

Since Duncan Watts and Steve Strogatz published “Collective Dynamics of Small-World Networks” in Nature in 1998, there has been an explosion of interest in mathematical models of large networks, leading to numerous research papers and books. These works have given us new measures of large networks including hub nodes, broker nodes, connection path length, small-world phenomena, six degrees of separation, power laws of connectivity, and scale-free networks. The National Research Council carried out a study evaluating the emergence of a new area called “network science,” which could provide the mathematics and experimental methods for characterizing, predicting, and designing networks.

The new area has its share of controversies. An example is whether the power law distribution of number of nodes of given connectivity leads to valid conclusions for real networks. The power law distribution predicts that the network will have hubs—a few nodes with high connectivity—and has led to the claim that such networks are vulnerable to attacks against the hubs. The Internet has been reported to follow power law connectivity but its design resists hub failures. How might we explain this anomaly?

David Alderson has become a leading advocate for formulating the foundations of network science so that its predictions can be applied to real networks. He is an assistant professor in the Operations Research Department at the Naval Postgraduate School in Monterey, Calif., where he conducts research with military officer-students on the operation, attack, and defense of network infrastructure systems. We interviewed him to find out what is going on.

Peter Denning
Editor

In Search of the Real Network Science: An Interview with David Alderson

UBIQUITY: Most people think a network is a bunch of nodes and connections. Is that how you see them?

DAVID ALDERSON: The nodes-and-connections idea comes from the mathematical definition of a *graph*, which we typically define as a set of *vertices* (also called *nodes*) and *edges* (also called *arcs* or *links*). In the three centuries since the time of Leonhard Euler, the study of graphs has been an important part of mathematics.

UBIQUITY: Can graphs represent networks?

ALDERSON: I think it's important to distinguish network from graph. A *network* consists of a graph *plus additional data* interpreting the nodes and arcs. These data are typically domain-specific and often critical to the definition of the system and its function. For example, a graph representing the connectivity of a communication network tells only part of the story—one would need additional data such as the bandwidths of the links and the queuing capacities of the nodes if one wanted to understand even the simplest behavior of the system. While a graph is always present in a network, it is not enough to define a network.

These differences allow us to distinguish an electric power grid different from a gene regulatory system or a social network. The network is a complete system. The connectivity depicted in the graph is not enough to tell us what the system does or to help us predict how the system will behave in the future.

UBIQUITY: Graphs might explain why mathematicians are interested in networks ... but why scientists?

ALDERSON: Scientists are always interested in models or representations that help them to understand, explain, and predict the world around us. The scientific approach complements and enables the engineering and mathematical approaches for complex systems. Mathematicians give us precise language for describing the laws of systems and deriving equations about the

future behavior of those systems. Engineers design and build systems within real constraints. All three perspectives are essential for the advancement of knowledge and technology.

UBIQUITY: So might I summarize that as Science models, Mathematics proves, and Engineering designs?

ALDERSON: That's right. They all need each other. For example, the growing size and complexity of the Internet in recent years has defied the mathematicians and engineers to predict its growth and evolution. Scientists have flocked to the study of the Internet and are trying to provide models that mathematicians make precise and engineers use for design. I am a scientist with an engineering background, so it appeals to me on more than one level.

UBIQUITY: Was there a "defining event" that told you to devote yourself to the study and understanding of networks?

ALDERSON: In the late 1990s as a graduate student, I had the opportunity to participate in workshops supporting the U.S. Presidential Commission on Critical Infrastructure Protection (PCCIP). This group included leaders from industry, academia, and government who were concerned that the growing interconnectivity of our critical infrastructures, and in particular their dependence on the Internet, could lead to new national security vulnerabilities. Throughout these meetings, there was a general recognition that we had insufficient scientific understanding of these systems.

UBIQUITY: Is that what motivated you to study critical infrastructure networks?

