

Scale-free model for spatio-temporal distribution of outbreaks of avian influenza

Michael Small, David M Walker, and Chi K. Tse

Electronic and Information Engineering
Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
Email: ensmall@polyu.edu.hk

Abstract—Traditional models for geographical spread of infectious agents invoke a diffusion mechanism. By observing the global distribution of avian influenza outbreaks among wild and domestic birds, we show that this model is not appropriate. We find that the outbreaks of avian influenza follow a scale-free distribution that can accurately model connectivity between outbreaks on a scale-free complex network. We explore possible mechanisms that can generate such complex transmission dynamics. Our results indicate that heterogeneity in both the human and animal populations is insufficient to explain our results. Only when we model the potential transmission pathways between flocks with a scale-free network do we obtain a scale-free distribution of connectivity between outbreaks.

1. Introduction

Although complex networks have been observed in a wide variety of physical and social systems [1, 12, 8, 7] and there has been substantial discussion that such structures may underlie transmission of infectious agents within various communities [9, 10, 11], there is currently no direct experimental evidence supporting this hypothesis. Nonetheless, a number of theoretical studies have shown that topological structures typical of complex networks lead to transmission dynamics markedly different from that predicted by standard disease transmission models. In this paper we examine the global spatio-temporal distribution of avian influenza cases in both wild and domestic birds and find that the network of outbreaks, and the links between them forms a scale-free network. We find that the exponent of this distribution is less than 2 and therefore the distribution has neither finite mean or variance. Consequently, in contrast to standard mathematical models of disease transmission [2, 6], the current avian influenza outbreak does not exhibit a positive threshold: the disease will continue to propagate even with a vanishingly small rate of transmission.

In this paper we are interested in the global spatio-temporal transmission of avian influenza. Each node in the network is a particular outbreak of avian influenza and links between nodes are determined based on spatio-temporal proximity. Hence, a node in our network could be a single

observed wild bird or an entire poultry farm. What matters for the model is identifying discrete locations in time and space, and linking them together. Analysis of this network shows that it is scale-free. That is, the number of links from a given node $k \geq 1$ has probability distribution

$$P(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

with $\gamma > 1$. The denominator $\zeta(\gamma)$ is Riemann's zeta function. If $1 < \gamma \leq 2$ this distribution does not have a finite mean. Even if $2 < \gamma \leq 3$ the variance of the number of links is infinite and therefore even with very small (but non-zero) rate of transmission, transmission will still persist [6].

The links between nodes are determined by proximity and represent potential transmission pathways. Two nodes are linked if transmission between them is possible. Ideally, we should treat the actual transmission pathways. But, that information is not available. We therefore assume that transmission can occur only over a local area (in both time and space). We have considered both the un-weighted network and a weighted version. The un-weighted version is equivalent to assigning a non-zero (but vanishing) probability of transmission between any two nodes in the network. For the weighted network we consider only the most plausible pathways. Although there is good reason to suppose that transmission of this virus between bird flocks may follow a scale-free distribution [5], it is currently not obvious that the traditional, and alternative, uniform mixing models are inadequate.

2. Turning the data into a network

The data are a compilation of all reported avian cases of avian influenza between 25 November 2003 and 10 March 2007 consisting of 3346 recorded cases. For each case, the date of the outbreak t_n and the location (longitude λ_n and latitude ϕ_n) are recorded. Individual cases may either be wild birds that are found post-mortem and determined to be infected with a strain of avian influenza or the detection of an avian influenza strain in a domestic flock. Data relating to the magnitude of each incident are also recorded. The data originally come from World Organisation for Animal Health alerts (see <http://www.oie.int/>) and

