

# Thinking outside the cube

The discovery that many complex systems are actually well-structured networks has not only changed the landscape of physics, but also how we visualize patterns in science, explains **César A Hidalgo**

**César A Hidalgo** is at the Center for International Development, Harvard University, Cambridge, US, e-mail cesar\_hidalgo@ksg.harvard.edu

To my colleagues in economics and political sciences at the Harvard Kennedy School, I am the guy who makes pretty pictures. Although trained as a physicist, I build and design beautiful maps that depict the internal architecture of complex systems in nature and society. These maps, called network visualizations, enable us to transform an amorphous mass of data into colourful and inhomogeneous structures of patterns and connections that are relatively easy for people to recognize and remember. Their complex structure appeals to our pattern-recognizing brains, in sharp contrast to the same data appearing in a spreadsheet or scatter plot.

At hi-tech places like the Massachusetts Institute of Technology's Media Lab or the Center for Complex Network Research at Northeastern University, only a few streets away from my office, network visualizations are considered business as usual, but at the Harvard Kennedy School they are looked at with awe by some and with scepticism by others. It is easy to see why, as network visualizations are far removed from the visual aids most common in economics, such as regression tables or pairs of lines that intersect at a point of equilibrium. But it turns out that such visualizations can reveal patterns that no other method can.

## Everything is linked

Network science, which is the basis of my “pretty pictures”, is a fast-growing segment of the highly interdisciplinary field of study known as complexity science. From a philosophical perspective, complexity science differs greatly from other disciplines, as its goal is to develop understanding not by using a reductionist “divide and conquer” analysis – in which natural systems are reduced to their components and analysed independently – but to study how the components of a system interact and give rise to emergent properties and behaviour. Network science studies how the different pieces of scientific puzzles, found by using reductionist methods, fit together.

Many network scientists are, like me, physicists. They have been highly influential in giving direction to this new scientific paradigm, by setting its agenda and honing its quantitative methods. Many of them were attracted to the field following the discovery, made in 1998 by Réka Albert and Albert-László Barabási while at the University of Notre Dame in Indiana, that most networks in nature and society follow a power-law distribution: meaning that the probability that a randomly chosen node in a network has  $k$  links connecting it to other nodes decays as an inverse power of  $k$ .

Barabási, Albert and colleagues first observed this power-law distribution on the World Wide Web: the probability that the number of webpages (nodes) with exactly  $k$  links is proportional to  $1/k^n$ , with  $n$  being

approximately 2. Together with the Web, many similar power-law-distributed “scale-free” networks were discovered: research collaborations; the Internet; protein networks; sexual relationships; and even Hollywood. In each case, the value of  $n$  always falls between 2 and 3.

Of course, physicists have been dealing with power laws ever since Newton discovered the law of gravitation. Power laws are ubiquitous in many physical systems and have been widely studied by statistical physicists: from fractals, in which power-laws define their dimensions; to critical phenomena, such as phase transitions. Following Barabási's discovery, it was only natural that physicists, proficient in calculating and simulating power-law distributions, should have spearheaded the subsequent research on networks.

