Scale-Free Distribution of Avian Influenza Outbreaks

Michael Small,* David M. Walker, and Chi Kong Tse

Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

(Received 6 March 2007; published 30 October 2007)

Using global case data for the period from 25 November 2003 to 10 March 2007, we construct a network of plausible transmission pathways for the spread of avian influenza among domestic and wild birds. The network structure we obtain is complex and exhibits scale-free (although not necessarily small-world) properties. Communities within this network are connected with a distribution of links with infinite variance. Hence, the disease transmission model does not exhibit a threshold and so the infection will continue to propagate even with very low transmissibility. Consequentially, eradication with methods applicable to locally homogeneous populations is not possible. Any control measure needs to focus explicitly on the hubs within this network structure.

DOI: 10.1103/PhysRevLett.99.188702 PACS numbers: 89.75.Hc, 05.45.Tp, 87.19.Xx, 89.75.Da

Complex networks have been observed in a wide variety of physical and social systems [1]. In particular, it has been proposed that such structures may underlie transmission of infectious agents within various communities [2]. Despite a lack of direct experimental evidence supporting this hypothesis, a number of theoretical studies have shown that topological structures typical of complex networks (in particular, scale-free and small-world topologies) lead to transmission dynamics markedly different from that predicted by standard disease transmission models. We examine the global spatiotemporal distribution of avian influenza cases in both wild and domestic birds and find that the network of outbreaks and the links between them form a scale-free network. That is, the connectivity k of outbreaks exhibits a distribution $P(k) = k^{-\gamma}/\zeta(\gamma)$. We find that the exponent of this distribution $\gamma < 2$ and therefore the distribution has neither finite mean or variance. Consequently, in contrast to standard mathematical models of disease transmission [3,4], 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. Hence, one can only hope to eradicate (or even control) this disease by specifically focussing on transmission at the hubs of this network structure [5,6].

Standard mathematical models of geographical transmission of an infectious agent assume that the terrain is locally homogeneous and that the pathogen will diffuse uniformly [7]. A natural consequence of this formulation is that if the transmissibility of the pathogen is lower than some threshold, the disease will terminate. Recent studies of infectious agents (usually either biological pathogens or computer viruses) in certain complex networks have shown that in these networks such a threshold does not exist. In particular, if the connectivity within a network follows a scale-free distribution and the transmissibility of the agent is positive, then an epidemic is inevitable [3,4]. For the case when recovery from the infected state confers immunity, an epidemic is inevitable only if the population is infinite [3] or if the system is not closed—which, since the life cycle of domestic poultry is relatively short, is the situation for avian influenza. Moreover, one can expect that the infection will persist indefinitely. Whereas, if a disease is spreading according to standard mathematical models with uniform mixing, then it can be eradicated by lowering the rate at which it is transmitted below a predetermined, but nonzero, threshold. However, if a disease is spreading on a scale-free network, then eradication of that disease is only possible if transmission is reduced to precisely zero.

In this Letter we consider the network consisting of individual communities (the network nodes) explicitly connected by possible infection pathways (the links). We then focus only on the communities which have become infected, but study all possible links between them. Our analysis of data from the geographical and temporal distribution of avian influenza outbreaks exhibits a network that is scale-free. That is, the number of links, $k \geq 1$, has probability distribution

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

with $\gamma > 1$. The denominator $\zeta(\gamma)$ is Riemann's zeta function and provides the appropriate normalization constant. Note that 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 nonzero) rate of transmission, transmission will still persist [3].

When considering disease transmission, we treat the nodes on the network as susceptible individuals or communities, and the links between them are 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 consider both the unweighted network (where all transmission pathways are considered as "possible") and a weighted version

(where we consider the plausible pathways and weight according to the number of incoming connections to a node). To address the limitations of our data we also test out method against a short time span late in the outbreak (to eliminate inconsistency between earlier reports of outbreaks) and large outbreaks (to minimize the effect of aggregation of reports). In all cases the results are both qualitatively and quantitatively similar.

To the best of our knowledge, there are no other data that confirm that any disease can be transmitted in this way. However, there is significant evidence that the communities which support various infections do exhibit scale-free structures. Experimental evidence shows that human travel (and therefore human contacts) exhibits scale-free structure [8] and that networks of sexual contacts (number of sexual partners) is also scale-free [9]. Computer simulations have shown that simulation in this manner is viable and suggested potential containment strategies [6]. In this Letter, we focus on transmission of the avian influenza virus in wild and domestic bird populations. There is good reason to suppose that transmission of this virus between bird flocks may follow a scale-free distribution [10]. Nonetheless, it is currently not obvious that the traditional, and alternative, uniform mixing models are inadequate.

