Optimal Surveillance Network Design: a Value of Information Model - Supplementary Materials

Matteo Convertino[1,2,3], **Yang Liu**[1], **Haejin Hwang**[1]

[1] School of Public Health, Environmental Health Sciences Division and Public Health Informatics Program, HumNat Lab, University of Minnesota, MN, USA

[2] Institute on the Environment, University of Minnesota, MN, USA

[3] Institute for Engineering in Medicine, University of Minnesota, MN, USA

October 16, 2014

Keywords: value of information, surveillance network, transmission network, network resilience, topology, uncertainty, biosafety, biodefense, outbreak sources

Corresponding author contact: Academic Health Center, Mayo Memorial Building, Suite 1232 12th fl., MMC807, 420 Delaware St. SE, Minneapolis, MN, 55455; email: matteoc@umn.edu

S1 Case-study Outbreaks: Specifics

S1.1 2004-2007 H5N1 Pandemic

The global spread of highly pathogenic H5N1 avian flu from birds in humans was considered a significant pandemic threat (1). Influenza A/H5N1, a severe strain of bird flu, was first found to infect humans during poultry outbreaks in Hong Kong in 1997. A mass culling of all poultry in Hong Kong may have prevented a human A/H5N1 pandemic (global epidemic) at that time. Ongoing outbreaks in poultry continued in Indonesia, Thailand and Vietnam through 2005. Human cases occurred in Thailand, Vietnam, and for the first time Cambodia. In April 2005, in the Qinghai Lake district of China, over 6,000 wild birds of many different species suddenly died from H5N1. This lake is a major stop for many migratory birds, and it is thought that this event contributed significantly to the apparent sudden subsequent spread of H5N1. Migratory birds, and possibly trade in poultry, have since spread the virus to parts of Russia, the Indian subcontinent, the Middle East, Europe and Africa. From inception through 2007, the total number of WHO-confirmed cases was 349, with 216 of those fatalities (as reported by the U.N. on January 15, 2008, confirming earlier deaths) reflecting a 62% fatality rate among WHO-confirmed cases through 2007. In 2005, when a markedly less-lethal strain in Northern Vietnam was responsible for most of the cases reported worldwide, only 42 of 97 people confirmed by the WHO to be infected with H5N1 died a 43% fatality rate peak. In 2006, the case fatality ratio was higher among the WHO-confirmed cases, with 79 deaths among 114 confirmed cases. In 2007, 59 of the 86 WHO-confirmed cases ended in death, again a 69% fatality rate. And 24 of the first 31 cases of 2008 (to April 30, 2008) have been fatal, or 77%. The incidence of human cases peaked, in each of the three years in which cases have occurred, during the period roughly corresponding to winter and spring in the northern hemisphere.

S1.2 2012 Salmonella Epidemic in USA

First reported in April 2012, an outbreak of salmonellosis caused by rare serotypes, Salmonella Bareilly and Salmonella Nchanga, was reported in 28 states, mostly in the Eastern U.S., having caused no deaths, but 425 cases of illness and 55 hospitalizations (2). This outbreak was linked to the consumption of raw scraped ground tuna product. The source was frozen raw yellowfin tuna product aka Nakaochi Scrape manufactured by Moon Marine USA Corporation. This product was voluntarily recalled after the CDC discovered strains of salmonella in the packages. This tuna scrape came from a single processing plant in India owned by Moon Marine International of Taiwan. Tuna are plentiful off the Indian coast, and the tuna pro-

cessing industry is expanding rapidly. India has dozens, perhaps hundreds, of fish processing facilities, but most are relatively small and their number, size and geographical dispersion make monitoring difficult.