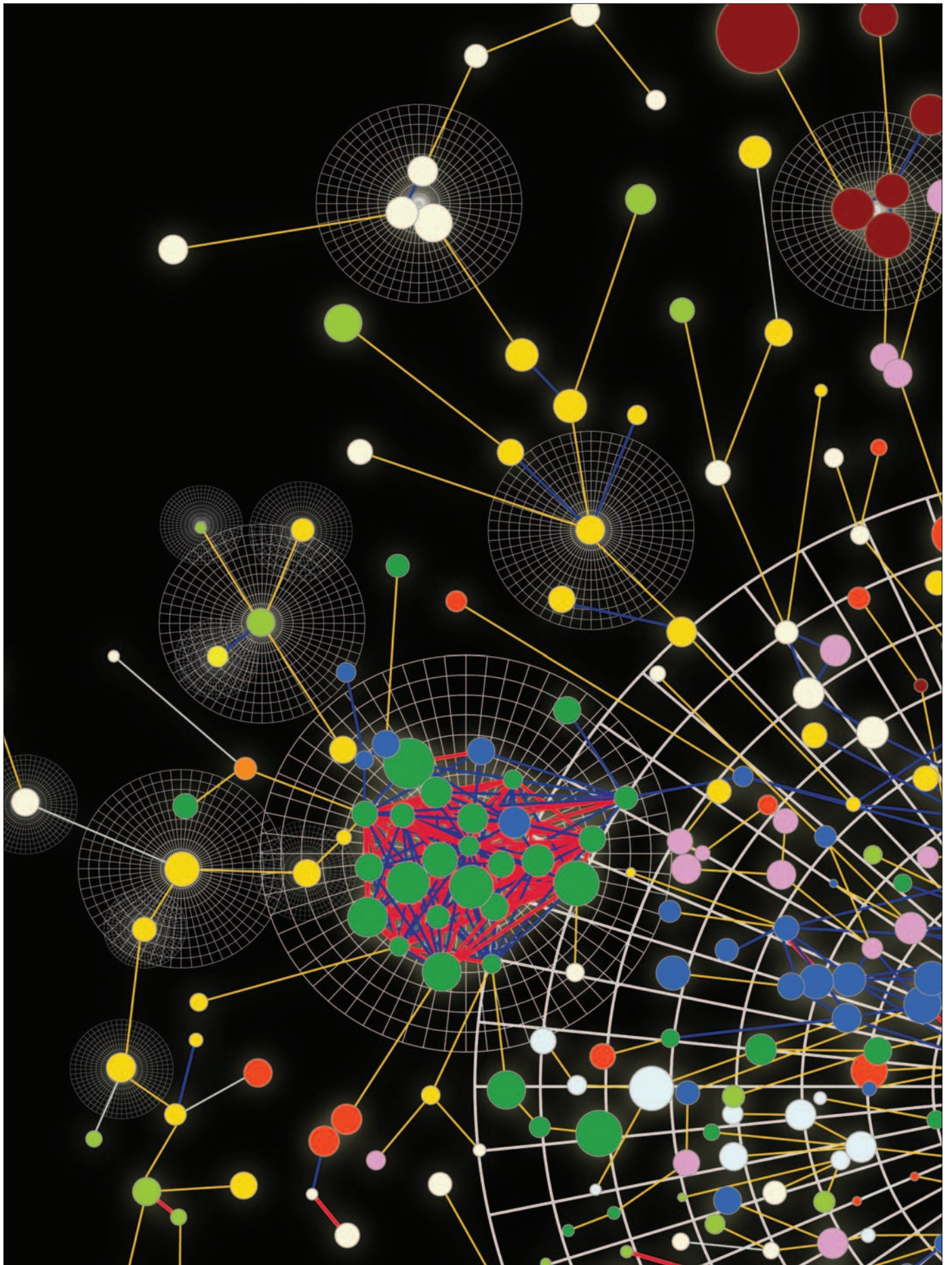
## Production links

A couple of years ago I began collaborating with a team of economists from the Harvard Kennedy School to create a map, called the Product Space, which links tradeable products (such as fruit, vehicles, chemicals, electronics and machinery) that tend to be exported by the same countries (see figure 1). The distance between any two products is proportional to the probability that a country exports both of them.

