

IBM Research Report

Modeling Propagation of Infectious Disease as Connected Network

Lianjun An, Young M. Lee
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598



Research Division

Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Modeling Propagation of Infectious Disease as Connected Network

Lianjun An and Young M. Lee

IBM T.J. Watson Research Center
Yorktown, New York 10598, U.S.A.

{alianjun,ymlee}@us.ibm.com

Abstract

Pandemic is likely to occur in the near future, and it could cause significant disruptions in society. Avian influenza is such an example as it can potentially evolve to be transmitted from human to human, and spread world-wide in a short period of time. In order to prepare for such disaster and to develop global mitigation strategies for society, government as well as enterprises, need to understand how fast diseases would spread and also the magnitude of infection. As world has become more global than ever, population, social network and transportation would make it much easier for diseases to spread than before. In this paper, we describe a disease spread model that combines compartmental epidemiological model with connected network of geographical locations and airports using system dynamics method. We also model how various mitigation actions would affect the spread of disease. The model is intended to be used for firms in studying possible impact of pandemic disease on their business. Various scenarios of disease spreads are simulated and presented.

Keywords: System Dynamics, Epidemiological Model, Connected Network, Infectious Disease

1. Introduction

Influenza pandemic has occurred in every few decades in history (1918, 1957 and 1968) (Congressional Budget Office 2005). It caused severe social and economic disruptions. Most people still had fresh memory on Severe Acute Respiratory Syndrome (SARS) epidemic outbreak in 2002-2003 in several Asian countries and Canada even though it did not develop into pandemic. During the last couple years, the world has experienced unprecedented threat from avian influenza (H5N1 virus) in poultry. The virus could evolve in such a way that allows for efficient human-to-human transmission. Potential pandemic outbreak in a visible future is quite likely.

Social and economic environments have changed substantially as we have seen from 1918 pandemic and 2002-2003 SARS epidemics. On one hand, the improved medical condition and communication system may reduce transmission of disease, improve awareness and prepare society in advance. On the other hand, increased travels, globally integrated business operations increased the chances for disease to propagate much faster than ever as people travel around more frequently via ground and air transportation. In fact, the 2002-2003 SARS, although it had just limited number of cases (about 8000 cases worldwide and 774 casualties), resulted in a meaningful influence on Asian countries and Canada's economy activity (Hsu et al. 2006).

Governments concern about the pandemic outbreak, and private enterprises should also be aware of it because many businesses would be impacted. Many businesses have their supply partners and markets in various parts of the world, and workforce are also shared from different geographical regions. The social and economic disruption resulting from pandemic will affect business operation in many different ways and will result in substantial loss of revenue for many enterprises. Therefore, it is important for enterprise to understand the scope and pattern of pandemic impact on economy and workforce and to develop strategic plan to reduce risk of disaster-related disruption of their business activities.

In order to estimate pandemic impact and understand disease propagation, we develop a model of disease propagation and mitigations. Specifically, we extend the single node epidemiological model, SEIR (Susceptible, Exposed, Infectious and Recovered) model (Kermack et al. 1927, Anderson et al. 1979), into a connected network model. To reflect the reality, we construct our parameters based on geographical and demographic data. Population density in each region is factored into the formulation of disease transmissibility. Information on adjacency of each region with neighboring regions is factored into the disease propagation rate. We have also used the information on air travel volume between paired airports for year 2005. The information is used to estimate the propagation rate between the regions where airports are located.

Government is interested in studying pandemic outbreak and establishes mitigation policies and prepares for the worst scenarios (Congressional Budget Office 2005). In our model, user can specify two categories of mitigation policies. For individual region, medical care type of mitigation policies can be specified, such as face mask, vaccine, quarantine, Tami flu, hygiene etc. For connected networks of regions, government level policies, that can be deployed to prevent disease spreading from one region to the other, such as border closing, county closing and airport closing, can be specified.

We develop a System Dynamics (SD) model to capture causal relationship and feedback loop. The SD model includes 6000 regions and 3600 airports worldwide. Government airport closing action based on threshold on infectious rate is also modeled through feedback loops captured in the model. The model also has capability to retrieve data required for the model such as population density, neighboring among regions and air passenger volume between regions from database. Simulation capability of SD model is vital for us to examine different scenarios and to evaluate different mitigation actions.