The data we use in this study are a compilation of all reported avian cases of avian influenza between 25 November 2003 and 10 March 2007. The data consist of 3346 recorded cases. For each case, the date of the outbreak and the location (longitude and latitude) are recorded. Individual cases may either be wild birds that are found (possibly *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 (most probably then followed by culling of that flock). Data relating to the magnitude of each incident are also recorded. Human cases of avian influenza have also been recorded in the same data set, but for this study, these are ignored. The entire data set is compiled from a variety of sources [11]; Fig. 1 depicts one snapshot.

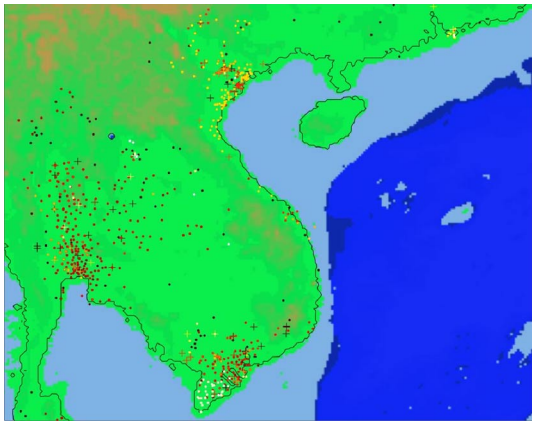

FIG. 1 (color online). *Avian influenza case data.*—Part of the data used in this study, overlaid against a crude map of the coastline of East Asia. Animal cases are marked with solid dots, human cases with crosses. Color coding is by date. The three large clusters correspond to the outbreaks in Cambodia and in the north and south of Vietnam (around Hanoi and Ho Chi Minh City), respectively. Hainan Island is marked in the northeast portion of the image and outbreaks in Hong Kong are shown in yellow in the far northeast corner.

The data consist of 3346 triples of the form (t_n, λ_n, ϕ_n) where t_n is the time (in days) since 25 November 2003 of the nth incident. The parameters λ_n and ϕ_n are the latitude and longitude of that case. 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

$$d(i, j) \leq (t_j - t_i)\mu \tag{1}$$

and

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

where $d(i, j)$ is the great circle distance between node i and node j in kilometers 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

$$d(i, j) = R \arctan\left\{\frac{\sqrt{[\cos\phi_j \sin\Delta\lambda]^2 + [\cos\phi_i \sin\phi_j - \sin\phi_i \cos\phi_j \cos\Delta\lambda]^2}}{\sin\phi_i \sin\phi_j + \cos\phi_i \cos\phi_j \cos\Delta\lambda}\right\}, \tag{3}$$

where $\Delta\lambda = \lambda_j - \lambda_i$ and the radius $R = 6372.795$ km.

The choice of the criterion (1) and (2) to determine connectivity is, of course, arbitrary. But, it is 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 (that is, within $t_j - t_i$ days) and sooner than $T_{\max}$ days. We 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. As one would expect, variation of these two parameters simply affects the density [12] within the network, without altering the basic network topology. Nonetheless, for the sake of brevity and concreteness, we choose to adopt specific values for this Letter. We take $T_{\max} = 10$ days and $\mu = 25$ km/day. The choice of 25 km/day is motivated by the apparent rate of spread of

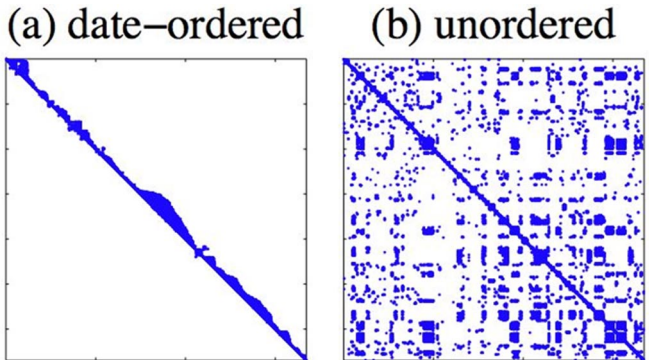

FIG. 2 (color online). *Network connectivity (adjacency) matrix.*—Both panels depict the connections present in the network deduced according to the criteria (1) and (2). If there exists a connection from node i (vertical axis) to node j (horizontal) then the point (j, i) is marked. In panel (a), the 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.

avian influenza cases in the early stage of the outbreak [13]. The choice of 10 days is only to provide more easily visible results. Larger values only make the network denser; smaller values make it more fragmented.