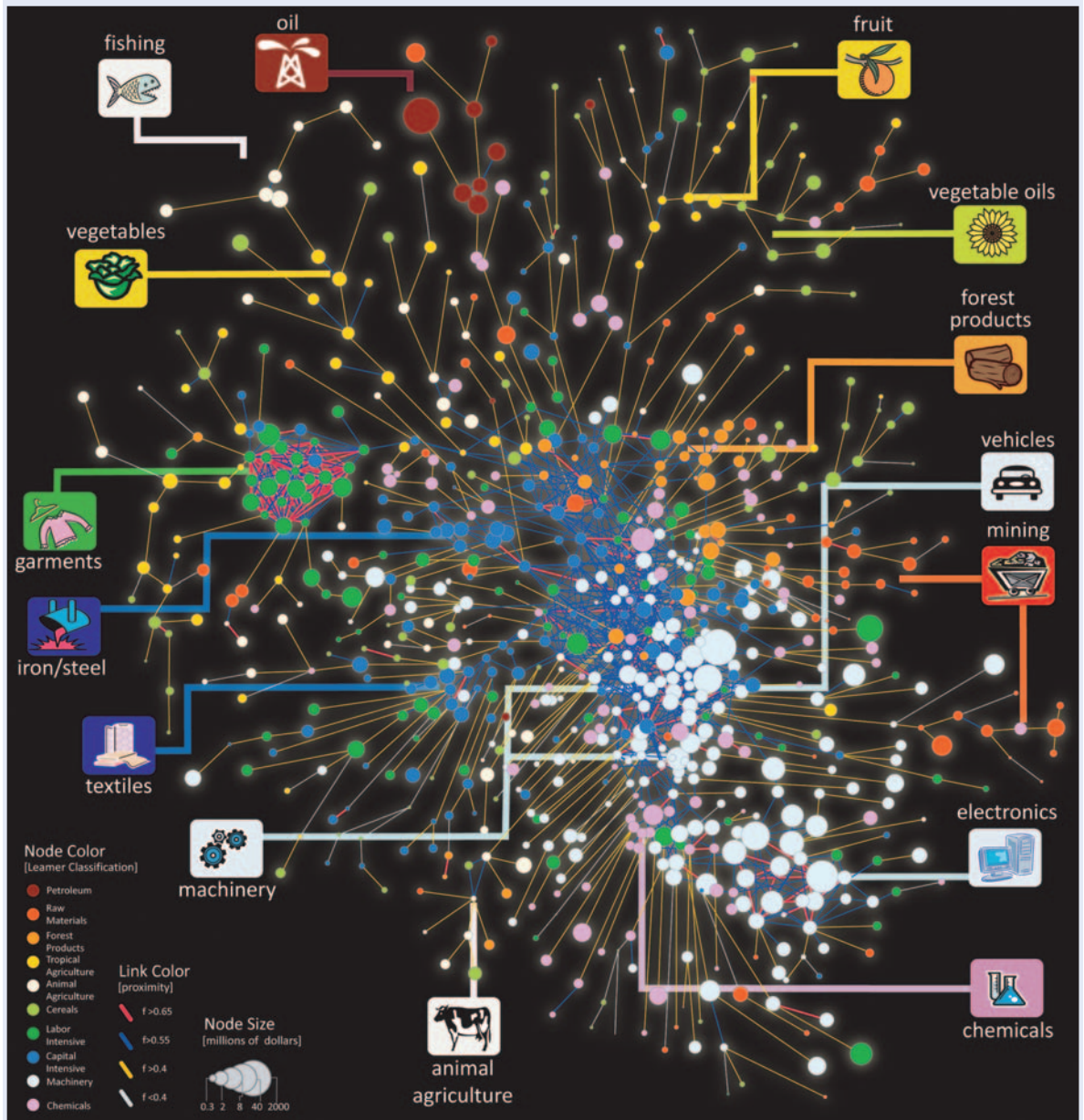
What the Product Space reveals is that some sets of products cluster very close to each other, while others are located further away in the space. This road map of development has important consequences for economic growth. For example, countries exporting products that lie mainly in the periphery of the Product Space will have serious difficulties in adapting their economy to new products that lie elsewhere in the diagram, resulting in a limited ability to upgrade their exports and grow.

