Predicting the 2014 Ebola Outbreak in West Africa using Network Analysis

Final Report

Shafi Bashar, Mike Percy, Romit Singhai

{*shafiab, mp81, romit*}*@stanford.edu*

Abstract

The current Ebola outbreak in West Africa is the worst in history. Most traditional epidemiological models are compartmental models that have a random-mixing assumption, calculating the effective reproductive rate of an outbreak. We survey three of these models: the classic SIR (Susceptible, Infectious, Recovered) model and two extensions used in Ebola research.

Network models allow for avoiding the random-mixing assumption inherent in compartmental models. This is done by assigning each individual a finite set of permanent contacts. We review generated contact network models for SARS, including an urban network, a random network, and a scale-free network. We then propose a worldwide network model that takes an alternate approach to modeling the flow of people based on economic trade data and attempt to predict the future of the Ebola outbreak.

I. Introduction/Motivation/Problem Definition (15%)

In this paper we explore predicting an Ebola outbreak using a worldwide network of trading countries.

II. Related Work

A. *Compartmental Epidemiological model*

The majority of research in epidemiological theory is based on the compartmental model. To model the progress of an epidemic in a large population, the individuals in the population are compartmentalized according to the state of the disease. The most widely used such model is the SIR model introduced in (Kermack and McKendrick, 1932):

- **S (Susceptible):** Individuals who have not yet caught the disease from contact with an infectious individual.
- **I (Infectious):** Individuals who have the disease. They have some probability of infecting susceptible people.
- **R (Recovered):** Individuals who have experienced the full infectious period, and are now non-infectious and immune.

The changes among these states over time are represented by a set of differential equations. In order to capture the dynamics of disease spread over time, a population-wide random mixing model is assumed, meaning that each individual has a small and equal chance of coming into contact with any other individual in the population. The basic reproductive number R_0 is defined as the average number of secondary cases generated by a primary case in a pool of mostly susceptible individuals, and is an estimate of epidemic growth at the start of an outbreak if everyone is susceptible. Almost all existing literatures (Chowell et al., 2004; Gomes et al., 2014; Legrand et al., 2007) on Ebola epidemic prediction are based on the modification of the basic SIR model.

In (Chowell et al., 2004), the authors model the effect of Ebola outbreaks in 1995 in Congo and in 2000 in Uganda using a compartmental model similar to the SIR model in Section II-A. However, a distinct feature of Ebola is that individuals exposed to the virus who become infectious do so after a mean incubation period. In order to reflect this feature, in the SIR model is extended with an additional "Exposed" compartment state. This SEIR model is reproduced in Figure 1:

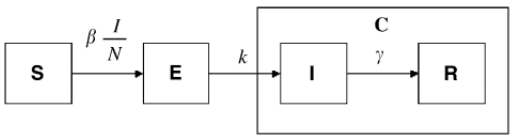


Fig. 1: SEIR model

In the SEIR model, susceptible (S) individuals in contact with the virus enter the exposed (E) state at a rate of $\beta I/N$. The exposed (E) individuals undergo an average incubation period of $1/k$ days before progressing to the infectious (I) state. The exposed state is assumed to be asymptomatic as well as uninfectious. Infectious (I) individuals move to the R state, either recovered or dead, at a rate of γ. The parameters are β, the transmission rate per person per day; N, the total effective population size; and I/N, the probability that contact is made with an infectious individual.

In (Legrand et al., 2007), the Ebola outbreaks in Congo in 1995 and Uganda in 2000 are also studied. However, they expand on the model from (Chowell et al., 2004) by modeling the spreading of Ebola in heterogeneous settings. In order to gain more insight into the epidemic dynamics, the infectious phase is subdivided into three stages: infection in a community setting (I), infection in a hospital setting (H), and infection after death assuming a traditional funeral (F). The resulting model is known as the SEIHFR model.

A CDC (Centers for Disease Control and Prevention) report (Meltzer et al., 2014) published in September 26, 2014 make use of the SEIHFR model to provide an estimate of the future number of cases of current Ebola epidemic. While SEIHFR model provides finer grained modeling of the behavior of Ebola, it is also complex and the link to a network-based model becomes tenuous. In addition, current 2014 Ebola epidemic is still in progress and we do not have the entire picture of the epidemic cycle. Given the limited number of data points, trying to fit a model will large number of parameters may lead to overfitting. Because of these reasons, we do not plan to use the SEIHFR model in our work.

B. Network based Epidemiological model

In the compartmental models described in previous section, the underlying assumption is random mixing among population. Therefore, an infected individual can infect any other individuals in the population. Even though (Legrand et al., 2007) modified the SEIR model to reflect the heterogeneity of infection states, the underlying assumption is still random mixing. However, contagious diseases like Ebola spread via networks formed by physical contacts among individuals. While an individual may have the same number of contacts per unit time in either a random mixing model or a network contact model, within a static network model the set of contacts is fixed, versus a random-mixing model wherein it is continually changing. A static network model thus captures the permanence of many human relationships.

In our search for existing literature on networks based model on spreading of Ebola, we haven't come across any previous work that precisely does this, possibly due to the lack of detailed data in the locations historically affected by the disease. In (Newman, 2002), the authors provide the relation between the compartmental SIR model and a random network model. The author introduces the concept of transmissibility in a network model. Transmissibility of a disease in a network model is defined as the average probability that an infectious individual will transmit the disease to a susceptible individual with whom they have contact. The authors also provide equations that connects the transmissibility and degree distribution of a random network with the basic reproductive number of an SIR model.