ALDERSON: It certainly reinforced it. Our infrastructure systems (electric power, telecommunications, transportation, etc.) are the fabric of our modern world. These ubiquitous "conveniences" are critical to our economic and social welfare. Most of these systems were developed and deployed in isolation, but they are now all interconnected in ways that we often don't appreciate until something goes wrong. I worry about a large-scale disruption caused by accidental failure, natural disaster, or intentional attack.

UBIQUITY: Isn't network science much bigger than a study of infrastructure systems?

ALDERSON: It sure is. We live in an era of Google, Facebook, and Twitter, a time of explosive connectivity. Massive data sets are now being integrated. Footprints of social relationships are everywhere. And we are finding networks in the structures and behaviors of living systems, from the smallest genetic systems to the largest ecosystems. Scientists are responding to the call to understand these extremely large systems.

UBIQUITY: And what have network scientists accomplished so far?

ALDERSON: Scientists look for recurring patterns that can be observed in the structure or behavior of systems and then derive and validate models based on those recurrences. Since all networks have an underlying graph structure, a natural starting point was to focus on observed patterns in graph connectivity and models that attempt to explain them. In the last decade, there have been a plethora of scientific and popular publications on this topic.

UBIQUITY: The term "power law" comes up a lot in discussions about networks. Why?

ALDERSON: Power laws in connectivity were one of the patterns observed repeatedly in recent empirical studies of networks of all kinds. In power law distributions, most nodes have very few connections while a few have orders of magnitude more.

UBIQUITY: Why all the fuss about power laws?

ALDERSON: Power laws are natural and ubiquitous, showing up in everything ranging from income distributions to earthquake magnitudes. Power laws had been studied in detail in the 1950s and 1960s and were rediscovered in networks during the last decade. In physics, power laws are closely associated with phase transitions, which are rather exotic phenomena. So some researchers interpreted the ubiquity of power laws in networks as evidence of a broader, universal force at work in complex systems. The study of power laws in large networks became a cottage industry.

UBIQUITY: Power laws are special!

ALDERSON: Actually, they aren't special at all. They can arise as natural consequences of aggregation of high variance data. You know from statistics that the Central Limit Theorem says distributions of data with limited variability tend to follow the Normal (bell-shaped, or Gaussian) curve. There is a less well-known version of the theorem that shows aggregation of high (or infinite) variance data leads to power laws. Thus, the bell curve is normal for low-variance data and the power law curve is normal for high-variance data. In many cases, I don't think anything deeper than that is going on.

In fact, lots of mechanisms can produce power laws, so the presence of a power law by itself does not imply anything about the process that led to it. And remember that we are talking about connectivity only (i.e., the graph) and not the full network as a system, so power laws provide only a crude description from the outset. I think power laws have been a big distraction.

UBIQUITY: You have been a critic of some of the contemporary network science research. Is this research leading us to false conclusions about designing resilient and dependable networks?

ALDERSON: Some of the early work in network science suggested that power laws yield rules of thumb such as "protect the parts with the most connections." In a graph, removal of highly connected nodes can break the graph into isolated pieces. We would not want that for the Internet or any critical infrastructure. In the real Internet, Google has a huge number of connections ... but if for some reason Google failed, the Internet would still function. The "pure graph" interpretation of the Internet leads to a false conclusion of fragility. There are loads of similar examples. That rule of thumb is misleading and potentially dangerous. If I were applying it to critical infrastructure, I would personally worry that I might be putting my limited protective resources in the wrong place.

UBIQUITY: Why? What's the problem?

ALDERSON: The basic conceptual problem is a failure to distinguish between graph and network. The graph may have nodes with lots of connections, often called "hubs," but the network may be designed so that the failure of hubs is not an issue. The backbone of the Internet is designed by engineers in a mesh structure that guarantees many redundant paths in case of a node or link failure. If there is a single point of failure, it typically results in a disruption to local connectivity only and does not affect the Internet as a whole.

UBIQUITY: But aren't these scientific studies based on connectivity data for real networks?

