

other nodes to fail as well, which may eventually lead to failure cascades and the breakdown of the system, denoted as systemic risk. This applies in particular to financial networks where links represent standing debts and claims between connected financial institutions. However, it is not well understood how the structure of a financial network affects the probability of a systemic failure. Although a topical subject, most theoretical and empirical methods are not suited to predicting cascading network effects. The mainstream view assumes that a denser network allows for a better diversification of the individual failure risk (26). However, systemic risk has been shown to increase, depending on the coupling strength between nodes (27). Furthermore, most stable dynamic network models account for only the addition or removal of a single agent to or from the network at each instance of time. However, the addition or removal of whole groups of agents to or from the network (e.g., as part of a systemic failure) may result in a larger, less predictable, and less stable system.

In summary, we anticipate a challenging research agenda in economic networks, built upon a methodology that strives to capture the rich process resulting from the interplay between agents' behavior and the dynamic interactions among them. To be effective, however, empirical studies providing insights into economic networks from massive data analysis, theory encompassing the

appropriate description of economic agents and their interactions, and a systemic perspective bestowing a new understanding of global effects as coming from varying network interactions are needed. We predict that such studies will create a more unified field of economic networks that advances our understanding and leads to further insight. We are still far from a satisfactory identification and integration of the many components, but the recent advances outlined suggest a promising start.

#### References and Notes

1. F. Vega-Redondo, *Complex Social Networks* (Econometric Society Monographs, Cambridge Univ. Press, Cambridge, 2007).
2. A. Barrat, M. Barthelemy, A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge Univ. Press, Cambridge, 2008).
3. M. O. Jackson, A. Wolinsky, *J. Econ. Theory* **71**, 44 (1996).
4. R. Albert, A.-L. Barabasi, *Rev. Mod. Phys.* **74**, 47 (2002).
5. J. Hagedoorn, *Res. Policy* **31**, 477 (2002).
6. M. Granovetter, *Getting a Job: A Study of Contacts and Careers* (Univ. of Chicago Press, Chicago, 1995).
7. V. Bala, S. Goyal, *Econometrica* **68**, 1181 (2000).
8. M. D. König, S. Battiston, M. Napoletano, F. Schweitzer, *Netw. Heterog. Media* **3**, 201 (2008).
9. M. Marsili, F. Vega-Redondo, F. Slanina, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1439 (2004).
10. S. P. Borgatti, A. Mehra, D. J. Brass, G. Labianca, *Science* **323**, 892 (2009).
11. R. M. May, S. A. Levin, G. Sugihara, *Nature* **451**, 893 (2008).
12. G. Iori, G. De Masi, O. Precup, G. Gabbi, G. Caldarelli, *J. Econ. Dyn. Control* **32**, 259 (2008).
13. M. Boss, H. Elsinger, M. Summer, S. Thurner, *Quant. Finance* **4**, 677 (2004).
14. G. Fagiolo, S. Schiavo, J. Reyes, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **79**, 036115 (2009).

