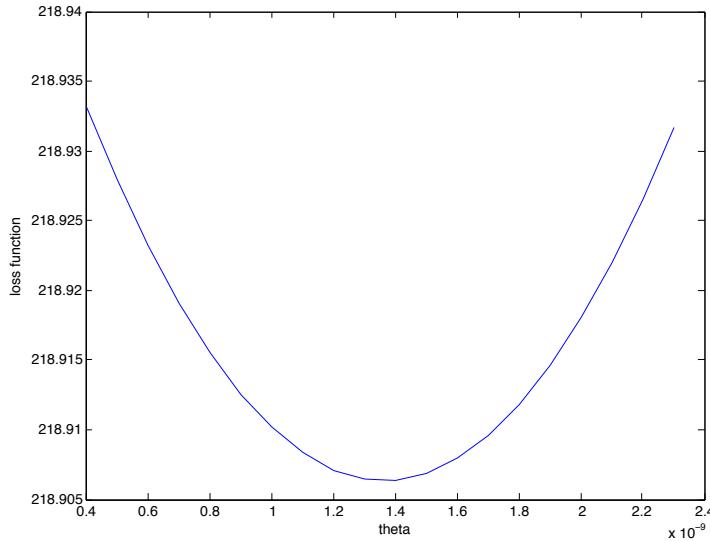


Figure 5: Estimated  $\theta$

## 3.4 Cascaded Poisson Process Model

### 3.4.1 Proposed model

Multi-dimensional Hawkes process[4] are used to model repeated events and influence between people. It provided inspiration for us to model the infections of disease as a inhomogeneous Poisson Process to track the temporal and spatial behaviour of disease transmissions.

From Fig. 3 we can see the new infections, we model the new infections as a diffusion process. We consider the diffusion process using discrete time steps.

One person being infected is a point process, so viewing it from a higher level, the infection is a Poisson Process.

The intensity of the Poisson Process is related to many factors.

The intensity in one city/area depends on two parts, one part comes from itself, and the second part comes from all other cities.

$$\mu_{i,t_k} = \lambda_1 \sum_{m=1}^{k-1} P_{i,t_m} e^{-\beta(k-1-t)} + \lambda_2 \sum_{j:j \neq i} \sum_{m=1}^{k-1} P_{j,t_m} \sigma_{j,i} e^{-\beta(k-1-t)}$$

We do a MLE to fit the data to the model and get  $\lambda_1$  and  $\lambda_2$ .

The likelihood function is:

$$L(\lambda_1, \lambda_2) = \prod_{i \in V} \prod_k \frac{(\mu_{i,t_k} \Delta t)^{P_{i,t_k}} e^{-\mu_{i,t_k} \Delta t}}{P_{i,t_k}!}$$

$$\log L(\lambda_1, \lambda_2) = \sum_{i \in V} \sum_k [P_{i,t_k} \log (\mu_{i,t_k} \Delta t) - \mu_{i,t_k} \Delta t - \log P_{i,t_k}]$$

$$\frac{\partial \log L(\lambda_1, \lambda_2)}{\partial \lambda_1} = \sum_{i \in V} \sum_k \left[ P_{i,t_k} \frac{\frac{\partial \mu_{i,t_k}}{\partial \lambda_1}}{\mu_{i,t_k}} - \frac{\partial \mu_{i,t_k}}{\partial \lambda_1} \Delta t \right]$$

$$\frac{\partial \log L(\lambda_1, \lambda_2)}{\partial \lambda_2} = \sum_{i \in V} \sum_k \left[ P_{i,t_k} \frac{\frac{\partial \mu_{i,t_k}}{\partial \lambda_2}}{\mu_{i,t_k}} - \frac{\partial \mu_{i,t_k}}{\partial \lambda_2} \Delta t \right]$$

Solve for equation 1:

$$\frac{\partial \log L(\lambda_1, \lambda_2)}{\partial \lambda_1} = \sum_{i \in V} \sum_k \left[ \frac{P_{i,t_k} \sum_{m=1}^{k-1} P_{i,t_m} e^{-\beta(k-1-t)}}{\lambda_1 \sum_{m=1}^{k-1} P_{i,t_m} e^{-\beta(k-1-t)} + \lambda_2 \sum_{j:j \neq i} \sum_{m=1}^{k-1} P_{j,t_m} \sigma_{j,i} e^{-\beta(k-1-t)}} \right. \\ \left. - \left( \sum_{m=1}^{k-1} P_{i,t_m} e^{-\beta(k-1-t)} \right) \Delta t \right]$$

solve for equation 2:

$$\frac{\partial \log L(\lambda_1, \lambda_2)}{\partial \lambda_2} = \sum_{i \in V} \sum_k \left[ \frac{P_{i,t_k} \sum_{j:j \neq i} \sum_{m=1}^{k-1} P_{j,t_m} \sigma_{j,i} e^{-\beta(k-1-t)}}{\lambda_1 \sum_{m=1}^{k-1} P_{i,t_m} e^{-\beta(k-1-t)} + \lambda_2 \sum_{j:j \neq i} \sum_{m=1}^{k-1} P_{j,t_m} \sigma_{j,i} e^{-\beta(k-1-t)}} \right. \\ \left. - \left( \sum_{j:j \neq i} \sum_{m=1}^{k-1} P_{j,t_m} \sigma_{j,i} e^{-\beta(k-1-t)} \right) \Delta t \right]$$

All the terms are positive. So use gradient descent to find the solution.

$$\lambda_1^* = \lambda_1 - \gamma \frac{\partial \log L(\lambda_1, \lambda_2)}{\partial \lambda_1}$$

$$\lambda_2^* = \lambda_2 - \gamma \frac{\partial \log L(\lambda_1, \lambda_2)}{\partial \lambda_2}$$

$\gamma$  is the step size.

We get  $\beta$  from statistical data.

$$\beta = 0.05$$

## 4 Disease Controlling

## Appendix

## References

- [1] S. Merler, M. Ajelli, L. Fumanelli, M. F. Gomes, A. P. y Piontti, L. Rossi, D. L. Chao, I. M. Longini, M. E. Halloran, and A. Vespignani, “Spatiotemporal spread of the 2014 outbreak of ebola virus disease in liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis,” *The Lancet Infectious Diseases*, 2015.
- [2] J. E. Anderson, “The gravity model,” tech. rep., National Bureau of Economic Research, 2010.
- [3] D. Karemera, V. I. Oguledo, and B. Davis, “A gravity model analysis of international migration to north america,” *Applied Economics*, vol. 32, no. 13, pp. 1745–1755, 2000.

- [4] K. Zhou, H. Zha, and L. Song, “Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes,” in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pp. 641–649, 2013.