In this paper, we present simulation results for several scenarios with different outbreak origin, extent of disease and with mitigation actions. The most of simulation results look intuitively correct. The result also provides some very interesting propagation patterns that might be not intuitively obvious. For instance, sometimes, airport closing could lead to a higher peak of infectious percentage in some region, or lead to last longer in infectious period in other region.

This work is done as a part of global pandemic strategy study to assess economic impact of the pandemic on the firm (Chen-Rotzo et al. 2007). It includes an end-to-end story about infrastructure, workforce, and economic impact of the pandemic, and then supply and demand impact on the firm based on the underlying business nature. The model about workforce change under the pandemic is described by Lee et al. (2007).

The paper is organized as the following. In section 2, we review literatures and previous work. Section 3 introduces the networked SEIR model. Section 4 demonstrates some simulation

results for certain disease spreading scenarios. Section 5 concludes the paper and discusses further research direction.

2. Literature Review and Previous Work

Epidemiological modeling work went back to 1927 when Kermack and McKendrick (1927) developed a model, with three groups of population; Susceptible population (S), Infectious population (I) and Recovered population (R). Such a model is long known as the SIR model and the mathematical equation is called Kermack-McKendrick equation. The SIR model has been used widely in epidemiology. The system describes the evolution process of disease affecting population. In this model, a part of susceptible population (S) becomes infectious population (I) as a result of social contacts, and a part of the infectious population (I) becomes recovered the recovered population (R). Note that some in the infectious population might die depending the mortality rate. Also the model typically assumes that a part of the recovered population would develop permanent immunity. For certain diseases, immunity assumption may not be valid, and the recovered person could get infected again.

The model can be extended to include an additional group of population called Exposed population (E) (Anderson et al. 1979). In this version of model, a part of susceptible population can become the exposed population first, and then it can become the infectious population. Such a model is called SEIR model, and one such model is shown in the middle part of Figure 1. The top and bottom parts will be explained later due to our extension. There is a positive feedback loop in the SEIR model (indicated by loop sign). Higher Infectious Population leads to higher infection rate through social contact. There are many varieties of the SEIR model. For instance, Hsu et al. (2006) extended the model to study SARS outbreak from November 2002 to July 2003, and added several other population groups such as quarantined and isolated population. They examined compound effects from intervention measures, including quarantine, and public response.

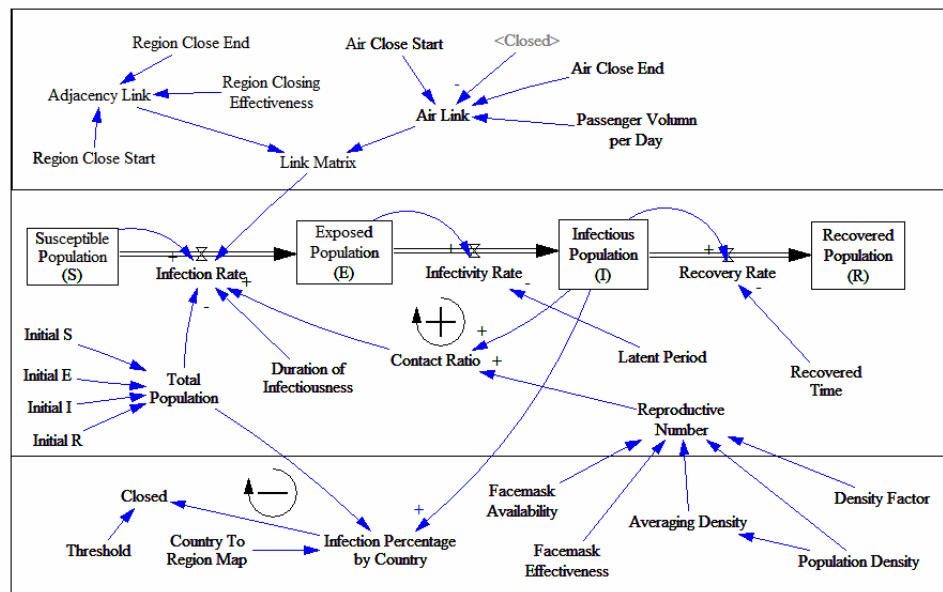


Figure 1: SEIR Disease Spread Model

The models mentioned above are deterministic models and they typically assume homogeneity of populations with respect to the four population categories. There have also been stochastic modeling approaches including the work by Germann et al. (2006) who introduced a large-scale stochastic agent-based simulation model (EpiCast) to predict US nationwide spread of disease and to assess various mitigation strategies. Ford et al. (2006) also created a software system and framework called STEM (Spatial and Temporal Epidemiological Model) for modeling multiple populations in geographically distributed locations. This framework is capable for putting additional model layer to represent airport connection network and specify different characteristic for each region. We develop a light-weighted networked model using a SD tool, and utilize available geographical and demographic data. Mitigation strategies are implemented feedback loop, and they can be specified differently by regions.