15. D. Garlaschelli, M. I. Loffredo, *Phys. Rev. Lett.* **93**, 188701 (2004).
16. S. Battiston, J. F. Rodrigues, H. Zeytinoglu, *Adv. Complex Syst.* **10**, 29 (2007).
17. J. Reyes, S. Schiavo, G. Fagiolo, *Adv. Complex Syst.* **11**, 685 (2008).
18. M. Kosfeld, *Rev. Netw. Econ.* **3**, 20 (2004).
19. S. Callander, C. Plott, *J. Public Econ.* **89**, 1469 (2005).
20. W. Powell, D. White, K. Koput, J. Owen-Smith, *Am. J. Sociol.* **110**, 1132 (2005).
21. B. Kogut, G. Walker, *Am. Sociol. Rev.* **66**, 317 (2001).
22. D. Sornette, F. Deschates, T. Gilbert, Y. Ageon, *Phys. Rev. Lett.* **93**, 228701 (2004).
23. T. A. Snijders, G. G. van de Bunt, C. E. Steglich, *Soc. Networks*, in press; published online 26 March 2009 (10.1016/j.socnet.2009.02.004).
24. J. Reichardt, D. White, *Eur. Phys. J. B* **60**, 217 (2007).
25. A. Kirman, *J. Evol. Econ.* **7**, 339 (1997).
26. F. Allen, D. Gale, *J. Polit. Econ.* **108**, 1 (2000).
27. S. Battiston, D. Delli Gatti, M. Gallegati, B. Greenwald, J. Stiglitz, *J. Econ. Dyn. Control* **31**, 2061 (2007).
28. We would like to thank M. König, C. J. Tessone, S. Battiston, and S. Vitali (ETH Zurich) for aid with figures and S. White for commentary on the text. Data for Fig. 2 were provided by Orbis Database (end of 2007), Bureau Van Dijk. F.S. and D.S. acknowledge financial support from the ETH Competence Center, "Coping with Crises in Complex Socio-Economic Systems" (CCSS), through ETH Research Grant CH1-01-08-2. F.V.-R. gratefully acknowledges financial support from the Spanish Ministry of Education under grant SEJ2007-62656. A.V. acknowledges funding from NIH, DTRA, the EC-FET program and the Lilly Foundation. D.W.'s work is supported by external faculty funding at the Santa Fe Institute and anonymous nonprofit contributions to the University of California at Irvine faculty group in Social Dynamics and Complexity.

10.1126/science.1173644

#### PERSPECTIVE

## Predicting the Behavior of Techno-Social Systems

Alessandro Vespignani

We live in an increasingly interconnected world of techno-social systems, in which infrastructures composed of different technological layers are interoperating within the social component that drives their use and development. Examples are provided by the Internet, the World Wide Web, WiFi communication technologies, and transportation and mobility infrastructures. The multiscale nature and complexity of these networks are crucial features in understanding and managing the networks. The accessibility of new data and the advances in the theory and modeling of complex networks are providing an integrated framework that brings us closer to achieving true predictive power of the behavior of techno-social systems.

Modern techno-social systems consist of large-scale physical infrastructures (such as transportation systems and power distribution grids) embedded in a dense web of communication and computing infrastructures whose dynamics and evolution are defined and

driven by human behavior. To predict the behavior of such systems, it is necessary to start with the mathematical description of patterns found in real-world data. These descriptions form the basis of models that can be used to anticipate trends, evaluate risks, and eventually manage future events. If fed with the right data, computational modeling approaches can provide the requested level of predictability in very complex settings. The most successful example is weather forecasting, in which sophisticated supercomputer infrastructures are used to integrate current data and

huge libraries of historical meteorological patterns into large-scale computational simulations. Although we often complain about the accuracy of daily weather forecasts, we must remember that numerical weather models and predictions allow us to project the path and intensity of hurricanes, storms, and other severe meteorological occurrences and, in many cases, to save thousand of lives by anticipating and preparing for these events.

Given the success that has been achieved in weather forecasting for decades, why haven't we achieved the same success in the quantitative prediction of the next pandemic spatio-temporal pattern or the effects over the next decade of connecting billions of people from China and India on Internet growth and stability? The basic difference is that forecasting phenomena in techno-social systems starts with our limited knowledge of society and human behavior rather than with the physical laws governing fluid and gas masses. In other words, though it is possible to produce satellite images of atmospheric turbulence, we do not yet have large-scale worldwide, quantitative knowledge of human mobility, the progression of risk perception in a population, or the tendency to adopt certain social behaviors. In recent years, however, tremendous progress has been made in data gathering, the development of new informatics tools, and increases in computational power. A huge flow of quantitative data that combine the demographic and behavioral

Center for Complex Networks and Systems Research, School of Informatics and Computing, and Pervasive Technology Institute, Indiana University, Bloomington, IN 47408, USA; and Institute for Scientific Interchange, Turin, Italy. E-mail: alexv@indiana.edu