S1.3 2010 Cholera Epidemic in Cameroon

Cholera is endemic to more than 50 countries, remaining as a major cause of morbidity and mortality worldwide. In 2010, one of the major cholera epidemic occurred in Central and West Africa, resulting in 2,466 deaths (3; 4; 5; 6). In the Far North Region of Cameroon, one of worst-hit region in West Africa, more than 9400 cases and 600 deaths were reported. The most recent data that from the Ministry of Health show that 7,247 cases occurred in Cameroon (6,000 of these in the Far North region), including 483 deaths. Other sources talk about more than 9400 cases and 600 deaths. Yet, the cholera burden in terms of death is relatively high (4) with a fatality rate of 12% (245 deaths) at the peak of infection that is one of the highest recorded in the world. The outbreak in 2010 in Cameroon was certainly the largest ever experienced in Cameroon in the last 20 years in terms of illnesses and deaths. Such outbreaks particularly hit the North rural region close to Lake Chad, the East mountain region close to Nigeria, and the urban area of the capital of the Far North Region.

S2 Supplementary Methods

S2.1 Radiation Model of Mobility Fluxes

We use a physical-based and data-verified numerical approximation of the food/human mobility with the purpose to detect the backbone of the food supply chain and human mobility network that explains the observed spatio-temporal patterns of outbreaks. The magnitude of a network flux (link weight) represents the total food/human flux from one country to another represented as nodes. For the food/human mobility network we use a recently introduced radiation model (7) versus the most classical gravity model that underestimates long-range trades. For this model, the connection probability of the average food/human mobility flux from community (i.e., node) i to j, is $Q_{ij} = \frac{H_i H_j}{(H_i+H_{ij})(H_i+H_j+H_{ij})}$, where H_i and H_j is the population of the source and destination community i and j, respectively, and H_{ij} is the total population within a radius centered in i excluding the source and destination populations. The average food/human mobility flux between communities is defined for any food/human mobility flux incoming in USA as $\langle F_{ij} \rangle = F_i \, Q_{ij}$ where F_i if the food/human flux from node i (i.e., an exporting country when considering external fluxes in USA). Yet, food/human mobility patterns are defined according to a connection matrix in which food

leaves a community i for a target community j with a trade rate γ.

S2.2 Surveillance Network Topologies

Network topologies are defined by different arrangements of nodes and links, thus connectivity and distance among nodes define the topology of networks. Connectivity and distance are defined in terms of network clustering, C, and average path length L. There are other network factors that are suitable to characterize network topologies in more detail (8; 9); however, here we focus only on C and L because of their simplicity, manageability and ability to characterize networks. The clustering coefficient C of a selected node is defined in a probabilistic sense as the probability that two randomly selected neighbors are connected to each other (8; 9; 10). Evidence suggests that in most real-world networks, and for example in social networks, nodes tend to create tightly knit groups characterized by a relatively high density of ties; this likelihood is greater than the average probability of a tie randomly established between two nodes (8). The average path length L is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. Both clustering and path length depends on the node degree k and the number of nodes N (8; 9). The degree of a node in a network is the number of connections or edges the node has to other nodes. The utility of both clustering and path length in term of rewiring probability varies as a function of k and N. The rewiring probability determines the probability that an edge gets rewired and defines the network topology according to the Watts-Strogatz model (11) from the regular to the random network topology. The rewiring probability is proportional to the velocity to respond to changes determined by both random and target attack. In a surveillance network perspective, an attack can imply the inability to report outbreak information. In the paper we reproduce the range of existence of all feasible network topologies by varying C and L as a function of four classes determined by the number of nodes and the node degree (Figure S1).

All feasible networks are built with the Watts-Strogatz model (11) that begins with a regular network structure. The regular network has a very high clustering coefficient along with high average path length (10). As the rewiring probability U increases the network topology changes. We investigated the three most important network topologies: regular, small world, and random. Within the small world network topology, there is the scale free network topology that is an "ultra small world" network (12). The rewiring creates shortcuts across the network, varies the mean path length L and the clustering C across the network from the regular to the random topology. As U increases the mean path length decreases rapidly, as shown in Figure S1 (right plots). The rewiring also decreases the clustering

coefficient C, but there is a large range of U where the clustering coefficient is still large and the path length is small, that is the small world network regime.

