Fighting Epidemics in the Information and Knowledge Age

Hai Zhuge and Xiaoqing Shi, Chinese Academy of Sciences

Computing professionals can play a key role in quickly containing outbreaks of biological viruses.

Since its initial outbreak, scientists have sought to control the severe acute respiratory syndrome epidemic. So far, though, these SARS researchers have paid too little attention to the complex relationships between the epidemic, the environment, and our Information and Knowledge Age society. Fortunately, information technology can help scientists investigate the epidemic's cause and unveil its relevant principles and rules.

VIRUSES AND THE ENVIRONMENT

Human beings, like all multicellular organisms, share their environment with versatile viruses. Evolution over billions of years has maintained a balance of energy and material flow between species, a balance essential to harmonious coexistence and continuing evolution.

A species can prosper when it can obtain enough energy and material from the environment, particularly from other species. Through energy and material flow, nature forms subtle interactions between the material's producers, consumers, and decomposers. Any unrecoverable change in the flow can threaten a species' prosperity.

We have simulated the spread of SARS and shown that isolation control measures had no significant effect on containing the epidemic's outbreaks (http://kg.ict.ac.cn)—an outcome counterintuitive to the thinking of many people and some governments, which sought to combat the disease by implementing measures more or less blindly.

Humanity's natural resilience to infection proved most important in curbing SARS. Natural resistance also plays a role at the society and ecosystem levels. A healthy society and ecosystem can maintain the harmony of existence by nurturing the resilience necessary to resist some invasions while evolving to resist others.

INDIVIDUAL, SOCIETY, AND ECOSYSTEM HEALTH

We used a hierarchical network model to investigate ecosystem health. The model's top level contains the ecosystem itself, which consists of several subsystems or communities. A subsystem can further possess several smaller subsystems. Each community contains populations and, in turn, each population contains individuals.

Each node in the hierarchical network has a health value. Each low-level node makes a health contribution to its parent node. Symmetrically, each parent node affects the health of its child nodes. The health of the top-level node—the ecosystem itself—depends on its subsystems' or communities' health. An internal node's health depends on the health of all its child nodes and the parent.

An individual's health depends on the health of that person's population and its intrinsic health value. A node's health-contribution value depends on its scale, health, and the amount of energy, material, and information it processes. The individual's health contribution value reflects that person's position and role in a population. Using this model, we investigated the dynamic relationship between ecosystem, communities, and individuals.

SOCIAL EFFECTS

Humans possess intelligence and have cultures that differ from region to region. People will adopt intelligent behavior to avoid infection if they know how to protect themselves and can accurately assess the current situation—the state of the epidemic, their environment, and society. Public response will tend to counteract an epidemic by reducing contact between people through, for example, canceling public gatherings. This kind of natural resilience prevents an epidemic from spiraling out of control.

However, public response to an epidemic depends on clear and timely dissemination of information about the initial outbreak. In human society, religious, cultural, political, and adminis-

Continued on page 114

Continued from page 116

trative factors affect that clarity and timeliness. Further, the nature of a society's evolution means that it develops different health standards at different stages. In the Information and Knowledge Ages, to keep its members healthy, a society should be stable, educated, politically democratic, diverse in religion and culture, effective and efficient in government, respectful of the law, cooperative in spirit, socially responsible, and generous in sharing information and knowledge.

INFORMATION AND KNOWLEDGE SHARING POWER

Information and knowledge sharing profoundly influenced the extent and duration of the SARS epidemic. At first, lack of information and knowledge sharing hampered China's efforts to research the virus and control the epidemic. SARS appeared initially in Guangdong province, but during the outbreak's early stages, the obtained experience with controlling and curing the disease was not available to health workers in other affected regions.

This lapse had three serious consequences. First, other regions' local governments could not implement proper control measures in the epidemic's early stages. Second, many of these regions' doctors and nurses became infected because they did not know how to protect themselves from SARS. Third, scientists failed to provide timely and effective advice based on useful research because the various regional research institutes managed their research resources independently, particularly virus samples.

Lack of clear and timely information has social effects as well, causing fear, destroying trust, and disrupting normal ties between people. Information and knowledge sharing can be a powerful tool for fighting an epidemic. Governments should promote and support information and knowledge sharing, make policies that encourage professional workers to share their experience, and set up an expediting authority to collect, organize, qualify, and confirm the information and knowledge being shared.

INTERNET-BASED INFORMATION AND KNOWLEDGE SHARING

People can use the World Wide Web to rapidly share information across regions. Indeed, the World Health Organization first learned of China's SARS epidemic via the Web. Currently, however, the Web cannot guarantee the accuracy and reliability of the data it holds. To overcome these limitations, scientists are exploring ways to reshape the Web. The Semantic Web (www.semanticWeb.org) and the Grid (www.gridforum.org) represent just two of these efforts.

Information and knowledge sharing profoundly influenced the extent and duration of the SARS epidemic.

The ideal Internet-based eco-environment management system would integrate ecological environment evaluation, data collection and publication, simulation, forecasting, and situation monitoring. It could help governments make decisions based on the current ecological and epidemiological situation, forecast developments, and evaluate the eco-environment. Professional workers in various regions could store and publish information in the system and thus make that data immediately available to people in other regions.

Because information sources reside in different regions, collected data can be both inconsistent and redundant. Some people may have several jobs or maintain homes in different regions. Some cases of infection may be reported more than once when, for example, a patient changes hospitals. A patient's records may not be updated promptly when that person's status changes. Ideally, the system should be able to maintain consistency and eliminate redundancy in all these cases.

