COMPLEX NETWORKS

From data to models

Data on the movement of people becomes ever more detailed, but robust models explaining the observed patterns are still needed. Mapping the problem onto a 'network of networks' could be a promising approach.

**Marta C. González and
Albert-László Barabási**

**are at the Center of Complex Network Research,
Department of Physics and Computer Sciences,
University of Notre Dame, Notre Dame, Indiana
46556, USA.**

e-mail: Marta.C.Gonzalez.119@nd.edu; alb@nd.edu

If you have a mobile phone, your carrier always knows your whereabouts, has a list of your friends and knows how often you have kept in touch with them lately. If misused, this record, together with datasets capturing your e-mail, web-browsing or buying habits collected by various companies could lead to significant intrusions into your privacy. However, these records represent a huge opportunity to science, offering access to patterns of human behaviour at a level and of detail previously unimaginable. Quantifying and understanding such patterns may help us to design better public transport and safer public spaces, or to control a disease outbreak. But how do we balance the inherent tension between the need for information and privacy? One way is to follow the framework developed by Colizza and colleagues[1], ignoring the individuals and focusing instead on a coarse-grained description of the system, using block or cell variables.

For many problems of significant importance, such as predicting a potential viral outbreak, the ultimate model and monitoring system would need to know the whereabouts of each individual in a country[2]. In industrial countries, with almost 100% mobile-phone penetration, such information is readily available to phone companies. Indeed, each phone communicates with the closest tower, leading to a natural partition of the country into distinct geographic cells (see Fig. 1). Given that calls are recorded for billing purposes, the movement of each mobile phone user can be reconstructed, as illustrated by the solid lines in Fig. 1. Yet, if we want to predict the spread of a contact-based viral infection, such as influenza or SARS, such detailed individual
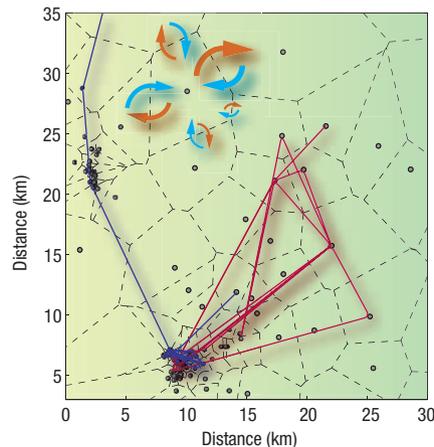

Figure 1 Individual and block-based representation of mobile communications. Each circle represents a mobile-phone tower and the dashed lines correspond to a Voronoi diagram that roughly delimits the main reception zone of each tower, partitioning the space into individual cells. The blue and red solid lines show the trajectory of two mobile-phone users, illustrating how the call activity helps us to track individual motion. For many applications, like modelling the spread of diseases, it may be sufficient to monitor the flux capturing the aggregated motion of individuals between different cells, as illustrated in the top of the figure, rather than individual trajectories.

information is not required. Knowing the number of people moving from one region to the other would be sufficient (Fig. 1). Such real-time monitoring of the aggregated motion of individuals, rather than individual trajectories, may be acceptable for privacy advocates and would be sufficient to develop public monitoring and alert systems[3]. Yet this approach poses some fundamental scientific challenges: how do we formulate the complex diffusion and spreading problem using such a coarse-grained description?

The difficulties involved in the transition are illustrated by the work of Colizza et al.[1], who have developed a general formalism to study a system in which a number of particles — be they individuals, chemical

species or pieces of information — can coexist in the same region, with each region being represented as a node in a network. Diffusion to nearby regions takes place along the network links. The network can be a regular lattice if it aims, for example, to capture reactions between diffusive reactants. However, if the problem at hand described a virus spreading by airline traffic between cities, it would have a so-called scale-free architecture[4] (as found in numerous real-world networks). For the mobile-phone-based system of Fig. 1, the network captures the traffic between regions determined by geographically neighbouring mobile-phone towers.

