A Heterogeneous Model on Information Epidemics in Social Networks

Chin-Chia Hsu
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan
Email:b00901161@ntu.edu.tw

Kwang-Cheng Chen
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan
Email:ckc@ntu.edu.tw

Abstract—Information epidemics have been shown useful in various networking and malware network security investigations. Through modern social network analysis looking into small world phenomenon, we therefore examine the mechanisms of cross-country epidemics composed of direct contagions via spatial social interactions and indirect propagations via carriers on transportation network with deep insights. An innovative model to analyze the mixture of the domestic ripple-based spreading and air-traveling is proposed, together considering the incubation period and screening level. Our heterogeneous network model enables quick analytical predictions on the intricate spreading behaviors of not only the epidemics but also potential information or malware dissemination.

Index Terms—Epidemiology, transportation network, incubation period, social networks

I. INTRODUCTION

With the advances of modern transportations, the risk of pandemics grows much higher since the pathogens can be carried globally in much shorter periods than ever, bumping up the probability of large-scale outbreaks. The mechanism of disease spreading can be perceived as two dynamics: the direct contagion in contact networks and the indirect carriage via transportation or migrating birds among geographical areas, like airplanes and countries together as typical representatives. A contagious virus between face-to-face persons turns out to leap over thousands of miles via flights. Famous examples as SARS in 2003, and Ebola in 2014 show that the transportation network leads to outbreaks in countries far away from the origin one. It suggests that the modern disease epidemic mechanism is similar to the malware propagation in heterogeneous network [1], which consists of two inherent interacting schemes:

- Diffusion: Inside a large geographical area like country, agents live, move and have contact with one another. In contrast to far-distance traveling, their mobilities can be regarded as static and thus the movement of agents is neglected. Patients are infectious and spread virus to agents in their proximity. We visualize this mechanism as pervading aroma and call it diffusion.
- Hopping: An infected-traveler can potentially carry pathogens to other geographical areas via transportation networks or by itself such as flying. For it’s long-distance movement in short time, we call it hopping.

In this paper, we are exploring the analytic model in [2], extending it together with social network analysis under the dynamics of heterogeneous network structure and aiming to investigate the prognostication of a possible cross-area outbreak by exploiting the diffusion and hopping embedded in the network. Looking into the characteristics of viruses, we introduce the vital phenomenon: incubation period, that is, the interval from the time one gets infected to the time it appears symptoms. Though the screening checks at the airports lower the risk of viruses traversing from one country to another, they depends on features of the underlying pathogens. With such intertwining components, our heterogeneous model, however, enables flexible modeling and subsequent analysis. Starting from a single patient, by coupling diffusion state curves, we fully describe epidemics and obtain the analytic predictions. Furthermore, our heterogeneous model allows various applications in evaluating other spreading phenomena and consequently serves as a reference to design advertisement or control strategies of information in complex networks as in [3]. For social network analysis, we propose the method of how the limited resources are distributed for optimally expediting the overall information spreading, considering peoples’ novelty of a frenzy and willingness to discuss that are analogous to the incubation period and screening level.

So far, many researchers explored on how diseases prevail under effects of different homogeneous networks like [4] and [5]. Viewing population in different scales enables heterogeneity among subpopulation. In [6], they utilized this concept along with traditional SIR model and spatial relationships to model the diffusion and hopping respectively. On the specific case in [7], we see the vitality of incubation period that gives rise to late identification of outbreaks, and obscures knowledge of the efficacy of control strategies. In [8], they emphasized the heterogeneity of the network of structure in transmission and investigated various epidemiical situations, including using incubation period in the network model. We inspiredly take it into account, but not in a probabilistic manner, to investigate how the incubation period and screening level may affect the probability and scale of pandemics. Utilizing the diffusion model [2] and small world model [9], we propose to superimpose the diffusions and model the transportation network having small-world features, discarding the distance
factor that no longer plays a leading role as the advances of transportation technology.

As pointed out in [8], the phenomenon of information spreading in networked societies on Internet resembles the epidemics. The effects of artificially structured on-line network on behavior diffusion were investigated in [10]. In [11], a study was presented on the spread of information spreading in networked societies on Internet resembles the epidemic of encountering a susceptible agent on the peripheral of the network. In other words, we obtain the diffusion state curves of transition from state $S$ to $E$.

### II. System Model

We construct a heterogeneous network as a whole picture depicted in Figure 1. Countries are nodes and they are linked if they have flights between. Diffusions of virus are happening inside each countries.

Our system is running in discrete-time fashion with the unit being the average interval between successive flights and incubation period is normalized depending on its relative length to this interval. It is configured such that exact one airplane flies for each link at each unit of time.

Before detailing the components of our heterogeneous model, we list the parameters used in the model.

**Nomenclature**

- $T$ Total time for one simulation
- $N$ Number of countries
- $P$ Population of a country
- $n$ Number of a country’s neighbor
- $p_{rw}$ Probability for rewiring network
- $S_k(t)$ Number of $S$ agents in the $k$th country at time $t$
- $E_k(t)$ Number of $E$ agents in the $k$th country at time $t$
- $I_k(t)$ Number of $I$ agents in the $k$th country at time $t$

$F_k(t)$ Diffusion state curve of $E$ in the $k$th country
$J_k(t)$ Diffusion state curve of $I$ in the $k$th country
$D$ Population density
$R_d$ Contagion range
$\eta$ Number of agents within $R_d$ of a given agent
$\lambda_d$ Infection probability if staying within range $R_d$
$\tau$ Incubation period
$C$ Airplane capacity
$p_{sc}$ Range level of screening failure
$p_{sc,k}$ Screening failure in the $k$th country
$r_{i,j}$ Risk on the direct flight from country $i$ to $j$

#### A. Spreading via Diffusion

People have three health states: healthy and susceptible ($S$); infected, infectious but no symptoms (in the incubation period) ($E$); infectious with symptoms ($I$). People in either state $E$ or $I$ can infect others. The incubation period $\tau$ is the required time of transition from state $E$ to $I$.

The diffusion process is based on the diffusion model proposed in [3]. In a single diffusion generated by an infectious agent from neighbor country at $r$, let $G(r,s)$ denote the number of agents infected, including $E$ and $I$, during time interval $[r,r+s]$. The agents of $E$ on the peripheral spread outward. Given the population density $D$, the radius $R_w(r,s)$ of the diffusion wave is approximated as

$$R_w(r,s) \approx \sqrt{\frac{G(r,s)}{\pi D}}$$  \hspace{1cm} (1)

With respect to an increase of $s$, its incremental spatial infection at time $r+s$ is

$$\Delta G(r,s) = \lambda_d \eta R_d \sqrt{\pi D} \sqrt{G(r,s)} \frac{S(r+s)}{P}$$  \hspace{1cm} (2)

In fact, $R_d \sqrt{\pi D} = \sqrt{\pi}$, so we later only set value to $\eta$, and do not specify $R_d, D$. We model how states transit in only one wave of diffusion in country $k$. In this particular case, the number of agents of the certain state versus time is called as a diffusion state curve and we care $F_k(t), J_k(t)$. The only one diffusion means the spreading prevails without the limit by factor $S(r+s)/P$ that originally models the chance of encountering a susceptible agent on the peripheral of the ripple. In other words, we obtain the diffusion state curves $F_k(t), J_k(t)$ by setting $S(r+s)/P = 1$.

Right before $t$, denoted as $t_-$, the newly-infected agents during $[t-1,t]$ contributes to the growth of state $E$.

$$\Delta F_k(t-) = \Delta G(0,t)$$  \hspace{1cm} (3)

The agents $E$ that got infected at $(t-\tau)_-$ transit to state $I$ at time $t$. Right after $t$, denoted as $t_+$, this transition decreases the number of agent $E$.

$$\Delta J_k(t) = \Delta F_k((t-\tau)_-)$$  \hspace{1cm} (4a)

$$\Delta F_k(t_+) = -\Delta J_k(t)$$  \hspace{1cm} (4b)

So far, we derive the state curves $F_k(t), J_k(t)$ that will show their usefulness in coupling diffusion and hopping.
B. Transportation Network

Before constructing flight network, we look into who takes airplanes for traveling. We suppose that everyone in the countries enjoy the equal chance and need for flight access, and thus the state distribution of the passengers buying flight tickets is exact the one in its origin country before the departure.

We assume that agents $S$ are surely able to board flights, while agents $I$ with symptoms cannot since the screening check absolutely detects the victims (they may too have no stamina). We also suppose that the screening check cannot completely differentiate $E$ agents from those in state $S$, but will block some of them from boarding because signs, e.g. contact history, will reveal that the agents might be $E$. Whether an $E$ agent can enter the flight and gain grant to visit is uncertain. The probability of the screening failure by the $k$th country, whether departure or arrival, is denoted as $p_{sc,k}$. Figure 3 sketches the process where both origin and destination countries perform screening though it’s not always true.

We have agents, countries but lack flight routes. If we observe how airlines arrange routes, we can find that most of them are bidirectional for convenience of tourists and too operations and maintenances of the airlines. The direct flights between two countries are modeled as two directed arc to catch the bi-direction feature. Each directed arc from country $i$ to $j$ is assigned a risk $r_{i,j}$ to capture how likely it is for an $E$ agent to travel via this arc.

$$r_{i,j} = p_{sc,i} \times p_{sc,j}$$

(5)

In our network, $p_{sc,k}$ is generated independently and uniformly from a configurable range level $p_{sc}$.

After modeling individual links, we turn to the flight network. Much examination on real data reveals that it is scale-free with small world traits. Since our current focus is not on the network but the method, simultaneously to mitigate the complexity of construction of network, we adopt the small world model [9] where the range of rewiring $p_{rw} \in [0.01, 0.1]$ corresponds to small-world networks. We manipulate $p_{rw} = 0.1$ to establish one flight network and stick to apply it for all analysis.

C. Spreading via Transportation Network

We utilize superposition to couple the diffusion effects with hoppings in flight network. We assume that the travelers rush to different spots with far distances between after arrival in the destination. It implies that in early stages, the scattered patients from neighbor countries won’t make circles of diffusion overlap and it validates the setting $S(t)/P = 1$ in diffusion state curves. It is crucial for our application of superposition: the domestic epidemic comprises delayed versions of the circles diffusions incident to visitors that arrives at different time.

Let the pathogen pop up in country 1 and $N(x)$ mean the set composed of neighbors of $x$. At time $t$, for every country, we can derive its current number of state $E$,

$$E_i(t) = \sum_{k \in N(i)} \sum_{m=0}^{t-1} \left[ r_{k,i} C_{k,m} E_k(m) \right] F_k(t-m), \forall i \neq 1$$

(6)

Given time $m < t$ and country $i$, the parentheses in double summation means expected value of $E$ agents from country $k \in N(i)$ since $C_{k,m} E_k(m)/P$ of buyers for the flight are agents $E$ and it’s expected that the screening fails with probability $r_{k,i}$. Such patients start circles of diffusion at $m$ and will contribute $F_k(t-m)$ of agents $E$ at time $t$. We sum up the contributions from all neighbors. For the original outbreak, it has one agent $E$ at time 0 so we add another term in this case.

Similarly we have equations for state $I$,

$$I_i(t) = J_i(t) + \sum_{k \in N(i)} \sum_{m=0}^{t-1} \left[ r_{k,i} C_{k,m} J_k(m) \right] J_k(t-m)$$

$$J_i(t) = \sum_{k \in N(i)} \sum_{m=0}^{t-1} \left[ r_{k,i} C_{k,m} J_k(m) \right] J_k(t-m), \forall i \neq 1$$

(7)

and by $S_i(t) + E_i(t) + I_i(t) = P, \forall i, t$, we obtain $S_i(t)$.

It’s worth noticing that the application of diffusion state curves and superposition heavily reduce the computation as we need not compute function $G(r,s)$ for every pair of $(r,s)$ and only calculate for the one $r = 0$ thanks to approximating $S(t) \approx P$. Another advantage is that usage of diffusion state curves makes the modeling on diffusion mechanism flexible.

III. Numerical Analysis

To focus on incubation period and screening level, we let all the countries be homogeneous in terms of their population and density, and therefore the same diffusion state curves $F_k(t), J_k(t), \forall k \in \{1 \ldots N\}$ (we denote it as $[1 : N]$ afterwards). The network settings are listed in Table I, primarily based on the flight network over capitals in West Africa. We certainly pick the 15th country as the first affected country.

We estimate when the countries will encounter an outbreak, under the manipulable variables incubation period $\tau$ and screening range $p_{sc}$. Given the first affected country, the
outbreak time when the epidemic breaks out with at least one
E agent in the $k$th country is defined as

$$B_k := \arg \min_t E_k(t) + I_k(t) \geq 1, \forall k \in [1 : N]$$  \hfill (8)

Given $\tau$ and range $p_{\text{sc}}$, we simulate 100 different sets of the $p_{\text{sc},k}$ uniformly drawn from $p_{\text{sc}}$, and for each country average the outbreak time $B_k$ in (8) over the 100 simulations. We visualize the spreading pattern in terms of averaged outbreak time for the case $\tau = 15$ in Figure 4, and scale the nodes according to its shortest hops to the 15th. Since equation (5) shows $r_{i,j} = r_{j,i}$, the flights between node $i$ and $j$ are linked as one undirected edge with the same risk for clarity. Let each link be distance 1 and hence the shortest distance is equivalent to shortest hops; we use interchangeably. It shows that the outbreak times of the countries with the same hops are quite the same. Also we observe that the screening failure has great effects on the outbreak time.