# Pushing Networks to the Limit

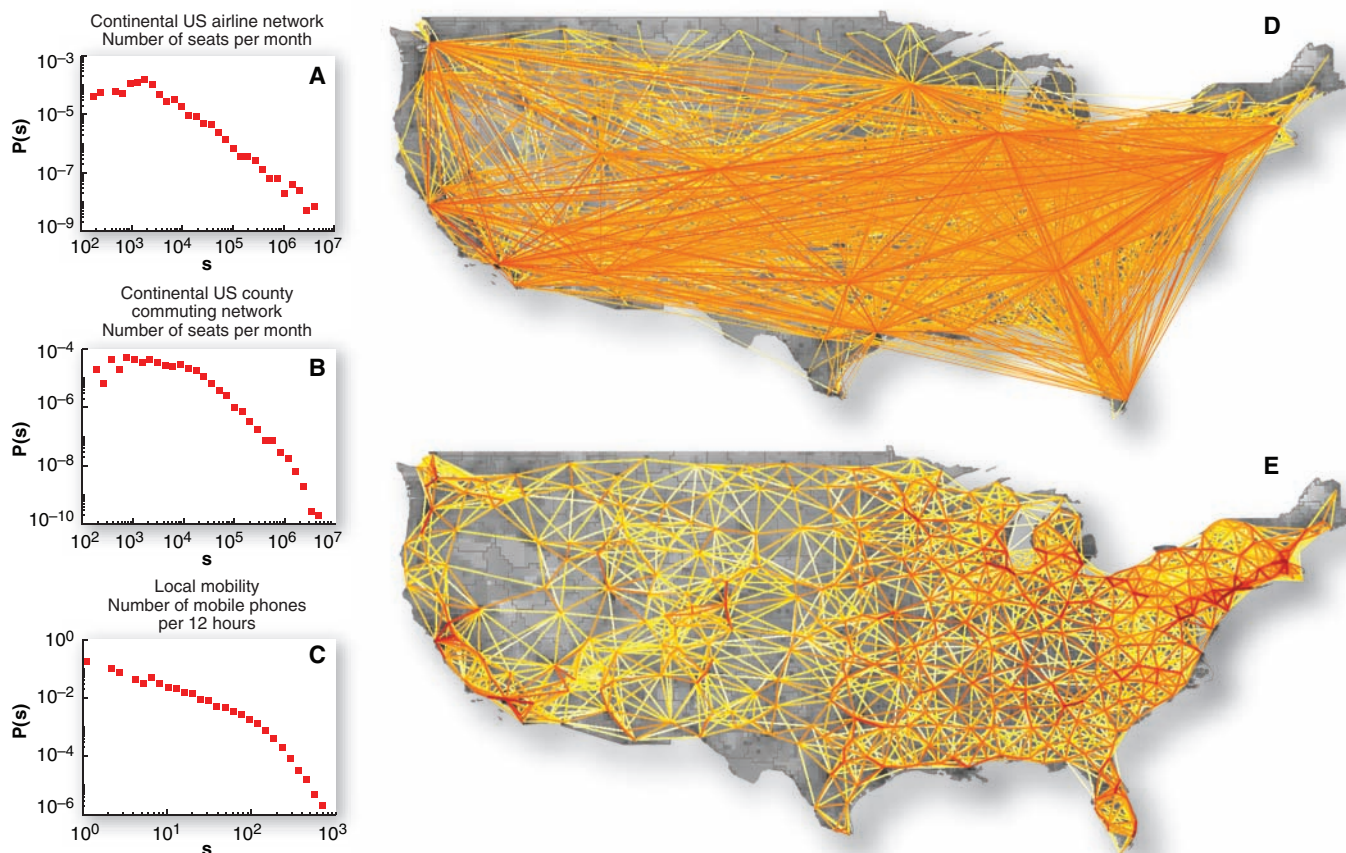
aspects of society with the infrastructural substrate is becoming available (1–6). Analogously to what happened in physics, we are finally in the position to move from the analysis of the “social atom” or “social molecules” (i.e., small social groups) to the quantitative analysis of social aggregate states, as envisioned by social scientists at the beginning of the past century (7). Here, I refer to “social aggregate states” as large-scale social systems consisting of millions of individuals that can be characterized in space (geographic and social) and time. The shift from the study of a small number of elements to the study of the behavior of large-scale aggregates is equivalent to the shift from atomic and molecular physics to the physics of matter. The understanding of how the same elements assembled in large number can give rise, according to the various forces and elements at play, to different macroscopic and dynamical behaviors opens the path to quantitative computational approaches and forecasting power. Yet at the same time, the study of social aggregate states present us with all the challenges already faced in