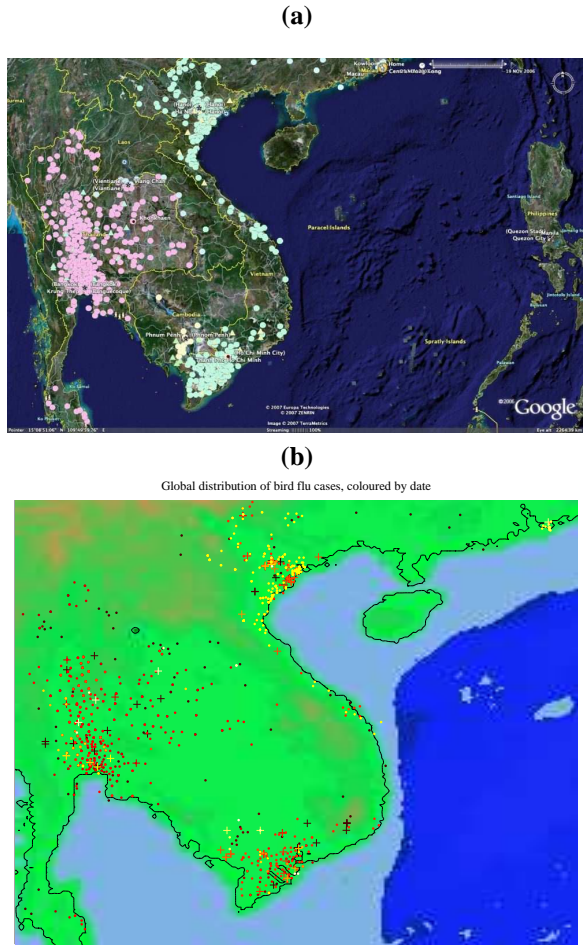


Figure 1: **Avian influenza case data.** (a) Raw data displayed within Google Earth (<http://earth.google.com/>). Circles indicate animal cases, triangles denote human cases. Colour coding is by date. (b) Part of the data used in this study, overlaid against a crude map of the coastline of East Asia. Human cases are marked with crosses, animal cases with solid circles. Colour coding is by date. The three large clusters correspond to the outbreaks in Cambodia and in the north and south of Vietnam (around Hanoi and Ho Chi Minh City), respectively. In panel b, the coast of Hainan island is marked in the north-east of the image and outbreaks in Hong Kong are shown in yellow in the extreme north-east corner.

has been manually entered using ArcGIS and converted to Keyhole Markup Language (KML) using Arc2Earth (<http://www.arc2earth.com/>). The data is available, in a format compatible with Google Earth (KML), from <http://www.declanbutler.info/Flumaps1/avianflu.html>. Fig. 1 depicts one snapshot of this data.

Each incident (t_n, λ_n, ϕ_n) corresponds to a node on the graph of infection links. We construct a directed link from node- i (t_i, λ_i, ϕ_i) to node- j (t_j, λ_j, ϕ_j) if

$$d(i, j) \leq (t_j - t_i)\mu \quad (1)$$

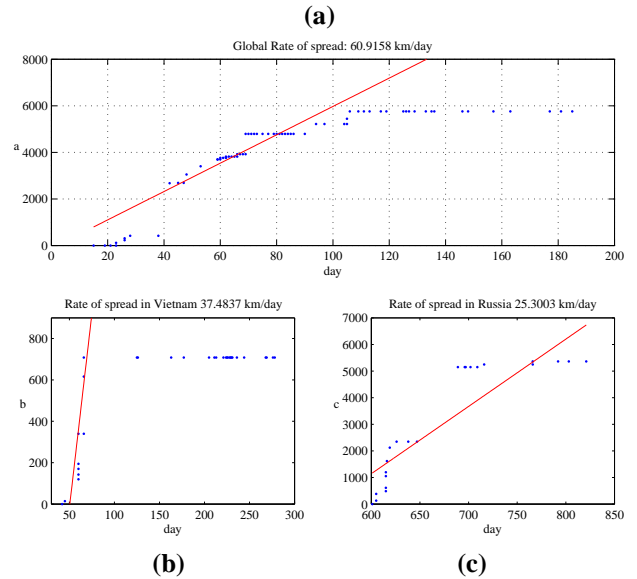


Figure 2: **Rate of transmission of avian influenza.** In each of (a,b,c) we computed the maximum distance between reported avian influenza cases as a function of time. (a) The global rate of transmission is depicted (all countries) along with an average rate of spread of about 61 km/day. In panel (b) and (c) we repeat the same calculation for cases reported in Vietnam and Russia. The average rate of spread within these two geographical regions is about 37 km/day in (b) and 25 km/day in (c). The rate of spread computed in panel (a) reflects the fact that transmission is in 2 dimensions (i.e. the *diameter* grows by about 60 km/day), whereas panel (b) and (c) reflect situations where the disease propagation is in one linear direction only (i.e. the growth of the *radius* is being measured). In the case of Vietnam (b) this is due to the unique geography of the country: two initial outbreaks centred in Hanoi and Ho Chi Minh City spread roughly south and north, respectively. In Russia (c) the reported cases represent the spread of avian influenza from Asia to Europe along the Kazakhstan border.

and

$$0 \leq (t_j - t_i) < T_{\max} \quad (2)$$

where $d(i, j)$ is the great circle distance between node- i and node- j in kilometres and μ is a positive constant (units of km/day) corresponding to the approximate geographical rate of transmission of the virus. Great circle distance is computed from longitude and latitude using standard spherical geometry and a value of $R = 6372.795$ km for the radius of the (assumed to be spherical) earth.