To quantify the results, instead of averaging $B_k$ over all nodes, we classify them by shortest hops and take an average within the same group to avoid the bias from the uneven numbers of countries of each group since the outbreak time ties the shortest hops as seen in Figure 4. Figure 5 shows in each configuration of $\tau$, the average outbreak time of three screening levels versus shortest hops. As $\tau$ increases with other variables fixed, we observe that the outbreak time is globally but a little advanced. If $\tau$ is fixed and $p_{\text{sc}}$ is increased from $[0, 0.05]$ to $[0.05, 1]$, say, more failures, we can see the salient margins between blue and yellow bars. However, as $[0, 0.05]$ moves to $[0.05, 1]$, the reduced margins turn out to be slight.

To explain why the screening level has more impact on outbreak time, we plot the diffusion state curves shown in Figure 6. It is interesting that the $F_k(t)$ curves come to linear and we inspect our diffusion model. A circle wave of diffusion starts spreading at time 0 and is expanding at $R_d$/unit. After $\tau$, E agents transit to I, and to envision, one can imagine another wave expanding at the origin and at the same speed. For $t \in [1, \tau]$, the number of agents in state $E$ in the circle,

$$E_S(t) \approx \lambda_d \pi (R_d t)^2 D = \lambda_d \pi t^2$$  \hfill (9)
For \( t \geq \tau + 1 \) and right after the transition occurs,

\[
E_S(t_+) = \lambda_d \eta C t^2 - \lambda_d \eta C (t - \tau)^2
= \lambda_d \eta C (2t\tau - \tau^2)
\]  \hspace{1cm} (10)

Thus, before \( \tau \), equation (9) says that \( F(t) \) grows as quadratic curves while thereafter climbs up in linearity as (10). Compared to equation (5) that states the risk is proportional to square of the screening failure probability, the slow increase of potential carriers \( E \) is not that important if the overall screening check is on the high levels.

Another importance is how much time the pathogens take to “hop” to the neighbors. The outbreak delay versus the shortest hop in plotted in Figure 7. The delays are increasing but sudden drop in the 4th hop whatever the case is. We point out that the delay is determined by the degree along with its neighbors. If a country ranks in \( y \) hops from origin and has many neighbors located in \( y - 1 \) or \( y \) hops, it will soon have an outbreak and a short delay. Without both high degrees and neighbors ranked in previous hop, it takes more time to have an outbreak in the country. Since in the small-world network model, nodes are endowed with the same degrees, whether many neighbors are in previous hop comes to be the key.

The neighbors to the origin form a cluster and it speeds up the spreading to countries at the 1st hop within after the first outbreak. As the epidemic approaches to other countries, these countries face a smaller cluster of affected countries at previous hops so that the reinforcement of prevalence is weaken, conducive to a longer delay. In reality, we observe that the areas in the proximity of the initial outbreak soon suffer the invasion of the disease. The time of hearing the next outbreak news is delayed for the clustering effect is reduced. In our transportation construction, the delay to the 4th hop is shortened because only 30 connected nodes are used and the reinforcement in clustering appears saliently at the final hop toward which the epidemics spread.

We articulate the conditions under which our model is valid and accurate in foretelling the epidemics. Recall that the circles of diffusions are supposed to non-overlap one another so that we can superimpose the diffusion state curves. The analysis of epidemic scale in each country is hence valid for the early stages but digress as time goes. Figure 6 provides information on this disparity. After the time 80, the patients (both \( E \) and \( I \)) accumulate to about 1/10 of the population \( P = 10^5 \). The analysis ends up with more deviations after 80 since the first diffusion in the country. However, the delays are lower than 60 in each case, which means the diffusion elapse in previous hop is not so long that the assumption of non-overlapping still works. Hence the model predicts the outbreak time well while fails to tell the epidemics in any country as simulation time goes large.

IV. EXTENSION TO ON-LINE SOCIAL NETWORKS

What our heterogeneous model enables us is to model the information dissemination through artificial social communities on the Internet, analytically investigate the speed of propagation of twittered news or gossips, and in turn develop an optimal strategy of quickly deluging the information flow given limited resources for activating the agents.