the physics of matter, from turbulence to multiscale behavior.

## Reality Mining and Proxy Networks

The level of information flow regarding technosocial systems is not just due to advances in number crunching power of modern computer processors. Insights into the nature of the interlinks between people and technology and the dissolution of boundaries between the cyberworld and our real-world social activities are changing our accessibility to data, leading to “reality-mining,” which has been defined as the collection of machine-sensed environmental data that are related to human social behavior (2). A prime example of the people/technology interlinkage can be found in the analysis of human mobility. In the past, approaches to human interactions and mobility have mostly relied on census and survey data, which were often incomplete and/or limited to a specific context. Despite advances in the study of human transport (8, 9), this lack of data has hindered the construction of a general framework

of human mobility based on dynamical principles at the individual level with the ability to bridge spatial scales, from small communities to large urban areas and countries, in a bottom-up perspective. However, in pioneering work, Brockmann *et al.* (4) showed that popular Web sites for currency tracking (such as <http://en.eurobilltracker.com> and [www.wheresgeorge.com](http://www.wheresgeorge.com)) collect a massive number of records on money dispersal that can be used as a proxy for human mobility. This work opened the path to the general exploitation of proxy data for human interaction and mobility (10). Analogously, modern mobile phones and personal digital assistants combine sophisticated technologies such as Bluetooth, Global Positioning System, and WiFi, constantly producing detailed traces on our daily activities (2, 11). For instance, in a recent study, Gonzalez *et al.* (6) used mobile phone data to track the movements of 100,000 people over a 6-month time span. Furthermore, it is now possible to use sensors and tags that produce data at the microscale of one-to-one interactions (1, 2).



**Fig. 1.** Multiscale properties of mobility networks. On the left, we report the probability distribution  $P(s)$  for the traffic, measured as the number of traveling individuals, on any given connection, of three different networks: (A) the continental U.S. airline network, (B) the continental U.S. county commuting network, and (C) the mobility among telephone tower cells in a major urban area. In all cases, the distributions are highly skewed and span from three to seven orders of magnitude.

On the right, we show the illustration of the continental U.S. airline network (D) and the commuting network (E) among major census areas. The color scale from yellow to dark red identifies the traffic flow magnitude in logarithmic scale. The airline network is made mostly by long-range connections as compared with a gridlike ordering of the commuting network. The daily average flow of the commuting network is one order of magnitude larger than that of the airline network.

Through confronting us with serious ethical and privacy questions, these kinds of data and the reduced cost of producing, accessing, and communicating information on techno-social systems are changing our understanding of a wide range of phenomena (12–17). The spatial dynamics of human infectious diseases are determined by the mobility of individuals who carry a disease into previously uninfected populations. Analogously, human migration and mobility mediate a large number of bioinvasions, defined as the introduction of previously unknown organisms in ecosystems. The evolution of languages and dialects is also driven by the mixing of populations and the merging and/or isolation of communities. Finally, the daily mobility of humans in Internet space defines our exploitation and foraging of information.