With these values of $T_{\max}$ and μ we construct a complex network of connectivity of avian influenza cases. The connectivity between individual nodes in that network is shown in Fig. 2. In Fig. 2(a) the diagonal structure of the matrix indicates connection between temporally adjacent nodes. The clustering in Fig. 2(b) is due to geographical localization. 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 span 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 may be entirely artificial. The geographical connectivity may well be small-world (as illustrated for humans in [8] and various livestock populations in [14]), 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. 3. Figure 3(a) illustrates that this network is composed of discrete clusters. The two main reasons for this disconnectedness are our initial assumptions concerning connectivity (1) and (2) and the inevitable incompleteness of available data.

In Fig. 3(b) we depict the link distribution and an estimate of the scale exponent. Following [15] we estimate the exponent γ using a maximum likelihood estimator which avoids statistical bias associated with a linear fit to the log-log plot [16]. 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

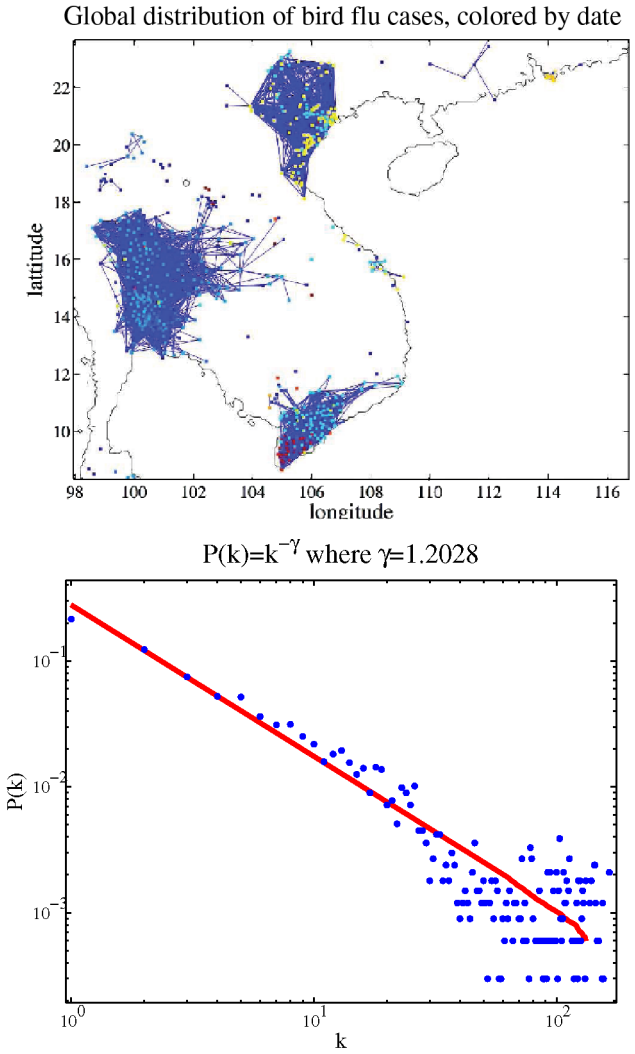

FIG. 3 (color online). *Network degree distribution.*—In the upper panel, the data from Fig. 1 are redrawn with the addition of network connections. Clearly, the entire network is not connected. Nonetheless, from this network we compute the degree distribution (lower panel) and display it on log-log scale. The data exhibit 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 are sampled from a power law distribution.

variance. Conversely, increasing this average number of connections or choosing a more complicated metric [rather than Eq. (3)] can increase the connectedness of the final network.

To examine the robustness of our result we also examined the networks obtained only from the most recent data (18 months, since October 2005) and from large outbreaks (more than 50 deaths). These tests resulted in a restriction of our data set to 1942 and 899 nodes, respectively. In both cases similar scale exponents (~1.2) and KS-test confidence levels were obtained. Finally, we also considered

the construction of a weighted network. Each node was weighted by the reciprocal number of incoming links, and the degree of each node was set to be the sum of these weights. The degree distribution we obtain in this way was also approximately scale-free ($\gamma \approx 2.51316$), but because of the high degree of clustering in the network the relationship only extended over one decade [17].