People pay attention to the latest news, but soon turn their focus to other things, leading to less discussion and fading of information dissemination. Considering the life cycles of products or propagandas, policy-makers are required to generate a frenzy as soon as possible, broadly over people of various personalities and tastes. We characterize people of heterogeneous preferences as social communities where they have strong ties...
to other members, along with the peoples’ novelty on a new topic. We state how to utilize our heterogeneous model on such social network analysis:

A. Configuration of the Heterogeneous Model

Agents have three informed states related to an upcoming product or latest news: have no interest or never obtain the information (S); have heard, adopt and willing to discuss with others (E); only adopt in mind but no longer talk to others about this (I). The corresponding term to the incubation period \( \tau \) is the activation period during which agents actively converse with others on the subject. By such definitions, we can derive the diffusion state curves \( F_k(t), J_k(t) \) in the same manner.

In social networks, communities are analogous to countries and the weak tie between communities is each assigned a probability by which one member of \( E \) may send information to influence one \( S \) member of another community, similar to the screening failure \( p_{sm} \). The communities interconnect like the transportation network. As discussed before, the underlying diffusion and hopping models are configurable as long as we can obtain diffusion state curves for coupling the mechanisms.

B. Analysis on Speed of Information Dissemination

In contrast to epidemics with uncertainty of the first outbreak, we can selectively place information sources on any communities in designing strategies of information spreading. Advertisements cost, and we express the budget constraint as the total \( M \in \mathbb{N} \) agents we can activate. In the \( j \)th of total \( N \) communities, at \( t = 0 \), we put \( x_j \in \{0 \cup \mathbb{N} \}, j \in [1 : N] \) sources of state \( E \) distant between, subject to

\[
\sum_{j=1}^{N} x_j \leq M \tag{11}
\]

Given the deployment \( \{x_1, \ldots, x_N\} \), we define the arrival time \( B_k, k \in [1 : N] \), same as the concept of outbreak time, to be the shortest time of at least one informed agent appearing in the \( k \)th community. We can directly compute by superposition of diffusion state curves \( F_k(t), J_k(t) \) similar to (6)(7). However, we need to run simulations over all possibilities inherent in the constraint (11) to determine the arrival time of each community and it takes mass computations. We solve the optimization in another approximated but efficient way.

For each deployment of information sources \( \{x_1, \ldots, x_N\} \), we say that the influence delay from \( j \) to \( k \) is \( B_k^{(j)}, j, k \in [1 : N] \), the shortest time of at least one influenced agent showing up in the \( k \)th community when at \( t = 0 \) only \( x_j \) sources lie in the \( j \)th communities with all the other communities in state \( S \). Set \( B_k^{(j)} = \infty \) if and only if \( x_j = 0 \). The overall arrival time is the minimum among the influence delays from the others, \( B_k = \min_{j \in [1 : N]} B_k^{(j)} \tag{12} \)

We face an optimization problem of minimizing the total arrival time over all deployments subject to the budget in (11),

\[
\min_{\{x_1, \ldots, x_N\}} \sum_{k=1}^{N} B_k \tag{13}
\]

The evaluation of \( B_k^{(j)} \) can be carried first before solving optimization problem since we place \( x_j \) sources only in the \( j \)th community with all other zeros. For each deployment, we use corresponding \( B_k^{(j)} \)'s and need not reevaluate.

The value of total arrival time in the optimal solution to (13) is not exactly the one that is obtained by direct evaluations. The reason lies in the reinforcement amid neighbored communities that accelerates the influence and advances the arrival time. Since the arrival time focuses on the very beginning of spreading in a community, the scale of prevalence is not large so that the reinforcement is of small impact, which suggests that our approach to the optimization problem pays off as compared to considerable computations.

V. Conclusion

Our model combines the mechanisms of diffusions and the network in spreading phenomena, and allows diverse configurations in heterogeneity of nodes and the network structure. We relieve the computation burden incident to various settings by utilizing the diffusion state curves and superposition that also allows flexible choices on the diffusion mechanism.

In our application, we modify the diffusion model to catch the impact of incubation period and screening check. We assess when the outbreak takes place in cross-country spreading. It shows the capability of our model to function as a quick guide for evaluating the propagation speed and prevalence of spreading phenomena inherent in social networks or migration network. Then we formulate an optimization problem of how to place the sources to achieve quick information dissemination over our heterogeneous social networks given the advertisement budget. These results can be exploited for the purpose of developing policies to avoid outbreaks or proposing strategies to advertise products.

REFERENCES