3. Networked SEIR Model

We develop our disease spread model using system dynamics method (Forrester 1961, Sterman 2000), and used a modeling tool called Vensim (<http://www.vensim.com>). Vensim has a capability of breaking each system variable into an array whose elements represent a sub-domain, and we separate the world into about 6000 regions using the array to differentiate disease spread in various regions of the world. Based on availability of data, the granularity of our model goes down to county level in some countries, state/province level for some country and even country level in some continents. We assume that each region has homogeneous population structure. However, transmissibility would differ from region to region due to different population density. Also spreading of disease from one region to another depends on geographic contact among the regions and air transportation infrastructure. We include 3600 airports around the world in our model.

It is important to understand severity of disease in multiple regions. For single node model, the severity of disease is typically measured in terms of infectious population over total population, duration of infectiousness and the mortality ratio among infectious population. Another factor that indicates the severity of a pandemic is the scope of spread, i.e., how many regions, countries, continents have been affected. The SARS outbreak in 2002-2003 had a relatively high mortality rate (about 10% infectious people died). However, due to quarantine measure deployed during the SARS breakout, the infected population was separated very effectively and the scope of the spread was not too high (mostly confined in Asian countries and a few rare occurrences in other places, totaling of 29 countries).

3.1. Networked Model

In order to model the propagation of pandemic around world, we extend a single node SEIR model to a connected network model with multiple nodes. Disease propagates along the connections in the network. In our model, the connections represent people traveling by road or by air from a region to another. Transmissibility of disease in a region would be different from another as population densities are different from one another. Figure 2 shows a networked model with two nodes. The rate in which the susceptible population becomes the exposed population is affected by infectious population from both regions, when both regions are connected through either road or air.

In addition to modeling spread of disease, we also model how mitigation actions affect the spread of disease. We include two types of mitigation actions in our model. One type of

mitigation action, shown on the left side of the Figure 2 as Mitigation Action I, could affect transmission rate of disease between two regions through the connection, because these mitigation actions, such as airport closing, border closing, and port closing, would reduce or block movement of people from one region to another. The other type of mitigation action, shown on the right side of the Figure 2 as Mitigation Actions II, would modify the transmissibility within a region, since those mitigation actions such as distributing medical supplies can be deployed independently for each region. This model of disease transmissibility shown here is applied to all the regions that are connected.

