Practical Models towards the Eradication of Ebola

1 Introduction

Ebola virus disease, or Ebola hemorrhagic fever, is a severe epidemic with a high risk of death, killing average 50 percent of those it infects [3]. The virus is receiving rapidly increasing attentions due to its recent unprecedented outbreak in West Africa, in which Ebola has infected over 22,000 people and killed over 9,000 people according to World Health Organization [4].

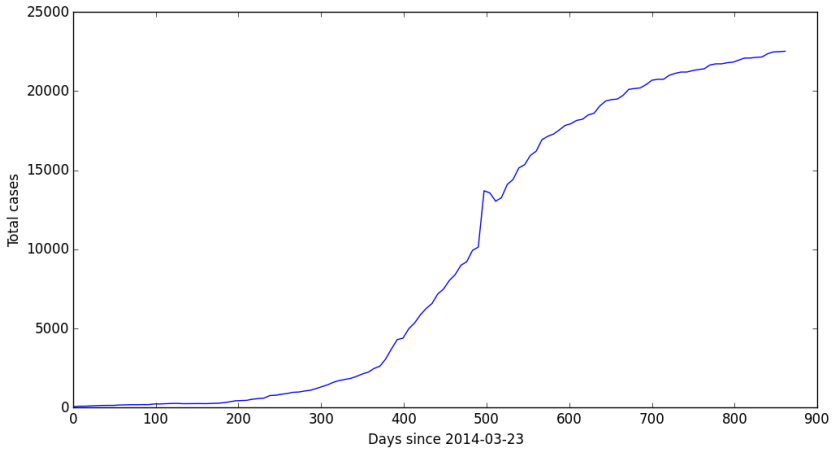


Figure 1: Total cases plotted using data from WHO situation reports

In this paper, we assume that a new medication for curing early stage of Ebola has been invented according to the problem statement, based on which we build our thesis basically in four steps. First, we render a deterministic epidemic model to study the pathological traits on how the virus propagates. Next, we find out a viable plan to construct medication distribution centers with the help of graphs modelling the traffic network among dozens of districts in epidemic area. Then we simulate the estimated efficiency of the medication based on the model we created previously, and after that, factors in the model are quantified by multi-variable linear regression algorithm to measure the influence of each.

2 Background

2.1 Restatement of Original Problem

A new medication announced by World Medication Association is said to be effective in curing patients in early stage of Ebola and a plan is required to distribute and evalutate the medication. The solution should consider following factors:

1. the spread of the disease (pathological traits)
2. the manufacturing of the medicine

3. fesible delivery systems including locations of delivery

The main goal for the plan is to eradicate Ebola from West Africa.

2.2 Assumptions

To abstract from reality and simplify the model we will discuss in following sections, a few reasonable assumptions need to be made to clarify implementation details:

- Ignore migration effect of the population.
- Ignore inter-influence among regions in Modified SEIR.
- Ignore natural birth rate and death rate due to the data was collected over a short period.
- Assume that people recovered from Ebola get permanet immunity.
- Assume that the smallest unit of transportation time is week.

3 Solution

3.1 Deterministic Epidemic Model

In this section, we will discuss Modified SEIR, a deterministic epidemic model for simulating the spread of the virus. The model divides population into 6 compartments, which is shown in Figure 2.

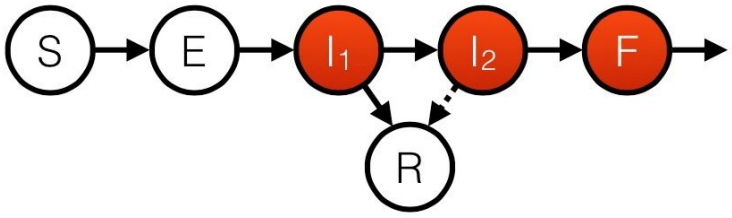


Figure 2: Modified SEIR epidemic model

According to previous research, Ebola has an average incubation period of 8-10 days during which it has nearly no infectivity [3], and this period is represented by compartment E. Then, the virus progresses to two broad stages [3][9], which are represented as compartment I1, I2 in Figure 2. The first stage comes with mild symptoms, such as fever and headache, while the syndrome of the second one is more severe. Recovery, represented by compartment R, is more likely to take place at stage I1, in contrast with a high mortality rate at stage I2. Compartment F indicates dead but not buried, which gradually fade out from the model.