S2.3 Integrated Network Resilience

The integrated network resilience (INR) (elsewhere called "index of network resilience" (13; 14)) is defined as a function of two metrics, the clustering coefficient C and the average path length L that depend on the network rewiring probability. The maximization of the rewiring probability maximizes the resilience of the system that is a function of the rapidity to detect and respond which translates into the minimization of number of epidemic cases. Such resilience is only considering the topology of networks assuming that surveillance function is working efficiently throughout all nodes. The contribution of C and L is combined together in a multiattribute model as in (13) and (14) to calculate INR. Here we think the rewiring probability as the "utility" or resilience of a network to maintain its topology after a random or targeted attack as a function of C and L (8; 9). More precisely, a network is identified by the index $n = (C(k, N), L(k, N))$, where k and N are the average node degree and the number of nodes that alter $U(C)$ and $U(L)$. k and N are typically design factors of networks and depends on the available budget. Because of the different variability of C and L with respect to the rewiring probability (Figure S1) we assume that INR is calculated as linear combination of the partial utilities $U(C)$ and $U(L)$ for both factors as:

$$INR(C(k, N), L(k, N)) = U(C)\, w_C + U(L)\, w_L \ , \tag{1}$$

where w_C and w_L are weights that assign relative preferences (or constraints) to the clustering and to the path length of a network according to the favored network design and boundary conditions. For example decision makers may want a more clustered network or a more widespread network with longer link length. In term of INR it is generally better to have a medium value for both C and L but this is not always achievable. Preferences can depend on the system considered and other socio-economic considerations. Here, we assume $w_C = w_L = 0.5$ for all systems. In general, INR depends on the system where it is computed because the location of nodes influence the probability of rewiring of the network.

The small world network is the network that preserves its topological features for the largest range of rewiring probability in comparison to all other network topologies and this contributes to the high structural resilience of the small world network (15). That means that given a change in the network (that can be caused by a random failure or a targeted attack) the network maintains the features of the small world topology and likely its function

(16). The scale free network is instead characterized by a very small range of rewiring of its links that preserves the scale free features. Thus, on average the resilience of scale free networks is smaller than small world networks (16).

Supplementary Figure Captions

Figure S1. Rewiring probability of network topologies as a function of clustering and average path length. The utility of a network is expressed by its rewiring probability that is conditional to the clustering coefficient, C, and the average path length L. The utility curves are calculated for the same networks for four classes of node abundance N and average degree k within the system. The green dotted lines define the range of existence of the four network topologies: regular, small world, scale free, and random network topology from the top to the bottom of each plot.

Figure S2. Integrated network resilience and error probability. Two measures are included in the same plot along the y-axis: from bottom to top the Integrated Network Resilience is shown; from top to bottom the distribution of error distances from correct outbreak sources is reported. The probability distribution of the distance from correct outbreak sources is related to the distribution of outbreak wave velocities. Both INR and pdf(distance) are calculated for the three surveillance network topologies and epidemic considered.

Figure S3. Outbreak waves and wave velocity error. The top plot shows the error in the outbreak velocity as a function of the value of information related to different surveillance network topologies for the outbreaks considered in this paper. The error is related to the misreporting of cases in space and time. The bottom plot is showing outbreak waves related to the epidemic considered; from left to right epidemic waves are shown for different instantiations from the beginning to the end of the epidemic spanning geographical distances (x-axis). j_n is the normalized number of infected dependent on time and distance. The outbreak source is at distance equal to zero.

References

[1] Kilpatrick AM, Chmura AA, Gibbons DW, Fleischer RC, Marra PP, et al. (2006) Predicting the global spread of h5n1 avian influenza. Proceedings of the National Academy of Sciences 103: 19368-19373.