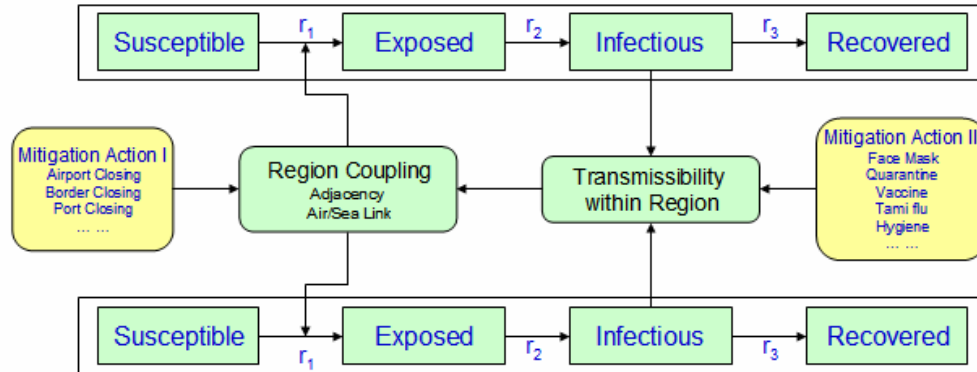


Figure 2: Two nodes SEIR model with mitigation

Similarly, the concept in the two nodes model applies to the general networked model. As shown in the figure 1, the middle part of the figure represents multiple nodes of SEIR process using arraying capability. The coupling among regions is shown in the top part; the link matrix (6,000 x 6,000) represents the quantified connection. For instance, if the entry at [i,j] is nonzero, it means there exists some movement of people between regions i and j through road and air transportation. We decompose the “link matrix” into “Adjacency Link” and “Air Link”. Mitigation policy, such as “region closing” and “port closing”, would reduce “Adjacency Link”. Mitigation policy, such as “airport closing”, would reduce “Air Link”. Mathematical formulation that describes the link will be given in Section 3.3. The bottom part of Figure 1 records the modification of “Reproductive Number”, which is related to population density, in each. Also medical preventive mitigation actions, like “Facemask” and “Vaccine”, affect “Reproductive Number” in that region. Mathematical formulation to quantify transmissibility change will be given in Section 3.2. In the bottom part of Figure 1, we also add additional negative feedback loop to automate “Airport Closing” mitigation action based on a specified threshold. First, we use the “Country to Region Map” to aggregate the infectious population and obtain “Infectious Percentage by Country”. Then If the value of “Infectious Percentage by Country” is greater then the specified “Threshold”, we close all airport going in and out of that country (the value of “Air Link” would be modified for some entries).

Since the size of array for representing geographical regions is huge, it is impossible to manually enter data on population and area of 6,000 regions and 3600 airports into the Vensim model. We first enter the geographic and demographic data into database, and then use programming model to retrieve such data and set value for the Vensim model before doing simulation. Similarly, SEIR data generated by simulation are retrieved from Vensim output file and are saved into database for further analysis.

3.2. Mathematical Models of Transmissibility

In modeling transmissibility within a region, we first formulate how the population density affects the transmissibility (through Reproductive Number in model) as

$$R_e[i] = R_o * \left(\frac{D[i]}{D^*} \right)^\lambda, \quad D^* = \frac{1}{N} \sum_{i \in S_l} D[i] \quad (\text{Equation 1})$$

where i is a subscript for different region; S_l is the set of all indexes of regions with size of N (6000 in our model, including counties, or states, or countries); $R_e[i]$ is the effective ‘‘Reproductive Number’’ at the region indexed by i ; R_o is the original reproductive number, which is defined as the average number of persons effectively contacted by each person during the infectious period; $D[i]$ is the ‘‘Population Density’’ in the region and D^* the ‘‘Average Density’’ in all considered regions. The exponential parameter $\lambda > 0$ (Density Factor in lower part of Figure 1) is used to adjust the sensitivity of density on transmissibility. When λ is close to zero, the value of $\left(\frac{D[i]}{D^*} \right)^\lambda$ would be close to one and the transmissibility would be less sensitive to population density change.

To account influence of the second type of mitigation actions (Mitigation Action II), the expression for R_e is further changed to be

$$R_e[i] = R_o * \left(\frac{D[i]}{D^*} \right)^\lambda * \prod_{j \in S_j} (1 - E_2[i, j] * C[i, j]) \quad (\text{Equation 2})$$

where j is the index for the second type mitigation set S_j ; $E_2[i, j]$ is for the effectiveness of the mitigation policy for region i for mitigation action j ; $C[i, j]$ represents percentage coverage (availability) of the mitigation action. $E_2[i, j]$ represents effectiveness of the mitigation action j for region i . Therefore, the third factor in the RHS of equation 2 counts for compounded effect of the second type mitigation actions. Note that, both $E_2[i, j]$ and $C[i, j]$ is indexed by i , since mitigation actions can be applied to each region differently from the other regions.

3.3. Mathematical Model for Network Connections

To account for the influence from different regions, the infectious rate r_1 (‘‘Infection Rate’’ in Figure 1) is adjusted by introducing a $N \times N$ ‘‘Link matrix’’ (see top part of Figure 1) M

$$r_1[i] = \frac{S[i] * \sum_{k \in S_l} M[i, k] * R_e[k] * I[k]}{T_{se} * P_o[i]} \quad (\text{Equation 3})$$

where k is an index for region; $I[k]$ represents infected population in region indexed by k ; $S[i]$ is the susceptible population in region indexed by i ; $P_{own}[i]$ is the population in region indexed by i ; $R_e[k]$ is ‘‘Reproductive Number’’ in region k ; T_{se} is ‘‘Duration of Infectiousness’’ i.e., the time taken for the susceptible population to become the exposed. Without the connection

influence (road or air), $M[i, k]$ would be an identity matrix, which contains value of one for diagonal entries, and value of zero for non-diagonal entries. Let $P_{con}[i]$ be the total population of regions connected to the region i . Then the probability, $p_r[i, k]$, of people traveling from region i to region k would be

$$p_r[i, k] = \frac{P_{own}[k]}{P_{con}[i]} \quad (\text{Equation 4})$$

Suppose that daily travel percentage with neighboring regions is $\beta[i]$. Then off-diagonal entries would be

$$M[i, k] = \frac{\beta[i] + \beta[k]}{2} * \frac{p_r[i, k] * P_{own}[i] + p_r[k, i] * P_{own}[k]}{P_{own}[i] + P_{own}[k]} \quad (\text{Equation 5})$$

