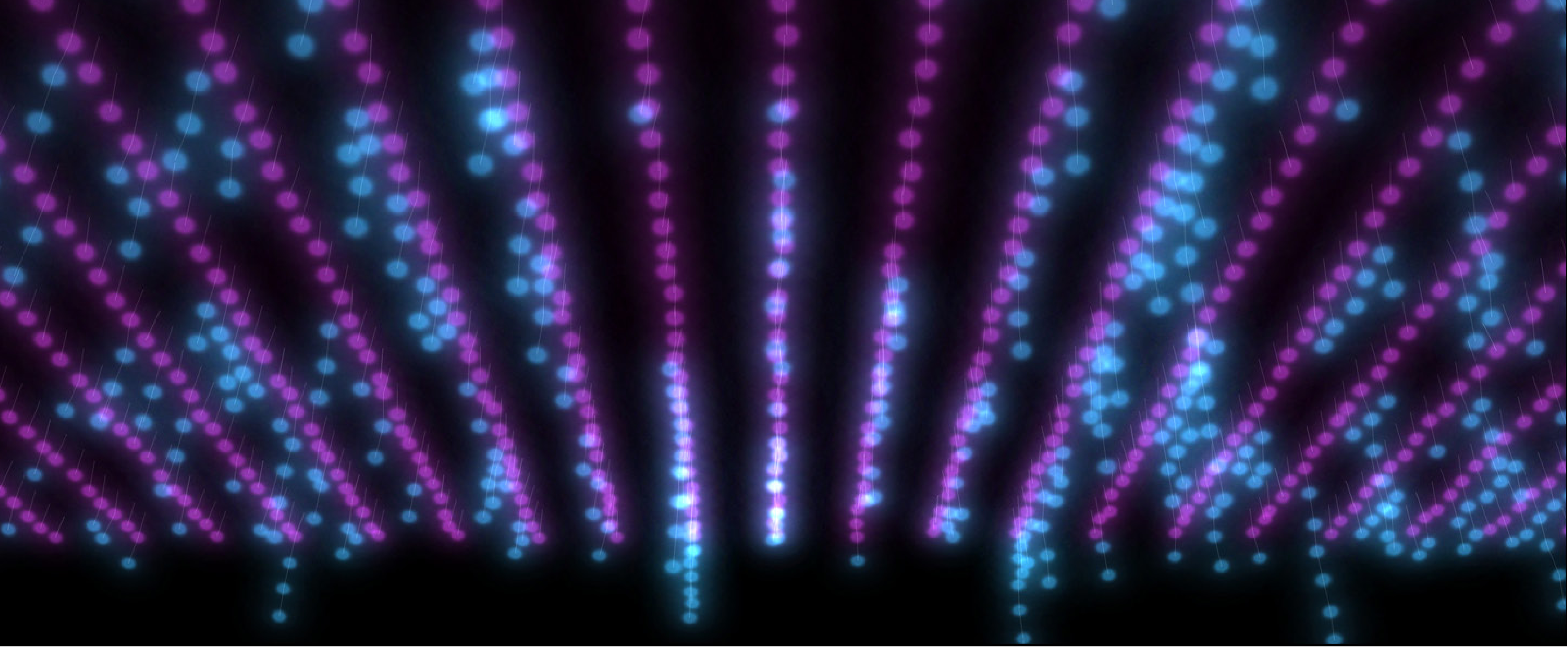
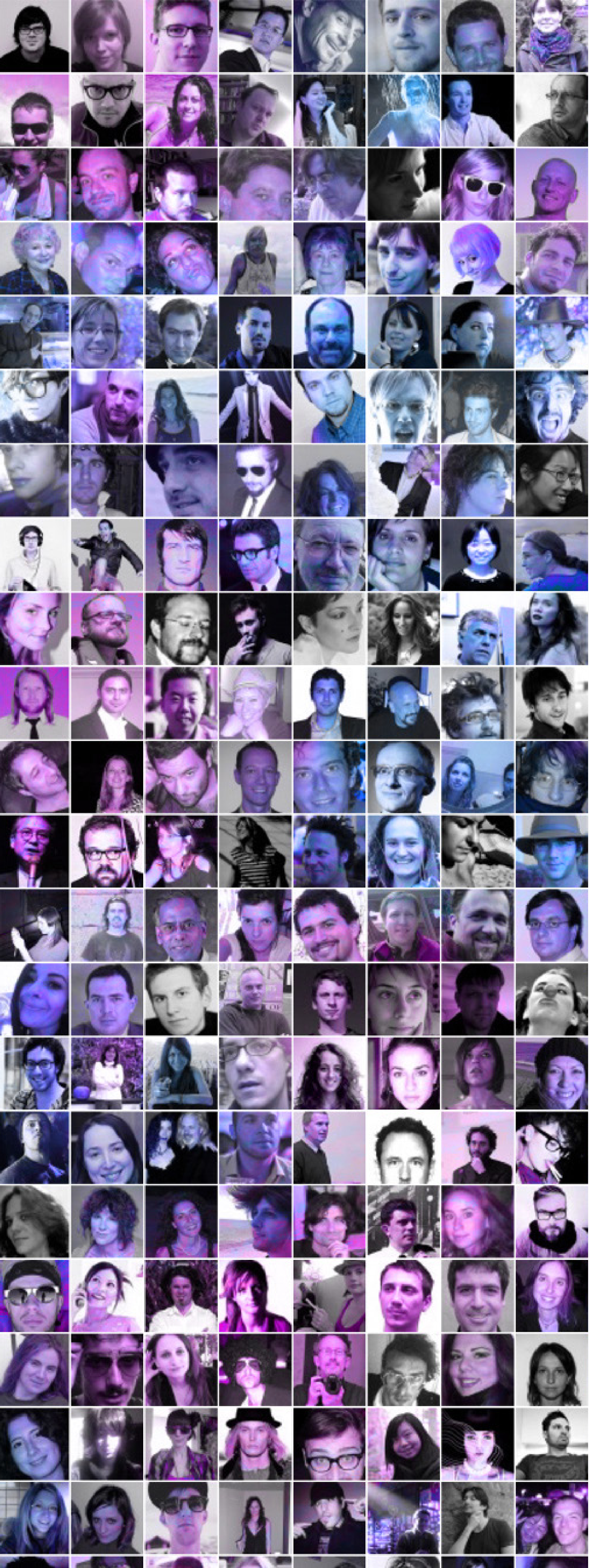


Network Science

Albert-László Barabási

Data Visualization by **Mauro Martino**
Data Analysis by **Márton Pósfai**

PDF Version:
November 2012



CHAPTER 1
INTRODUCTION

INTRODUCTION

This book aims to help teach network science to an interdisciplinary audience. Many of the choices I made in presenting the material were guided by the desire to offer an enjoyable, yet systematic introduction to the field. I kept in mind that those entering the field are just as interested in learning about the genesis of the concepts network science introduced, as the tools they can use to study real networks and interpret the obtained results.

Several over-arching themes are present in this book, helping to offer an effective introduction:

(i) Given the empirical roots of network science, there is strong emphasis on empirical data. We have therefore assembled a set of 'canonic' databases, representing networks that are frequently analyzed in network science to test various network characteristics. Whenever possible, we use these datasets to illustrate the tools we introduce.

(ii) Given the potential diversity of the students interested in the field that may be familiar with one domain of inquiry but not other, we devote special sections to each dataset. The goal is to offer some degree of familiarity with the range of datasets explored in network science, and through this diversity to learn about the issues pertaining to data collection and curation.

This book is not a finished product but a work in progress. Hence we continue to update it, adding additional chapters as they are finished.

There is a dedicated website to this project (Image 1.1),

http://barabasilab.com/networksciencebook

that contains not only the chapters, but also the slides I used in my classes to teach the material. Those who are interested in teaching any part of the book are welcome to use these slides. The website also offers tools to provide feedback on the material, from comments to suggestions for improvement.

