

The Big Idea Behind Big Data

November 17, 2017 11:02 AM ET

[ADAM FRANK](#)



Matjaz Slanic/Getty Images/iStockphoto

In the Spring of 2009, the [H1N1/09](#) virus — dubbed "swine flu" — made the jump from pigs to people and began claiming its first victims.

Fearing the beginning of a global swine flu pandemic, terrified health officials began planning for the worst. Shutting down the world's major airports became the nuclear option of their arsenal — the last hope for halting the virus from reaching unstoppable thresholds of contagion.

That, though, was before two Italian scientists demonstrated that shuttering the air-transport system could delay the dreaded epidemic threshold by, at most, a few weeks (while also leading to economic chaos).

The researchers providing this crucial insight were not doctors but physicists. More importantly, the math they deployed was the same as that used daily by researchers at "Big Data" giants like Facebook, Google — and the NSA. It was also the same math used by scientists around the world studying the human genome, the efficiencies of green power grids, the economics of world trade and a hundred other

applications. It was the math of a new kind of science whose promise had to be weighed against the darker side of Big Data, with all its implications for surveillance and control.

It's a science called Network Theory.

For more than 400 years, science has transformed the world again and again by discovering new additions to its census of "things." The discoveries of microscopic germs, of electromagnetic fields, of genes, and of quarks each introduced us to new, previously unimagined players on the cosmic stage.

As we find our way in a world shaped by Big Data, it's not the reams of information we gather, but the networks they illuminate that are the newest addition to science's index of things. That is what makes networks the big idea behind Big Data.

But what are networks and how do they fit into our data-driven culture? A network is just an entity where the *connections* between parts matters more than the parts themselves. We all have intimate experience with networks in the web of personal relations expanding from you to your friends to their friends and so on — the social network Facebook made so explicit. But the genes on which life depends also form a network where the expression of one gene depends on the activity of others. Inter-linked food chains in ecology are networks of multiple predators feeding on multiple layers of prey. The chain of world commerce is also a network where trade represents connections between businesses and nations.

Networks are everywhere in nature and society. But before *computers* granted us the power to collect, store and analyze astronomical amounts of data — Big Data — we were blind to their pervasiveness and their power to shape the world.

The spaghetti-like maze of air-routes found in the back of your in-flight magazine gives you a visual clue to why networks (in this case the air transport network) are difficult to study. They embody a property scientists call *complexity*. When everything is connected to everything else, that's complexity — and it makes for wickedly hard problems.

Networks and their complexity defy science's penchant for reductionism, for breaking phenomena down into smaller parts. But by seeing networks in terms of their own essentials, scientists have still found a way to embrace their inherent holism. No matter how complex, every network is just a collection of nodes (i.e. dots) and connections (lines).

The power of this abstraction is something every city dweller already understands via transit maps, like those for the London Tube. Stations become dots linked by lines representing different train routes. In 1931, [Henry Beck](#) developed the first abstracted maps of the Tube network after many failed attempts to combine a representation of the subway system with the actual geography of London. Beck finally realized geography didn't matter to riders. In trying to get from Liverpool St. to Gloucester Rd., only the architecture of the links *between* stations is important.

Geography matters less than "netography." This is a fundamental lesson for all networks and the first step in recognizing them as a distinct and new kind entity for science to explore.

Kevin Bacon understands netography. In the 1990s, the game "Six-Degrees of Kevin Bacon" linked the actor to any other Hollywood player based on movies in which he appeared — and links to actors in-common from the films. The Bacon game originated from famous, early experiments with social networks showing you are connected to any other U.S. citizen by just six hops.

Explaining this "six degrees of separation" phenomenon was one of Network Theory's early victories. Using mathematics to represent nodes and connections, scientists like [Steven Strogatz](#) discovered how just a few shortcuts (like people randomly meeting on a train) allows networks to become *small worlds* (yes, that's a technical term). These shortcuts vastly reduce the number of hops it takes to go from anywhere in the network to anywhere else. In that sense "it's a small world" is not just a colloquialism. It's a universal law of these "things" called networks, just as gravity is a universal law of things called matter.

Real world networks are so complex, however, that abstracted mathematical models alone only go so far. This is where Big Data enters the story, with the rapid growth of the Internet in the 1990s. With every click between webpages, hundreds of millions of people began laying down tracks scientists could use to finally map the real behavior of complex networks. In exploring these huge data sets, researchers found astonishing new laws for how real-world networks behave. One key was the discovery of super-connected nodes they called "hubs." O'Hare airport is a super-connected hub in the air-transportation network, just as Andy Warhol was a hub in the New York City art network.

Using the Internet's first generation of Big Data, researchers like physicist [Albert-László Barabási](#) of Northeastern discovered hubs controlling the behavior of all large networks from protein regulation to webpages.

"Nature evolved the metabolic network for cells over 4 billion years," says Barabasi, "but that same architecture emerged in the World Wide Web after just a decade."

Different networks, same laws.

The essential allure of science has always been the chance to understand the world at a fundamental level. For physicists like myself, this has meant studying the world's most basic entities, like subatomic particles. But climb higher on the world's ladder of structures — from molecules to cells to societies — and the "things" become more complex and eventually impossibly difficult to describe in this manner. No one is crazy enough to try predicting the social response to an earthquake by describing the atoms in its victims. Reducing the whole to its parts has limits. Network Theory offers a different path — and that is what makes it so thrilling.

Seeing the world through the lens of network theory offers scientists a powerful top-down perspective. It holds the promise that we might find elegant, mathematical laws for domains like the behavior of the brain or the movement of society, domains that used to be off-limits to such descriptions. Network Theory promises insights of a kind that were impossible before — and already it's delivering on that promise in ways that can offer real assistance to a climate-plagued world of 7 billion plus inhabitants.

It is Network Science that is allowing [Alessandro Vespignani](#), one of the researchers who conducted the groundbreaking H1N1 study, to meet the Centers for Disease Control challenge of accurately predicting the annual sweep of flu outbreaks. Such month-by-month flu predictions, impossible in the past, can allow the CDC to better time the production and distribution of vaccines. Political scientist [David Lazer of Harvard](#) is using similar ideas to develop tools that monitor the behavior of cell phone networks in the aftermath of emergencies like the Boston Marathon bombing. By watching how the network lights up during a disaster, it may be possible to cut through the fog of confusion and pinpoint exactly where first responders should target their resources. The rapidly developing understanding of networks has also allowed biologists like [Neo Martinez](#) at University of Arizona to map the response of an entire ecosystem to the collapse of a single species. Martinez's results holds the promise of accurately managing fisheries in the face of climate change and resource depletion with an acuity that can bring order to their wicked complexity.

As a scientist, it's impossible to ignore the truly revolutionary perspective Network Theory opens on not just the natural world (the traditional domain of physicists) but also the world *we have created* — the social order and the technologies on which it depends. But as a citizen, I am deeply troubled by the darker implications inherent in Big Data resources this new science relies upon. We are gaining tools that can easily make the world not only less humane — but less human.

Balancing such dualities in a scientific revolution is not, however, new. At the turn of the 20th century, we knew nothing of atoms or their constituents. Penetrating the atomic world led us to the daily electronic marvels we now take for granted. It also allowed us to build nuclear weapons of inconceivable destructive power whose legacy still haunts us.

Network Theory is perhaps the first true, universal science to emerge from our digital revolution. It has given us a fundamentally new way to understand how the world is built. It is showing us that all life exists in and through networks, whose web of connections flash, flicker and pulse with energy and information.

As that knowledge is expanded and its implications understood, fear must be balanced with an informed discussion.