END NOTE

The infection tree of global epidemics

ANA PASTORE Y PIONTTI, MARCELO FERREIRA DA
COSTA GOMES, NICOLE SAMAY, NICOLA PERRA and
ALESSANDRO VESPIGNANI

*Laboratory for the Modeling of Biological and Socio-Technical Systems, Northeastern University, Boston,
MA 02115, USA*
(e-mails: {`anapastore,marfcg,nsamay,nicolaperra,alexves`}`@gmail.com`*)*

The spreading of transmissible infectious diseases is inevitably entangled with the dynamics of human population. Humans are the carrier of the pathogen, and the large-scale travel and commuting patterns that govern the mobility of modern societies are defining how epidemics and pandemics travel across the world. For a long time, the development of quantitative spatially explicit models able to shed light on the global dynamics of pandemic has been limited by the lack of detailed data on human mobility. In the last 10 years, however, these limits have been lifted by the increasing availability of data generated by new information technologies, thus triggering the development of computational (microsimulation) models working at a level of single individuals in spatially extended regions of the world. Microsimulations can provide information at very detailed spatial resolutions and down to the level of single individuals. In addition, computational implementations explicitly account for stochasticity, allowing the study of multiple realizations of epidemics with the same parameters' distribution. While on the one hand these capabilities represent the richness of microsimulation methods, on the other hand they face us with a huge amount of information that requires the use of specific data reduction methods and visual analytics.

In the study of the global spreading of epidemics, it is now possible to develop models that keep track of single individuals flying on specific airline connections. This approach allows the construction of networks characterizing the spreading of the infection from one area of the world to another (generally, urban or census areas). These networks identify in each microsimulation the "infector" city of each "infected" city (Balcan et al., 2009a; Eggo et al., 2009). In other words, every time an individual carries the disease from a city i to city j, we draw a link that shows that the infection propagated along that connection. On its turn, each "infectee" city can be the "infector" of another city, giving rise to a tree that characterizes the spatial spreading of the disease at the city level, as measured from the movement of single individuals.

Keeping track of the infection process produces a considerable amount of data in each single microsimulation; data which are then multiplied by the number of stochastic realizations needed to extract meaningful statistical patterns. While in each realization one single invasion tree emerges, by repeating the microsimulation process, each city j can have a single infector source i or several possible infector

Fig. 1. Examples of (a) an infection network, and (b) infection tree. On the network, each directed edge weight represents the probability of the source node being responsible for introducing the disease on the target one. The infection tree is built by taking into account the paths with higher probability. (color online)

sources. The average of p_{ij} over all realizations will give us the resulting probability that the infection will propagate along the connection linking city i to city j. These averages are then used as the weight of the edge between the corresponding nodes on the network, as shown in Figure 1(a).

Generally, the resulting networks are extremely complex as the infection of each city j might follow alternative routes in each realization. For this reason, it is convenient to use data reduction strategies for the geographical infection trees that consider only the most likely transmission routes. One possible technique, for a set of simulations sharing the same source of the outbreak, is the extraction of the Directed Minimum Spanning Tree using the Chiu–Liu/Edmonds algorithm (Chu & Liu, 1965; Edmonds, 1967). In this case, the cost of each edge w_{ij} is given by the probability $1 - p_{ij}$. This procedure results in a directed spanning tree where every node j has no more than one incoming edge, corresponding to the most likely infection path that leads to the outbreak at each infected node (Figure 1(b)).

Here we consider the microsimulations of global epidemic spreading generated by the Global Epidemic and Mobility (GLEAM) model (Balcan et al., 2009a; Balcan et al., 2010). GLEAM is an individual-based stochastic model that considers more than 3,300 census areas in 220 countries of the world and their population on a resolution of 15×15 minutes of arc. It integrates official airline data and commuting patterns among all census areas. Within each subpopulation, a compartmental structure models the disease spread among individuals. Individuals can move from one subpopulation to another along the mobility network; in this way, an outbreak originated in a seed subpopulation can lead to a global scale epidemic. For each realization, the model tracks the infection tree at the single individuals level and allows the construction of the epidemic tree. The detailed algorithms for the construction of the epidemic tree in the GLEAM model can be found in Balcan et al. (2009a). The GLEAM model has been used in the real time analysis of the H1N1 2009 pandemic, and the calibration and the results obtained with the model are reported in Balcan et al. (2009b) and Tizzoni et al. (2012).