In (Meyers et al., 2005), the authors model the spread of the 2002-2003 outbreak of SARS in Hong Kong and Canada using a network model. A contact network model attempts to characterize every interpersonal contact that can potentially lead to disease transmission in the community. Each person in the community is represented as a node in the network and each contact between two people is represented as an edge connecting them. In (Meyers et al., 2005), the authors presented three different contact network models for SARS - *(a) Urban Network:* A plausible urban-setting contact network generated using simulation based on data from City of Vancouver, Canada. Households were randomly chosen and their members given ages, schools, occupations, hospital beds as patients, and caregivers according to statistics from public data. This model offers a high degree of realism, but is complex. *(b) Random Network:* A random network with Poisson degree distribution in which individuals connect to others independently and uniformly at random. *(c) Scale-free Network:* A truncated power law degree distributed network can model individuals, called "superspreaders" with unusually large numbers of contacts or "supershedders" who are unusually effective at excreting the virus into the environment they share with others.

Neither (Meyers et al., 2005; Newman, 2002) captured the temporal progression of the epidemic using network model. Instead, the authors provide the final number of infected nodes in a network for a given value of T. Compare to the compartmental model, there are two major drawbacks in the network based model

(a) Compartmental model is well-established model and can generally capture the temporal progression of disease spread in a localized population fairly well. In contrast, with a contact network, it is hard to model the spread of disease over time. In general, percolation theory is used in conjunction with a contact network to predict the final number of individuals that can get affected in a given network. However, it is generally hard to predict the stage of the disease at a given time with a network model.

(b) In general, it is almost impossible to model the actual contacts among individuals. Therefore, in most network model, additional assumptions are made to model the contact network.

C. Combination of Compartmental and Network based Epidemiological model

As discussed in previous sections, a compartmental model is convenient in analyzing the temporal progression of a contagious disease in a localized population. However, due to the underlying assumption of random mixing among populations, such model is not suitable to analyze large scale outbreak of contagious diseases in a world-wide scale. A much attractive alternative solution in analyzing the progression of disease in a world-wide scale is a combination of compartmental and network model. In such formulation, compartmental model is used on a local scale to track the progression of disease in individual countries (cities or continents). To capture the dynamics of inter-country (city or continent) spread of disease, some form of network data can be used. Such network data are usually converted to inter-country (city or continent) transfer of population per unit time.

In (Balcan et al., 2010) the authors model the inter-country and inter-city spread of population using airport network data and commuting network data. The authors then use such model in conjunction of stochastic compartmental model to analyze the 2001-2002 seasonal influenza spreading. A similar works (Gomes et al., 2014) attempts to predict the spread of Ebola to different parts of the world based on a model that incorporates both the compartmental approach and the use of world-wide air traffic flows. In this model, the world is divided into geographcal regions defining a subpopulation network, modeling connections among subpopulations representing traffic flows due to transportation infrastructure.

III. Modeling localized epidemic spread using a random mixing model

In first phase of the project, to calculate the basic reproductive number of the current Ebola epidemic spread in different countries, we performed model fitting on the data we gathered from Rivers (2014). We considered the SEIR model described in Chowell et al. (2004). The following set of differential equations are used to represent this model Chowell et al. (2004):

$$\frac{dS}{dt} = \frac{-\beta SI}{N}, \quad \frac{dE}{dt} = \frac{\beta SI}{N} - kE, \quad \frac{dI}{dt} = kE - \gamma I, \quad \frac{dR}{dt} = \gamma I, \quad \frac{dC}{dt} = kE. \tag{1}$$

Here, S, E, I, and R denote the number of susceptible, exposed, infectious and removed individuals at time t. C is not an epidemiological state, however it is useful to keep track of the cumulative number of infected cases from the time of the onset of the outbreak. In order to model the effect of intervention on the spread of the disease, in the above model, the transmission rate β is modeled as a function of time. At the initial phase of the outbreak, before intervention, β is parameterized by β_0. After intervention, the value of β transitions from β_0 to β_1, $\beta_0 > \beta_1$ as follows:

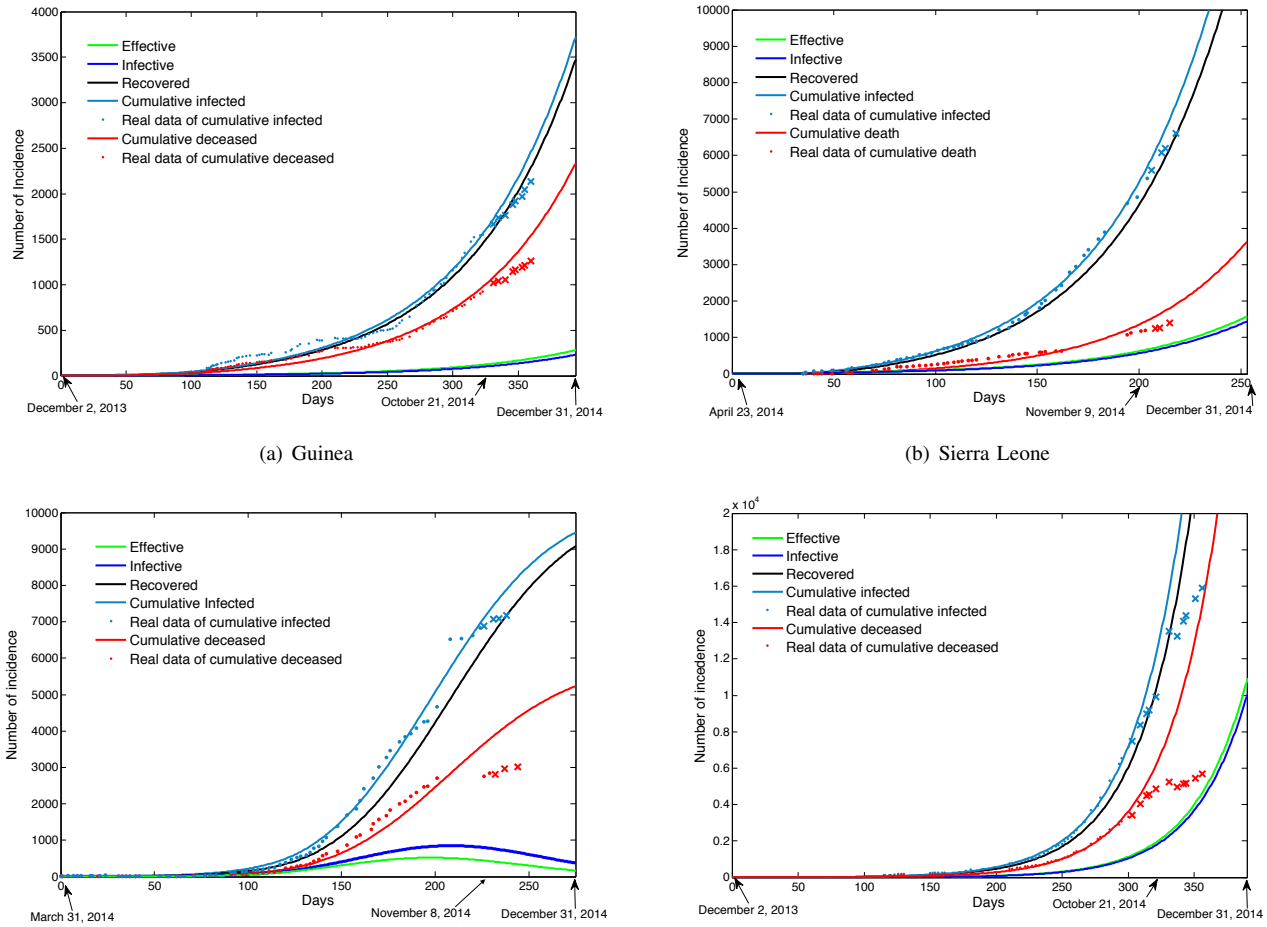
$$\beta(t) = \begin{cases} \beta_0 & t < \tau \\ \beta_1 + (\beta_0 - \beta_1)\exp\left(-q\,(t-\tau)\right) & t \geq \tau \end{cases} \tag{2}$$