We examine the data of suspected cases in each district reported by World Health Organization [4]. To resolve the overall characteristics of the disease, the cases of all countries combined are analyzed to estimate the principal parameters for further use.

	Definition	Units	Range
β_1	Transmission parameter for first stage of illness	$\text{people}^{-1}\text{days}^{-1}$	Estimated
$\frac{\beta_2}{\beta_1}$	Ratio of infectiveness in first vs. second stage of illness	unitless	1.5-5
$\frac{\beta_F}{\beta_1}$	Ratio of infectiveness in first stage vs. funeral transmission	unitless	1.5-5
α^{-1}	Average incubation period	days	8-10
γ_1^{-1}	Length of first stage of illness	days	5-7
γ_2^{-1}	Length of second stage of illness	days	1-2
γ_F^{-1}	Time from death until burial	days	1-3
γ_R^{-1}	Duration of ETU bed occupancy after recovery	days	5-15
δ	Overall mortality rate (fraction)	unitless	Estimated
δ_2	Mortality rate (fraction) among those in second stage of illness	unitless	0.9-1
k	Fraction of the population at risk, symptomatic fraction, and reporting rate	unitless	0.001-1

Table 1: Notations used in discussion

3.1.1 Notations

Several notations and definitions used in further discussions are listed in Table 1.

3.1.2 Structure of The Model

As described above, Ebola often progresses over multiple stages of illness an initial infectious stage in which symptoms tend to be milder often progressing to diarrhea, vomiting, and a second, more intense stage during which the more advanced symptoms (such as hemorrhaging and multi-organ failure) manifest [9], with a typical time to progression of approximately 5-7 days [1]. Most recoveries occur from the first stage, while the second stage is typically fatal [7]. The corresponding equations for each compartment are shown below:

$$\frac{dS}{dt} = -(\beta_1 I_1 + \beta_2 I_2 + \beta_F F)S \tag{1}$$

$$\frac{dE}{dt} = (\beta_1 I_1 + \beta_2 I_2 + \beta_F F)S - \alpha E \tag{2}$$

$$\frac{dI_1}{dt} = \alpha E - \gamma_1 I_1 \tag{3}$$

$$\frac{dI_2}{dt} = \delta_1 \gamma_1 I_1 - \gamma_2 I_2 \tag{4}$$

$$\frac{dF}{dt} = \delta_2 \gamma_2 I_2 - \gamma_F F \tag{5}$$

$$\frac{dR}{dt} = (1-\delta_1)\gamma_1 I_1 + (1-\delta_2)\gamma_2 I_2 - \gamma_R R \tag{6}$$

S describes the fraction of people who are susceptible here. E represents the fraction of people who are in the incubation period. I1 represents the people who are in the first stage of infectious while I2 represents the second one. F is the fraction of people who are recently killed by the disease, which means that they can still be the sources of infection. Finally, let R represent the people who are recovered from Ebola.

The recovered individuals still have some chance of transmission after recovery, but here we do not include this fact for simplicity. Whats more, because the data we use is of a relative short period of time, the births and deaths of population are also ignored.

The first equation shows that the susceptible people are infected by the infected people as well as the recently dead ones due to some ratio. The negative multinomial in the second equation shows the natural process of Ebola, or the people in the incubation period are transformed to the first stage of infectious. The equation set is established by this principle.

To apply the model, we adopt following equations:

$$y_C = kN \int_0^t \alpha E dt \tag{7}$$

$$y_D = kN \int_0^t \delta_2 \gamma_2 I_2 dt, \tag{8}$$

where y_C represents the cumulative cases, and y_D cumulative deaths. Parameter k is a combination of several factors.

3.1.3 Parameter Estimation

As the model has been set up, we need estimate the parameters to uncover the underlying of the transmission of Ebola. Let $\delta = \delta_1 \delta_2$ be the general mortality rate.

