Epidemic modeling using an effective air-traffic network

Focus paper:

Team members and contributions

- **Akanksha**: Report, Presentation
- **Daniel**: Report, Presentation. Attempted to model the H1N1 infection using an SIR model
- **Dileep**: Construction of the connectivity network, Plot of infection vs. time and calculation of arrival times, Calculation of shortest paths - geographical distances and effective distances. Linear regression plots of arrival times vs. distance metrics. Report, Presentation
- **Lingyu**: Time-lapse of H1N1 infection and its spread in the EU. Report, Presentation

Background and Motivation

Reaction diffusion models conventionally used to study the global spread of epidemics like Black Death have been rendered inaccurate by the global-mobility network which is drastically reshaped by modern air traffic. Long distance travel has led to complex and spatially incoherent spreading patterns. As a result it has become a major challenge of the 21st century to construct an accurate model for global disease dynamics. Such a model is essential to deploying effective containment strategies and potentially saving many lives as well as reducing the socio-economic impact of epidemics.

The key idea pursued in the paper is effective as opposed to geographic distance. It incorporates the concept that two cities or nodes can be spatially separated by large distance but the effective distance, in terms of air-traffic connectivity, can be significantly smaller. The spread of contagion can be effectively modeled if the connectivity matrix is considered. For instance, two major cities that are separated by large geographical distance but are strongly connected with considerable number of flights are more vulnerable to spread of contagion rather than the suburbs of the city that might have comparatively far less intercity traffic even after being closer spatially. The complexities of these spatio-temporal patterns can be simplified enough by replacing conventional geographic distance with effective distances derived from the underlying air-traffic network, so as to be able to define the epidemic wavefront, predict time of arrivals at distant location and the pace of propagation.

Methods

Data:

The global air-traffic flow network which the authors has utilized for their work is proprietary. Since we did not have access to this data, we constrained to model H1N1 epidemic spreading only within countries in the EU for which air-traffic data was publicly available. [1]

The global H1N1 epidemic data was retrieved from WHO [2]. The data contained all H1N1 infection cases from 2009 to present for every major country.

H1N1 that originated in Mexico (around week 10-11) spread to European countries by week 18. Snapshots of the spread throughout the EU is shown in figure 1. Affected countries included Netherlands (1 case), Italy (4 cases), Denmark (1 case), Germany (9 cases), Switzerland (1 case) and Ireland (2 cases). In the meantime, number of recorded H1N1 cases exceeded 2000 in Mexico. This raises a potential concern, since the data relies on nodes that are located outside of the used network.
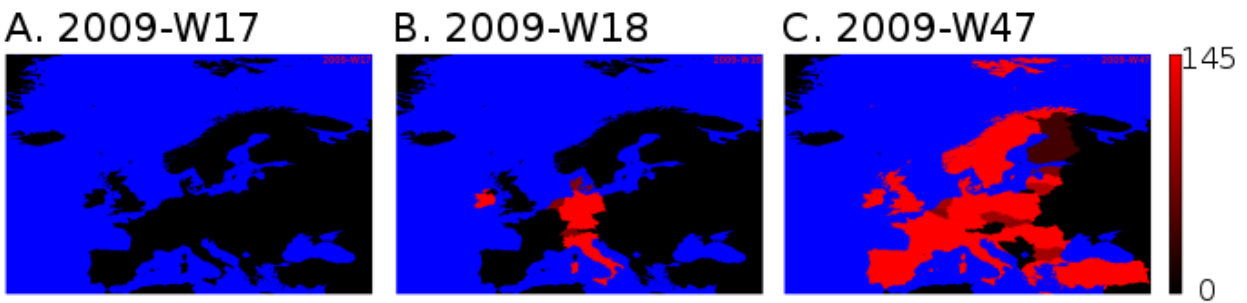
Figure 1. Snapshots of the spread of H1N1 across Europe in 2009. **A)** Week 17 of 2009 **B)** Week 18. **C)** Week 47.