We consider a hypothetical pandemic with the same parameters of the H1N1 2009 pandemic (Balcan et al., 2009b), starting in Hanoi, Vietnam. The infection tree obtained is shown with the nodes in their corresponding geographical location in Figure 2, along with a color code indicating the disease arrival time at each node. From Vietnam the epidemic spreads to the main hubs in Asia – Singapore, Bangkok, Tokyo, and Hong Kong. From those hubs the epidemic finds its way to the American Continent, Europe, and Australia. It is interesting however to note that different
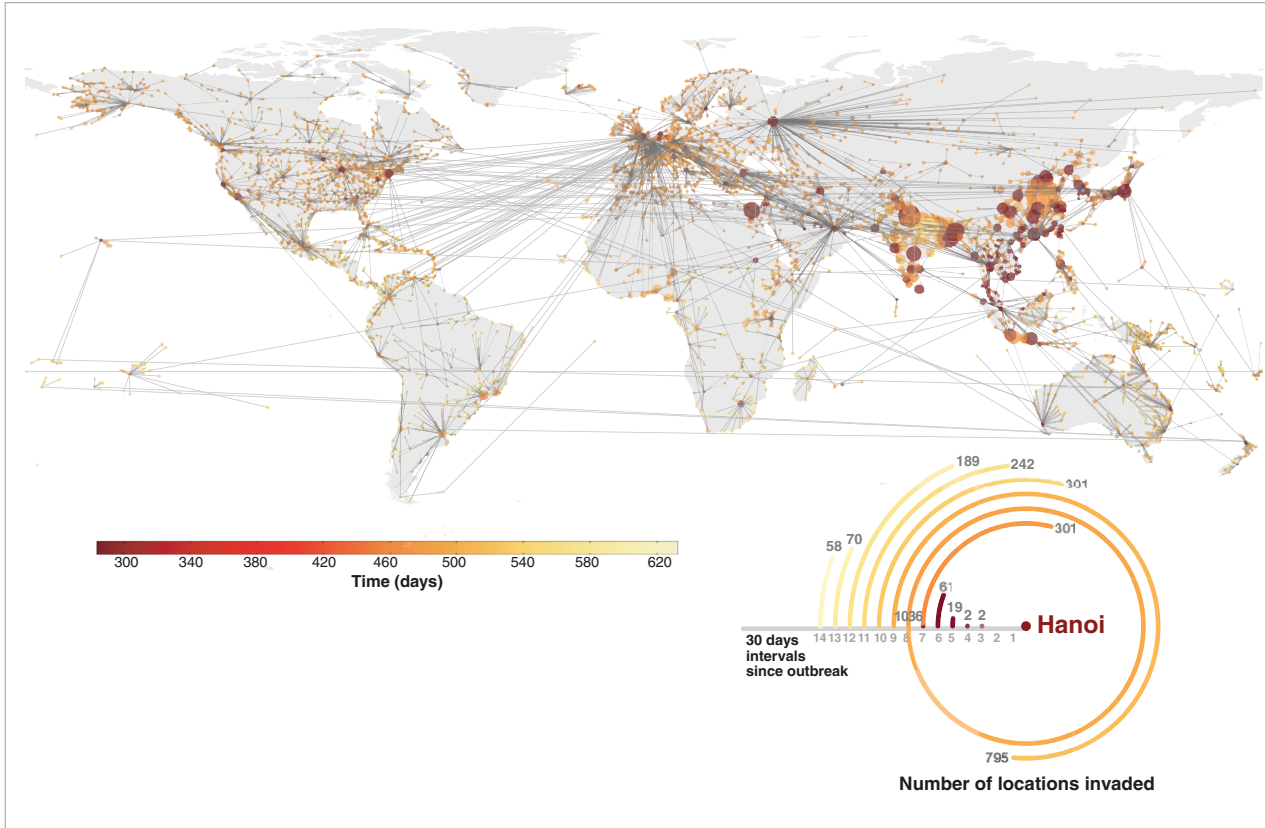
Fig. 2. Infection tree for an outbreak from Hanoi, Vietnam, over 1,000 simulations. The size of each node is proportional to the population, and the color corresponds to the time of disease arrival at that node from dark red (earlier) to light