[2] GATS (2014) Multistate outbreak of salmonella bareilly and salmonella nchanga infections associated with a raw scraped ground tuna product. Technical report. http://www.cdc.gov/salmonella/bareilly-04-12/index.html?s_cid=cs_654; Accessed April 2014.

[3] Guevart E, Noeske J, Solle J, Essomba J, Edjenguele M, et al. (2006) Factors contributing to endemic cholera in douala, cameroon. Med Trop (Mars) 66.

[4] Njoh, ME (2010) The cholera epidemic and barriers to healthy hygiene and sanitation in cameroon: A protocol study. Technical report, Umea University. http://bvs.per.paho.org/texcom/colera/MENjoh.pdf.

[5] Tatah A, Pulcherie K, Mande N, Akum N (2012) Investigation of water sources as reservoirs of vibrio cholerae in bepanda, douala and determination of physico-chemical factors maintaining its endemicity. Onderstepoort J Vet Res 79.

[6] Convertino M, S L, M A, Morris S (2014) Importance, interaction, and scale-dependence of cholera outbreak drivers: Metamodeling predictions. BMC Infectious Diseases .

[7] Simini F, González MC, Maritan A, Barabási AL (2012) A universal model for mobility and migration patterns. Nature 484: 96–100.

[8] Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74: 47–97.

[9] Newman M (2003) The structure and function of complex networks. SIAM Review 45: 167-256.

[10] Emmert-Streib F, Dehmer M (2012) Exploring statistical and population aspects of network complexity. PLoS ONE 7: e34523.

[11] Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393: 440–442.

[12] Cohen R, Havlin S (2003) Scale-free networks are ultrasmall. Phys Rev Lett 90: 058701.

[13] Halpern BS, Longo C, Hardy D, McLeod KL, Samhouri JF, et al. (2012) An index to assess the health and benefits of the global ocean. Nature 488: 615620.

[14] Pandit A, Crittenden J (2012) Index of network resilience (inr) for urban water distribution systems. URL http://www.tisp.org/index.cfm?cdid=12519&pid=10261.

[15] Sinha S (2005) Complexity vs. stability in small-world networks. Physica A: Statistical Mechanics and its Applications 346: 147 - 153.

[16] Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. Nature 406: 378–382.