The choice of criterion to determine connectivity is arbitrary but also natural. If we assume that the geographical rate of transmission of the virus is uniform and equal to μ , then node- i is deemed to be connected to node- j if the virus at node- i can travel as far as node- j before the outbreak is observed to occur at node- j and sooner than T_{\max} days. We

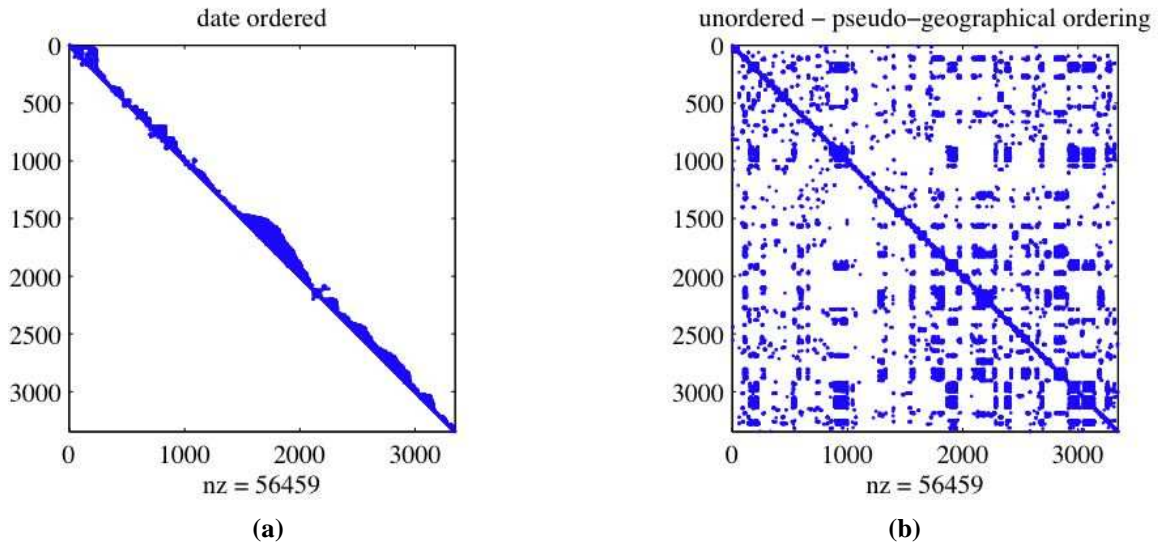


Figure 3: **Network connectivity (adjacency) matrix.** Both panels depict the connections present in the network deduced according to the criteria (1) and (2). If there exists a connection from node i (vertical axis) to node j (horizontal) then the point (j, i) is marked. In panel (a) the nodes are ordered according to time (that is, $i < j$ only if $t_i \leq t_j$). In panel (b) the points are ordered approximately geographically.

have varied both parameters μ and T_{\max} over a wide range of values ($3 < \mu < 50$ km/day and $5 < T_{\max} < 30$ days) and have not found significant qualitative variation in the results. For the sake of brevity and concreteness the rest of this report focuses on the specific values $T_{\max} = 10$ days and $\mu = 25$ km/day. The choice of 25 km/day is motivated by the apparent rate of spread of avian influenza cases in the early stage of the outbreak as depicted in Figure 2. The choice of 10 days is only to provide more easily visible results.

3. Properties of that network

Figure 3 depicts the adjacency matrix for this network. In Fig. 3 (a) the diagonal structure of the matrix indicates connection between temporally adjacent nodes. The clustering in Fig. 3 (b) is due to geographical localisation. The *sample* average number of connections from a given node is (a relatively large) 16.8 and because of our criterion for selecting connectivity, nodes are connected only if they are separated by no more than 10 days. Hence, the fact that the data spans 1203 days indicates that the shortest path between random nodes can be very large: hence, this is not a small-world network. However, the reason for this is entirely artificial. The geographical connectivity may well be small-world, but because we constrain nodes in time, this feature is suppressed. However, the available data make it impossible to resolve this issue. Nonetheless, the resultant network is scale-free. This is evident from Fig. 4. Figure 4 (a) illustrates that this network is composed of discrete clusters. The two main reasons for this disconnectedness is

our initial assumptions concerning connectivity (1) and (2) and the inevitable incompleteness of available data.

In Fig. 4 (b) we depict the link distribution and an estimate of the scale exponent. We observe that by altering T_{\max} or μ we can change γ , but changing these parameters does not affect our main result: the network is scale-free and has infinite mean and variance. Conversely, increasing this average number of connections or choosing a more complicated metric (rather than great circle distance) can increase the connectedness of the final network.

