

# It's a small world

James J. Collins and Carson C. Chow

**The concept of Six Degrees of Separation has been formalized in so-called 'small-world networks'. The principles involved could be of use in settings as diverse as improving networks of cellular phones and understanding the spread of infections.**

A few years ago, on American campuses, it was popular to play Six Degrees of Kevin Bacon. In this game, participants attempt to link the actor Kevin Bacon to any other actor through as few common films and co-stars as possible. Links are formed directly between Bacon and another actor if they appeared in the same film or indirectly through a chain of co-stars in different films (Fig. 1).

In the world of mathematics, a similar amusement involves assessing one's Erdős number, which measures the number of links needed to connect one to the prolific mathematician Paul Erdős through jointly authored papers. For example, individuals have an Erdős number of 1 if they co-authored a paper with Erdős. If one of their co-authors wrote a paper with Erdős, then they have an Erdős number of 2, and so forth. It has been pointed out<sup>1</sup> that Dan Kleitman has a combined Erdős/Bacon number of 3 because he wrote a paper with Erdős and appeared in *Good Will Hunting* with Minnie Driver, who appeared with Bacon in *Sleepers*.

These games are related to the popular concept of Six Degrees of Separation<sup>2</sup>, which is based on the notion that everyone in the world is connected to everyone else through a chain of at most six mutual acquaintances. If two people have one mutual acquaintance, then they have one degree of separation. The estimate of six degrees of separation, which is related to the small-world phenomenon<sup>3,4</sup>, arises from pioneering empirical work by Milgram<sup>3</sup> and can be understood heuristically from a somewhat unrealistic assumption of random connectivity. That is, if each person knows about one hundred individuals, and given that there are about a billion people on the Earth, then seven connections or six degrees of separation are enough to link everyone together.

On page 440 of this issue<sup>5</sup>, Watts and Strogatz formalize this idea in what they call small-world networks. They demonstrate through numerical simulations that a network need not be very random to get this small-world effect. They consider a connected network with nodes and links. In the friendship analogy, each node represents a person and each link represents a single connection to an acquaintance. They then define

two measures. The first is a characteristic path length. This is the smallest number of links it takes to connect one node to another, averaged over all pairs of nodes in the network. The second measure is the clustering coefficient. This measures the amount of cliquishness of the network, that is, the fraction of neighbouring nodes that are also connected to one another. For example, in an all-to-all connected network, the clustering coefficient is one.

An example of a large-world network is one that is regularly and locally connected like a crystalline lattice. Such a network is highly clustered and the characteristic path length is large, scaling with the typical linear dimension of the network. On the other hand, a completely random network is poorly clustered and the characteristic path

length is short, scaling logarithmically with the size of the network.

What Watts and Strogatz<sup>5</sup> do is to shift gradually from a regular network to a random network by increasing the probability of making random connections from 0 to 1 (see Fig. 1, page 441). They then measure the characteristic path length and the amount of clustering of the network as a function of the amount of randomness. They find that path length and clustering depend differently on the amount of randomness in the network. The characteristic path length drops quickly, whereas the amount of clustering drops rather slowly. This leads to a small-world network in which the amount of clustering is high and the characteristic path length is short. So a small world can exist even when the cliquishness is imperceptibly different from that of a large world.

The explanation for this effect is that it only takes a few short cuts between cliques to turn a large world into a small world. In the friendship analogy, it only takes a small number of well-connected people to make a world small. The interesting and surprising thing is that it is impossible to determine whether or not you live in a small world or a large world from local information alone. The average person (node) is not directly associated with the key people (the clique-linkers).

Small-world connectivity has consequences that could be good or bad,



Figure 1 Three degrees. Because Kevin Bacon has appeared in many films, most actors have low Bacon numbers and the game Six Degrees of Kevin Bacon has declined in popularity. It is possible to centre the game around a newer star such as Leonardo DiCaprio. These film stills, running clockwise, show that in this case there are at most three degrees of separation between DiCaprio and Helena Bonham-Carter, through Kate Winslet (*Titanic*, Columbia TriStar; *Sense and Sensibility*, Columbia TriStar), Emma Thompson (*Sense and Sensibility*; *Much Ado About Nothing*, Entertainment Films) and Kenneth Branagh (*Much Ado About Nothing*; *Frankenstein*; Columbia TriStar). Short cuts between cliques could be created in this game through some of DiCaprio's well-connected co-stars such as Sharon Stone (*The Quick and the Dead*; TriStar; not shown).

depending on the system and circumstances. In a simple model of disease spreading, for example, Watts and Strogatz show that the time for global infection behaves much like the characteristic path length. So it takes only a few short cuts to increase the spreading of disease significantly. Clearly, in this case, small-world coupling is problematic.

The findings of Watts and Strogatz could be put to good use, however, particularly in the case of existing networks. For instance, it may be possible to improve the performance of cellular-phone networks by deliberately introducing a few random connections between cells (nodes). Such a change could improve traffic flow around the network, without requiring the creation of an entirely new set of relay stations. Similar dynamics could also be exploited to improve the flow of information throughout the Internet. Although we are not in a position to redesign the Internet from scratch, it is possible

to introduce a few random links between nodes along the backbone of the Internet. These small-world modifications could substantially reduce the time needed to send a message by electronic mail or find a particular Web site, as well as improve the reliability of the overall network. Strategies for determining and achieving optimal small-world connectivity remain to be developed. □

James J. Collins and Carson C. Chow are in the Center for BioDynamics, Department of Biomedical Engineering and Department of Mathematics, Boston University, 44 Cummings Street, Boston, Massachusetts 02215, USA.  
e-mail: collins@enga.bu.edu

1. <http://www.acs.oakland.edu/~grossman/Erdoshp.html>
2. Guare, J. *Six Degrees of Separation: A Play* (Vintage, New York, 1990).
3. Milgram, S. *Psychol. Today* 2, 60–67 (1967).
4. Kochen, M. (ed.) *The Small World* (Ablex, Norwood, NJ, 1989).
5. Watts, D. J. & Strogatz, S. H. *Nature* 393, 440–442 (1998).