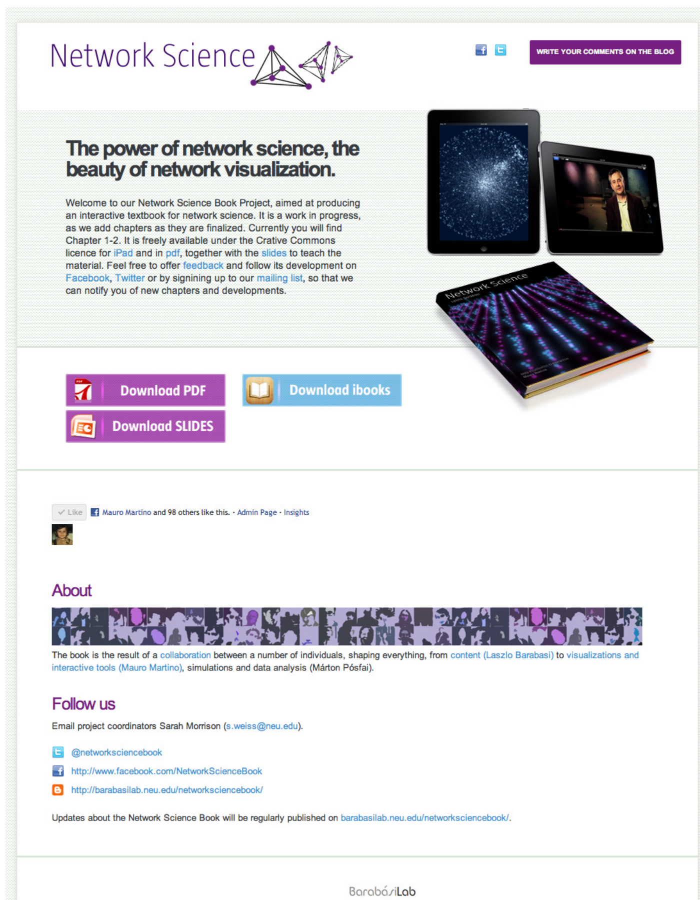


Image 1.1 http://barabasilab.com/networksciencebook

SECTION 2

FROM SADDAM HUSSEIN TO NETWORK THEORY

American forces encountered relatively little military resistance as they took control of Iraq during the invasion that started in March 19, 2003. Yet, many of the regime's high ranking officials, including Saddam Hussein, avoided capture.
Hussein was last spotted kissing a baby in Baghdad some time in April 2003, and then his trace went cold. To aid awareness of the officials they sought, the coalition forces designed a deck of cards, each card engraved with the image of one of the 55 most wanted. It worked. By May 1st 15 men on the cards were captured and by the end of the month another 12 were under custody. Yet, the ace of spades (Image 1.2a), i.e. Hussein himself, remained at large.

Intelligence officials hoped that some of the high ranking officials would surely know Hussein's whereabouts. Yet, it was not to be. This became painfully obvious after the capture of Saddam's trusted personal secretary and the ace of diamonds. Newspapers trumpeted his mid-June capture as the war's biggest feat, as this could lead to Saddam's whereabouts. Yet, the dictator parted ways with his ally soon after the invasion, sending a clear signal to the investigators: relying on the traditional lines of power was of little help in trying to find him. Instead, they decided to turn to a tool that had little presence in military thinking before: network theory [1].

In 2003 network theory was an already burgeoning research field, but the soldiers in the war zone had little access to the exploding advances in this area. Instead, they arrived to it through a healthy dose of common sense and intuition. Col. James Hickey, in charge of a series of raids known as *Operation Desert Scorpion*, wanted to know the relationship between everyone killed or captured. The task fell to Lt. Col. Steve Russell, who was in direct charge of the raids, and Brian Reed, the operations officer under Hickey, who was exposed to social networks during his studies at West Point. Reed started to systematically reconstruct the social network of Saddam's inner circle. He did not rely on government documents and decrees, but rather gossip and family trees. As they meticulously pieced together an extensive diagram of who is related to whom in the Tikrit region, where Saddam was from, they started to use network diagrams to guide the raids. In one of those raids they found over $8 million in US currency, about $1 million in Iraqi currency, jewelry worth over $2 million, rifles, and ammunition. Yet, the biggest prize was Saddam's family photo album, providing the faces of those that the family

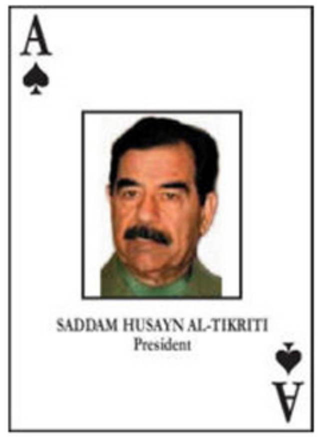


Image 1.2a
The network of Saddam Hussein.

Ace of Spades. One of the 55 cards the US military has handed out to the coalition forces in Iraq, each listing a top official to be captured following the country's 2003 invasion. The card shows the ace of spades, with the image of Saddam Hussein, Iraq's deposed president and dictator, the top prize of the hunt.

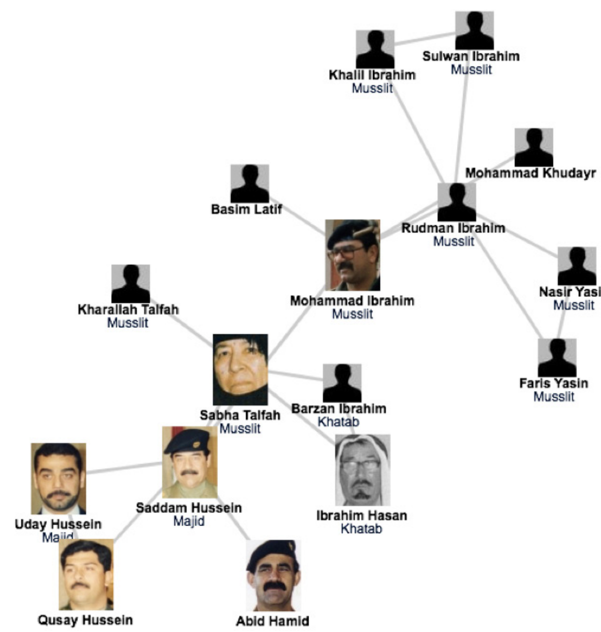


Image 1.2b
The network of Saddam Hussein.

The Social Network. A small region of the social network reconstructed by the US forces in the process of searching for Saddam Hussein. The map represents the relationship between individuals in Saddam's inner circle.

trusted, filling with intimate details of their growing network diagram.

The maps consistently pointed to two individuals, Rudman Ibrahim and Mohammed Ibrahim (Image 1.2b). Not high in the government hierarchy, they were Saddam's second-level bodyguards, serving as his driver, cook, or mechanic. Yet, Rudman had a heart attack and died within a few hours of his capture, without having a chance to reveal his secrets. Next the investigators turned to their network diagram to identify individuals who could know the whereabouts of Mohammad, dubbed the fat man. He was not a major player in the regime's power structure, hence while Saddam's whereabouts were handled with fear, Mohammed's social ties were not as protected. Sure enough, once they found someone to turn Mohammad Ibrahim in, he revealed the spider hole that hid the dictator at a farm near the Tigris river. The capture of Saddam Hussein illustrates many issues that we will encounter as we delve into network theory:

• It shows the predictive power of networks, allowing even non experts to extract crucial information from them, as the soldiers did using Saddam's social network.

• It underlines the need for accurate maps of the networks we study, and the often heroic difficulties encountered during the mapping process.

• It demonstrates the remarkable stability of these networks: the capture of Hussein was not based on fresh intelligence, but rather on his pre-invasion social links, unearthed from old photos stacked in his family album.

• It shows that the choice of network we focus on makes a huge difference: it took months for the military to realize that the hierarchical network that described the official organization of the Iraqi government was of no use when it came to Saddam Hussein's whereabouts.

In many ways the network building exercise by the US military, deployed to capture Saddam Hussein, was a primitive one driven more by intuition and guesswork than hard science. The purpose of this book is to turn these insights into a robust theory and methodology, so that we can fully and repeatedly unleash their predictive power.

SECTION 3

VULNERABILITY DUE TO INTERCONNECTIVITY

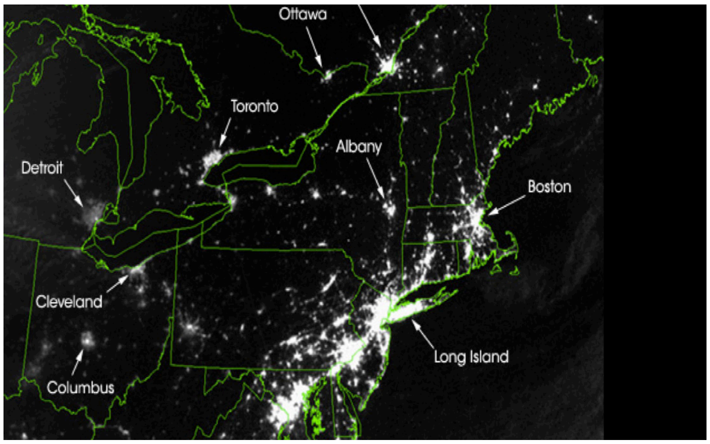
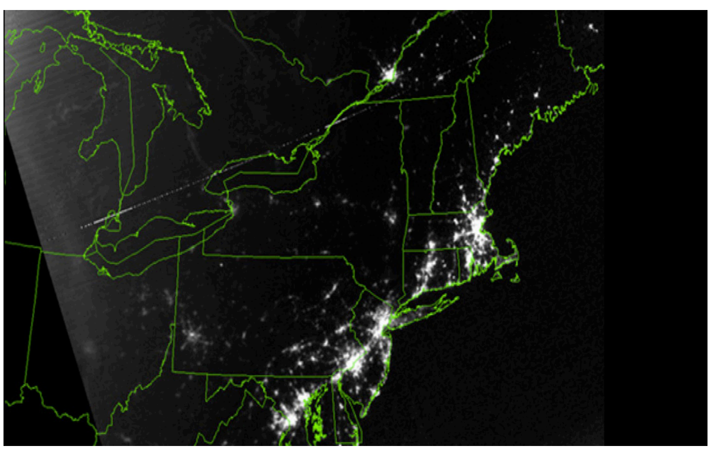


Image 1.3a, 1.3b

2003 North American blackout.

Uper Panel
Satellite image of August 13, 2003: 9:29pm EDT 20 hours before.
Lower Panel
Satellite image of August 14, 2003: 9:14pm EDT 5 hours after.



At a first look the two satellite maps of Image 1.3a/b are indistinguishable: lights shining brightly in highly populated areas, and dark spaces marking vast uninhabited forests and oceans. Yet, upon closer inspection something strange becomes apparent. The light in several regions, Toronto, Detroit, Cleveland, Columbus, Long Island have simply disappeared. This is not a doctored shot from the next Armageddon movie but represents a real image of the US Northeast on August 14, 2003, the night of a blackout that left an estimated 45 million people in eight US states and another 10 million in Ontario without power. It illustrates a much ignored aspect of networks, one that will be an important theme in this book: vulnerability due to interconnectivity.

The 2003 blackout is a typical example of a cascading failure. When a network acts as a transportation system, a local failure shifts loads to other nodes. If the extra load is negligible, the rest of the system can seamlessly absorb it, and the failure remains effectively unnoticed. If the extra load is too much for the neighboring nodes to carry, they will either tip or redistribute the load to their neighbors. Either way, we are faced with a cascading failure, the magnitude of which depends on the network position and capacity of the nodes that have been removed in the first and subsequent rounds. Case in point is electricity: as it cannot be stored, when a line goes down, its power must be shifted to other lines. Most of the time, the neighboring lines have no difficulty carrying the extra load. If they do, they will also tip and redistribute their increased load to their neighbors.

Cascading failures can occur in most complex systems. They take place on the Internet, when traffic is rerouted to bypass malfunctioning routers, occasionally creating denial of service attacks on routers that do not have the capacity to handle extra traffic. We witnessed one in 1997, when the International Monetary Fund pressured the central banks of several Pacific nations to limit their credit. There was a cascading failure behind the 2009-2011 financial meltdown, when the US credit crisis paralyzed the economy of the globe, leaving behind scores of failed banks, corporations, and even bankrupt states. Cascading failures are occasionally our ally, however. The world wide effort to dry up the money supply of terrorist organizations is aimed at crippling terrorist networks, and doctors and researchers hope to induce cascading failures to kill cancer cells.
The Northeast blackout illustrates an important theme of this book: we must understand how the network structure affects the robustness of a complex system. We will therefore develop quantitative tools to assess the interplay between network structure and dynamical processes on networks and their impact on failures. Although such failures may appear chaotic and unpredictable, we will learn that they follow rather reproducible laws that can be quantified and even predicted using the tools of network science.

SECTION 4

NETWORKS AT THE HEART OF COMPLEX SYSTEMS

"I think the next century will be the century of complexity."

Stephen Hawking

We are surrounded by systems that are hopelessly complicated, from the society, whose seamless functioning requires cooperation between billions of individuals, to communications infrastructures that integrate billions of cell phones with computers and satellites. Our ability to reason and comprehend the world around us is guaranteed by the coherent activity of billions of neurons in our brain. Our very existence is rooted in seamless interactions between thousands of genes and metabolites within our cells. These systems are collectively called complex systems. Given the important role they play in our life, in science and economy, the understanding, mathematical description, prediction, and eventually the control of such complex systems is one of the major intellectual and scientific challenges of the 21*st* century.

The emergence of network theory, at the dawn of the 21st century is a vivid demonstration that science can live up to this challenge. Indeed, *behind each complex system, there is an intricate network that encodes the interactions between the system's components*:

- The network describing the interactions between genes, proteins, and metabolites integrates the processes behind living cells.
- The wiring diagram capturing the connections between neural cells holds the key to our understanding of brain functions.
- The sum of all professional, friendship, and family ties is the fabric of the society.
- The network describing which communication devices interact with each other, capturing internet connections or wireless links, is the heart of the mod-

Box 1.1

com.plex
[adj., v. kuh m-pleks, kom-pleks; n. kom-pleks]

1) composed of many interconnected parts; compound; composite: a complex highway system

2) characterized by a very complicated or involved arrangement of parts, units, etc.: complex machinery

3) so complicated or intricate as to be hard to understand or deal with: a complex problem

Source: Dictionary.com

Image 1.4
The subtle networks behind the economy.

A credit card, selected as the 99th object in the popular exhibition by the British Museum, entitled The History of the World in 100 Objects. This card is a vivid demonstration of the interconnected nature of the modern economy, creating subtle linkages that one normally does not even think of. The card was issued in the United Arab Emirates in 2009 by the Hong Kong and Shanghai Banking Corporation, commonly known HSBC, a London based bank. The card functions through protocols provided by VISA, an USA based credit association. Yet, the card adheres to Islamic banking principles, which operates in accordance with Fiqhal-Muamalat (Islamic rules of transactions), most notably eliminating interest or riba. The card is not limited to muslims in the United Arab Emirates, but it is also offered to Muslim minorities in non-Muslim countries, and is used by many non-Muslims who agree with its strict ethical guidelines.

ern communication system.

- The power grid, a network of generators and transmission lines, supplies with energy virtually all modern technology.

- Trade networks maintain our ability to exchange goods and services, being responsible for the material prosperity that an increasing fraction of the world has enjoyed since WWII (Image 1.4). They also play a key role in the spread of financial and economic crises.

Networks are at the heart of some of the most revolutionary technologies of the 21st century, empowering everything from Google to Facebook, CISCO, and Twitter. At the end, networks permeate science, technology, and nature to a much higher degree than may be evident upon a casual inspection. Consequently, it is increasingly clear that *we will never understand complex systems unless we gain a deep understanding of the networks behind them.*

The scientific explosion that network science experienced during the first decade of the 21st century is rooted in the discovery that *despite the apparent differences, the emergence and evolution of different networks is driven by a common set of fundamental laws and reproducible mechanism.* Hence despite the amazing diversity in form, size, nature, age, and scope characterizing real networks, most networks observed in nature, society, and technology are driven by common organizing principles. In other words, once we disregard the nature of the components and their interactions, the obtained networks are more similar than different from each other. In the following sections, we discuss the forces that have led to the emergence of this new research field and its impact on science, technology, and society.

SECTION 5

TWO FORCES HELPED THE EMERGENCE OF NETWORK SCIENCE

Why didn't network science emerge two hundred years earlier? The networks it explores are by no means new: metabolic networks date back to the origins of life, with a history of four billion years, and the Internet is over four decades old. Furthermore, many disciplines, from biochemistry to sociology, and brain science, have been dealing with their notion of networks. Graph theory, a prolific subfield of mathematics, has focused on networks since 1735. Why do we dare 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 (Image 1.5). To understand why this happened only now, and not two hundred years earlier, we need to discuss the forces that have contributed to the emergence of network science.

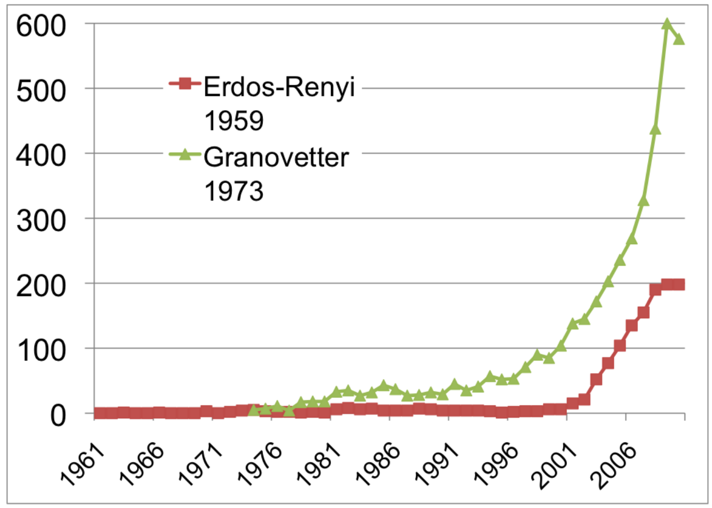


Image 1.5

The emergence of network science.

While the study of networks has a long history from graph theory to sociology, the modern chapter of network science emerged only during the first decade of the 21st century, following the publication of two seminal papers in 1998 [2] and 1999 [3]. The explosive interest in network science is well documented by the citation pattern of two classic network papers, the 1959 paper by Paul Erdős and Alfréd Rényi that marks the beginning of the study of random networks in graph theory [4] and the 1973 paper by Mark Granovetter, the most cited social network paper [5]. Both papers were hardly or only moderately cited before 2000. The explosive growth of citations to these papers in the 21st century documents the emergence of network science, drawing a new, interdisciplinary audience to these classic publications.

The emergence of network maps: To describe the behavior of a system consisting of hundreds to billions of interacting components, we first need a map of the system's wiring diagram. In a social system, this would require knowing the 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, this corresponds to a detailed list of binding interactions and reactions that the genes, proteins, and metabolites participate in. In the past, we either lacked the tools to map these networks out, or it was difficult to keep track of the huge amount of data behind these maps. The emergence of the Internet, offering effective and fast data sharing methods, together with cheap digital storage, fundamentally changed this, allows us to collect, assemble, share, and analyze data pertaining to real networks.

While many of the canonical maps studied today in network science were not collected with the purpose of studying networks (Box 2), we witnessed an explosion of map making at the end of the 1990s. These offered detailed maps of the networks behind numerous complex system, from cell to the economy. Examples include the CAIDA or DIMES project aimed at obtaining an accurate map of the Internet [8]; the hundreds of millions of dollars spent by biologists to systematically map out protein-protein interactions in human cells [6], or the Connectome project of the US National Institute of Health that aims to trace the neural connection in mammalian brains [7].

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 quantum mechanics; the nodes of the WWW are web documents and the links are URLs maintained by computer algorithms; the nodes of the social network are individuals, the links representing family, professionals, friendship, and acquaintance ties. The processes that shape these networks also differ greatly: metabolic networks are shaped by billions of years of evolution; WWW is collectively built by the actions of millions of individuals; 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. Yet, a key discovery of network science is that the architecture and the evolution of networks emerging in various domains of science, nature, and technology are rather similar to each other, allowing us to 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 we will aim to understand its origins, encoding the laws that shape network evolution, as well as its consequences in understanding network behavior.

The origins of network maps

Many of the maps studied today by network scientists were not generated with the purpose of studying networks:

- *The list of chemical reactions that take place in a cell were discovered over a 150 year period by biochemists and biologists. In the 1990s they were collected in central databases, offering the first chance to assemble the networks behind a cell.*

- *The list of actors that play in each movie were traditionally scattered in books and encyclopedias. With the advent of the Internet, these disparate data were assembled into a central database by imdb.com, mainly to feed the curiosity of movie aficionados. The database offered the first chance for network scientists to explore the structure of the affiliation network behind Hollywood.*

- *The detailed list of authors of millions of research papers were traditionally scattered in the table of content of thousands of journals, but recently the Web of Science, Google Scholar, and other sites assembled them into comprehensive databases, easing the search for scientific information.*

In the hands of network scientists these databases turned into the first science collaboration maps. Hence, much of the early history of network science relied on the investigators' ingenuity to recognize and extract the networks from existing datasets. Network science changed that: today well-funded research collaborations focus on map making from biology to the Internet.

Box 1.2

SECTION 6

THE CHARACTERISTICS OF NETWORK SCIENCE

Network science is distinguished, not only by its subject matter, but also by its methodology. In the following we briefly discuss the key characteristics of the approach network science adopted to understand complex systems, helping us better understand the domain we are about to embark on.

Interdisciplinary nature: Network science offers a language through which different disciplines can seamlessly interact with each other. Indeed, cell biologists 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 the need to understand their systems' robustness to failures or deliberate attacks. To be sure, each discipline brings along a different set of technical details and challenges, which are important on their own. Yet, the common character of the many issues various fields struggle with have 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; algorithms developed by computer scientists for graph partitioning have found novel applications in cell biology.

Empirical, data driven nature: The tools 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 and utility. As we will see in the coming chapters, we will never be satisfied with developing the abstract mathematical tools to describe a certain network property. Each tool we develop will be tested on real data and its value will be judged by the insights it offers about a system's structure or evolution.

Quantitative and mathematical nature: To contribute to the development of network science, it is essential to master the mathematical tools behind it. The tools of 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, control and information theory, statistics and data mining, helping us extract information from incomplete and noisy datasets.

Computational nature: Finally, given the size of many of the networks we explore, and the exceptional amount of data behind them, network science offers 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 help practitioners with diverse computational skills analyze networks.

THE IMPACT OF NETWORK SCIENCE

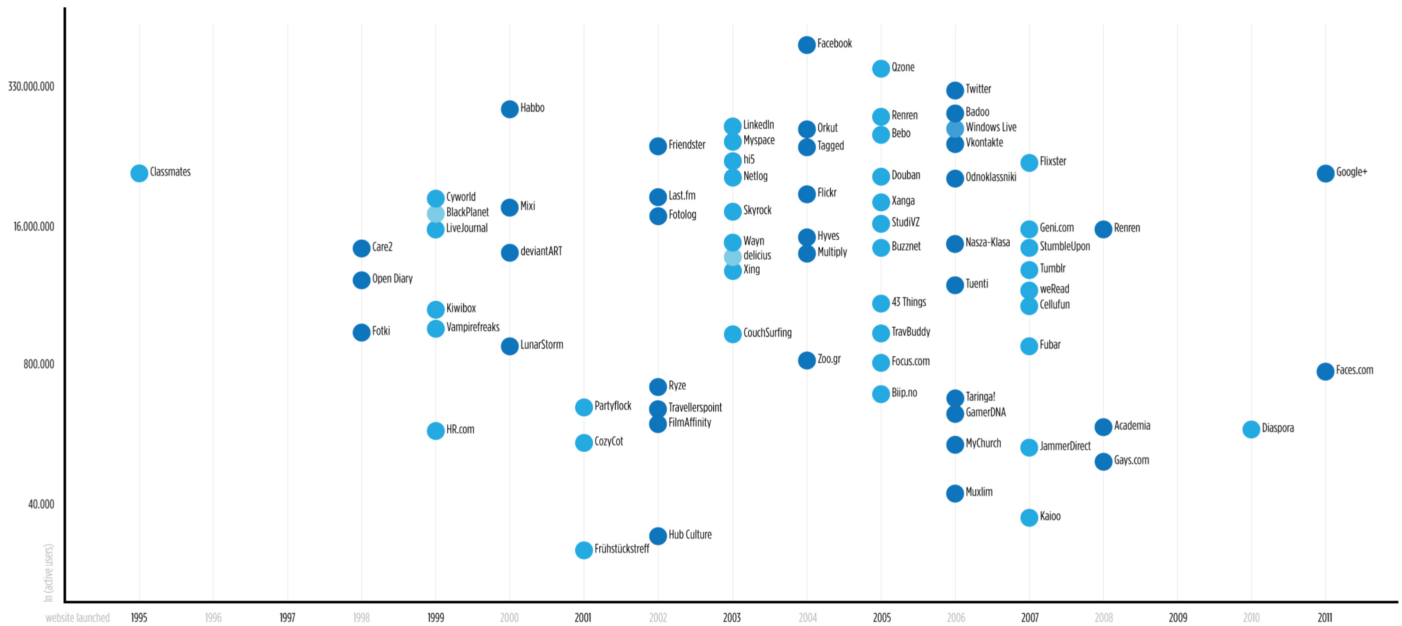


Image 1.6

The rise of social networking.

The popularity of the best known social networks, in terms of the number of users they attracted by the end of 2011 (vertical axis) shown as a function of their founding year (horizontal axis).

The impact of a new research field is measured both by its intellectual achievements as well as by the reach and the potential of its applications. While network science is a young field, its impact is everywhere around us, as we discuss below.

Economic Impact: From web search to social networking.

Some of the most successful companies of the 21st century, from *Google* to *Facebook*, from *Cisco* to *Apple* and *Akamai*, base their technology and business model on networks. Indeed, Google is not only the biggest network mapping operation, building a comprehensive map of the WWW, but its search technology relies on the network characteristics of the Web. Networks have gained particular popularity with the emergence of Facebook, the company with the oft-emphasized ambition to map out the social network of the whole planet. While Facebook was not the first social networking site, it is likely also not the last: an extensive ecosystem of social networking tools, from *Twitter* to *Orkut*, are attracting an impressive number of users (Image 1.6). The tools developed by network science fuel these sites, aiding everything from friend recommendation to advertising.

Health: From drug design to metabolic engineering.

The human genome project, completed in 2001, offered the first comprehensive list of all human genes [9, 10]. Yet, to fully understand how our cells function, and the origin of disease, we need accurate maps that tell us how these

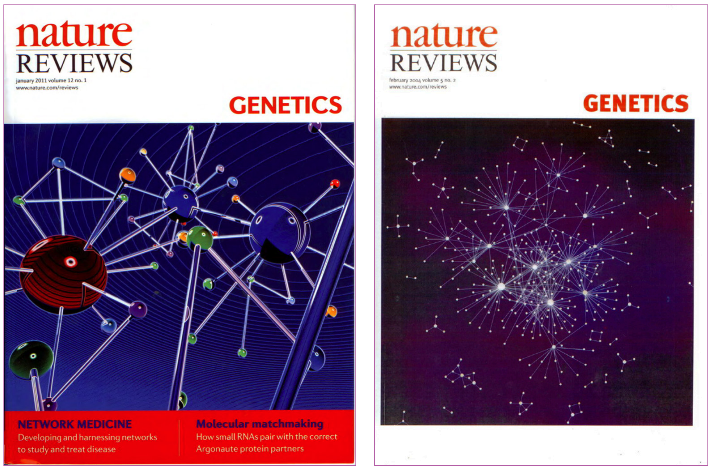


Image 1.7a, 1.7b

Networks in biology and medicine.

a) The cover of two issues of *Nature Reviews Genetics*, the top review journal in genetics. The cover from 2004, focuses on network biology [11], the cover from 2011 discuses network medicine [12].

b) The prominent role networks play in both cell biology and medical research is illustrated by the fact that the 2004 article on network biology is the second most cited article in the history of Nature Reviews Genetics.

genes and other cellular components interact with each other. Most cellular processes, from the processing of food by our cells to sensing changes in the environment, rely on molecular networks. The breakdown of these networks is responsible for most human diseases. This has led to the emergence of network biology, a new subfield of biology that aims to understand the behavior of cellular networks. A parallel movement within medicine, called network medicine, aims to uncover the role of networks in human disease (Image 1.7a/b). Networks are particularly important in drug development. The ultimate goal of network pharmacology is to develop drugs that can cure diseases without significant side effects. This goal is pursued at many levels, from millions of dollars invested to map out cellular networks to the development of tools and databases to store, curate, and analyze patient and genetic data. Several new companies take advantage of these opportunities, from *GeneGo* that aims to collect accurate maps of cellular interactions from scientific literature to *Genomatica* that uses the predictive power behind metabolic networks to identify drug targets in bacteria and humans. Recently most major pharmaceutical companies have made signifi-

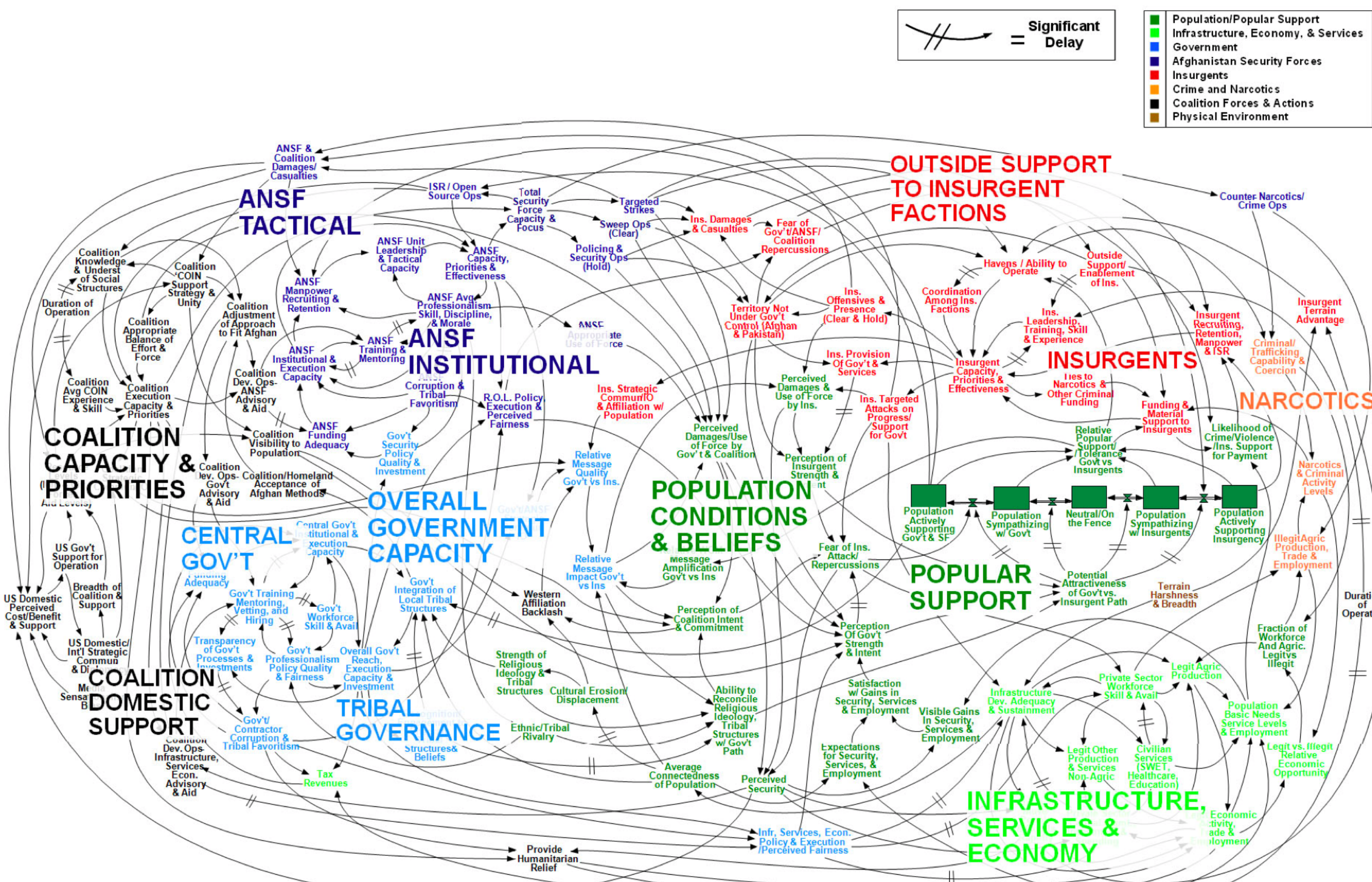


Image 1.8

The network behind a military engagement.

This diagram was designed during the Afghan war to portray the American strategy in Afghanistan. While it has been occasionally ridiculed in the press, it portrays well the complexities and the interconnected nature of a military's engagement. *(Image from New York Times)* cant investments in network and systems medicine, seeing it as the path towards future drugs.

Security: Fighting Terrorism.

Terrorism is one of the maladies of the 21st century, absorbing significant resources to combat it worldwide. Network thinking is increasingly present in the arsenal of various law enforcement agencies in charge of limiting terrorist activities. It is used to disrupt the financial network of terrorist organizations, to map terrorist networks, and to uncover the role of their members and their capabilities. While much of the work in this area is classified, several success stories have surfaced. Examples include the use of social networks to capture Saddam Hussein or the capture of the individuals behind the March 11, 2004 Madrid train bombings through the examination of the mobile call network. Network concepts have impacted military doctrine as well, leading to the concept of net-war, aimed at fighting low intensity conflicts and crime waged by terrorist and criminal networks that employ decentralized flexible network structures [13]. One of the first network science programs at the college level was started at West Point, the US Army's military academy. In 2009 the Army Research Lab and the Department of Defense devoted over $300 million to support network science centers across the US.

Epidemics: From forecasting to halting deadly viruses.

While the H1N1 pandemic was not as devastating as it was feared at the beginning of the outbreak in 2009, it gained

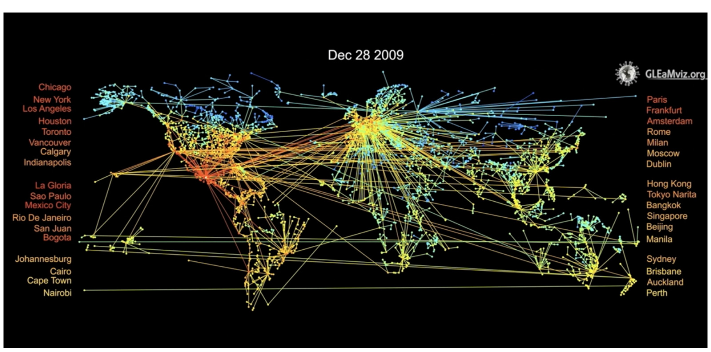


Image 1.9
Predicting the H1N1 epidemic.

The predicted spread of the H1N1 epidemics during 2009, representing the first successful prediction of a pandemic. The project, relying on the details of the worldwide transportation networks, foresee that H1N1 will peak out in October 2009, in contrast with the normal January-February peaks of influenza. This meant that the vaccines planned for November 2009 were too late, which was indeed the case. The success of this project shows the power of network science in facilitating advances in areas affected by networks.

Movie by D.Balcom, B.Gonçalves, H.Hu, and A.Vespignani.

a special role in the history of epidemics: it was the first pandemic whose course and time evolution was accurately predicted months before the pandemic reached its peak (Image 1.9) [14]. This was possible thanks to fundamental advances in understanding the role of networks in the spread of viruses. Indeed, before 2000 epidemic modeling was dominated by compartment models, assuming that everyone can infect everyone else one word the same socio-physical compartment. The emergence of a network-based framework has fundamentally changed this, offering a new level of predictability in epidemic phenomena.

Today epidemic prediction is one of the most active applications of network science [15, 16]. It is the source several fundamental results, covered in this book, that are used to predict the spread of both biological and electronic viruses. The impact of these advances are felt beyond biological viruses. In January 2010 network science tools have predicted the conditions necessary for the emergence of viruses spreading through mobile phones [17]. The first major mobile epidemic outbreak that started in the fall of 2010 in China, infecting over 300,000 phones each day, closely followed the predicted scenario.

Brain Research: Mapping neural network.

The human brain, consisting of hundreds of billions of interlinked neurons, is one of the least understood networks from the perspective of network science. The reason is simple: we lack maps telling us which neurons link to each other. The only fully mapped neural map available for research is that of the *C.Elegans* worm, with only 300 neurons. Should detailed maps of mammalian brains become available, brain research could become the most prolific application area of network science. Driven by the potential impact of such maps, in 2010 the National Institutes of Health has initiated the *Connectome* project, aimed at developing the technologies that could provide an accurate neuron-level map of mammalian brains.

Management: Uncovering the internal structure of an organization.

While traditionally management uses the official chain of command to understand the inner structure of an organization, it is increasingly evident that the informal network, capturing who really communicates with whom, matters even more for the success of a company. Accurate maps of this network can expose lack of communication between key units, can identify individuals who play an outsize role in bringing different departments and products together, and help higher management diagnose diverse organizational issues. Furthermore, there is increasing evidence in the management literature that the position of an employee within this network correlates with his/her productivity [18].

Therefore, several dozen consulting companies have emerged with expertise to map out the true structure of an organization. Established consulting firms, from *IBM* to *SAP*, have added social networking capabilities to their consulting business. These companies offer a host of services, from identifying opinion leaders to preventing employee churn and from identifying optimal groups for a task to modeling product diffusion (Image 1.10a/b/c/d). Hence lately network science tools are increasingly indispensable in management and business, enhancing productivity and boosting innovation within an organization.

Network science can therefore offer a microscope for higher management, helping them improve the company's effectiveness by uncovering the true network behind any organization.

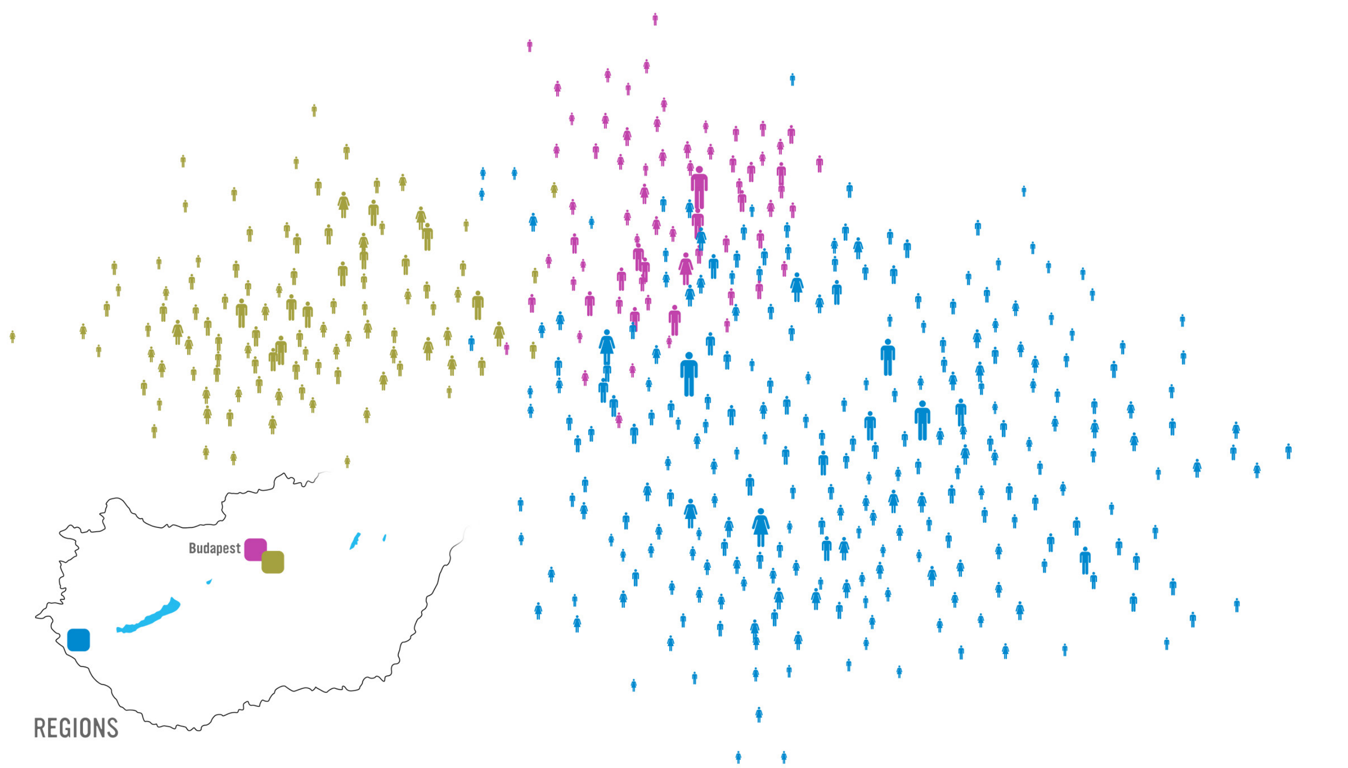


Image 1.10a

Understanding the inner workings of an organization.

The workforce of a Hungarian company with three main locations, one on Budapest, whose employees are shown in purple, and two manufacturing sites outside of the city, shown in yellow and blue. The company had a major internal communication problem: information that reached the workers about the intentions of the higher management often had nothing do to with the management's real plans. Seeking to understand the source of this discrepancy, and looking for ways to embrace information flow within the company, the management turned to Maven 7, a social networking consulting company that applies network science in diverse organizational setting.

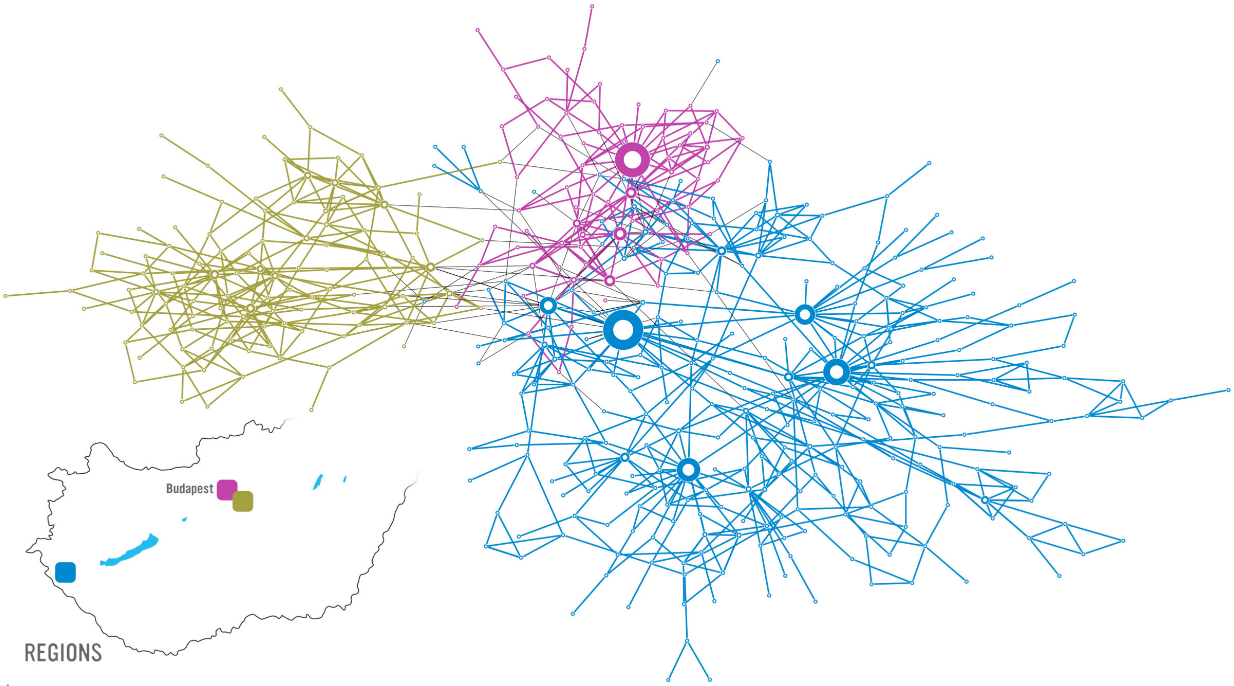


Image 1.10b

Understanding the inner workings of an organization.

Having the list of the workers and their role in the company, together with the official hierarchy is not sufficient to understand how an organization works. For that we need to know who listens to whom, who is asking for advice from whom, eventually uncovering the paths through which knowledge and information travels within the organization. Hence Maven 7 developed an online platform to ask each employee whom do they turn to for advice when it comes to decisions impacting the company, from restructuring to advancement. This allowed them to build the map shown above, where two individuals are connected if one nominated the other as his/her source of information on organizational and professional issues.
The map identifies several highly influential individuals that are the hubs of the organization. The problem was that none of the hubs were part of the leadership.

Image 1.10c

Understanding the inner workings of an organization.

The position of the leadership within the company's informal network is illustrated on this map, where we colored the nodes based on their company rank within the company. None of the company directors, including the CEO, shown in red, are hubs. Nor are the top managers, shown in blue. The hubs are managers, group leaders and associates, or workers. The biggest hub, hence the most influential individual, is an associate, shown as a gray node in the center.

Image 1.10d

Understanding the inner workings of an organization.

The image indicates that a significant fraction of employees are one to two links from the biggest hub. It turns out that he is the safety and environmental expert in the company, whose job is to visit each location and talk with most employees. There is only one part of the company he has no links to: the directors or the top management. With little access to the management and their intentions, he passes on information that he collects along his trail, effectively running a gossip center.

How does one remedy this situation? Fire the biggest hub? He is not the problem and firing him would probably make the problem even more acute. The real issue is that higher management failed to put in place proper channels of communication, leaving behind a structural hole, which was naturally filled by the environmental and safety manager. Offering him and the few other hubs access to the true information can fundamentally change the reliability of information within the company. Network science can therefore offer a potent microscope for higher management, helping them improve the company's effectiveness by uncovering the true network behind an organization.

SCIENTIFIC IMPACT

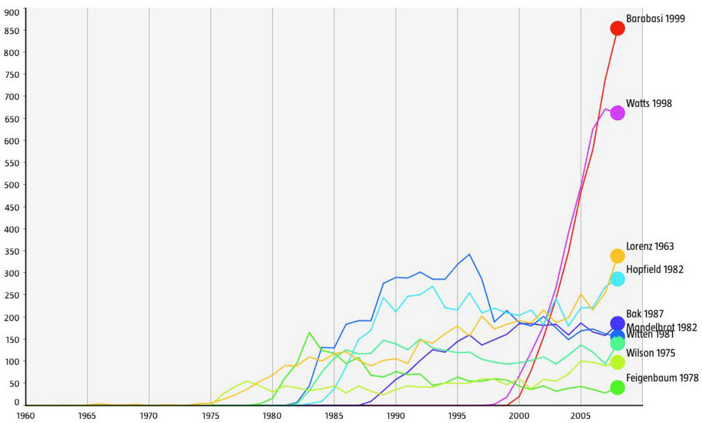


Image 1.11

Complexity and network science.

The impact of network science can be put into perspective by looking at the citation patterns of the most cited papers in complexity. The study of complex systems in the 70s and 80s was dominated by Edward Lorenz's 1963 classic work on chaos [19], Kenneth G. Wilson's renormalization group [20], and Mitchell Feigenbaum's discovery of the bifurcation diagram [21]. In the 1980s the community has shifted its focus on pattern formation, following Benoit Mandelbrot's book on fractals [22] and Thomas Witten and Len Sander's introduction of the diffusion limited aggregation mode [23]. Equally influential was John Hopfield's paper on neural networks [24] and Per Bak, Chao Tang and Kurt Wiesenfeld's paper on self-organized criticality [25]. These papers are continuing to define our understanding of complex systems, each of them writing a separate chapter in modern statistical mechanics. The video compare their citation pattern with the citations of the two most cited papers in this area [2,3].

Nowhere is the impact of network thinking more evident than in the scientific community. The most prominent scientific journals, from *Nature* and *Science* to *Cell* and *PNAS*, have devoted special issues, reviews, or editorials addressing the impact of networks on various topics from biology to social sciences. During the past decade, each year several dozen international conferences, workshops, summer and winter schools have focused exclusively on network science. A successful network science meeting series, called *NetSci*, attracts the field's practitioners since 2005. Several general-interest books, making the bestseller lists in many countries, have brought network science to the public. Most major universities offer network science courses, attracting a diverse student body. Finally, *Science Magazine* has devoted a special issue to networks, marking the ten-year anniversary of the paper that reported the discovery of scale-free networks [3] (Image 1.12).

The relative impact of network science can be put into perspective by looking at the citation patterns of the most cited papers in the area of complex systems (Image 1.11). Each of these papers are citation classics, cumulatively amassing anywhere between 2,000 and 5,000 citations, continuing to gather anywhere between 50 to 300 citations a year. To see how the interest in network science compares to these classic discoveries, in Movie 3 we also show the citation patterns of the two most cited network science papers: the 1998 paper on small-world phenomena by Duncan Watts and Steve Strogatz [2] and the 1999 Science paper reporting the discovery of scale-free networks by Albert-László Barabási and Réka Albert [3]. As one can see, the growth in citations to these papers unparalleled in the area of complex systems.

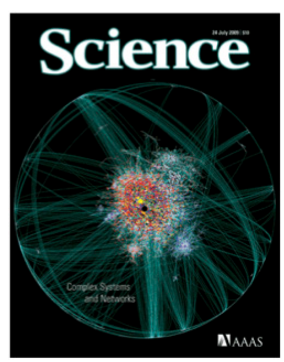


Image 1.12

Complex systems and networks.

Special issue of Science magazine on Complex Systems and Networks, published on July 24, 2009, marking the 10th anniversary of the 1999 discovery of scale-free networks [3].

Several other metrics indicate that network science is impacting in a defining manner particular disciplines. For example, several research fields witnessed network papers become some of the most cited papers in their leading journals:

- The 1998 paper by Watts and Strogatz in *Nature* on small world phenomena [2] and the 1999 paper by

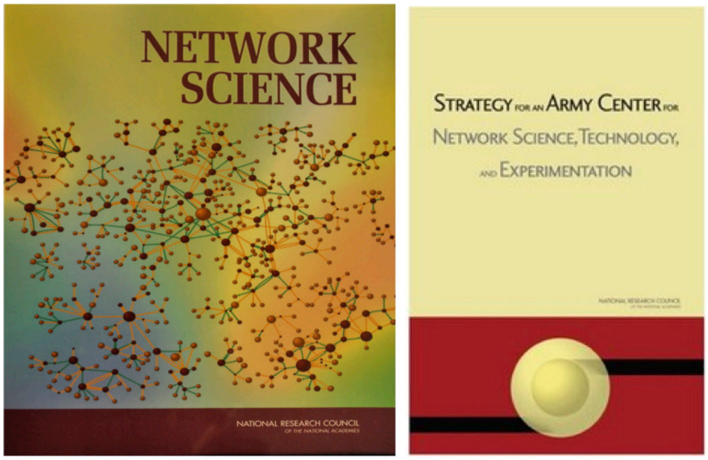


Image 1.13
National Research Council Reports.

The two National Research Council Reports on network science have not only documented the emergence of a new discipline, but have also explained their long-term impact on a number of research fields, as well as national competitiveness and the military. They have urged dedicated support for the field, leading to the establishment of a series of network science centers in US and the network science program within NSF.

Barabási and Albert in Science on scale-free networks [3] were identified by ISI as the top ten most cited papers in physics during the decade after their publications. Furthermore, currently (2011) the Watts-Strogatz paper is the second most cited of all papers published by *Nature* in 1998, and the Barabási-Albert paper is the most cited paper among all papers published in Science in 1999.

- Four years after its publication, the *SIAM* review of Mark Newman on network science became the most cited paper of any journal published by the *Society of Industrial Mathematics* [26].

- *Reviews of Modern Physics*, published continuously since 1929, is the physics journal with the highest impact factor. Currently the most cited paper of the journal is Chandrasekhar classic 1944 review that summarized the author's work that led to his Nobel in physics, entitled *Stochastic Problems in Physics and Astronomy* [27]. During over 60 years since its publication, the paper gathered over 5,000 citations. Yet, it will soon be taken over by a paper published only in 2001 entitled *Statistical Mechanics of Complex Networks*, the first review of network science [28].

- The paper leading to the discovery that in scale-free networks the epidemic threshold is zero, by Pastor-Satorras and Vespignani [29], is the most cited paper among the papers published in 2001 by *Physical Review Letters*, a position the paper is sharing with a paper on quantum computing.

- The paper by Michelle Girvan and Mark Newman on community discovery in networks [30] is the most cited paper published in 2002 by *Proceedings of the National Academy of Sciences*.

- The 2004 review entitled *Network Biology*, by Barabási and Oltvai [11], is the second most cited paper in the history of *Nature Reviews Genetics*, the top review journal in genetics.

Given this extraordinary response by the scientific community, network science was examined by the National Research Council (NRC), the arm of the US National Academies in charge of offering policy recommendation to the US government. NRC has assembled two panels, resulting in two publications [31], defining the field of network science (Image 1.13). They not only document the emergence of a new research field, but highlight the field's vital importance to national competitiveness and security. Following these reports, the National Science Foundation (NSF) in the US established a network science directorate and a series of network science centers were established by the Army Research Labs.

General Audience

The results of network science have excited the public as well. This was fueled partly by the success of several general audience books, like *Linked: The New Science of Networks* by Albert-László Barabási, *Nexus* by Mark Buchanan, and *Six Degrees* by Duncan Watts, each being translated in many of languages. Newer books, like *Connected* by Nicholas Christakis and James Fowler, were also exceptionally successful (Image 1.15). An award-winning documentary, *Connected*, by Australian filmmaker Annamaria Talas, has brought the field to our TV screen, being broadcasted all over the world and winning several prestigious prizes (Image 1.14). Networks have inspired artists as well, leading to a wide range of network science research inspired art-project, and even an annual symposium series that

Image 1.14
Connected.

The trailer of the award winner document *Connected*, directed by Annamaria Talas, focusing on network science.

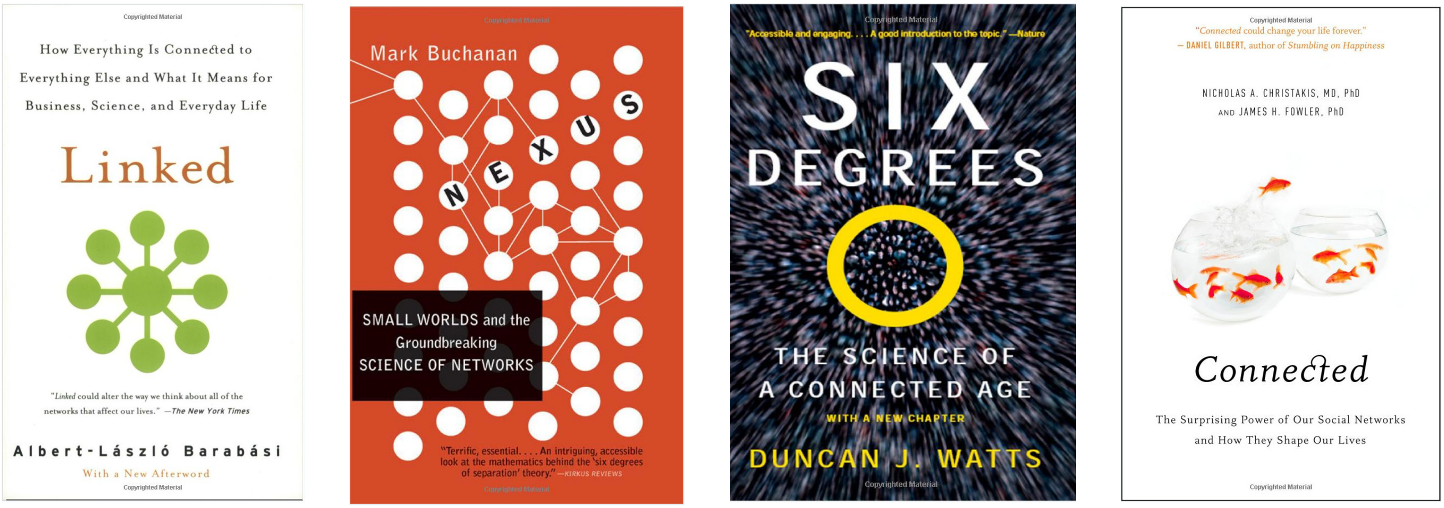


Image 1.15

Wide impact.

Four widely read books are bringing network science to the public.

brings together, on a yearly basis, artists and scientists [32]. Fueled by successful movies like *The Social Network*, and a series of novels and short stories, from science fiction to novels exploiting the network paradigm, today networks have permeated popular culture.

SUMMARY

While the emergence of the scientific interest in networks was rather sudden, the enthusiasm for the field was responding to the emergence of a wider social awareness of the importance of networks. This is illustrated in Image 1.16, where we show the usage frequency of the words that represent two important scientific revolutions of the past two centuries: evolution, capturing the most common term to refer to Darwin's theory of *evolution*, and *quantum*, the most frequently used term when one refers to *quantum* mechanics. The use of evolution increases only after the 1859 publication of Darwin's *On the Origins of Species*. The word *quantum*, first used in 1902, is virtually absent until the 1920s, when quantum mechanics gains prominence. The use of the word network has increased dramatically following the 1980s. While the word has many uses (as do evolution and *quantum*), its dramatic rise captures the extraordinary awareness of networks in the society at large. Indeed, evolution and quantum mechanics are just as important as core scientific fields, as they are as enabling platforms: the current revolution in genetics is built on evolutionary theory, and quantum mechanics offers a platform for a wide range of advances in contemporary science, from chemistry to wireless communications. In a similar fashion, network science is an enabling science, offering new tools and perspective for a wide range of scientific fields from social networking to drug design. Given the wide importance and impact of networks, we need to develop the tools to study and quantify them. The rest of this book is devoted to this worthy subject.

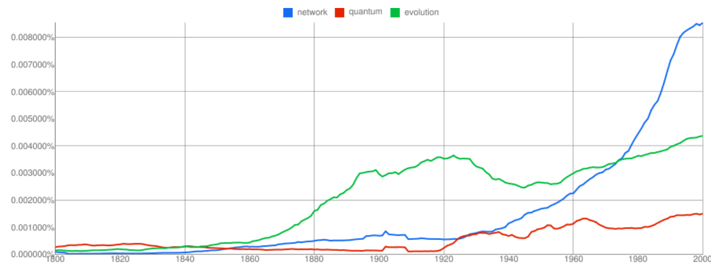


Image 1.16
The rise of networks.

The frequency of the use of the words *evolution* and *quantum* represents the major scientific advances of the 19th and 20th century, namely Darwin's theory of evolution and quantum mechanics. The plot indicates the exploding awareness of networks in the last decades of the 20th century, preparing a fertile ground for the emergence of network science. The plots were generated by using the ngram platform of Google: http://books.google.com/ngrams.

BIBLIOGRAPHY

[1] C. Wilson. *Searching for Saddam: a five-part series on how the US military used social networking to capture the Iraqi dictator.* 2010. www.slate.com/id/2245228/.

[2] D. J. Watts and S .H. Strogatz. *Collective dynamics of 'small-world' networks*. Nature, 393 (440), 1998.

[3] A.-L. Barabási and R. Albert. *Emergence of scaling in random networks*, Science, 286 (509), 1999.

[4] P. Erdős and A. Rényi. *On random graphs*. Publicationes Mathematicae, 6 (290), 1959.

[5] M. S. Granovetter. *The strength of weak ties.* American Journal of Sociology, 78 (1360), 1973.

[6] K. Venkatesan, J.-F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, M. Zenkner, X. Xin, K.-I. Goh, M. A. Yildirim, N. Simonis, J. M. Sahalie, S. Cevik, C. Simon, A.-S. de Smet, E. Dann, A. Smolyar, A. Vinayagam, H. Yu, D. Szeto, H. Borick, A. Dricot, N. Klitgord, R. R. Murray, C. Lin, M. Lalowski, and J. Timm. *An empirical framework for binary interactome mapping*. Nature Methods, 6 (83), 2009.

[7] O. Sporns, G. Tononi, and R. Kötter. *The Human Connectome: A Structural Description of the Human Brain*. PLoS Comput. Biol., 1 (4), 2005.

[8] http://www.caida.org/ http://www.netdimes.org/

[9] International Human Genome Sequencing Consortium. *Initial sequencing and analysis of the human genome*. Nature, 409 (6822), 2001.

[10] J. C. Venter et al.,*The Sequence of the Human Genome*, Science, 291 (1304), 2001.

[11] Z. N. Oltvai and A.-L. Barabási. *Understanding the cell's functional organization*. Nature Reviews Genetics, 5 (101), 2004.

[12] N. Gulbahce, A.-L. Barabási, and J. Loscalzo. *Network medicine: a network-based approach to human disease*. Nature Reviews Genetics, 12 (56), 2011.

[13] J. Arquilla and D. Ronfeldt, *Networks and Netwars: The Future of Terror*, Crime, and Militancy (RAND: Santa Monica, CA), 2001.

[14] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, and A. Vespignani. *Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility*. BMC Medicine, 7 (45), 2009.

[15] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, and J. J. Ramasco, A. Vespignani, *Multiscale mobility networks and the spatial spreading of infectious diseases.* Proc. Natl. Acad. Sci., 106 (21484) 2009.

[16] L. Hufnagel, D. Brockmann, and T. Geisel, *Forecast and control of epidemics in a globalized world.* Proc. Natl. Acad. Sci., 101 (15124), 2004.

[17] P. Wang, M. Gonzalez, C. A. Hidalgo, and A.-L. Barabási. *Understanding the spreading patterns of mobile phone viruses*. Science, 324 (1071), 2009.

[18] L. Wu , B. N. Waber, S. Aral, E. Brynjolfsson, and A. Pentland, *Mining Face-to-Face Interaction Networks using Sociometric Badges: Predicting Productivity in an IT Configuration Task*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1130251

[19] E. N. Lorenz, *Deterministic Non periodic Flow*. J. Atmos. Sci., 20 (130), 1963.

[20] K. G. Wilson, *The renormalization group: Critical phenomena and the Kondo problem*, Rev. Mod. Phys. 47 (773), 1975.

[21] M. J. Feigenbaum, *Quantitative Universality for a Class of Non-Linear Transformations*. J. Stat. Phys. 19 (25), 1978.

[22] B. B. Mandelbrot, *The Fractal Geometry of Nature*. W.H. Freeman and Company. 1982

[23] T. Witten, Jr. and L. M. Sander, *Diffusion-Limited Aggregation*, a Kinetic Critical Phenomenon. Phys. Rev. Lett., 47 (1400), 1981.

[24] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci., 79 (2554), 1982.

[25] P. Bak, C. Tang, and K. Wiesenfeld. *Self-organized criticality: an explanation of 1/f noise*. Phys. Rev. Lett., 59 (4), 1987.

[26] M. E. J. Newman. *The structure and function of complex networks*, SIAM Review. 45 (167), 2003.

[27] S. Chandrasekhar. *Stochastic Problems in Physics and Astronomy*, Rev. Mod. Phys., 15 (1), 1943.

[28] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Mod. Phys., 74 (47), 2002.

[29] R. Pastor-Satorras and A. Vespignani. *Epidemic spreading in scale-free networks*. Phys. Rev. Lett., 86 (3200), 2001.

[30] M. Girvan and M. E. J. Newman. *Community structure in social and biological networks*. Proc. Natl. Acad. Sci., 99 (7821), 2002.

[31] National Research Council, *Network Science*. Washington, DC: The National Academies Press, 2005.

National Research Council. Strategy for an Army Center for Network Science, Technology, and Experimentation . Washington, DC: The National Academies Press, 2007.

[32] M. Schich, R. Malina, and I. Meirelles (Editors), *Arts, Humanities, and Complex Networks* [Kindle Edition], 2012.