The Semantic Grid will provide an interconnected environment that could provide such a solution (H. Zhuge, "Clustering Soft-Devices in the Semantic Grid," *IEEE Computing in Science and Engineering*, vol. 4, no. 6, 2002, pp. 60-62). This future incarnation of the Web will be a virtual social environment in which people can immerse themselves in work, obtain intelligent services, and effectively share information and knowledge.

VIRUS SPREAD MODEL

Our research shows that SARS only spreads through a small subset of the threatened society. Simulation results tell us that the number of people infected is sensitive to population density. Survey results from Beijing show that SARS spread mainly among close-contact medical staffs, a very small but dense part of the overall population. The Beijing survey shows that the largest close-contact infection tree covered only 37 people.

This characteristic makes analyzing the SARS spread network feasible. We can think of it as a network of cliques—small, close-contact spread subnetworks. People within the same clique share some common social roles—belonging, for example, to the same family or workplace—and are more likely to infect each other than to infect people belonging to different cliques.

We use five variables to characterize the spread network:

- *Contact frequency* reflects how often network nodes contact each other.
- Nodes with more and higher contact-frequency arcs have a higher *contact rank*.
- *Infected ratio* reflects the probability of a node being infected. A node's infected ratio is the sum of the products of the infected ratio of all directly linked nodes and their contact ranks.
- *Infect ratio* reflects the probability of a node infecting other nodes. A node's infect ratio is the prod-

uct of its infected ratio and its contact rank.
- We determine a node's *infect rank* from the product of its contact rank and infect ratio.

Drawing an analogy between the spread network and the Web's hyperlink network reveals some similarities. The two entities' structures resemble each other. Further, the spread network's contact rank variable closely resembles the Web's page rank variable (J. Kleinberg and S. Lawrence, "The Structure of the Web," *Science*, vol. 294, no. 30, 2001, pp. 1849-1850). But the spread network's nodes and arcs constitute a more complex structure than the Web's structures. This relationship lets us view the hyperlink network as a special case of the spread network.

Research on Web page distribution shows that the page rank obeys the power-law distribution (L.A. Adamic and B.A. Huberman, "Power-Law Distribution of the World Wide Web," *Science*, vol. 287, no. 24, 2000, p. 2115): The high-rank nodes account for only a minute proportion of all the network's nodes. This result also applies in the spread network.

Thus, effective epidemic-control measures must prevent the high-rank nodes from being infected. Each person usually plays more than one role in society and thus can participate in several nodes that belong to different cliques. All people who participate in higher-rank nodes should be monitored at a higher priority.

The number of infected people is a function of time. Further, if we consider the spread network across time, the five variables are also functions of time. Surveys in Beijing show that infected people are not highly contagious in the latent period, so we can filter out a smaller spread-network clique by determining the effective contact frequency according to the epidemic's development.

Current Web technologies support automatic analysis of peoples' professional and contact relationships. We can develop software that will perform this analysis by classifying homepages, extracting interrelated keywords, analyzing hyperlinks, and tracing the Web pages that these people and their family members, colleagues, and employers commonly access.

Using semantic markup languages to replace HTML-based Web pages would make this automatic analysis more effective. Accomplishing this would require an eco-environment management service system to update personal information and monitor the epidemic.

The Web has played an important role in controlling the spread of biological viruses.

ECO-ENVIRONMENT MANAGEMENT

Establishing an eco-environment management service system would help collect and manage accurate social, epidemic, and eco-environment information; eliminate redundant data; facilitate resource sharing; help analyze epidemiological data; and assist in conducting collaborative research by integrating local systems. Connecting nationwide systems through the Internet could create a worldwide eco-environment management service system that would help with collaborative monitoring, simulation, research, management, and control of epidemic situations.

Such a system should include globally distributed resources—devices, information, knowledge, and services—that could dynamically and intelligently collaborate to provide effective just-in-time services on demand to help manage an emerging crisis.

COMPUTING PROFESSIONALS

The Web has played an important role in controlling the spread of biological viruses. However, the Internet community itself still encounters versatile and persistent e-viruses. Drawing an analogy between e-virus epidemics and viruses like SARS suggests that computing professionals

- can acquire a good understanding of how a virus spreads and how such a spread should be controlled,
- are uniquely qualified to develop methods and strategies for controlling an e-virus epidemic that can be used to help professionals in other fields,
- should seek to help professionals in other fields through collaboration that makes the most of the different professions' combined experiences, and
- will be instrumental in ensuring that tomorrow's Internet provides healthy support for interdisciplinary collaboration.

The intertwining of computer and biological epidemics, the eco-environment, society, and information technologies makes it obvious why controlling SARS required the cooperation of many professions, not least the computing profession. As we learn how to cooperate more effectively, we will be better able to deal with future crises and plan for sustainable development, environmental improvement, social prosperity, and professional success. ■

***Hai Zhuge** is executive director and a professor at the Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences. Contact him at zhuge@ict.ac.cn.*

***Xiaoqing Shi** is an assistant professor at the Key Lab of Systems Ecology, Chinese Academy of Sciences. Contact her at shixq@mail.rcees.ac.cn.*

Editor: Neville Holmes, School of Computing, University of Tasmania, Locked Bag 1-359, Launceston 7250; neville.holmes@utas.edu.au