yellow (later). Each concentric arc on the circular plot on the bottom right is proportional to the number of locations invaded at 30 days intervals since the outbreak. The full circle representing 1,036 locations, which occurred in the interval between 210 and 240 days since the outbreak in Hanoi. (color online)
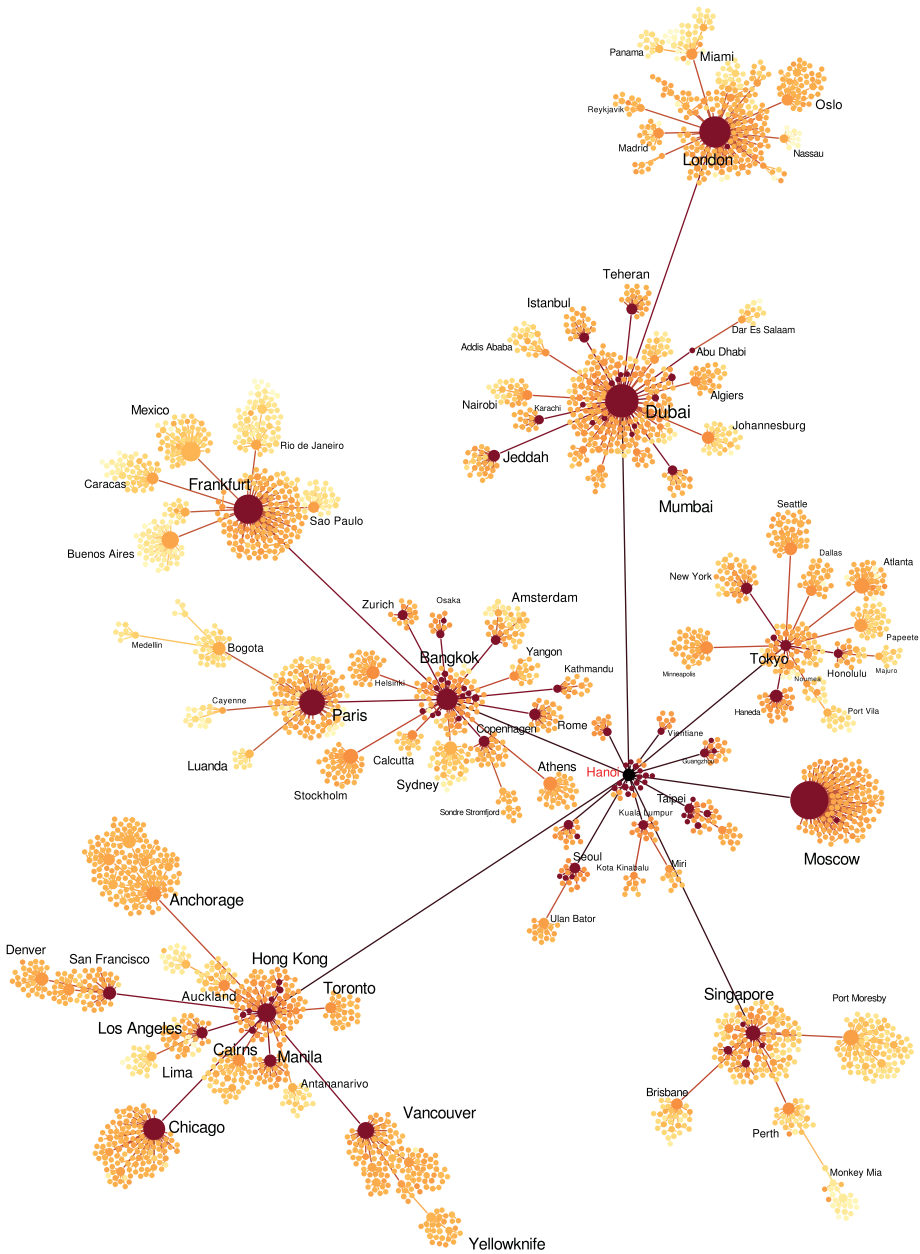
Fig. 3. Network representation of the infection tree. The origin of the outbreak is the city of Hanoi, Vietnam, colored in black. The color code for the nodes follows the same relation as in Figure 2, with dark red meaning early arrival of the disease and light yellow meaning a wider time gap between the outbreak in Hanoi and the arrival on the corresponding node. The size of each node is proportional to the degree. The labels are related to the hub at the corresponding branch. (color online)