Network Analysis:

The EU air-traffic network was constructed using networkX in python. The flow rates (passenger flux) between the nodes (countries) were calculated as an average of the last 10 years. The connectivity matrix was then added to its transpose and averaged in order to make the matrix symmetric. This was necessary in order to equate influx and outflux at every node (equilibrium). Geographical and effective distances between the nodes were then calculated. The geographical distances between countries was the great-circle distance between the centroids of airports of each country. The effective distance was the shortest distance between nodes in the most probable path tree.

Probable path between m and n,
$$P_{mn} = F_{mn}/F_n,$$

Directed length dmn,
$$d_{mn} = (1 - \log P_{mn}) \geq 1$$

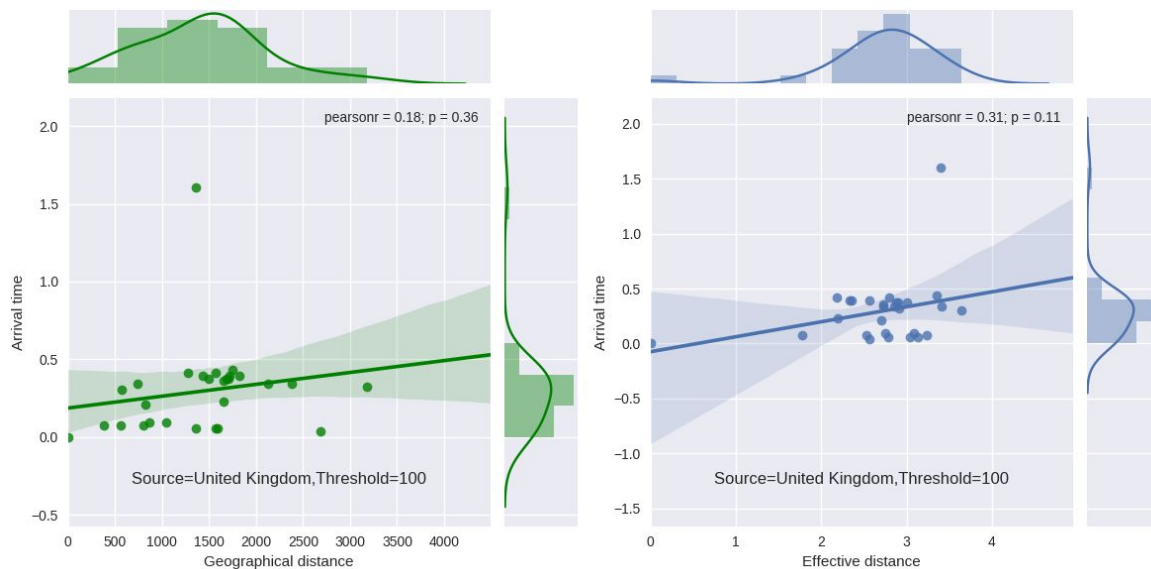Effective distance,
$$d(m|n) = \min_{\Gamma} \lambda(\Gamma)$$

All the data, scripts and results of this study are available on github. [3]

Results:

A connectivity network was constructed using effective distance and a visual representation of the network is shown in Figure 2.

Figure 2. Connectivity network for European countries. Each node is a country and the size of the circle is proportional to the population of the country. The thickness and color of the edges are the flow rates between the nodes.

Figure 3. A linear regression plot of arrival times vs. various distance metrics. Source: UK.

The results indicate that effective distance is not a much better metric to study disease arrival times as compared to the geographical distances.

Discussion and future prospects

- One of the key variables in this model is the origin location of the infection. Since the disease did not originate in any of the nodes in this network it is difficult to pinpoint the exact source of the infection - different thresholds lead to identification of different sources.
- The other major problem that we encountered was the fact that this network is not isolated and is connected to various countries that are not part of the network and as a result infected individuals can enter any node in the network from outside.
- The other point of disagreement with the results of the focus paper is the fact that the EU is also very highly connected through good land transport networks.