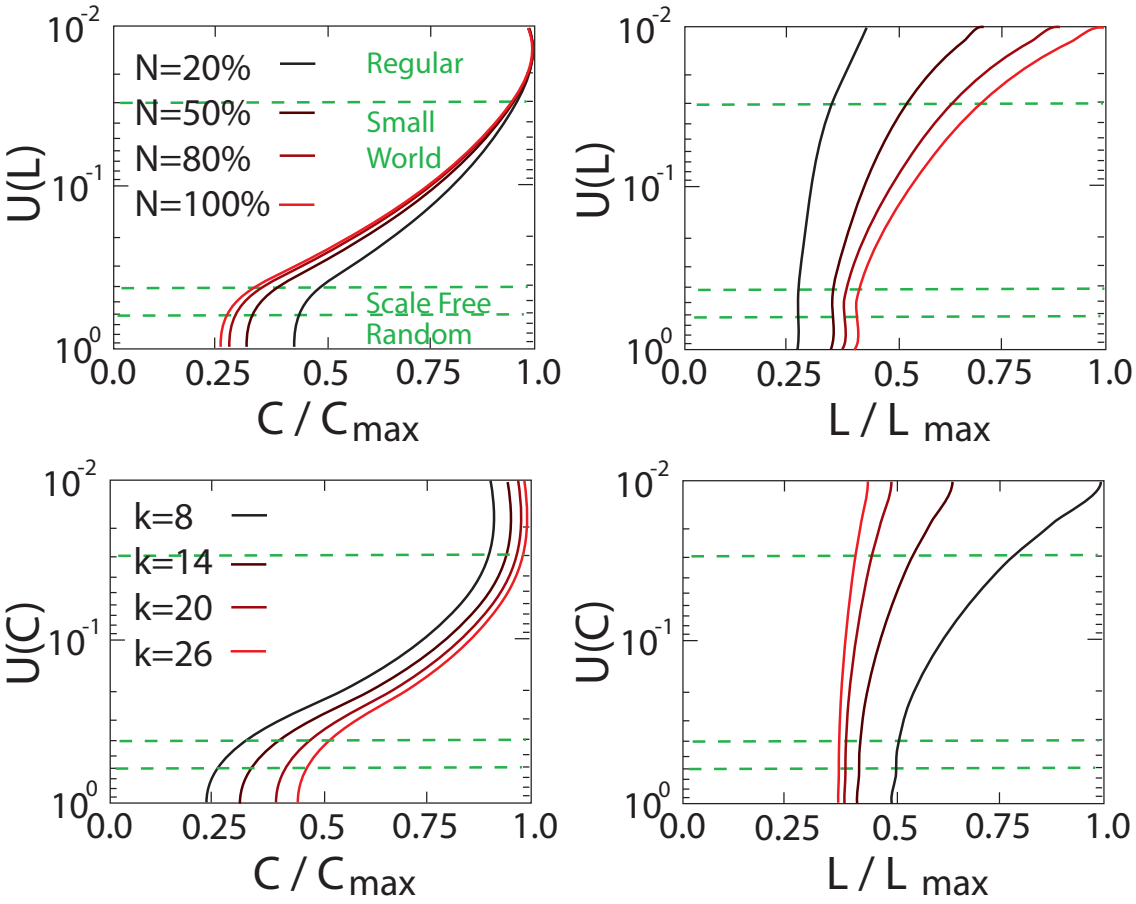

N=20%
N=50%
N=80%
N=100%
Regular
Small
World
Scale Free
Random
U(L)
C / C max
L / L max
k=8
k=14
k=20
k=26
U(C)

Figure S1:

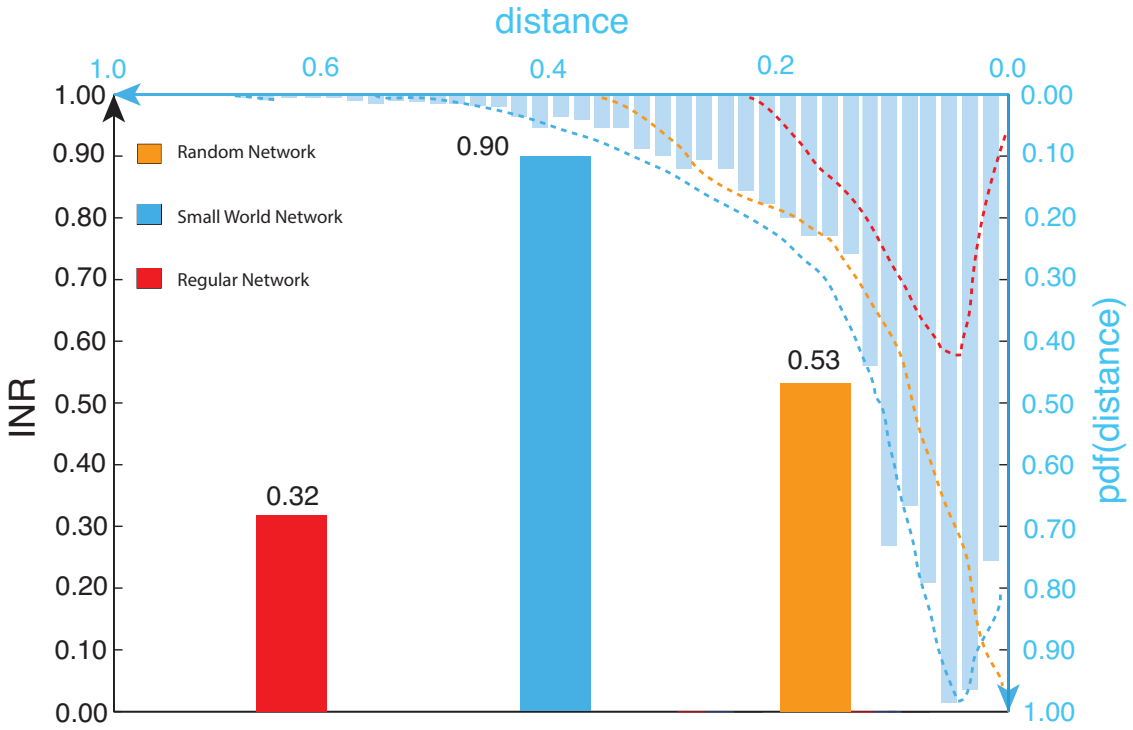
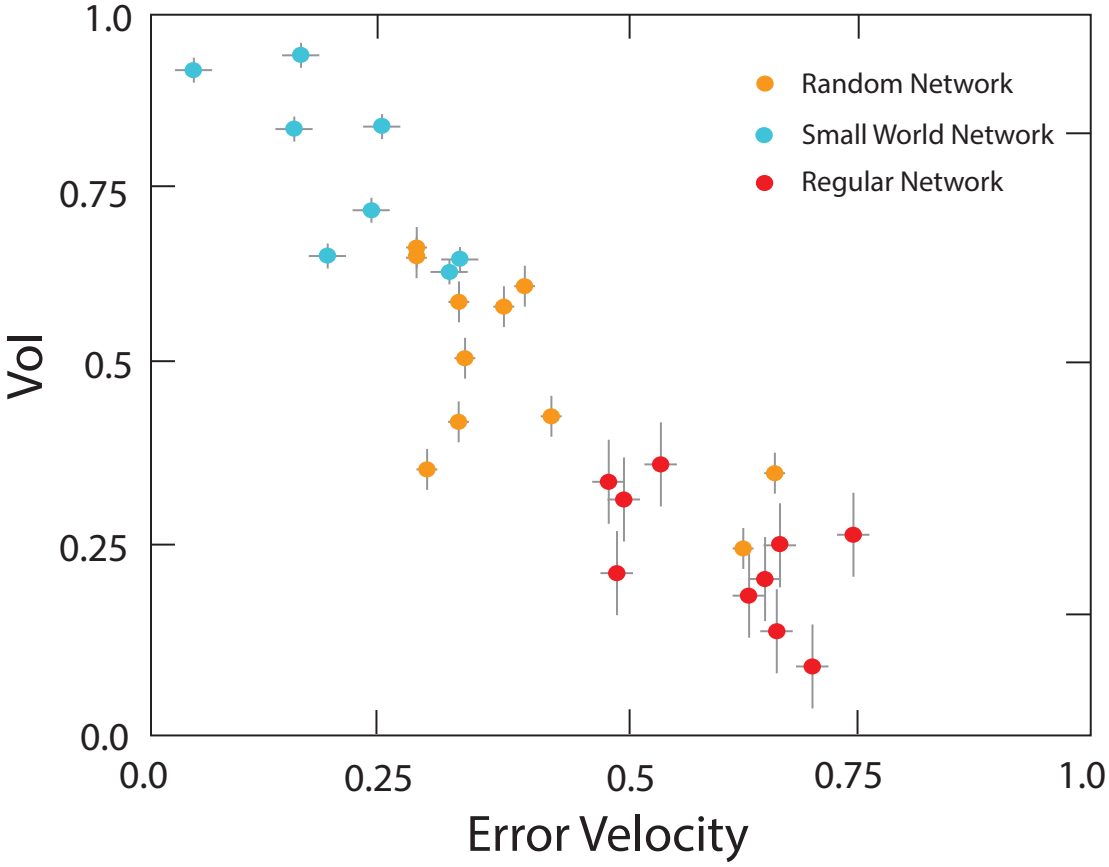

distance
1.0
0.6
0.4
0.2
0.0
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
INR
Random Network
Small World Network
Regular Network
0.32
0.90
0.53
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
pdf(distance)