pathways are used by the epidemic to reach different parts of the United States or Europe (see Fig. 3). For instance, the United Kingdom and the Southeastern US are reached through Dubai that acts as the Middle East gateway. Noticeably, this infection tree is indeed extremely different from the one modeled and observed in the data concerning the actual H1N1 pandemic starting from Mexico. This implies a sensitivity to the initial conditions of the disease that makes the study of infection tree extremely relevant. One possibility is to create a catalog of possible infection trees that could be used to chart the next pandemics at their very early stages of unfolding.

It is worth stressing that by construction the minimum spanning tree algorithm provides for each city the most likely path from the seed, but it neglects alternative routes that might have probabilities of occurrence just slightly smaller than those selected by the algorithm. Other algorithms, weighting scheme, and backbone filtering schemes can thus be used to provide alternative definition of the infection tree (Eggo et al., 2009; Serrano et al., 2009; Brockmann & Helbing, 2014). The infection tree also lends itself to further analysis in which it is possible to evaluate the role of each node or group of nodes in the spread as well as shed light on the global circulation of seasonal influenza. In summary, the use of the infection tree provides a meaningful representation of the epidemic spreading, statistically aggregating the microsimulation results, thus proving this data reduction technique a valuable visual tool in the mapping of the progression of epidemics in the case of scenario analysis and pandemic preparation plans.

Acknowledgments

References

Balcan, D., Colizza, V., Goncalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, **106**(51), 21484–21489, doi:10.1073/pnas.0906910106.

Balcan, D., Goncalves, B., Hu, H., Ramasco, J. J., Colizza, V., & Vespignani, A. (2010). Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of Computational Science*, **1**, 132–145.

Balcan, D., Hu, H., Goncalves, B., Bajardi, P., Poletto, C., Ramasco, J. J., Vespignani, A. (2009). Seasonal transmission potential and activity peaks of the new influenza A(H1N1): A Monte Carlo likelihood analysis based on human mobility. *BMC Medicine*, **7**, 45, doi:10.1186/1741-7015-7-45.

Brockmann, D., & Helbing, D. (2014). The hidden geometry of complex, network-driven contagion phenomena. *Science*, **342**, 1337–1342.

Chu, Y. J., & Liu, T. H. (1965), On the shortest arborescence of a directed graph. *Science Sinica*, **14**, 1396–1400.

Edmonds, J. (1976). Optimum branchings. *Journal Research of the National Bureau of Standards*, **71B**, 233–240.

Eggo, R. M., Cauchemez, S., & Ferguson, N. M. (2011). Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States. *Journal of the Royal Society Interface*, **8**, 233–243.

Serrano, M. A., Boguna, M., & Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, **106**, 6483–6488.

Tizzoni, M., Bajardi, P., Poletto, C., Ramasco, J. J., Balcan, D., Goncalves, B., Vespignani, A. (2012), Real-time numerical forecast of global epidemic spreading: Case study of 2009 A/H1N1pdm. *BMC Medicine*, **10**, 165, doi:10.1186/1741-7015-10-165.