Where τ is the time when interventions begin and q controls how quickly the rate of transmission changes from β_0 to β_1. The SEIR model under consideration is a non-linear model with six parameters. The current Ebola epidemic is still spreading and depending on the preventative measures taken, the underlying dynamics of the spread can change drastically any time. The Ebola data for three most affected countries in West Africa - Guinea, Sierra Leone and Libera were represented as (t_i, y_i), $i = 1, 2, \ldots, n$ where t_i represents ith reporting time and y_i the cumulative number of infectious cases from the beginning of the outbreak of to time t_i. The model parameters $\Theta = (\beta_0, \beta_1, k, q, \gamma, \tau)$ for these three countries were estimated using a non-linear least-square procedure by fitting these data to the cumulative number of cases $C(t, \Theta)$ in Equations (1),(2). In addition, we performed model fitting for the West Africa region by adding up the data from these three countries. The paramter estimates are presented in Table II, III in Appendix.

Given the limited number of data available, instead of fitting all six parameters to the model, we decided to fix some of the parameters based on studies on previous Ebola epidemics. In Chowell et al. (2004), the incubation time of the Ebola virus $1/k$ is found to vary between 1 and 21 days, with a mean time of 6.3 days for previous Ebola outbreaks. For ease of data fitting, we set this parameter value to the mean value of 6.3 days. We note that the dynamics of the current epidemic may differ from previous ones, and therefore fixing a value based on the prior estimate may lead to some inaccuracy. The choice of selecting initial outbreak time t_0 and intervention time τ is a difficult problem. To this end, we looked into different sources like WHO (World Health Organization) and CDC websites to learn more about the timeline of the spread. In Guinea, a 2-year-old boy died on December 2, 2013, later diagnosed as an Ebola patient. We consider this incident as the index case for Guinea and set t_0 to December 2. On March 2, the Government of Guinea informed WHO regarding the possibility of an Ebola epidemic and declared a national heath emergency. We considered this date as the intervention date and set τ to 110. In Sierra Leone, one person died on April 2014. In June 12, 2014 the country declared an emergency and closed its borders with neighboring Guinea and Liberia. We consider the first date as t_0 and second date as the intervention time, therefore set τ to 50. In Liberia, on March 31, 2014, there was official confirmation of two people infected with Ebola. We set this date to t_0 and set I_0 and C_0 in our model as 2. The government of Liberia shut down all schools on July 30, 2014. We consider this date as the date of intervention in Liberia and set τ to 120.

The optimization problem involving the model fitting of SEIR model is a non-linear least square regression problem and contains a large number of local minima. Therefore, the choice of initial parameter estimate is an important consideration to get the globally optimal solution. In order to find a good initial choice of the parameters as input to the non-linear least squares solver, we first perform a Latin hypercube sampling on the 4-dimensional parameter space. We grid up the hypercube with a number of grid points in each dimension. We then choose the sample that minimizes the least squares error as the initial input. In order to calculate the 95% confidence interval of the estimated parameter, we performed bootstrapping based on residual error.

The estimated parameters for the SEIR model of Guinea, Sierra Leone, Liberia, and the overall West Africa region is presented in the Appendix. In Figure 2, we presented the number of incidents at different compartments of the SEIR model, as well as the cumulative number of infectious cases over time. In our project milestone report, our model fit was based on



(a) Guinea

(b) Sierra Leone

(c) Liberia

(d) West Africa

Fig. 2: SEIR model fit results for 2014 Ebola epidemic data

the last data we collected on Oct. 21, 2014. At the time of writing this final report, we have additional data points available. Based on data collected till November 9, 2014, we have updated our parameter estimation for Sierra Leone and Liberia. We, however, kept the same model for Guinea as in our milestone report, since we have the most amount of data available for Guinea already due to an earlier outbreak date in Guinea (Dec. 2, 2013). We extrapolated the graph up to Dec. 31, 2014. We note that forecasting future cases may not be accurate as the underlying factors of the epidemic are changing rapidly with the increase in safety measures. The unused data for Guinea (Oct. 21 - Dec. 5), Sierra Leone (Nov. 10 - Dec. 5) and Liberia (Nov. 10 - Dec. 5) are used for calculating test error. These values are also plotted in Figure 2 (points represented with '.' sign are used for training, points represented with 'x' sign are used for testing). We observe that the predictions of the model to the cases up to Dec. 5 are mostly on par with the observed data for Guinea, Sierra Leone and Liberia. For the Guinea epidemic, we have data for the longest range of days and the estimated model, therefore, captures most of the dynamics of the spread in contrast to the other cases.