The contribution from air transportation is formulated in the similar way. In the case of passenger volume information, $V[i, k]$, are available between regions i and k for a year, we use the real data to populate the matrix as $M[i, k] = V[i, k] / 365$. For air connection, we assume full mesh connection among hub airports, and non-hub airports are connected only to the hubs in the same country.

To count for the first type mitigation actions (Mitigation Type I), the link matrix M is adjusted by an additional factor as,

$$M[i, k] = \frac{\beta[i] + \beta[k]}{2} * \frac{p_r[i, k] * P_{own}[i] + p_r[k, i] * P_{own}[k]}{P_{own}[i] + P_{own}[k]} * (1 - E_1[i, k]) \quad (\text{Equation 6})$$

where $E_1[i, k]$ represents the effectiveness of road or air closing between regions i and k . The value $E_1[i, k]$ varies from 0 to 1, with zero representing no closing at all, and one for fully effective closing.

4. Simulation Scenarios and Results

We setup simulation scenarios which may aid a firm in understanding the propagation of a pandemic disease and impact of mitigation actions. We simulated a disease starting from a region, and some government mitigation policies being applied in certain regions.

Due to global integrated economy and outsourcing, many companies are interested in the worldwide propagation pattern of pandemic disease. In particular, many companies expands financial invest in Asian market, such as China and India. But Asian countries are more vulnerable to disease outbreak compared with Europe and the United States. Based on such concern, we choose Vietnam as the disease origin country and investigate the spread along network connection via ground and air transportation.

First scenario is for a mild case of pandemic in which the unmodified reproductive number is relatively small. We also assume that all airports are open at all time during the pandemic. Figure 3 shows changes of infectious population over time in several regions around world. The infectiousness in various countries outside Vietnam occurs in later time depending on the proximity of those countries from Vietnam. The simulation result shows that the magnitude and duration of the infectiousness in various countries is different from country to country because of difference in population density. For Thailand, we simulate in the county level and aggregate

data to get country level. The curve has wider time span due to fact that the peak in different regions might not happen at the same time.

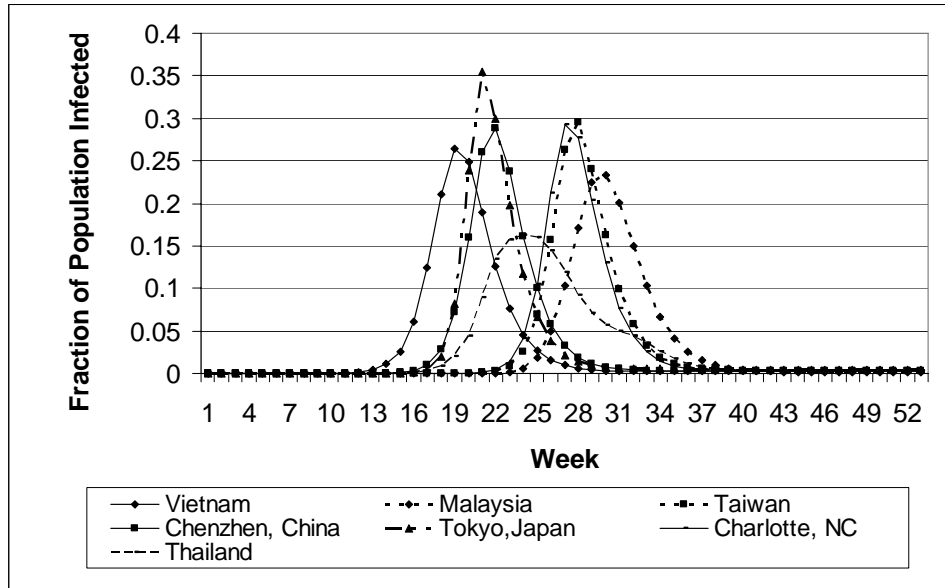


Figure 3. Scenario 1: Mild Disease Seeded in Vietnam, All Airports Open

Note that spread along adjacent link and along air link have different patterns. For two regions that are far apart from one another, disease propagation from one to the other might take a while since it passes through several intermediate regions in sequential pattern. For two regions that are connected with air link, the spread would be very fast and it would spread around the world rapidly. It can be seen from Figure 3 that the growth rate in Tokyo is faster than the growth rate in Chenzhen, China, since the latter mainly be affected by adjacent link.