Due to the wide range of parameter values can be used to generate the same initial epidemic growth rate, there might be different points in time at which the epidemic dynamics begin to deviate from exponential growth. Thus, we adopted a Latin Hypercube (LH) algorithm in which we sample most of our parameters across realistic ranges given in Table 1. Our result is shown in Table 2.

α	0.1	$\frac{\beta_2}{\beta_1}$	3.38
$\frac{\beta_F}{\beta_1}$	3.44	δ_2	0.99
γ_F	0.36	γ_2	0.53
γ_1	0.19	k	0.0026

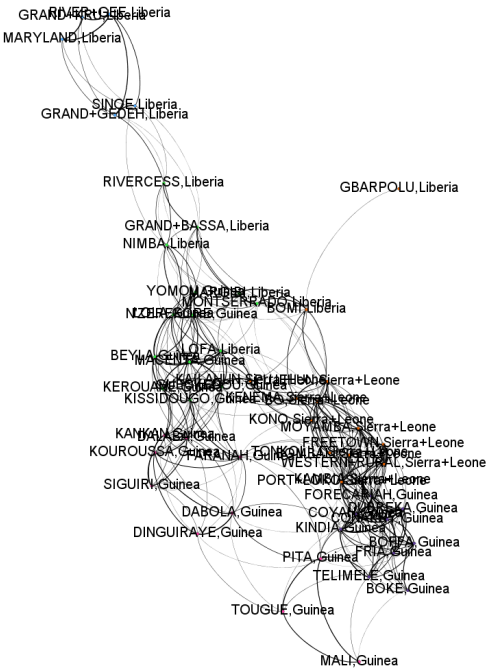
Table 2: Estimated parameters

3.2 Delivery of New Medication

In this section, we focus on the network analysis among 52 districts included in sitation reports [4] announced weekly by World Health Organization across three countries, Guinea, Sierra Leone and Liberia. To model a realistic traffic network, we first gathered the information on estimated duration spent on driving from one district to another from Google Distance Matrix API [2], which is done via Python. Partial data is listed in Table 3.

	Mali	Boke	Conakry
Mali	1 min	4 hours 55 mins	6 hours 48 mins
Boke	4 hours 53 mins	1 min	4 hours 8 mins
Conakry	6 hours 48 mins	4 hours 9 mins	1min

Table 3: Estimated driving durations among Mali, Boke, Conakry



From the data we collected, we derived a undirected complete graph $G = (V, E)$ modelling the traffic network in three countries, where V is a collection of districts and E contains edges connecting every pair of nodes in V. Weight $w(e)$ of edge e are normalized from durations $d(u, v)$ between districts u and v:

$$w(e) = 1 - \frac{(d(u,v) + d(v,u))}{2L},$$

where L is the limitation of maximum duration. In this model, we assume $L = 18000$, which is 5 hours considering the fatigue of long-time driving.

Since Google Distance Matrix API has provided the duration in numerical format, we can store the network graph in a adjacency matrix. The graph is plotted on the left image using Gephi [5]. Thickness of edges indicates the distance between two nodes connected, and negative weighted edges are omitted.

3.2.1 Case Locality

While the cause of Ebola virus disease still remains uncovered, it is possible that cases exists in which the patient is infected by external factors we did not take into consideration, such as the migration of the infected animals. To exclude external factors and do further analysis in next section, we need verify whether the cases are geographically related, that is, one case located closely to one another.

For a case c, let $l(c)$ be the district (or a node in the graph) where the case occurs, $P(c)$ be the collection of cases occured a week before case c. Define locality $L(c)$ for a case c:

$$L(c) = \begin{cases} 1, & \text{if any case occured in the same district} \\ \sum_{p \in P(c)} \frac{w(l(p),l(c))}{||P(c)||}, & \text{otherwise} \end{cases}$$

where $||P(c)||$ is the number of element in set $P(c)$.

To reduce the effect of constant part (that is, the place where plague is widely spreading), we assume 100 cases occured in the location regardless of the number of cases reported. A randomly-

distributed curve are also added for comparison, in which cases from new locations are randomly chosen while locality of cases from old locations still counts 1. Thus we aggregate and plot data in comparison with total cases and random-distributed cases, which is shown in Figure 3.