Conceptually the idea of a hidden geometry of pandemic spread by hitchhiking global air traffic networks is not difficult to understand. However, several factors make it difficult to reproduce.
A large portion of this project was made highly difficult to replicate due to the fact that the original flow network was commercially owned and hence was not available for free use. The network used in this project was for Europe only, whereas the original paper used a global air traffic flow network. A good first step to improving this project would be to obtain the global network and then test the existing code on that network. Moreover, simulations of the infection in an isolated network could help avoid the problem of entry of infected individuals from the outside.

Despite that, it is possible to consider future developments for this project. Once a more rigorous model has been built and tested it is possible to investigate possible counter epidemic actions. For example, simulating an epidemic and observing which nodes (airports) are most essential to the spread of the epidemic. This can be done by restricting flow through nodes or by imposing a quarantine method. The quarantine method would work with a certain efficiency of accurately detecting the infection in a passenger, meaning some infected passengers would still be able to pass undetected. A complementary investigation could be conducted to investigate a possible relationship between the infectivity and lethality of a disease and how many nodes need to have quarantine zones to successfully curtail the outbreak.
This study can be potentially used in not only predicting how an outbreak progresses but can also give insight into the most effective way to minimize or completely prevent the spread of the epidemic.
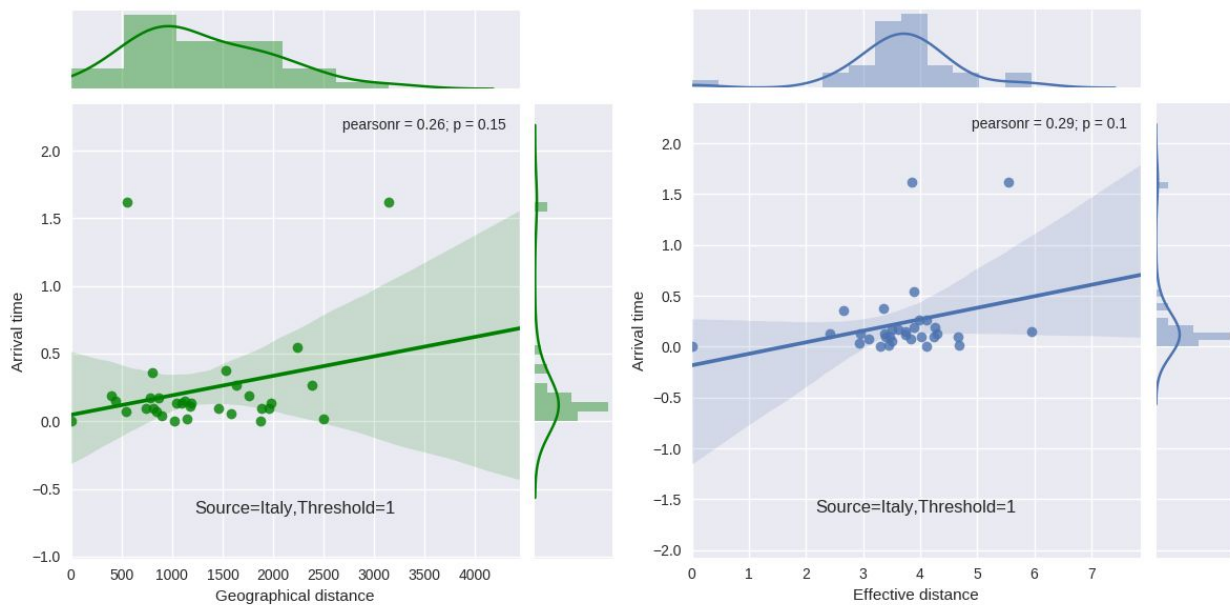
References