### Network Thinking

The Internet and virtual worlds are networks that we navigate and explore every day (17–19). Human-interaction models are based on social networks in which nodes represent individual interacting agents and the links are potential interactions (20). Mobility, ecological, and epidemiological models rely on metapopulation networks that consist of entire populations interlinked by virtue of the exchanges between groups of individuals (21). A large body of work has shown that most real-world networks exhibit dynamic self-organization (that is, they become more complicated over time without the intervention of outside forces) and are statistically very heterogeneous; these characteristics are typical hallmarks of complex systems (22–24). The various statistical distributions characterizing these networks (including

the probabilities of node connection and the intensities of the connecting links) are generally heavy tailed and skewed, and they vary over several orders of magnitude (25). The foremost challenge offered by complex networks therefore resides in their interconnectedness (networks of networks) and multiscale nature. Figure 1 depicts three networks that exemplify human mobility at different scales, ranging from cross-continental airline travel to within-city mobility among mobile phone cell towers. Ideally, to make predictions about the processes driven by human mobility, we need to integrate this data, with its wide-ranging granularities (from a few hundred meters and a few hours to thousands of kilometers and several days), into a huge multiscale network.

Thus, the complexity of techno-social systems calls for a “network” mindset. A simple example is provided by the large-scale description of epidemic spreading. The spread of the plague epidemic in the 14th century (the Black Death) (26) was mainly a spatial diffusion phenomenon. Historical studies have established that the disease propagation followed a simple pattern that can be adequately described mathematically within the framework of continuous differential equations with terms that describe diffusion. As anticipated in 1933 (27), the large-scale and geographical impact of infectious diseases [such as the SARS epidemic (28) or the current swine flu epidemic] on populations in the modern world is mainly due to commercial air travel. An epidemic that starts in Southeast Asia will rapidly reach North America and Europe (Fig. 2). This picture, therefore, cannot be simply described in terms of diffusive phenomena; rather, it must incorporate the spatial structure of modern

transportation networks. For instance, it is the heavy-tailed nature of the airline traffic network that explains why travel restrictions alone are ineffective in containing a global epidemic unless the global mobility rate is reduced at least by one order of magnitude (29–31).

Another crucial aspect of modern network thinking is the dynamical self-organization that gives rise to large-scale infrastructure patterns independent of human planning and engineering of the system. The prime example of a dynamical self-organizing system may be the Internet, but most communication infrastructures, road and transportation systems, supply networks, and power distribution grids are also dynamically growing networks. Road construction, for instance, is obviously planned, and it is not surprising that considerations of optimization of cost, efficiency, and utility inform the planning effort. As a consequence, one could generally expect road networks to exhibit a high degree of regularity. Yet everyday experience suggests that this is not the case, especially in towns that have grown over a long period of time. For this reason, researchers have formulated simple road-formation models (32) that try to capture the tension between the notion of optimality that inspires planners and the limited time and spatial horizons that inform their decisions.

However, the biggest challenge in providing a holistic description of multiscale networks is the necessity of simultaneously dealing with multiple time and length scales. The final system’s dynamical behavior at any scale is the product of the events taking place on all scales. The single agent spreading a disease or single node of the Internet that fails are apparently not affected by



**Fig. 2.** Epidemic invasion tree obtained from the simulations of a pandemic originating in Hanoi, Vietnam. The nodes identify 3200 populations worldwide, and the directed links indicate the path along which the epidemic has moved

from one population to the other. The color map from dark red to dark blue is according to the time ordering of the epidemic invasion. Simulations obtained with the worldwide epidemic and mobility model from (38).

# Pushing Networks to the Limit

the multiscale nature of the network, just as single molecules do not care about the multiscale nature of turbulent fluids. However, the collective dynamical behavior and our ability to conduct mathematical and/or computational analyses of techno-social systems are constrained by the multiscale characteristic of the system. In the context of networks and techno-social systems, the multiscale challenge is making its appearance now because of the availability of large-scale data sets. Thus, we have to develop appropriate formalisms and techniques, as researchers studying multiscale physical systems (fluids, solids, distribution of masses in the universe, etc.) have done in the past (33). To achieve analytical understanding of techno-social systems and approach them computationally, we must find different strategies to deal with dynamical behavior and/or equations that work at very different characteristic scales but still influence each other. Such methods will finally allow the definition of layered computational approaches in which different modeling assumptions and granularities can be used consistently in the context of a general multiscale framework.