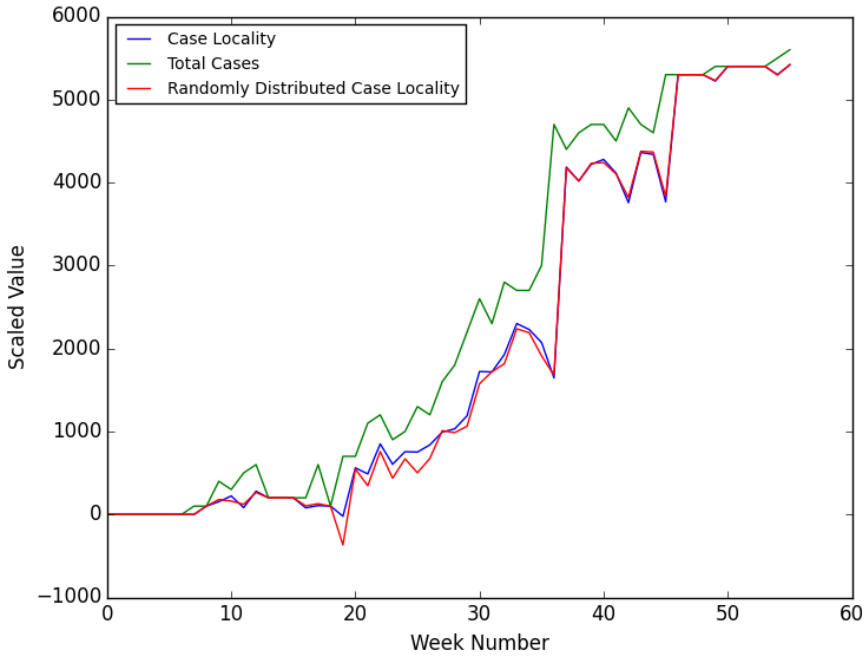


Figure 3: Case locality in comparison with total cases and random-distributed cases

From the graph we can observe that the curve of case locality resembles to the curve of total cases in a large part, and its estimation is significantly better than randomly distributed cases, especially when the plague is spreading among cities, indicating the geographical distribution of cases is intense and the dominating factor of epidemic is contributed by human contact.

3.2.2 Locating Distribution Centers

To eradicate the plague using new medication, we must determine which place or places become main medication distribution centers, covering their areas for the fastest medication transportation. To achieve this, community structure of the network is extracted using Blondel's method [6].

Modularity of a network [10] is defined as follows:

$$Q = \frac{1}{2m}\sum_{vw}(A_{vw} - \frac{k_v * k_w}{2m})\frac{s_v s_w + 1}{2},$$

where s is a membership variable (that is, $s_u = 1$ if u belongs to community 1 or $s_u = -1$ if u belongs to community 2), k_u is the degree of node u, and A is the adjacent matrix of the graph. Algorithm is based on the fact that the gain in modularity ΔQ obtained by moving an isolated node i into a community C can be computed by:

$$\Delta Q = (\frac{\sum_{in} + k_{i,in}}{2m} - (\frac{\sum_{tot} + k_i}{2m})^2) - (\frac{\sum_{in}}{2m} - (\frac{\sum_{tot}}{2m})^2 - (\frac{k_i}{2m})^2)$$

For detailed explanation of the formula, please refer to [6]. With 1.0 resolution [8] and weight-based detection, we got 5 communities with 0.506 modularity, part of whose members are listed in Figure 4. Colored graph representing different communities is shown in Figure 4.

#1	#2	#3	#4	#5
KENEMA	RIVER GEE	BOKE	MONTSERRADO	MALI
FREETOWN	GRAND GEDEH	CONAKRY	KISSIDOUGO	TOUGUE
MOYAMBA	SINOE	BOFFA	LOFA	FARANAH
TONKOLILI	GRAND KRU	DUBREKA	MARGIBI	PITA
PUJEHUN	MARYLAND	FORECARIAH	N'ZEREKORE	DALABA
BOMI		KAMBIA	GRAND BASSA	SIGUIRI
BOMBALI		COYAH	KEROUANE	DABOLA
KONO		TELIMELE	NIMBA	DINGUIRAYE
...		...	...	...