Marine ecology

## Microbial microdiversity

Jed A. Fuhrman and Lisa Campbell

The bacterial group *Prochlorococcus*, discovered only a decade ago, may be the most abundant component of phytoplankton in the sea. These tiny (0.6 μm) organisms uniquely contain the photosynthetic pigments divinyl chlorophyll *a* and *b*, and are major primary producers in tropical and subtropical waters (that is, some 75% of the world's oceans). They contribute between 10% and 80% of total local primary production<sup>1,2</sup>.

One apparent reason for their success is their ability to grow well over a wide range of light conditions. So, how do they do it? As reported by Moore *et al.* on page 464 of this issue<sup>3</sup>, it turns out that the feat is accomplished by the coexistence of closely related populations that are genetically adapted to remarkably different light levels. Close relatives from the same sample can be so different in their light response that one may grow optimally in light that is bright enough to inhibit the other completely. Together with the results of Urbach *et al.*<sup>4</sup>, these observations not only explain why *Prochlorococcus* is so successful, but they also bear on issues that are central to the measurement and interpretation of microbial diversity.

To understand these broader implications, one must first realize how such studies have recently changed. Classical identification of a microbe requires cultivation of the organism concerned, which is difficult or impossible for most organisms from nature (perhaps 99%). In consequence, although as many as 5,000 bacterial and archaeal species are officially recognized, the true number is probably in the millions<sup>5</sup>. There is simply no

practical way to carry out a comprehensive survey of natural microbial diversity by classical techniques alone.

Norman Pace and colleagues set out to find a way around this limitation. They showed<sup>6</sup> that one could clone the genes for 16S ribosomal RNA from a naturally occurring, mixed, microbial biomass, sequence them, and then identify the organisms from which the genes came by comparing the sequences to the large database of 16S rRNA sequences. Phylogenetic analysis places the clone sequences on a tree, from which relationships to known organisms or other environmental clones can be seen. By eliminating the need for cultivation, this approach has been a remarkably fruitful way to survey diversity. It has led to the discovery of major groups of organisms, still uncultured and without formal names, such as the low-temperature archaea<sup>7–9</sup>. In the past year alone, more than 20 new divisions of bacteria have been reported<sup>9–11</sup> at the phylum — or even possibly the kingdom — level.

The cloning approach has also yielded several clusters of closely related sequences, which may indicate that diversity is common on much smaller scales. Examples of marine clusters include SAR 11 (ref. 12) and Marine Group I archaea<sup>7–9</sup> (Fig. 1), both of which are distributed widely around the world yet are evolutionarily distant from all known cultures. Typical within-cluster differences in 16S rRNA sequences are <1–10%, and clades within clusters are also apparent<sup>9,12</sup>. But it is difficult to know what the within-cluster diversity means without having living cultures to examine. Are relatively small differ-

ences in sequences (termed microdiversity by Moore *et al.*<sup>3</sup>) the result of genetic drift or microevolution in physiologically similar clonal populations, or do they represent macroevolution of distinct populations with genetically determined adaptations and different niches? The answer to this question is crucial, because it tells us whether or not it is appropriate to lump together the members of the cluster for the purposes of diversity analysis. The difference could multiply our perception of natural diversity by several times.

Moore and colleagues' study<sup>3</sup> is exciting because it concentrates on a cluster that has been detected repeatedly in open ocean environments by the cloning approach<sup>13</sup>, but which can also be studied in nature by flow cytometry or in laboratory culture. This group has been called the Marine Picophytoplankton Clade, and includes *Prochlorococcus* and *Synechococcus* cyanobacteria<sup>4</sup>. These organisms do not have the same light-harvesting components (*Synechococcus* possess true chlorophyll *a* and phycobilisomes, and lack divinyl chlorophyll *a* and *b*), but they are about 96–98% identical in their 16S rRNA sequences. Phylogenetic analysis shows identifiable sub-lineages within this group, with one being the clade of *Prochlorococcus* that is adapted to high light intensities<sup>3,4</sup>. Members of this clade have about 97–98% 16S rRNA sequence identity with *Prochlorococcus* adapted to low light levels, but the two types coexist<sup>3</sup>.

Moore *et al.*, then, show how differences of only about 2% in 16S rRNA sequence correspond to ecologically significant physiological diversity. So if lumping together clusters with a 2% difference would be a mistake, what about even smaller differences? Examination of the 16S rRNA database from cultures shows many examples where there are notable physiological differences despite tiny (<0.5%) sequence differences (for example, the anthrax organism *Bacillus anthracis* and the insect pathogen *B. thuringiensis*<sup>14</sup>); and with plasmids or other mobile genetic elements often coding for important properties such as virulence, there are instances where organisms with even completely identical 16S rRNA sequences occupy different niches<sup>14</sup>.

Estimates of microbial diversity based on 16S rRNA probably therefore represent a minimum of the true ecological situation. There are other factors to be considered, however. For instance, a small part of the apparent diversity in cloned clusters is possibly due to variations in rRNA-coding gene sequences between the multiple copies typically present in many organisms. Or a small part may be due to experimental errors in carrying out the polymerase chain reaction, and cloning and sequencing<sup>12</sup>. Although these points are not new, they need to be