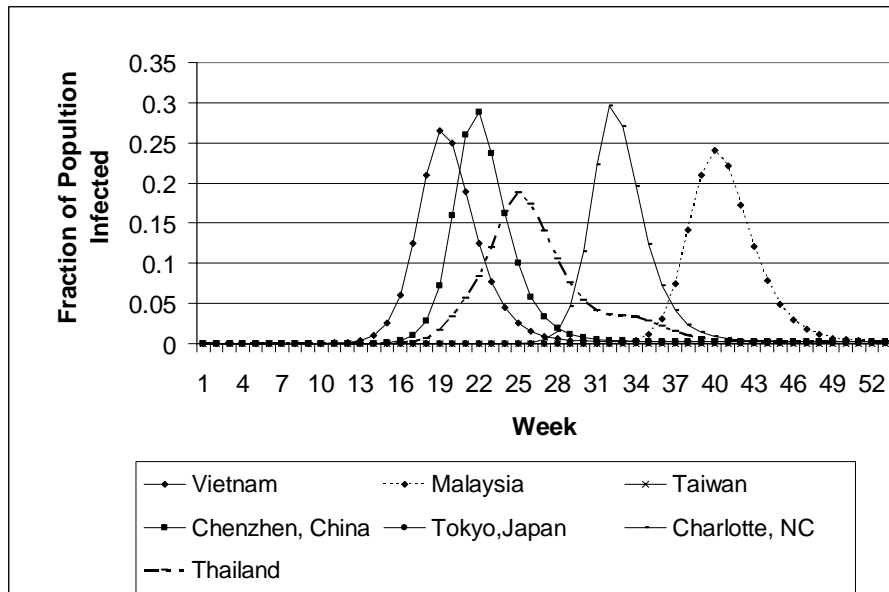


Figure 4. Scenario 2: Mild Case with SEA, Japan and China Airports Closed

Second scenario is also for a mild case but with airports in southeastern Asian, Japan and China closed. Figure 4 shows the simulation result. Comparing this scenario with the first

scenario, the infectiousness curves are very similar in shape and magnitude. However, the mitigation action, airport closing, slows down propagation of the disease from one region to another. The magnitude of peak is almost same, but the curve will shift to the right indicating slower propagation of the disease. Assuming everything else is same, the disease gets to other regions slower because of airport closings, but as long as it gets there, the disease still causes infectious population in that region and the magnitude of the infectiousness is almost the same.

Certain things that we observed from the simulation in above two scenarios are intuitive. And they are:

1. Infectiousness in Chenzhen of China does not change too much as airports are closed since traveling between Vietnam and Chenzhen is mainly via ground transportation.
2. Japan and Taiwan do not have infectiousness anymore after closing these airports because air transportation is only mean of transportation between Vietnam and these counties. In fact, there are no land connections from those regions to Vietnam and traveling by sea is excluded from the model.
3. Infectiousness in US and Malaysia occurs in later time because even without air transportation, there are still land transportation available between Vietnam and these countries. Therefore, closing of airports delays the propagation of the disease to those counties.

We notice some interesting simulation results that are rather unexpected for Thailand. The profile of infectiousness for Thailand looks quite different from other regions and it represents the influence effect of connected network structure. Its peak is not too high but lasts longer. It has a re-bouncing pattern during its decreasing period. The low peak of infectiousness in Thailand is not necessarily related to the lower population density there. The magnitude of infectiousness is higher when airports are closed than when they are open. After examining the infectiousness profile in its connected regions, we were able explain the simulation results. In fact, formation of these propagating patterns is related to synchronizing and non-synchronizing effects of infectiousness in all the regions that are connected to the region, Thailand. If the peaks of the infectiousness profiles of all the regions that are connected a particular region occur at around the same time, the peak of the infectiousness curve of the particular region, Thailand in this case, would occur around the same time and higher magnitude as the connected regions. However, if the peaks of the infectiousness profile of all the regions that are connected to a particular region occur in different times, then the infectious period would last longer. It is also possible that the infectiousness curve could have multiple peaks. After closing some airports, the propagation of a disease from one region to other is altered. Some regions are affected through ground connection only, and the others are affected by both ground and air. However, the arrival time would change since traveling path is different. Therefore, depending on how a region is connected to surrounding regions, the profile of infectiousness of a particular region with respect to surrounding regions would be in three different shapes; (1) infectiousness profile is same as the surrounding regions with certain delay, (2) infectiousness profile is synchronized with surrounding regions and amplified, (3) infectious profile is not synchronized with surrounding regions and is weakened with several peaks.