For the West Africa region plot, our prediction number is much higher than the actual data. This is understandable, as the underlying assumption behind the SEIR model is random mixing. Within a country without any movement restriction, this model is somewhat appropriate. However, for the aggregated model among multiple countries, this assumption is no longer valid due to stricter movement regulations between country borders. Therefore, a model with a random mixing assumption will overestimate the number of infectious cases.

TABLE I: Estimated number of total infected people (by December 31, 2014) and root mean square error (RMSE) of prediction

Country	Estimated Number of Infected individuals	Cross-validation Error	Test Error
Guinea	3724	100.96	218.9
Sierra Leone	14040	170.14	546.2
Liberia	9500	350.3	504.24

In Table I, we presented our estimate of the total number of people who could get infected by December 31, 2014 as well

as root mean square cross-validation error of our model fit and test error on the unused data.

IV. Mapping the compartmental model to the network based model

Approach

Since the compartmental models used in predicting spread of epidemics have a random mixing assumption which may not model the real world appropriately. In this section we made an attempt to map the compartmental model to network based model using percolation theory to model the spread of Ebola in West Africa.

The mapping between a compartment mode and a network based model is defined in the paper by Meyers et al. (2005).The transmissibility T of a disease which is defined as the average probability that an infectious individual will transmit the disease to an individual with whom they have contact. Reproductive number R_0 is calculated using the compartmental model and Transmissibility(T) is calculated using the relationship between the basic reproductive number and the transmissibility defined by Meyers as:

$$R_0 = T\frac{\langle k^2\rangle}{\langle k\rangle - 1}$$

We also calculate the value of epidemic threshold T_c, which is defined by Meyers as:

$$T_c = \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}$$

These numbers are used in modeling and analysis of our contact networks.

Data Preparation

For simulating a contact network for Ebola in various infected countries we used a data set from a social networking site available on the UCI Machine Learning Repository (2008). This undirected network has 4,846,609 nodes and 42,851,237 edges and average degree of 8.841. It is clearly a scale-free network because the degree distribution follows a power law after a degree of approximately 50.

We calculated the transmissibility values for each of the countries Liberia, Guinea and Sierra Leone in West Africa using the formulas described above.

Country	R_0	Transmissibility(T)
Liberia	0.001	0.001
Guinea	1.145	0.1148
Sierra Leone	1.24	0.1243

We also generated three other datasets for a preferential attachment network, random network and a small-world network with approximately same number of nodes and average degree as the real network for comparison purposes. We calculated the epidemic threshold $T_c = 0.1275$ using k=8.841 and used these values for running the simulations on the network datasets.

Simulation

Since our datasets were large and simulation using a percolation model is a compute intensive process so it was not possible to run the simulations in the reasonable time on a local machine so we ran our simulations on extra large EC2 instance (c3.8xlarge). Our simulations needed to run for multiple values of transmissibility so we developed a test harness written in python using GNU Parallel (2014) to leverage multiple cores of the EC2 instance. Our simulation using all four datasets and 20 different values of T completed in about 7 hrs for 200 iterations. We also ran another simulation to see the impact of minimum degree for patient zero for a given transmissibility value for two different networks.

Results and Conclusion

In Figure 3, we show the fraction of network infected for different values of transmissibility for various network types. The simulation captures the overall fraction of network infected but does not capture any temporal progression for Ebola. Above the epidemic threshold the rate at which the fraction of the network gets infected increases for all network types and the rate is higher for other networks in comparison to power law networks which leads to a conclusion that Ebola outbreak is less likely to become an epidemic in these networks . This is contrary to the compartmental model assumption that an outbreak will always become an epidemic when reproductive number is greater than 1. For preferential attachment, small-world and random networks for value of T in the range of 0.35-0.40 nearly the entire network gets infected but not the real network.

In Figure 4, we show that for networks with power law degree distribution[also know as scale free networks], the fraction of the network getting infected is not entirely dependent on the minimum degree of the initial node[patient zero] but also on