1. http://ec.europa.eu/eurostat/statistics-explained/index.php/Air_transport_statistics
2. http://apps.who.int/globalatlas/dataQuery/default.asp
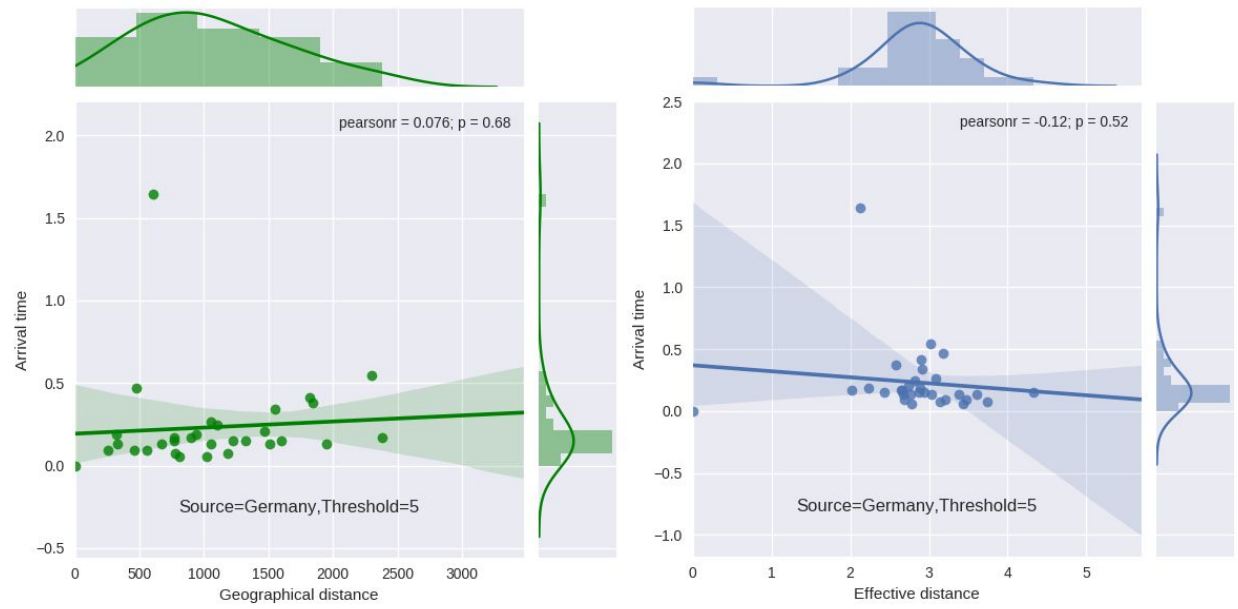3. https://github.com/kdileep1994/networks_project

Appendix
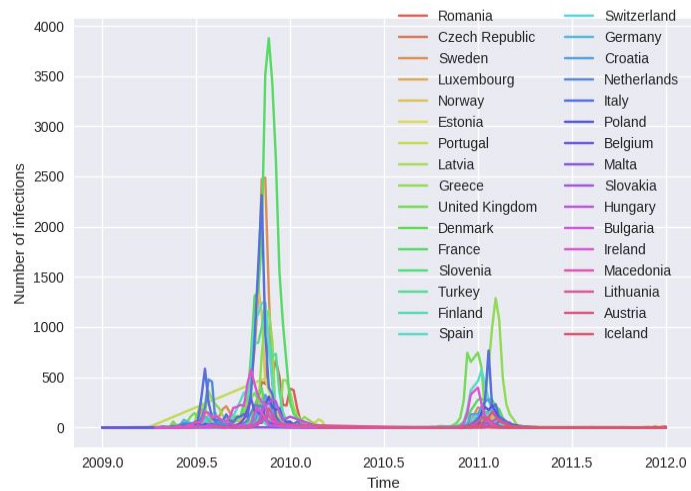
Figure 4. A plot of Number of infections as a function of time in years.

Figure 5. A linear regression plot of arrival times vs. various distance metrics. Source country: Italy

Figure 6. A linear regression plot of arrival times vs. various distance metrics.
Source country: Germany