## Taking Advantage of Multiscale Networks

Knowledge of network characteristics opens the path to the discovery and understanding of new statistical and dynamical laws governing large infrastructural systems coupled to social systems. Furthermore, the massive interconnectivity of spatially distributed populations and the complexity and strong heterogeneity of multiscale networks are the keys to the construction of ab-initio computational models, in which the behavior of the system can be understood in a bottom-up perspective, as opposed to the traditional mean-field or top-down strategies. This happens in a wide array of contexts ranging from urban planning (34) to epidemic modeling (35–38). Notable examples are the TRANSIM and EPISIMS projects (35), in which agent-based models, including millions of individuals, are used to simulate the dynamics and traffic of entire cities and the spread of biological agents, respectively.

In some cases, the understanding of complex networks provides counterintuitive and surprising approaches to the engineering and management of complex techno-social systems. For example, in power grids and other flow-carrying networks, the failure of a single node or line can trigger a domino effect (“cascading failure”), in which the overload induced by the flow redistribution may generate a global failure of the network. By taking advantage of the heterogeneity of the flow carried on the links of multiscale networks, A. E. Motter (39) has proposed an adaptive defense mechanism that is actually based on the removal of a certain number of nodes to induce intentional failures. Although this mechanism might appear counterintuitive, the intentional failure of appropriately chosen nodes does not amplify the cascade process and, on the contrary, is able to mitigate the final damage. In other words, we

now can provide a rationale for understanding the emerging tipping points and nonlinear properties that often underpin the most interesting characteristics of a techno-social system’s behavior.

## The Toughest Challenge

Although many basic conceptual questions remain unresolved, the major roadblock in defining the fundamental predictability limits for techno-social systems is their sensitivity and dependence on social adaptive behavior. In the absence of a stress on the system, a stationary state is reached in which the feedback between the social behavior and the physical infrastructure determines the details of how the network behavior and the dynamical process of interest play out. We can imagine using steady-state data to forecast system behavior under such “normal” conditions. However, in the case of catastrophic events (for instance, the disruption of social order during emergencies such as pandemics or major natural disasters), the behavior of techno-social networks is driven out of equilibrium into unknown territory.

An interesting and ethically challenging aspect of predicting and managing the unfolding of catastrophic events in techno-social networks is the system’s adaptation to predictions when they are made publicly available. Social behaviors react and adapt to knowledge of predictions. Contrary to what happens in physical systems, the predictions themselves are part of the system dynamic. In addition, predictions may point to unethical control and anticipation strategies favoring specific demographic sectors of the society. Finally, the risk of erroneous predictions may lead to costly or unethical social control mechanisms with no actual benefits. Whereas some of the above issues may find a partial solution through improvements in the accuracy and reliability of models, it is clear that social adaptation to predictions presents us with new methodological and ethical problems.

Addressing these problems involves tackling three major scientific challenges. The first is the gathering of large-scale data on information spread and social reactions that occur during periods of crisis. This is not presently out of reach, via large-scale mobile communication databases (such as mobile telephones, Twitter logs, and social Web tools) operating at the moment of specific disaster or crisis events. The second challenge is the formulation of formal models that make it possible to quantify the effect of risk perception and awareness phenomena of individuals on the techno-social network structure and dynamics. The third challenge concerns the deployment of monitoring infrastructures capable of informing computational models in real time. Complex systems and networks theory, mathematical biology, statistics, nonequilibrium statistical physics, and computer science all play a key role in the effort to meet these challenges. Although such an integrated approach might still be in its infancy, it now seems possible to imagine the creation of

computational forecasting infrastructures that will help us design better energy-distribution systems, plan for traffic-free cities, anticipate the demands of Internet connectivity, or manage the deployment of resources during health emergencies.