The colourful and intricate patterns that appear in the Product Space not only illustrate the very strong aesthetic appeal of network visualizations, but also show that networks can pack in enormous amounts of information and can be used to simultaneously represent data in three or even more dimensions. Hence, network visualizations allow us to literally think outside the cube, as they tap into an aspect of our cognition that existing methods are unable to reach. The Product Space is also a simple way to visualize the economic development of nations and so improve the dialogue between academics, policy-makers and government officials. The Product Space allow us to understand economic development from a perspective in which products are central to, rather than absent from, the discussion. It reveals an underlying economic pattern that may have been hitherto only hinted at by economists, but not formally quantified in this way.

Good visualizations, like good theories and models, spark our imagination and help ignite the understanding of systems and their consequences



1 The Product Space



In this network visualisation every node represents a different product, such as crude oil, passenger vehicles and cotton underwear. The proximity between a pair of products is proportional to the probability of any country exporting that pair of goods, given that it exports one good or other. Node size is proportional to the revenue generated by the trade of that product worldwide. The different colours represent a predefined classification of the products. An important aspect of the Product Space, which is easy to see from this visualization, is that the density of products is not uniform across the map. This observation has important implications, as it has been shown that countries tend to diversify their economies by jumping towards products that are nearby in the Product Space. Hence, the economic opportunities for countries with products located in peripheral parts of the map differ greatly from those countries with products located in its more densely connected regions.

**The patterns of diseases**

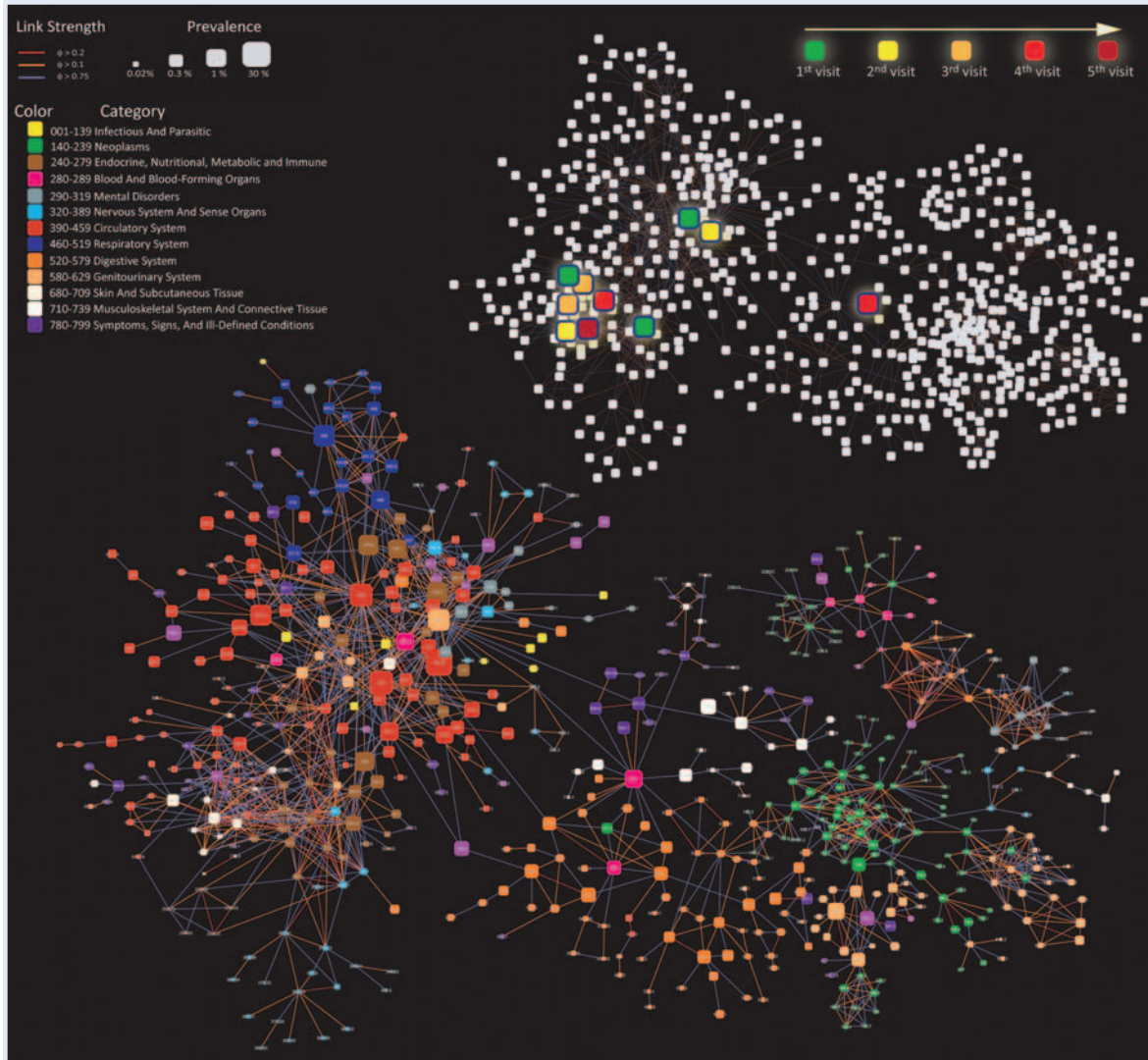
Another example of how the physics of networks can be applied to other disciplines comes from medical science. In a recent collaboration between Northeastern and Harvard universities, we used a database of medical records from a large population of elderly US citizens to build a “disease network” that links those diseases that tend to significantly affect the same individuals. Each node in the network is a particular disease, with the nodes representing hypertension and type 2 diabetes, for example, being very close to each other because the likelihood that a patient is affected