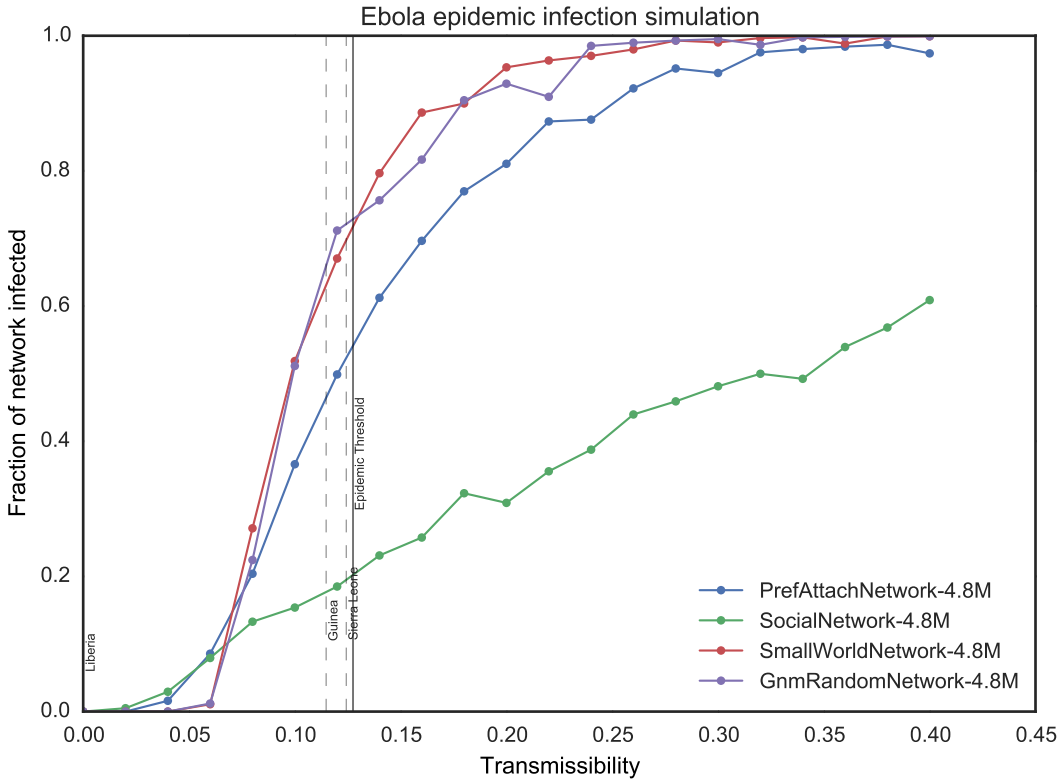


Fig. 3: Network Simulation for Transmissibility

transmissibility value. This is owning of the fact that the power law networks have majority of nodes with few edges or low degrees and small minority of nodes with high degrees or super spreaders. Since these high degree nodes are rare, outbreak at low transmissibility values may fail to reach these nodes. At higher transmissibility values above epidemic threshold, probability of reaching the super spreaders is higher leading to higher fraction of network getting impacted.The fraction become constant after a certain degree distribution, in our simulation around 50 since our real network follows the power law distribution approximately at degree 50.

We did have access to the contact network for the countries severely impacted by Ebola in 2014 to compare our result with actual values. It will be very interesting to infer the hidden network using the cascade diffusion model based on the data available. Running simulations on the inferred network and comparing with the actual results can provide better insights into the spread of current epidemic. The transmission probabilities and other heuristics developed can be used for future analysis.

V. Predicting worldwide spread

Approach

In order to simulate how the Ebola epidemic might spread across the world, we have assembled a worldwide network based on international trade. Because the trade numbers are in US dollars, we have assumed a linear relationship between exports in dollars and travelers going abroad. This is a strong assumption, and we are considering more sophisticated mappings. We are using this as a long-range network representing trade-based population movement and connecting localized subnetworks with these trade-driven edges. We are developing a simulation framework to run in discrete time steps and predict the spread of Ebola across this worldwide network.

Trade network data preparation

The trade network dataset we are using is based on data retrieved from the web service API to the UN Comtrade database (United Nations, 2014) via their web service API. The data queried for were country-to-country SITC-1 exports from the latest data available for each country.

Several problems with this raw dataset immediately became apparent which required working around. Many war-torn countries such as Liberia have not reported detailed export data to the UN in decades (Liberia last reported in 1984). As a result of this, the export totals are incorrect for the present day. In addition, this old data reports exports to countries that

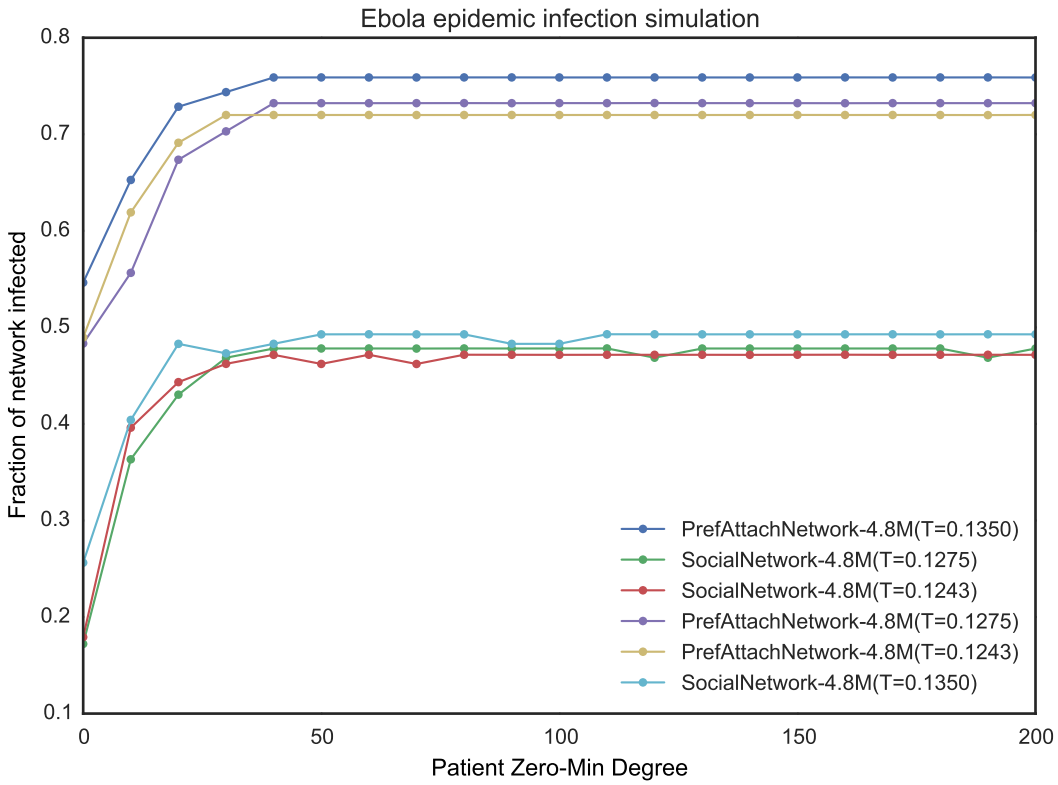


Fig. 4: Network Simulation for Patient Zero Minimum Degree

no longer exist, including East Germany and Yugoslavia. In order to make the dataset usable for modern-day predictions, we performed the following transformations on the data (using Perl scripts):

- Manually remapped non-existent countries to their modern equivalents. In a few cases, we had to do a best-effort mapping. For example, Yugoslavia split into many states, so we assign exports to historical Yugoslavia to modern-day Slovakia, since Slovakia currently has the largest economy of the states that once comprised Yugoslavia.
- Removed exports from states that no longer exist, preferring export data from modern equivalents.