Fourth scenario is for a severe case with all airports open. In that case, we choose a higher reproductive number, different recovery time, latent period. Figure 5 shows higher magnitude of infectious population in the same regions comparing with Figure 3.

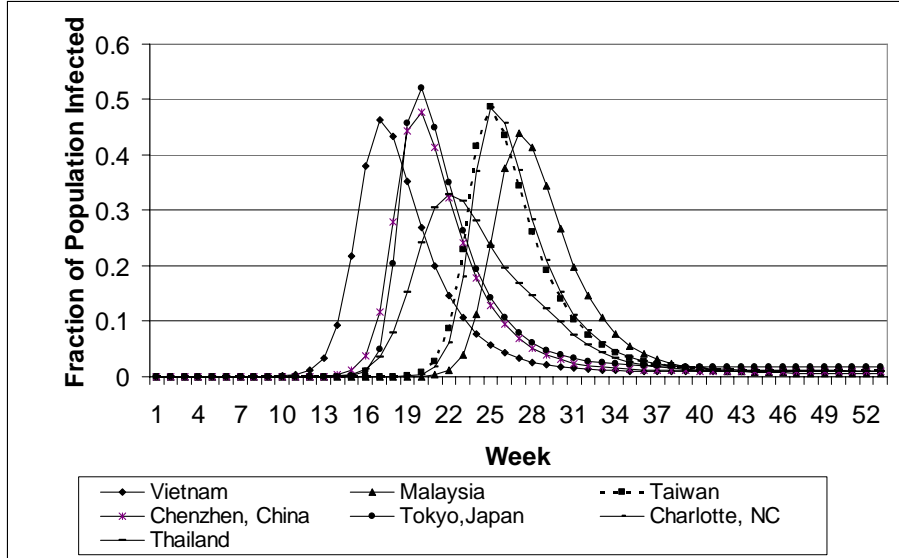


Figure 5. Scenario 3: Severe Disease Seeded in Vietnam with all Airports Open

Fifth Scenario is also for a severe case with airports in southeastern Asian, Japan and China closed. Figure 6 shows result similar to Figure 4 but with higher infectious ratio.

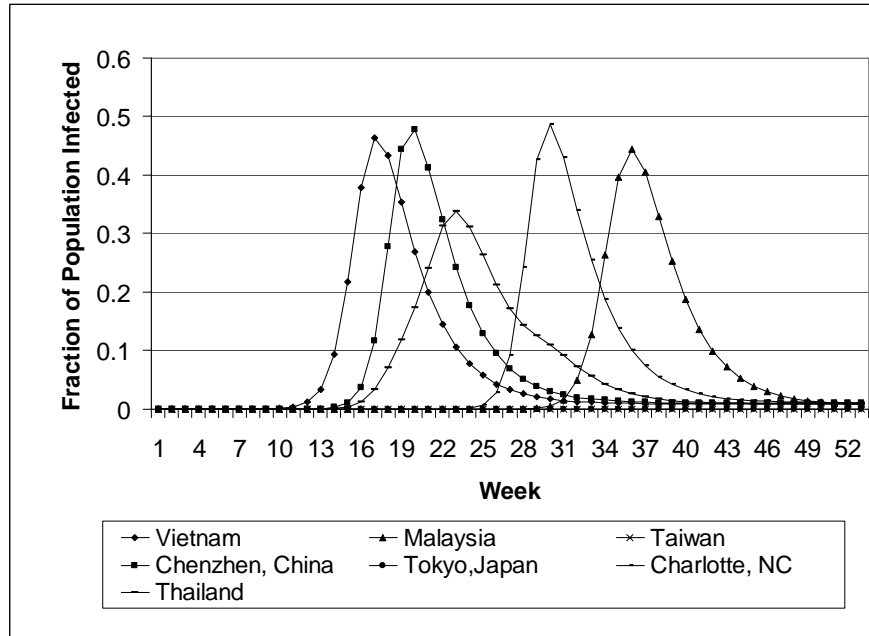


Figure 6. Scenario 4: Severe Case with SEA, Japan and China Airports Closed

We also test a mitigation policy of closing airports only when the fraction of infected population reaches a certain threshold. It was found out that it is difficult to simulate the conditional closing of airport in SD modeling. When disease is propagated to some countries before closing airport, the local infectious growth rate is mainly determined by social contact there. Unless the infected persons previously reached there are quickly identified and quarantined, the disease would quickly spread in the community even if airport is closed. However, in the SD modeling, if transmissibility inside a region is the smallest possible number

e.g., a single person, the transmissibility of another region which is in contact with the region by air link can be much smaller number e.g. smaller than a person, which nevertheless would grow and infect other people. This is not realistic situation but since we use a continuous model for connection between regions, disease can still spread to another region from a small fraction of a person. It may be more realistic to model the connection by stochastic or discrete modeling approach, for which it is possible to model the conditional closing of airport.

5. Conclusion and Discussion

We develop a networked disease propagation model that combines compartmental epidemiological model with connected network of geographical locations and airports using system dynamics method. The heterogeneity among various regions is taken into account through population density. The mathematical model incorporates two types of mitigation actions that would affect either transmissibility in individual region or propagating speed of disease from one region to the other. This modeling work is a part of a corporate pandemic awareness project (Chen-Ritzo et al. 2007) that is intended to assess business impact of pandemic. The disease propagation model allows the firm to simulate pattern and magnitude of disease spread and effectiveness of mitigation actions. The simulation results of the model have been used in determining financial impact of possible pandemic occurrence on the firm. The model can be further turned, for instance, to include heterogeneity with respect to time. In fact, transmissibility of a disease can change with time. It is also possible incorporate infrastructure difference of various regions into the model to study how pandemic risk index (<http://maps.maplecroft.com/downloads/PANDEM/index.pdf>) would affect the disease propagation. Also the connection between regions can be modeled through stochastic approach. Government mitigation policies can be further assessed through simulation.

Acknowledgement

The authors would like to thank Steve Buckley and other members of the Pandemic Business Impact Modeler Project of IBM Research, Ching-Hua Chen-Ritzo, Pawan Chowdhary, Igor Frolow, Naveen Lamba, Dharmashankar Subramanian and Tom Ervolina, for many useful discussions and suggestions.

References

Anderson, R. M. and May, R. M. (1979), "Population Biology of Infectious Diseases: Part I." *Nature* 280, 361-367.

Chen-Ritzo, C-H., An, L., Buckley, S., Chowdhary, P., Ervolina, E., Lamba, N., Lee, Y., and Subramanian, D., (2007), "Pandemic Business Impact Modeler", Proceedings of the 2007 INFORMS Simulation Society Research Workshop.

Congressional Budget Office, (2005), “A Potential Influence Pandemic: Possible Macroeconomic Effects and Policy Issues”.

Cooper, S., and Coxe, D., (2005) “A Investor’s Guide to Avian Flu”, BMO Nesbitt Burns Research.

Forrester, J. W., (1961), *Industry Dynamics*, MIT Press, Cambridge, MA.

Ford, D., Kaufman, J., and Eiron, I., (2006). “An extensible spatial and temporal epidemiological modeling”, *International Journal of Health Geographics*, vol. 5.

Lee, Y, An, L. (2007). “Modeling Dynamics of Workforce Absenteeism and Effectiveness of Mitigation Actions During Pandemics”, *System Dynamics Conference 2007* (submitted).

Germann, T., Kadau, K., Longini, Jr. I., , and Macken, C., (2006). “Mitigation Strategies for Pandemic Influenza in the United States”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 5935-5940.

Hsu, S. B., and Hsieh, Y. H., (2006), “Modeling Intervention Measures and Severity-Dependent Public Response during Severe Acute Respiratory Syndrome Outbreak” *SIAM J. Appl. Math.*, 66-2, 627-647.

Kermack, W. O. and McKendrick, A. G., (1927), “A Contribution to the Mathematical Theory of Epidemics”, *Proc. Roy. Soc. Lond. A* 115, 700-721.

Sterman, J. D., (2000), *Business Dynamics: System Thinking and Modeling for a Complex World*, Irwin McGraw-Hill, Boston.