Heterogeneity and predictability of global epidemics

Vittoria Colizza[1]*, Alain Barrat[2], Marc Barthélemy[1], and Alessandro Vespignani[1]

August 30, 2005

[1] School of Informatics and Biocomplexity Center, Indiana University, Bloomington, 47406, IN, USA

[2] LPT and UMR du CNRS 8627, Bâtiment 210, Université de Paris-Sud, 91405 ORSAY Cedex - France

Abstract

We investigate the role of the large scale properties of the airline transportation network in determining the global diffusion pattern of emerging disease. We present a stochastic computational framework for the forecast of global epidemics that considers the complete International Air Transport Association 2002 database complemented with census population data. We adopt an information theory approach to analyze quantitatively the level of heterogeneity and predictability of the epidemic pattern and its relation with the network's structure. The level of spatio-temporal heterogeneity of the spreading pattern is globally characterized and found to be a direct consequence of the network statistical complexity. The epidemic pattern predictability is quantitatively determined and traced back to the occurrence of epidemic pathways defining a backbone of dominant connections in the disease spreading. The presented results provide a general framework for the analysis of containment policies and epidemic risk forecast.

Keywords: Complex networks, Epidemiology, Noise and fluctuations.

The mathematical modeling of epidemics has often dealt with the problem of an appropriate description of real populations with complicated age, social and spatial structures and with heterogeneous

*School of Informatics, Indiana University, 901 E. 10th St., Bloomington, IN 47408-3912, USA, tel.+1-812-856-2226, fax +1-812-856-4764, vcolizza@indiana.edu

patterns in the contact network [1, 2, 3, 4, 5, 6]. Recently, the availability of unprecedented computer power has led to numerical approaches relying on agent-based modeling that simulate entire populations and their dynamics at the scale of the single individual and on a minute-by-minute basis [7, 8]. On the other hand, the inherent complex features and emerging properties [9, 10, 11] of the network in which epidemics occur are not the mere juxtaposition of complicated elements used for increased realism in sophisticated epidemic modeling [12]. Indeed, networks' complex properties often imply statistical fluctuations extending over several orders of magnitude and the breakdown of standard homogeneous approaches and results [5, 6].

These considerations are particularly relevant in the study of the geographical spread of epidemics where the various long-range heterogeneous connections typical of modern transportation networks naturally give rise to a very complicated evolution of epidemics characterized by heterogeneous and seeemingly erratic outbreaks [13, 14]. In this context, air-transportation represents a major channel of epidemic diffusion as recently documented for the SARS outbreak [15]. The modeling of global epidemic diffusion via the air transportation network dates back to the seminal paper of Rvachev and Longini [16] capitalizing on previous studies on the russian network [17]. Similar modeling approaches, even if limited by a very partial knowledge of the world-wide transportation network, have been used to study specific outbreaks such as pandemic influenza [18, 19, 20], HIV [21], and SARS [22]. The availability of the complete world-wide airport network dataset (*WAN*) and the recent extensive studies of its topology [23, 24] are finally allowing a full scale computational study of global epidemics. Here, we use the International Air Transport Association (IATA) database [25] containing the world list of airport pairs connected by direct flights and the number of available seats on any given connection for the year 2002. The resulting air-transportation network is therefore a weighted graph, comprising $V = 3,880$ vertices denoting airports and $E = 18,810$ edges whose weight $w_{j\ell}$ represents the passenger flow between airports j and ℓ. This dataset has been complemented by the population N_j of the metropolitan area served by the airport j as obtained by different sources. The final network dataset contains the $3,100$ largest airports, $17,182$ edges (accounting for 99% of the worldwide traffic) and the respective urban population data. The obtained network is highly heterogeneous both in the connectivity pattern and the traffic capacities. In particular the presence of broad statistical distributions and non-linear associations among the various quantities, contrary to linear relations used so far, indicate a possible major impact in the ensuing disease spreading pattern.

In this work we will consider for the first time a global stochastic epidemic model including the full International Air Transport Association (IATA) [25] database, aiming at a detailed study of the interplay among the network structure, the stochastic features and the infection dynamics in defining the global spreading pattern of epidemics [26]. While previous studies have in general focused on the *a-posteriori* analysis of real case studies of global epidemics, the large scale modeling allowed by the IATA database enables us to address general theoretical issues such as (i) the spatio-temporal statistical properties of the epidemic pattern, (ii) their relation with the complex features of the underlying transportation network and (iii) the reliability of forecasts and outbreak scenarios with respect to the intrinsic stochasticity of disease transmission and traffic flows. The model is analyzed by using an information theory approach that allows the quantitative characterization of the heterogeneity level of the spreading pattern and its predictability in presence of stochastic fluctuations. Results provide a general computational framework for the analysis of containment policies and risk forecast of global epidemic outbreaks. Simulations and reproductions of case studies of real epidemics are also presented.

References

[1] Anderson RM, May RM, *Infectious diseases in humans* (Oxford University Press, Oxford 1992).

[2] Hethcote HW, Yorke JA, *Lect. Notes Biomath.* (Berlin, Springer-Verlag, 1984), vol. 56.

[3] Kretzschmar M, Morris M, *Math. Biosci.*, **133**, 165(1996).

[4] Keeling M, *Proc. R. Soc. Lond. B* **266**, 859(1999).

[5] Pastor-Satorras R, Vespignani A, *Phys. Rev. Lett.* **86**, 3200(2001).

[6] Lloyd AL, May RM, *Science* **292**, 1316(2001).

[7] Chowell G, Hyman JM, Eubank S, Castillo-Chavez C, *Phys. Rev. E* **68**, 066102(2003).

[8] Eubank S, Guclu H, Anil Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, Wang N, *Nature* **429**, 180(2004).

[9] Albert R, Barabási A-L, *Rev. Mod. Phys.* **74**, 47(2000).

[10] Dorogovtsev SN, Mendes JFF, *Evolution of networks: From biological nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).

[11] Pastor-Satorras R, Vespignani A, *Evolution and structure of the Internet: A statistical physics approach* (Cambridge University Press, Cambridge, 2003).

[12] Ferguson NM *et al*, *Nature* **425**, 681(2003).

[13] Cohen ML, *Nature* **406**, 762(2000).

[14] Cliff A, Haggett P, *British Medical Bulletin*, **69**. 87(2004).

[15] http://www.who.int/csr/sars/en

[16] Rvachev LA, Longini IM, *Mathematical Biosciences* **75**, 3(1985).

[17] Baroyan OV, Genchikov LA, Rvachev LA, Shashkov VA, *Bull. Internat. Epidemiol. Assoc.* **18**, 22(1969).

[18] Longini IM, *Mathematical Biosciences* **90**, 367(1988).

[19] Grais RF, Hugh Ellis J, Glass GE, *European Journal of Epidemiology* **18**, 1065(2003).

[20] Grais RF, Hugh Ellis J, Kress A, Glass GE, *Health Care Management Science*, **7**, 127(2004).

[21] Flahault A, Valleron A-J, *Math. Pop. Studies* **3**, 1(1991).

[22] Hufnagel L, Brockmann D, Geisel T, *Proc. Natl. Acad. Sci. USA* **101**, 15124(2004).

[23] Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A, *Proc. Natl. Acad. Sci. USA* **101**, 3747(2004).

[24] Guimerà R, Amaral LAN, *Eur. Phys. J. B* **38**, 381(2004).

[25] http://www.iata.org

[26] Colizza V, Barrat A, Barthélemy M, Vespignani A, *q-bio.OT/0507029* (2005).