It is worth noting that we observe a fairly low scale exponent $\gamma \approx 1.2$. This is lower than the oft-cited “typical” range of $2 < \gamma \leq 3$, but of the same order of magnitude as experimental results for human travel [3] ($\gamma \approx 1.6$) and similar to the network scaling ($\gamma \approx 1.8$) reported for e-mail collaborative networks [4]. In this study, we have not traced the actual infection pathways. Instead, we take the observed data for outbreaks of avian influenza and construct a network that *contains* part of the underlying transmission paths. We assume that the virus propagates at a constant and relatively modest rate, and related events must be relatively close in time. Certainly, delays in detection and reporting of cases, and long-distances transmission (for example, via migratory birds) would violate these assumptions. Hence the network we construct is inevitably only an approximation. Nonetheless, the basic result remains valid: the spatial-temporal connectedness (defined by (1) and (2)) is scale-free.

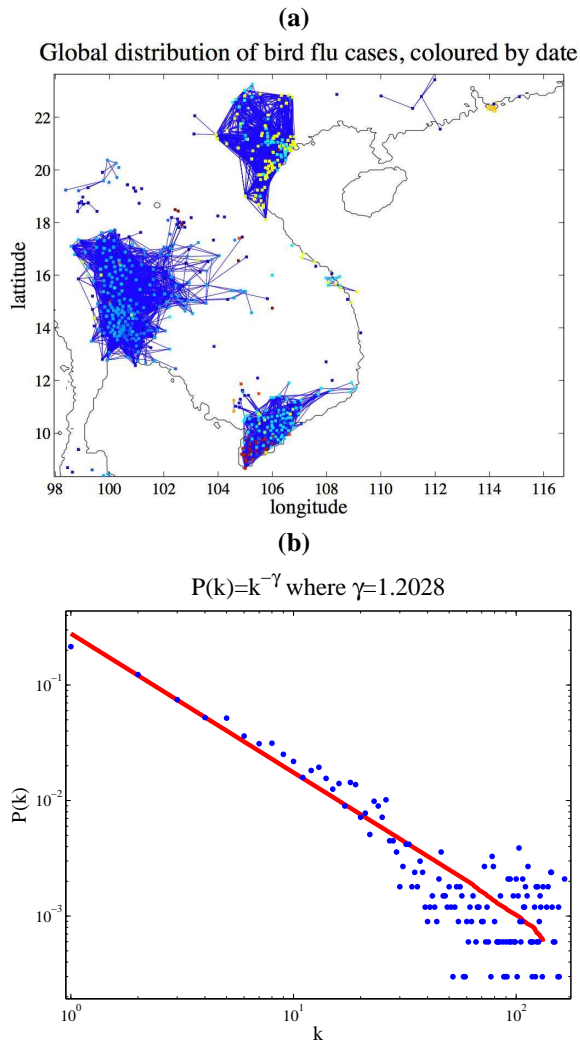


Figure 4: **Network degree distribution.** In the upper panel, the data from Figure 1 is re-drawn with the addition of network connections. Clearly, the entire network is not connected. Nonetheless, from this network we compute the degree distribution (lower panel) and display it on log-log scale. The data exhibits a scale-free distribution with estimated scale exponent of $\gamma \approx 1.2028$. Kolmogorov-Smirnov (KS) goodness-of-fit test indicates a value within the 90% confidence interval given that the underlying data is sampled from a power law distribution.

Acknowledgments

This work was supported by a grant from the Health, Welfare and Food Bureau of the Hong Kong SAR Government. We are indebted to Declan Butler for compiling and then providing the data used in this study.

References

- [1] A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [2] M. Boguna, R. Pastor-Satorra, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Physical Review Letters*, 90:028701, 2003.
- [3] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439:462–465, 2006.
- [4] H. Ebel, L.-I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. *Physical Review E*, 66:035103, 2002.
- [5] W. B. Karesh, R. A. Cook, E. L. Bennett, and J. Newcomb. Wildlife trade and global disease emergence. *Emerging Infectious Diseases*, 11:1000–1002, 2005.
- [6] R. M. May and A. L. Loyd. Infection dynamics on scale-free networks. *Physical Review E*, 64:066112, 2001.
- [7] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303:1538–1542, 2004.
- [8] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.
- [9] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86:3200–3203, 2001.
- [10] A. Schneeberger, C. H. Mercer, S. A. J. Gregson, N. M. Ferguson, C. A. Nyamukapa, R. M. Anderson, A. M. Johnson, and G. P. Garnett. Scale-free networks and sexually transmitted diseases — A description of observed patterns of sexual contacts in Britain and Zimbabwe. *Sexually Transmitted Diseases*, 31:380–387, 2004.
- [11] M. Small, C. K. Tse, and D. M. Walker. Super-spreaders and the rate of transmission of the SARS virus. *Physica D*, 215:146–158, 2006.
- [12] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442, 1998.