ALDERSON: Yes, but how one defines a "connection" in a network may not be unique. For example, much of the work in network science about the Internet is based on "Traceroute data." Traceroute is an Internet program that sends probe packets and reports a list of the Internet Protocol (IP) addresses that they visit. Addresses adjacent in the list need not have a physical connection between them. The network defined by IP connectivity is a virtual network. Therefore, the connectivity observed by Traceroute gives us limited, and sometimes misleading, information about the physical structure of connections among routers. What appears to be a hub in the traceroute graph may not be a hub in the real network.

The same thing is true with data about connections among web pages. It appears that Google (for example) is a huge hub. But in reality, Google has implemented a worldwide “cloud” of servers that present the Google interface to users. The “cloud” is a carefully designed, highly redundant, fail-safe network of servers. It is only an illusion that Google is a single hub. For the same reason, some reports of power law connectivity patterns in the Internet were also an illusion.

UBIQUITY: The “cloud” architecture clouds our understanding of what is actually vulnerable?

ALDERSON: Correct. Almost every large Internet service is implemented with a “cloud.” A lot of connections within the cloud, which are measured and recorded in graphs, are virtual, not real. The graph we are measuring bears no resemblance to the network system behind it. This should not surprise us. It was an explicit objective of the Internet’s architecture to be able to support a diversity of virtual topologies independent of their physical implementation.

UBIQUITY: So most of the connectivity we see from data is virtual and does not tell us much about the physical connectivity and its vulnerabilities?

ALDERSON: Exactly. A simple graph representation of that connectivity is particularly misleading because it typically omits most of the architectural features governing the behavior of the system.

UBIQUITY: And the architecture of a network is more than its connectivity?

ALDERSON: Absolutely. The architecture of the Internet consists of rules, implemented as hardware and software protocols, that define a control system for communication. What makes the Internet robust or vulnerable really comes from those protocols, not from its connectivity. Most of the “big problems” facing the Internet—such as email spam, viruses and worms, denial of service attacks, etc.—come from hijacking either these protocols or other mechanisms that make the Internet work in the first place. Connectivity patterns of the network do not play a role.

UBIQUITY: How can such failures of understanding be so widespread?

ALDERSON: Because many researchers do not know about the quality of the data they use. It sounds crazy to say it that way, but there it is. And the Internet has exacerbated the problem by making it so easy to share data.

UBIQUITY: The Internet has made data sharing worse?

ALDERSON: Let me explain. Getting good data is hard work and requires carefully designed and administered experiments. Scientists that gather data put them up on the Internet for others to use. One data set can attract many researchers who do not want to do the data collection themselves. If more people had an appreciation for the distinction between a graph and a network, and if it were easier to assess the idiosyncrasies and limitations of an individual data set, then others might be more careful when using them.

UBIQUITY: Some of those data sets are pretty huge, no?

ALDERSON: Huge doesn’t even begin to describe their size. Many of the phenomena under study involve petabytes (10 to the 15 power) or more. It’s really hard to collect such data and then to use them wisely for solid scientific conclusions. The current academic enterprise values the analysis of data far more than its collection and maintenance. And yet, without detailed “meta information” about how the data were collected and why, determining appropriate use will be difficult. The meta information will connect the data to the network and distinguish it from the graph.

UBIQUITY: What are some of the big questions on the research agenda of network science?

ALDERSON: There's still a lot of work to do to connect the science with the mathematics and engineering, and that's happening slowly. One largely unaddressed topic is how to model a network that includes a mix of automated and human processes in its control loop. A second big topic is how to model a network that needs to take action urgently.

UBIQUITY: What's on your research agenda?

ALDERSON: I want to know how should we design, build, and manage complex systems to avoid "rare, yet catastrophic" failures. With critical infrastructures, we need designs that make them more resilient and less susceptible to disruption, both accidental and intentional. I want to make sure we are investing our resources wisely.