We then needed to make the per-country exports sum up to the latest available export data. The United States Central Intelligence Agency (2013) World Factbook contains up-to-date export totals for most countries in the world. Using the above data set, on the (admittedly strong) assumption that the export distribution from country-to-country in the UN Comtrade data set has remained the same between the last year a country has reported detailed export data and the latest totals from the CIA, we linearly renormalized each outgoing edge in our data set so that the total sum of exports equalled the latest CIA data for each country. This required a manual step of mapping country names in the UN dataset to those used in the CIA dataset.

Once we had a "renormalized" dataset including detailed export edges and totals matching the latest available data, we needed to map these numbers to theoretical international travelers. We could not find complete, or even nearly complete, publicly available data on the number of international outgoing travelers per country. While inbound tourism numbers are available from The World Bank (2014b), for the outbound tourism numbers from The World Bank (2014c), many countries are missing, especially the African nations that we care so much about from an Ebola outbreak perspective. Some data are available from the UN World Tourism Organization (2014), however they appear to be behind a paywall. Data for total air travellers are available from The World Bank (2014a), however this total includes domestic flights and so is less useful for our purposes.

Our current approach to map outbound dollars to outbound travelers is to assume a linear relationship between these numbers, and also to assume that the same linear relationship holds for imports to inbound tourists to a country. We currently use the United States as a model and use the ratio of imports to the United States per year to the number of tourists visiting the United States per year. Based on data from the US Office of Travel & Tourism Industries (2013), 69.77 million people visited the United States in 2013. According to our renormalized data set, total imports into the United States were \$2.21 trillion during the same period. Dividing imports by visitors (an admittedly simplistic approach) gives us a scaling factor of approx. $31,665$. Therefore we have applied this scaling factor to all edges in our international exports network, giving us some approximation of outbound travellers from country to country based on export numbers.

Supplementing the trade network

Using a trade network has its flaws, even discounting errors stemming from poor approximations. For one, a trade-centric network ignores activity with low economic impact, such as a vegetable farmer traveling to a nearby country to sell his products at the market, or someone driving across the border to visit a family member. Relative to the economic impact of the industrial diamond or rubber trade (exports of Liberia) for example, these potentially disease-spreading behaviors simply are not represented equitably if at all in the model. On the other hand, economic trade data is widely accessible and is likely to be fairly accurate, and certainly it is safe to assume that that trade activity correlates with travel between two countries.

Due to the above drawbacks, we plan to supplement our trade network with a generated network that captures local commute-related movement as well. One approach to generating such a network is the gravity law, used for example to study traffic on Korean highways by Jung et al. (2008). The gravity law claims that the number of trips from point A to point B is inversely proportional to the square of the distance between them.

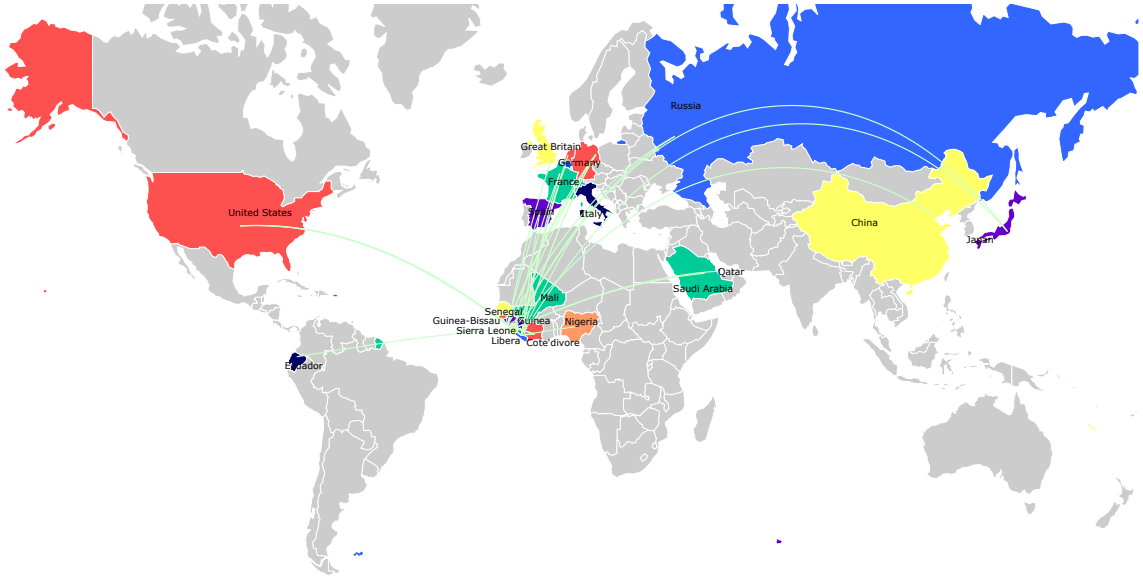


Fig. 5: World-wide trade network from Guinea, Sierra Leone and Liberia

Simulation model

To incorporate inter-country spreading behavior, we modified the SEIR differential equations in (1) to include a transportation operator Ω term. Similar approaches have been used in previous literatures on epidemiolgoical theory (Balcan et al., 2010; Grais et al., 2003). We assume that an individuals who is in susceptible or effective stages of Ebola can travel. Due to the severe nature of Ebola, an individual who is in the infected stage of the disease will most likely not be able to travel. Even if such individual travel to another country, due to stricter regulations and monitoring, such individual will most likely to be quarantined, and therefore will not likely to spread disease among the population of the country traveled. We therefore, do not include transportation operator in the differential equation corresponding to the infected stage. We define $\sigma_{i,j}$ as the number of people travelled from country i to country j everyday. We acquire the exact value of $\sigma_{i,j}$ from the trade network described in previous section. Using this value, the modified differential equations are presented below:

$$\begin{aligned}
\frac{dS_i}{dt} &= \frac{-\beta S_i I_i}{N_i} + \underbrace{\sum_{j=1, j\neq i}^{K} \left(\frac{\sigma_{j,i}}{N_j} S_j - \frac{\sigma_{i,j}}{N_i} S_i \right)}_{\Omega}, \\
\frac{dE_i}{dt} &= \frac{\beta S_i I_i}{N_i} - kE_i + \underbrace{\sum_{j=1, j\neq i}^{K} \left(\frac{\sigma_{j,i}}{N_j} E_j - \frac{\sigma_{i,j}}{N_i} E_i \right)}_{\Omega}, \\
\frac{dI_i}{dt} &= kE_i - \gamma I_i, \quad \frac{dR_i}{dt} = \gamma I_i, \quad \frac{dC_i}{dt} = kE_i, \text{ for } i = 1, \ldots, K
\end{aligned} \tag{3}$$

In Figure 6, we presented the world-wide network under consideration. Here, for each of the three major Ebola affected countries in West Africa - Guinea, Sierra Leone and Liberia, we have an weighted edge to the selected countries with most

amount of population movement. Although not shown in the figure, in our simulation, we also consider the weighted edges between the countries which are reachable from any of the three countries - Guinea, Sierra Leone, Liberia. The parameter values $\Theta = (\beta_0, \beta_1, k, q, \gamma, \tau)$ for Guinea, Sierra Leone and Libera are set to the estimates from Phase 1 result from Table II, III in Appendix. The parameters for the rest of the countries are unknown. For these countries, we reuse the estimated value of West Africa from Table III. We set t_0 to December 2, 2013 - the initial outbreak date in Guinea and initial number of infected individual at Guinea to one and to the rest of the countries to zero. We then run the simulation based on the differential equations in (3) for the duration up to December 31, 2014.

In Figure 6, we presented the potential number of infected people by December 31, 2014 in countries other than Guinea, Sierra Leone and Liberia based on our world-wide simulations. The number for Guinea, Sierra Leone and Liberia are similar to the result presented in Table I in Phase 1 as expected. We note that, the modified trade network data that we used is a simplistic realization of the very complex human migration pattern and does not take into account intricacies involving border restriction, enhanced monitoring etc. Therefore, the result obtained from our simulation result will rather represent an upper bound on the potential outbreak of the disease. Alternatively, we can interpret the results as a measure of relative risk of disease spreads in different countries. It is, however, interesting to see that the result obtained from such simplistic model represents the reality fairly well. So far, in addition to Guinea, Sierra Leone and Liberia, Ebola has spread to Mali, Senegal, Nigeria, United States and Spain. Our simulation result identified all of these countries. In our simulation, Cote d'Ivoire is identified as the country with the highest risk of spread. However, so far Cote d'Ivoire is Ebola free, even though it shares border will all of the three countries - Guinea, Sierra Leone and Liberia. The exact reason of Ebola not spreading in Cote d'Ivoire in not clear. Several speculations e.g. very strict border patrol, trade ban etc. are credited in the media. Our simulation result indicated 3 people to be infected in Nigeria, whereas 20 peoples got infected in reality. The reason behind such small number is the lack of trade as well as no shared border between Guinea, Sierra Leone, Liberia and Nigeria. However, Nigerian airport is the busiest hub in West Africa - therefore, disease may spread through airport transporation, which was not captured in our model.

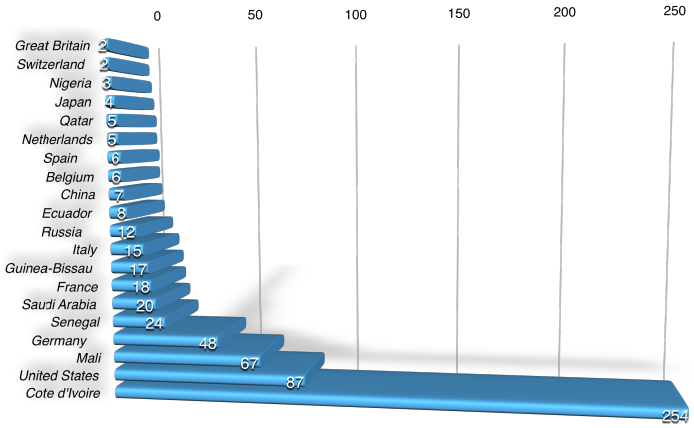


Fig. 6: Potential countries that could get infected and potential number of infected by December 31, 2014

REFERENCES

Duygu Balcan, Bruno Gonçalves, Hao Hu, José J Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of computational science*, 1 (3):132–145, 2010.

Central Intelligence Agency. The World Factbook 2013-14: Country Comparison: Exports, 2013. URL https://www.cia.gov/library/publications/the-world-factbook/rankorder/2078rank.html. [Online; accessed 6-November-2014].

Gerardo Chowell, Nick W Hengartner, Carlos Castillo-Chavez, Paul W Fenimore, and JM Hyman. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of Theoretical Biology*, 229(1):119–126, 2004.

GNU Parallel . GNU Parallel . online, 2014. URL http://www.gnu.org/software/parallel/.

MF Gomes, AP Piontti, Luca Rossi, Dennis Chao, Ira Longini, M Elizabeth Halloran, and Alessandro Vespignani. Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *PLoS Currents Outbreaks*, 2014.