Figure S2:


Random Network
Small World Network
Regular Network
Vol
Error Velocity
0.0
0.25
0.5
0.75
1.0

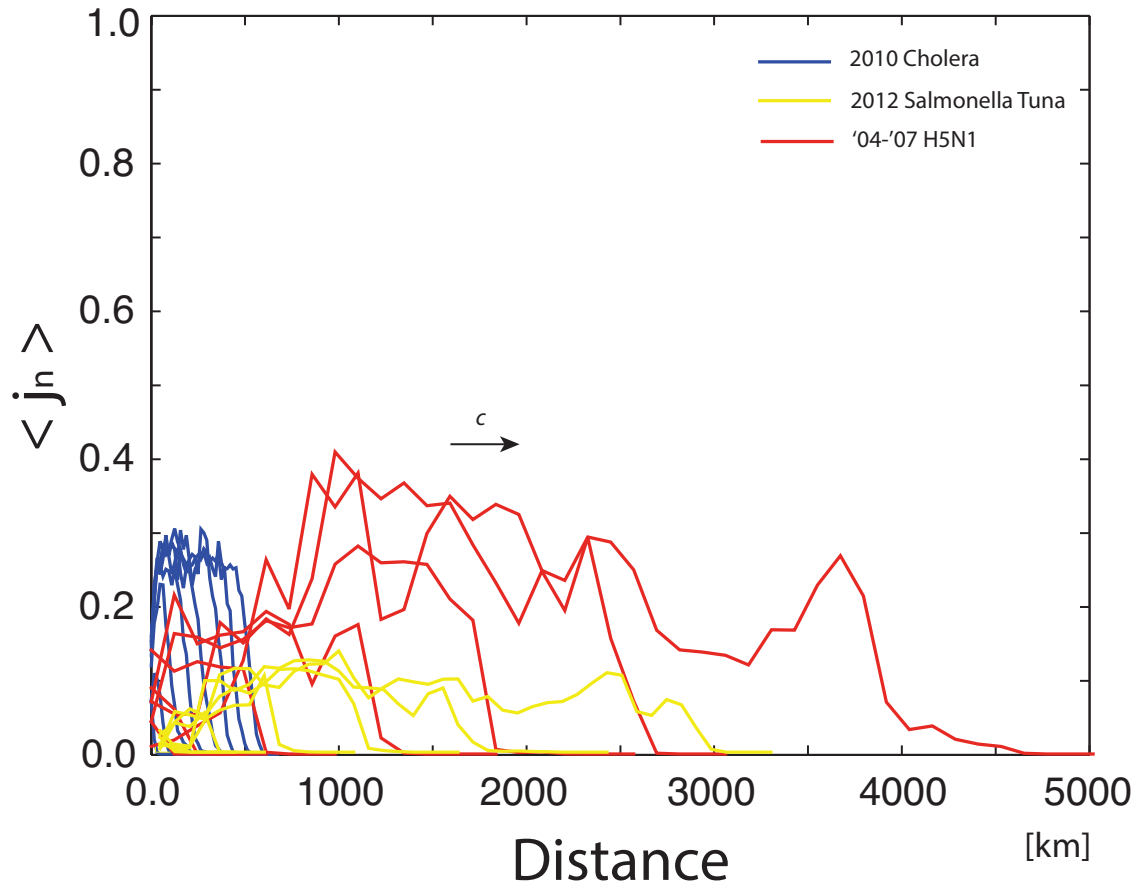

2010 Cholera
2012 Salmonella Tuna
'04-'07 H5N1
c
< jn >
Distance
[km]
0.0
0.2
0.4
0.6
0.8
1.0
1000
2000
3000
4000
5000

Figure S3: