TWO FORCES THAT HELPED NETWORK SCIENCE

Network science is a new discipline. One may debate its precise beginning, but by all accounts the field has emerged as a separate discipline only in the 21st century.

Why didn't we have network science two hundred years earlier? After all many of the networks that the field explores are by no means new: metabolic networks date back to the origins of life, with a history of four billion years, and the social network is as old as humanity. Furthermore, many disciplines, from biochemistry to sociology and brain science, have been dealing with their own networks for decades. Graph theory, a prolific subfield of mathematics, has explored graphs since 1735. Is there a reason, therefore, to call network science *the science of the 21st century*?

Something special happened at the dawn of the 21st century that transcended individual research fields and catalyzed the emergence of a new discipline (Figure 1.3). To understand why this happened now and not two hundred years earlier, we need to discuss the two forces that have contributed to the emergence of network science.

THE EMERGENCE OF NETWORK MAPS

To describe the detailed behavior of a system consisting of hundreds to billions of interacting components, we need a map of the system's wiring diagram. In a social system this would require an accurate list of your friends, your friends' friends, and so on. In the WWW this map tells us which webpages link to each other. In the cell the map corresponds to a detailed list of binding interactions and chemical reactions involving genes, proteins, and metabolites.

In the past, we lacked the tools to map these networks. It was equally difficult to keep track of the huge amount of data behind them. The Internet revolution, offering effective and fast data sharing methods and cheap digital storage, fundamentally changed our ability to collect, assemble, share, and analyze data pertaining to real networks.

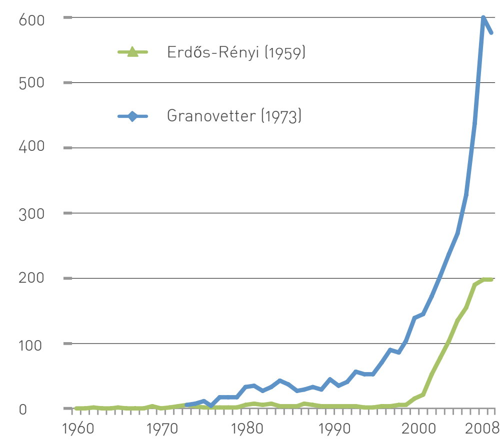

Figure 1.3
The Emergence of Network Science

While the study of networks has a long history, with roots in graph theory and sociology, the modern chapter of network science emerged only during the first decade of the 21st century.

The explosive interest in networks is well documented by the citation pattern of two classic papers, the 1959 paper by Paul Erdős and Alfréd Rényi that marks the beginning of the study of random networks in graph theory [2] and the 1973 paper by Mark Granovetter, the most cited social network paper [3]. The figure shows the yearly citations each paper acquired since their publication. Both papers were highly regarded within their discipline, but had only limited impact outside their field. The explosive growth of citations to these papers in the 21st century is a consequence of the emergence of network science, drawing a new, interdisciplinary attention to these classic publications.

Thanks to these technological advances, at the turn of the millenium we witnessed an explosion of map making (BOX 1.2). Examples range from the CAIDA or DIMES projects that offered the first large-scale maps of the Internet; to the hundreds of millions of dollars spent by biologists to experimentally map out protein-protein interactions in human cells; the efforts made by social network companies, like Facebook, Twitter, or LinkedIn, to develop accurate depositories of our friendships and professional ties; the Connectome project of the US National Institute of Health that aims to systematically trace the neural connections in mammalian brains. The sudden availability of these maps at the end of the 20th century has catalyzed the emergence of network science.

THE UNIVERSALITY OF NETWORK CHARACTERISTICS

It is easy to list the differences between the various networks we encounter in nature or society: the nodes of the metabolic network are tiny molecules and the links are chemical reactions governed by the laws of chemistry and quantum mechanics; the nodes of the WWW are web documents and the links are URLs guaranteed by computer algorithms; the nodes of the social network are individuals and the links represent family, professional, friendship, and acquaintance ties.

The processes that generated these networks also differ greatly: metabolic networks were shaped by billions of years of evolution; the WWW is built by the collective actions of millions of individuals and organizations; social networks are shaped by social norms whose roots go back thousands of years. Given this diversity in size, nature, scope, history, and evolution, one would not be surprised if the networks behind these systems would differ greatly.

A key discovery of network science is that *the architecture of networks emerging in various domains of science, nature, and technology are similar to each other, a consequence of being governed by the same organizing principles. Consequently we can use a common set of mathematical tools to explore these systems*.

This universality is one of the guiding principle of this book: we will not only seek to uncover specific network properties, but each time we ask how widely they apply. We will also aim to understand their origins, uncovering the laws that shape network evolution and their consequences on network behavior.

In summary, while many disciplines have made the important contributions to network science, the emergence of a new field was partly made possible by data availability, offering accurate maps of networks encountered in different disciplines. These diverse maps allowed network scientists to identify the universal properties of various network characteristics. This universality offers the foundation of the new discipline of network science.

BOX 1.2

THE ORIGINS OF NETWORK MAPS

A few of the maps studied today by network scientists were generated with the purpose of studying networks. Most are the byproduct of other projects and morphed into maps only in the hands of network scientists.

(a) The list of chemical reactions in a cell were discovered one-by-one over a 150 year period by biochemists. In the 1990s they were collected in central databases, offering the first chance to assemble the biochemical networks within a cell.

(b) The list of actors that play in each movie were traditionally scattered in newspapers, books and encyclopedias. With the advent of the Internet, these data were assembled into central databases, like imdb.com, feeding the curiosity of movie aficionados. The database allowed network scientists to reconstruct the affiliation network behind Hollywood.

(c) The list of authors of millions of research papers were traditionally scattered in the table of content of thousands of journals. Recently Web of Science, Google Scholar, and other services have assembled them into comprehensive databases, allowing network scientists to reconstruct accurate maps of scientific collaboration networks.

Much of the early history of network science relied on the investigators' ingenuity to recognize and extract networks from preexisting databases. Network science changed that: Today well-funded research collaborations focus on map making, capturing accurate wiring diagrams of biological, communication and social systems.

THE CHARACTERISTICS OF NETWORK SCIENCE

Network science is defined not only by its subject matter, but also by its methodology. In this section we discuss the key characteristics of the approach network science adopted to understand complex systems.

INTERDISCIPLINARY NATURE

Network science offers a language through which different disciplines can seamlessly interact with each other. Indeed, cell biologists, brain scientists (Figure 1.4) and computer scientists alike are faced with the task of characterizing the wiring diagram behind their system, extracting information from incomplete and noisy datasets, and understanding their systems' robustness to failures or attacks.

To be sure, each discipline brings a different set of goals, technical details and challenges, which are important on their own. Yet, the common nature of many issues these fields struggle with has led to a cross-disciplinary fertilization of tools and ideas. For example, the concept of betweenness centrality that emerged in the social network literature in the 1970s, today plays a key role in identifying high traffic nodes on the Internet. Similarly algorithms developed by computer scientists for graph partitioning have found novel applications in identifying disease modules in medicine or detecting communities within large social networks.

EMPIRICAL, DATA DRIVEN NATURE

Several key concepts of network science have their roots in graph theory, a fertile field of mathematics. What distinguishes network science from graph theory is its empirical nature, i.e. its focus on data, function and utility. As we will see in the coming chapters, in network science we are never satisfied with developing abstract mathematical tools to describe a certain network property. Each tool we develop is tested on real data and its value is judged by the insights it offers about a system's properties and behavior.

QUANTITATIVE AND MATHEMATICAL NATURE

To contribute to the development of network science and to properly use its tools, it is essential to master the mathematical formalism behind

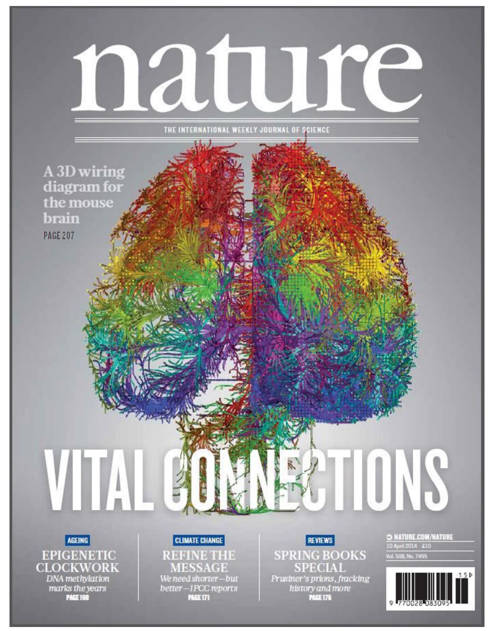

Figure 1.4
Mapping the Brain

An exploding application area for network science is brain research. The wiring diagram of a complete nervous system has long been available for C. elegans, a small roundworm, but neuronal connectivity data for larger animals has been missing until recently. That is changing thanks to major efforts by the scientific community to develop technologies that can map out the brain's wiring diagram. The image shows the cover of the April 10, 2014 issue of *Nature*, reporting an extensive map of the laboratory mouse [4] generated by researchers at the Allen Institute in Seattle.

it. Network science borrowed the formalism to deal with graphs from graph theory and the conceptual framework to deal with randomness and seek universal organizing principles from statistical physics. Lately, the field is benefiting from concepts borrowed from engineering, like control and information theory, allowing us to understand the control principles of networks, and from statistics, helping us extract information from incomplete and noisy datasets.

The development of network analysis software has made the tools of network science available to a wider community, even those who may not be familiar with the intellectual foundations and the full mathematical depths of the discipline. Yet, to further the field and to efficiently use its tools, we neet to master its theoretical formalism.

COMPUTATIONAL NATURE

Given the size of many of the networks of practical interest, and the exceptional amount of auxiliary data behind them, network scientists are regularly confronted by a series of formidable computational challenges. Hence, the field has a strong computational character, actively borrowing from algorithms, database management and data mining. A series of software tools are available to address these computational problems, enabling practitioners with diverse computational skills to analyze the networks of interest to them.

In summary, a mastery of network science requires familiarity with each of these aspects of the field. It is their combination that offers the multi-faceted tools and perspectives necessary to understand the properties of real networks.