## References and Notes

1. D. Lazer *et al.*, *Science* **323**, 721 (2009).
2. A. Pentland, in *The Global Information Technology Report 2008-2009* (World Economic Forum, Geneva, 2009), p. 75.
3. J. P. Onnela *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332 (2007).
4. D. Brockmann, L. Hufnagel, T. Geisel, *Nature* **439**, 462 (2006).
5. D. J. Watts, *Nature* **445**, 489 (2007).
6. M. C. Gonzalez, C. A. Hidalgo, A.-L. Barabasi, *Nature* **453**, 779 (2008).
7. G. A. Lundberg, M. Lawnsing, *Am. Sociol. Rev.* **2**, 318 (1937).
8. S. Erlander, N. F. Stewart, *The Gravity Model in Transportation Analysis* (VSP, Utrecht, Netherlands, 1990).
9. J. de Dios Ortuzar, L. G. Willumsen, *Modelling Transport* (Wiley, New York, 2001).
10. D. Brockmann, F. Theis, *Pervasive Comput.* **7**, 28 (2008).
11. F. Giannotti, D. Pedretti, *Mobility Data, Mining and Privacy* (Springer, Berlin, 2008).
12. S. Riley, *Science* **316**, 1298 (2007).
13. C. Viboud *et al.*, *Science* **312**, 447 (2006); published online 29 March 2006 (10.1126/science.1125237).
14. L. Hufnagel, D. Brockmann, T. Geisel, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15124 (2004).
15. V. Colizza *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2015 (2006).
16. G. M. Ruiz *et al.*, *Nature* **408**, 49 (2000).
17. B. A. Huberman, L. Adamic, *Lecture Notes in Physics* (Springer, Heidelberg, Germany, 2003).
18. R. Pastor-Satorras, A. Vespignani, *Evolution and Structure of the Internet* (Cambridge Univ. Press, Cambridge, 2004).
19. M. E. Crovella, B. Krishnamurthy, *Internet Measurements: Infrastructure, Traffic and Applications* (Wiley, Chichester, UK, 2006).
20. S. Wasserman, K. Faust, *Social Network Analysis* (Cambridge Univ. Press, Cambridge, 1994).
21. I. A. Hanski, O. E. Gaggiotti, *Ecology, Genetics and Evolution of Metapopulations* (Academic Press, San Diego, CA, 2004).
22. A.-L. Barabasi, R. Albert, *Science* **286**, 509 (1999).
23. M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
24. A. Barrat, M. Barthélemy, A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge Univ. Press, Cambridge, 2008).
25. A. Barrat *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747 (2004).
26. J. D. Murray, *Mathematical Biology* (Springer, New York, 1993).
27. A. Massey, *Epidemiology in Relation to Air Travel* (H. K. Lewis, London, 1933).
28. J. S. M. Peiris, K. Y. Yuen, K. Stohr, *N. Engl. J. Med.* **349**, 2431 (2003).
29. T. D. Hollingsworth, N. M. Ferguson, R. M. Anderson, *Nat. Med.* **12**, 497 (2006).
30. J. M. Epstein *et al.*, *PLoS One* **2**, e401 (2007).
31. V. Colizza, A. Vespignani, *J. Theor. Biol.* **251**, 450 (2008).
32. M. Barthélemy, A. Flammini, *Phys. Rev. Lett.* **100**, 138702 (2008).
33. G. I. Pavliotis, A. M. Stuart, *Multiscale Methods: Averaging and Homogenization* (Springer, Heidelberg, Germany, 2008).
34. M. Batty, *Science* **319**, 769 (2008).
35. S. Eubank *et al.*, *Nature* **429**, 180 (2004).
36. N. M. Ferguson *et al.*, *Nature* **442**, 448 (2006).
37. T. C. Germann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5935 (2006).
38. V. Colizza *et al.*, *PLoS Med.* **4**, e95 (2007).
39. A. E. Motter, *Phys. Rev. Lett.* **93**, 098701 (2004).
40. I thank V. Colizza, D. Balcan, B. Gonçalves, M. Gonzalez, and H. Hu for help with the figures and M. Gonzalez for the data used in Fig. 1C. I am partially supported by NIH, NSF, the Defense Threat Reduction Agency, the Lilly Endowment Foundation, and the Future Emerging Technologies projects Epiwork and Dynanets.

10.1126/science.1171990