Table 4: Partial member list for communities

From above we can conclude that at least 5 main medication distribution centers should be built for each community in the graph to achieve best transportation efficiency. To further locate distribution centers precisely, graphs for each community are extracted from old graph and parsed in order to find the central node.

For each graph $G'(V', E')$ extracted, let α be the factor of how important the number of cases occured in districts matters and $T(u)$ be the total cases occured in district u, we develop a function $M(u)$ for measuring penalty in creating a distribution center in district u:

$$M(u) = \sum_{v \in V'} d(u, v) * (5 - \exp(-10^{-\alpha} * T(v)))$$

The penalty includes two part: a) the distance between u and all other districts in graph G', b) multipliers max up to 5 punishing delayed delivery to districts having many cases.

Compute $M(u)$ for every district in the individual community graphs, and result of central points is shown in Table 5. Locations of those distribution centers are plotted using Google Maps service in Figure 5.

BO, Sierra Leone	RIVER GEE, Liberia	COYAH, Guinea
N'ZEREKORE, Guinea	KOUROUSSA, Guinea	

Table 5: Best locations of distribution centers for communities

3.3 Simulation of Applying Medication

In this section, based on the pathological characteristics and the map we get in the previous parts, we are going to use the Monte Carlo method(MC method) to simulate what circumstance will appear if we put medication in this area.

Further more, we will apply Multi-variable linear regression equation to analyze the affect of each factor to give a general guidance for the eradication of ebola with the data we record.

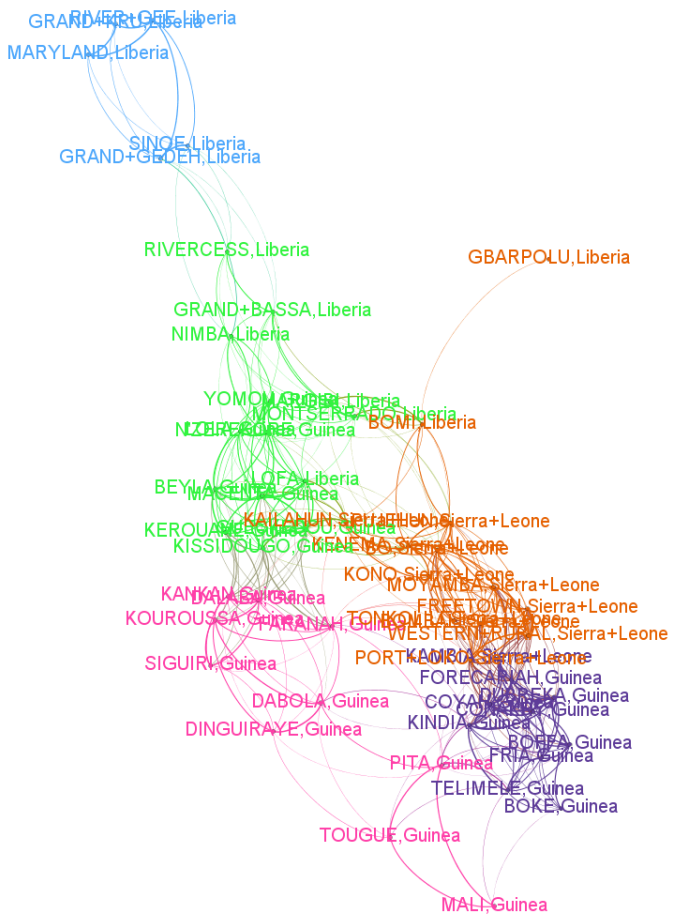


Figure 4: Colored graph based on community detection

3.3.1 Structure of Model

We suppose that the medication is produced in a certain speed for a long time, and the medication is delivered to the epidemic area every week. Let V be the amount of medication delivered per week.

According to the delivery system built in previous section, it is reasonable to believe that the medication can delivered to all the districts nearly with a minimum delay. Label all the district, let V_i be the medication each district can get per week.