by both is very high. A visualization of the network of phenotypic-disease associations is shown in figure 2.

Most importantly, this network shows how various disease associations are distributed, with densely connected groups of conditions condensed in some regions and others being more sparsely distributed in the outer, isolated parts of the map. For instance, heart diseases and diabetes are part of a densely connected cluster of other related diseases, whereas disorders like hypothyroidism are very weakly connected in the network.

One of the most important applications of the dis-

## 2 The phenotypic-disease network



This network (bottom) consists of connections between pairs of diseases that are observed to co-occur significantly in the same individuals. Nodes represent individual diseases and the size of the nodes is proportional to their prevalence in the population. Proximity between nodes is related to the degree of co-occurrence in patients. The different colours represent different disease categories. The data are from historical health records. We can use the network to represent the disease history of a patient (top) by highlighting which diseases the patient has been diagnosed with on previous visits to the doctor (coloured squares). The use of network visualizations, like the one presented here, can allow healthcare providers to look simultaneously at several dimensions of an individual's medical record.

ease network comes from its use for individual patients. Given the disease history of a patient, we can highlight in the map those conditions that affect them now and those that have done so in the past. The example in the figure shows the disease history of a diabetic patient who suffered subsequent complications that unfortunately led to gangrene in one of his lower limbs, which then had to be amputated. The structure of the map not only improves the speed at which doctors can look and compare the disease history of patients, but can also quickly inform them about other dimensions characterizing diseases, for example how often they occur. Most importantly, the phenotypic map shows us the diseases that are most closely associated with a given patient, thereby quickly highlighting some of the most serious potential health risks.

The production of high-quality visualizations has exploded over recent years due to the abundance of mas-

sive datasets and the growing demand for people to interact with them. Network visualizations are at the heart of this informatics revolution that intertwines researchers from physics, biology, computer science, engineering, social sciences and the arts. This is because network visualizations allow researchers to explore massive amounts of data in a way whereby several dimensions can be seen simultaneously, and highlighting the structure and correlations present in these large datasets in ways that appeal to our brains.

Good visualizations, like good theories and models, spark our imagination and help ignite the understanding of systems and their consequences. As the world becomes more interconnected and data increasingly come in petabyte batches, our means of communication needs to evolve to pack in more information and be more graphically appealing. Network visualizations are opening a new way to transform data into insight. ■