Viruses have an important property: if they spread too slowly, they may die out. If their spreading rate exceeds a critical threshold, however, an outbreak will occur, which could potentially reach a considerable fraction of the population. In 2001, Pastor-Satorras and Vespignani[5] established a now classic result: the epidemic threshold vanishes if the network on which the virus spreads is scale-free. Given the evidence that both sexual[6,7] and e-mail networks[8] may have scale-free characteristics, this means that even weakly virulent biological or e-mail viruses have the potential to spread — a prediction that has not only renewed interest in the interplay between the network structure and spreading processes, but has also initiated a vigorous debate on the subject.

But what happens with this vanishing threshold when we move from individual to block variables? Attempting to answer this question, Colizza and colleagues hand us a puzzle: when each individual interacts with only a finite number of other individuals within a block (which is a realistic assumption), the threshold remains, even if the network is scale-free. So something is lost in the translation. Indeed, descriptions based on block variables predict that weakly spreading viruses will easily die out and that traditional epidemic control measures aimed at decreasing the spreading rate could succeed in stopping an epidemic. Interestingly, the limit where the

threshold does vanish — when an individual interacts with every other individual within the block — is not realistic when direct-contact processes, such as e-mail or sexual interaction, are responsible for the spread.

This fundamental puzzle reflects the challenges in moving from individual to block variables, highlighting our limited understanding of how individuals behave within a block, or how they move between them. For example, there is a fine-scale network describing the interactions between individuals found within each cell, ultimately forcing us to address a multiscale problem, describing a network of networks. Money-tracking measurements indicate that individual travel patterns between blocks may be heavy tailed[9]. Similarly, all measured times between consecutive human-driven events, such as library visits or e-mails, seem to be described by heavy-tailed processes[10,11], which challenges the traditional Poisson-process-based modelling framework[11]. Therefore,

the model studied by Colizza et al.[1] explores logical extreme cases rather than experiment-based local mixing patterns. Yet, simple models rooted in statistical physics, such as this, go a long way in sharpening our understanding of the key features, difficulties and paradoxes involved in modelling human-driven processes.

Technology, as well as boosting our communication and monitoring capabilities, has inundated us with huge amounts of information about human activity patterns. This flood of data has the power to revolutionize our understanding of human behaviour, with applications from urban and transportation planning to emergency response and crime investigation. In the ensuing journey from data to models, statistical physics concepts have played a key role, offering a framework to quantify the highly stochastic human-driven processes. Such data-driven opportunities have challenged the physics community as well, forcing us to explore

both the limitations and the potential of our tools. This is a win–win situation, a vivid example of the changing nature of physics in the twenty-first century, which is taking us into areas where we did not dare, or could not venture before.

References

1. Colizza, V., Pastor-Satorras, R. & Vespignani, A. *Nature Phys.* **3,** 276–282 (2007).
2. Eubank, S. *et al. Nature* **429,** 180–184 (2004).
3. Madey, G., Szabo, G. & Barabási, A.-L. in *Lecture Notes in Computer Science* Vol. 3993 (eds Alexandrov, V. N., van Albada, G. D., Sloot, P. M. A. & Dongarra, J.) 417–424 (Springer, Berlin, 2006).
4. Barabási A.-L. & Albert, R. *Science* **286,** 509–512 (1999).
5. Pastor-Satorras, R. & Vespignani, A. *Phys. Rev. Lett.* **86,** 3200–3203 (2001).
6. Schneeberger, A. *et al. Sex. Transm. Dis.* **31,** 380–387 (2004).
7. Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. & Åberg, Y. *Nature* **411,** 907–908 (2001).
8. Ebel, H., Mielsch, L.-I. & Bornholdt, S. *Phys. Rev. E* **66,** 035103 (2002).
9. Brockmann, D., Hufnagel, L. & Geisel, T. *Nature* **439,** 462–465 (2006).
10. Barabási, A.-L. *Nature* **435,** 207–211 (2005).
11. Vazquez, A., Balázs, R., András, L. & Barabási A.-L. *Phys. Rev. Lett.* (in the press); preprint at <http://arxiv.org/abs/physics/0609184> (2006).