Yet there's no widely accepted medication in use now. We make the efficiency of the drug as variable. Suppose the medication can make a slightly infected individual (individual who is in state E or stateI_1) cured with the possibility of p_1 and make a individual badly infected cured with the possibility of p_2.

Let $s_{i,0}$,$e_{i,0}$,$i_{1_{i,0}}$,$i_{2_{i,0}}$,$f_{i,0}$ be the number of people at different state in the district with label i right before the first week of medication applying, and let $c_i$be the cumulative case in district i, n_i be the population in this area. Set $f_{i,0} = 0$ as we have no data to trace it, and we ignore the influence of the number recovered people because it doesn't appear other equations. The initial

Figure 5: Map plotted for distribution centers

number of these variables in district i can be settled as:

$$e_{i,0} = c_i * \sigma * \frac{1}{\alpha}/(\frac{1}{\gamma_1} + \frac{1}{\gamma_1} + \frac{1}{\gamma_2}) \tag{9}$$

$$i_{1_{i,0}} = c_i * \sigma * \frac{1}{\gamma_1}/(\frac{1}{\gamma_1} + \frac{1}{\gamma_1} + \frac{1}{\gamma_2}) \tag{10}$$

$$i_{2_{i,0}} = c_i * \sigma * \frac{1}{\gamma_2}/(\frac{1}{\gamma_1} + \frac{1}{\gamma_1} + \frac{1}{\gamma_2}) \tag{11}$$

$$s_{i,0} = n_i - c * (1 - \sigma) \tag{12}$$

The fraction of these variable are updated at the mean time. We use the fraction of people who are at the incubation period (who are at state E) in district i as example:

$$E_{i,k} = \frac{e_{i,k}}{n_i}$$

Suppose the medication is first provided to the people at state I_2 and I_1 and then randomly distributed.Let t_0 be the time duration of a week. The dynamic change of every symbol can be

described as:

$$i_{1_{i,k+1}} = [i_{1_{i,k}} + (\alpha * E_{i,k} - \gamma_1 * I_{1_{i_k}}) * t_0] * (1 - p_1) \tag{13}$$

$$i_{2_{i,k+1}} = [i_{2_{i,k}} + (\sigma_1 * \gamma_1 * I_{1_{i,k}} - \gamma_2 * I_{2_{i,k}}) * t_0] * (1 - p_2) \tag{14}$$

$$f_{i,k+1} = fi, k + (\sigma_2 * \gamma_2 * I_{2_{i,k}} - \gamma_F * F_{i,k}) * t_0 \tag{15}$$

$$e_{i,k+1} = ei, k + [(\beta_1 * I_{1_{i,k}} + \beta_2 * I_{2_{i,k}} + \beta_F * F_{i,k}) * S_{i,k} - \alpha * E_{i,k}] * t_0 \tag{16}$$

$$s_{i,k+1} = s_{i,k} - (\beta_1 * I_{1_{i,k}} + \beta_2 * I_{2_{i,k}} + \beta_F * F_{i,k}) * S_{i,k} * t_0 \tag{17}$$

Using the MC method, we can get the condition with Ebola, applying different strategy. Further more we can get a most optimal strategy when some parameters are settled.

3.3.2 Standard of Disease Control

An important problem to solve is to determine the standard to measure the condition of the disease. Inspired by the basic reproduction number, the relative difference between the number of newly infected cases and previous infected ones can be used as a measurement, we notice that the relative difference between the number of newly infected cases and previous infected ones can be used as a measurement. If the relative difference is less than 1, it means that the disease is going to fade, and the relative difference can be described as:

$$r = \frac{e_{k+1} + i_{1_{k+1}} + i_{2_{k+1}} - (e_k + i_{1_k} + i_{2_k})}{e_k + i_{1_k} + i_{2_k}}$$

The disease is under the process of eradication if r ¡ 0, and its current strain is held back if r = 1.

3.3.3 Multi-variable linear regression equation

Assume that every affecting factors has linear relation with Y, let X1, X2, X3, X4, X5, X6, X7 be the amount of drugs to restrain Ebola Virus, susceptible age structure, delivery system, delivery position, the speed of drug production, the speed of spread and other factors respectively. Thus, the multi-variable linear regression model is:

$$Y_t = \alpha_1 * x_1 + \alpha_2 * x_2 + \alpha_3 * x_3 + ... + \alpha_7 * x_7 + \epsilon$$

Using SPSS to analyze, we can see correlation between Y and each X_i, by which we can find the key factor of the process of eradication.