Rebecca F Grais, J Hugh Ellis, and Gregory E Glass. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European journal of epidemiology*, 18(11):1065–1072, 2003.

Woo-Sung Jung, Fengzhong Wang, and H Eugene Stanley. Gravity model in the korean highway. *EPL (Europhysics Letters)*, 81(4):48005, 2008.

William O Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. II. The problem of endemicity. *Proceedings of the Royal society of London. Series A*, 138(834):55–83, 1932.

J Legrand, RF Grais, PY Boelle, AJ Valleron, and A Flahault. Understanding the dynamics of Ebola epidemics. *Epidemiology and infection*, 135(04):610–621, 2007.

Martin I Meltzer, Charisma Y Atkins, Scott Santibanez, Barbara Knust, Brett W Petersen, Elizabeth D Ervin, Stuart T Nichol, Inger K Damon, and Michael L Washington. Estimating the future number of cases in the Ebola epidemic - Liberia and Sierra Leone, 2014-2015. *Morb Mortal Wkly Rep*, 63:1–14, 2014.

Lauren Ancel Meyers, Babak Pourbohloul, Mark EJ Newman, Danuta M Skowronski, and Robert C Brunham. Network theory and SARS: predicting outbreak diversity. *Journal of theoretical biology*, 232(1):71–81, 2005.

Mark EJ Newman. The spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

Caitlin Rivers. Data for the 2014 Ebola outbeak in West Africa, 2014. URL https://github.com/cmrivers/ebola. [Online; accessed 15-October-2014].

The World Bank. Air transport, passengers carried, 2014a. URL http://data.worldbank.org/indicator/IS.AIR.PSGR. [Online; accessed 12-November-2014].

The World Bank. International tourism, number of arrivals, 2014b. URL http://data.worldbank.org/indicator/ST.INT.ARVL. [Online; accessed 11-November-2014].

The World Bank. International tourism, number of departures, 2014c. URL http://data.worldbank.org/indicator/ST.INT.DPRT. [Online; accessed 12-November-2014].

UCI Machine Learning Repository. Social Network from TopCoder's Epidemic Contest, 2008. URL http://networkdata.ics.uci.edu/data.php?id=108. [Online; accessed 12-November-2014].

United Nations. United Nations Commodity Trade Statistics Database, 2014. URL http://comtrade.un.org/data/. [Online; accessed 6-November-2014].

UN World Tourism Organization. Outbound tourism data (calculated on the basis of arrivals data in destination countries), 2014. URL http://statistics.unwto.org/content/outbound-tourism-data-calculated-basis-arrivals-data-destination-countries. [Online; accessed 12-November-2014].

US Office of Travel & Tourism Industries. 2013 Monthly Tourism Statistics, 2013. URL http://travel.trade.gov/view/m-2013-I-001/table1.html. [Online; accessed 6-November-2014].

APPENDIX

TABLE II: Parameter estimation for Ebola SEIR model (Guinea & Sierra Leone)

	Guinea			**Sierra Leone**		
Incidence Dependent Parameters	*Value*		*Comments*	*Value*		*Comments*
Initial Case t_0	December 2, 2013		one person fell ill	April 23, 2014		one person fell ill
S_0	0		-	0		-
E_0	0		-	0		-
I_0	1		-	1		-
R_0	0		-	0		-
C_0	1		-	1		-
Intervention time	March 2, 2014		Gov. of Guinea informed WHO	June 12, 2014		Country declared emergency
τ	110		-	50		-
Estimated Parameters	*Value*	*95% CI*	*Comments*	*Value*	*95% CI*	*Comments*
Incubation Time $1/k$	6.3	-	based on previous works	6.3	-	based on previous works
Infection Time $1/\gamma$	5.4957	[5.43, 5.545]	-	6.386	[6.2112, 6.4733]	-
β_0	0.2407	[0.2374, 0.244]	-	0.356	[0.3391, 0.3643]	-
β_1	0.2084	[0.2033, 0.2135]	-	0.195	[0.1926, 0.1988]	-
q	32	[0.1, 100]	-	0.47	[0.1, 7.07]	-
Fatality Rate	0.67	-	-	0.289	-	-
R_0	1.323	[1.295, 1.341]	-	2.27	-	-
R_1	1.145	-	-	1.24	-	-

Contribution

Romit Singhai: Initial project idea and research, contribution to project proposal and milestone report specifically for mapping the compartmental model to network based model. Exploring the MCMC approach for compartmental model and developing code. I worked on data collection, coding the simulation model and running the simulations on EC2, analyzing results and plotting the graphs. I also wrote the part to map the compartmental model to contact network of the final paper including results and conclusions.

TABLE III: Parameter estimation for Ebola SEIR model (Liberia & West Africa overall)

	Liberia			**West Africa**		
Incidenced Dependent Parameters	*Value*		*Comments*	*Value*		*Comments*
Initial Case t_0	March 31, 2014		official confirmation two infected	December 2, 2013		one person fell ill in Guinea
S_0	0		-	0		-
E_0	0		-	0		-
I_0	2		-	1		-
R_0	0		-	0		-
C_0	2		-	1		-
Intervention time	July 30, 2014		School shutdown	March 2, 2014		Gov. of Guinea informed WHO
τ	120		-	110		-
Estimated Parameters	*Value*	*95% CI*	*Comments*	*Value*	*95% CI*	*Comments*
Incubation Time $1/k$	6.3	-	based on previous works	6.3	-	based on previous works
Infection Time $1/\gamma$	10.5	[8.32,10.7]	-	6.8	-	-
β_0	0.1697	[0.167, 0.199]	-	0.2	-	-
β_1	0.0001	[0.0001, 0.097]	-	0	-	-
q	0.0068	[0.0059-0.0187]	-	0	-	-
Fatality Rate	0.575	-	-		-	-
R_0	1.78	-	-	1.36	-	-
R_1	0.001	-	-	-	-	-