It is worth noting that the distance between outbreaks, with $d(i, j)$, $d(i, j) + \mu(t_i - t_j)$, or $|d(i, j) - \mu(t_i - t_j)|$ as a norm is not scale-free: it is multimodal and decays approximately exponentially. Hence, the scale-free features we observe are not due to the spatial (or temporal) distribution of outbreaks, but rather depend on the large variability in infectious pathways in the complex network topology. We also 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 [8] ($\gamma \approx 1.6$) and similar to the network scaling ($\gamma \approx 1.8$) reported for email collaborative networks [18].

In this study, we do not trace 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, and the inclusion of these additional factors could result in a more connected and more realistic network structure. Nonetheless, to do this would introduce many more parameters and cloud the basic result: the spatial-temporal connectedness [defined by (1) and (2)] is scale-free.

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 (Nature) for compiling and then providing the data used in this study.

*ensmall@polyu.edu.hk

[1] A. Barabasi and R. Albert, Science **286**, 509 (1999); R. Milo *et al.*, Science **303**, 1538 (2004); **298**, 824 (2002); D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[2] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001); A. Schneeberger *et al.*, Sex. Transm. Dis. **31**, 380 (2004); M. Small, D. M. Walker, and C. K. Tse, Physica (Amsterdam) **215D**, 146 (2006).

[3] R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).

[4] M. Boguna, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).

[5] Z. Dezső and A.-L. Barabási, Phys. Rev. E **65**, 055103(R) (2002).

[6] S. Eubank, H. Guclu, V. V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczakai, and N. Wang, Nature (London) **429**, 180 (2004).

[7] J. D. Murray, *Mathematical Biology* (Springer, New York, 1993), 2nd ed.

[8] D. Brockmann, L. Hufnagel, and T. Geisel, Nature (London) **439**, 462 (2006).

[9] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, Nature (London) **411**, 907 (2001).

[10] W. B. Karesh, R. A. Cook, E. L. Bennett, and J. Newcomb, Emerging Infectious Diseases **11**, 1000 (2005).

[11] The data originally come from World Organization for Animal Health alerts (see http://www.oie.int/) and World Health Organization case reports, and they are all manually entered using ArcGIS and converted to Keyhole Markup Language (KML) using Arc2Earth (http://www.arc2earth.com/). The data are available, in a format compatible with Google Earth (KML), from Astrophysical Journal Supplement.

[12] That is, the average number of connections; with uniform sampling of μ and $T_{\max}$ in the above range we find that γ varies according to a Gamma distribution with a mean of 1.35, standard deviation of 0.13, and support $[1, \infty)$.

[13] We computed the maximum distance between reported avian influenza cases as a function of time in various geographically limited locations (specific countries). The average rate of spread within Vietnam and Russia is about 37 km/day and 25 km/day, respectively. In each case, the disease propagation is in one linear direction only (i.e., the growth of the *radius* is being measured). In the case of Vietnam this is due to the unique geography of the country: two initial outbreaks centered in Hanoi and Ho Chi Minh City spread roughly south and north, respectively. In Russia the reported cases represent the spread of avian influenza from Asia to Europe along the Kazakhstan border.

[14] R. M. Christley and N. P. French, Equine Veterinary Journal **35**, 586 (2003); I. Z. Kiss, D. M. Green, and R. R. Kao, J. R. Soc. Interface **3**, 669 (2006).

[15] M. L. Goldstein, S. A. Morris, and G. G. Yen, Eur. Phys. J. B **41**, 255 (2004).

[16] In [15], the authors show that the maximum likelihood best value of γ is the solution to

$$\frac{\zeta'(\gamma)}{\zeta(\gamma)} = -\frac{1}{N}\sum_{i=1}^{N}\log(x_i),$$

where x_i is the number of links associated with the ith node.

[17] We have also used these data to construct a fully connected graph, first by eliminating singleton nodes, and then by connecting discrete clusters (closest first). The resulting graph is still scale-free ($\gamma \approx +1.32$) and exhibits strong clustering and high assortativity.

[18] H. Ebel, L.-I. Mielsch, and S. Bornholdt, Phys. Rev. E **66**, 035103(R) (2002).