4 Model Evaluation and Improvements

4.1 Accuracy

In the simulation part, we use the pathological characteristics of all the countries combined to simulate the condition in a small area. The accuracy of this could be improved by calculating the pathological characteristics for ebola in each area. Yet we still face the problem of lacking of necessary data, as the death cases are not listed for a specific area.

We have to make a compromise of accuracy and the demands for data. In another word, we calculate the transmit characteristics in a relative large district, and apply different parameters to different area.

4.2 Authenticity

The authenticity of the simulation is the key factor to ensure the reliability of out model. The problem here is the growth of the variable such as e_i doesn't happen in the order we stipulate. To decrease the influence of this drawback. We randomly proceed the change of each variable. The authenticity can be improved by this simple method.

References

[1] Ebola hemorrhagic fever — cdc. `www.cdc.gov/vhf/ebola/`. Accessed: 2015-02-09.

[2] Google distance matrix api. `https://developers.google.com/maps/documentation/distancematrix/`. Accessed: 2015-02-09.

[3] Who — ebola virus disease fact sheet. `http://www.who.int/mediacentre/factsheets/fs103/en/`. Accessed: 2015-02-09.

[4] Who — situation reports with epidemiological data: archive. `http://www.who.int/csr/disease/ebola/situation-reports/archive/en/`. Accessed: 2015-02-09.

[5] Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy, et al. Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362, 2009.

[6] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[7] Marisa C Eisenberg, Joseph NS Eisenberg, Jeremy P D'Silva, Eden V Wells, Sarah Cherng, Yu-Han Kao, and Rafael Meza. Modeling surveillance and interventions in the 2014 ebola epidemic. *arXiv preprint arXiv:1501.05555*, 2015.

[8] Renaud Lambiotte, J-C Delvenne, and Mauricio Barahona. Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*, 2008.

[9] Roger Ndambi, Philippe Akamituna, Marie-Jo Bonnet, Anicet Mazaya Tukadila, Jean-Jacques Muyembe-Tamfum, and Robert Colebunders. Epidemiologic and clinical aspects of the ebola virus epidemic in mosango, democratic republic of the congo, 1995. *Journal of Infectious Diseases*, 179(Supplement 1):S8–S10, 1999.

[10] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

A Non-technical Letter

Ebola hemorrhagic fever is a severe epidemic caused by ebolaviruses, which spreads by direct contact with body fluids of infected individuals. It quickly becomes a public concern due to a unprecedented outbreak in West Africa dated back to 2013. Since the last year, more than 22,000 individuals has been infected, over 9,000 of whom were dead. However, the situation now hopefully comes to an end. Recently scientists from World Medical Association has discovered a new medicine, which can cure Ebola in its early stage.

Through clinical experiments we find out that the performance of the new medicine is amazingly satisfied. As soon as the medicine is evaluated and approved in an accelerated process by FDA, which elevates testing possible treatments for Ebola as its first priority, the medicine can reach every corner of United States within 24 hours.

At the meantime, our factories are prepared to massively produce the medicine to ease and regain control of the horrible situation in West Africa. World Health Organization has helped us set up 5 distribution center for medicine in following location: Bo, Sierra Leone; River Gee, Liberia; Coyah, Guinea; N'Zerekore, Guinea; Kouroussa, Guinea. This will ensure the shortest transportation time in comparsion to the price reaching every district requiring for help across epidemic area.

We also urge residents that there is no need to panic, since only the signs and symptoms typically starts, could the viruses spreads from prime patients. Most of the cases of Ebola infections in Guinea during the 2014 outbreak are believed to have been contracted via unprotected (or unsuitably protected) contact with infected corpses during certain Guinean burial rituals. If you are experiencing some of following symptoms combined, for example, a fever, sore throat, muscle pain, and headaches at first, followed by vomiting, diarrhea and rash, do not hesitate and go for professional medical assistance at once.