

ALBERT-LÁSZLÓ BARABÁSI

# NETWORK SCIENCE

## PERSONAL INTRODUCTION



### ACKNOWLEDGEMENTS

MÁRTON PÓSFAI  
GABRIELE MUSELLA  
MAURO MARTINO  
NICOLE SAMAY

SARAH MORRISON  
AMAL HUSSEINI  
PHILIPP HOEVEL

Figure 1.0 (cover image)

**Chiharu Shiota: In Silence**

Chiharu Shiota is a Japanese installation artist working in Berlin. She creates interwoven webs of wool that engulf ordinary objects. The apparent randomness of these networks embrace the inherent tension between order and disorder, the very topic of network science.



This book is licensed under a  
Creative Commons: CC BY-NC-SA 2.0.  
PDF V26, 05.09.2014

# INTRODUCTION

Today, when each year a dozen or so conferences, workshops and schools focus on networks, when over a hundred books and three journals are devoted to the field, when most universities offer network science courses and one can get a PhD in network science on three continents, and when funding agencies have earmarked hundreds of millions of dollars to the subject, it is tempting to see this decade-old field's evolution as a straight path to success. But blinded by this cumulative impact, we may miss the most fascinating question: How could the field grow up this fast?

I call this chapter a *personal introduction* for the simple reason that I have no intention of offering an unbiased answer to this question. On the contrary, I plan to recall the emergence of network science from the perspective of a participant whose story I best know, which happens to be me. This is not a victory march, but my goal is to recall the winding and convoluted journey that I experienced, with its numerous setbacks and bursts. Instead of a bird's eye perspective, I will focus on those hard to forget trees that I repeatedly bumped into as I attempted to cross the forest. It is a reminder that scientific discovery is not as straightforward and smooth as textbooks, like the one following these pages, may occasionally insinuate.

## MY FIRST NETWORK PAPER (1994)

My fascination with networks started in December 1994, a few months into my brief postdoctoral position at IBM's legendary T. J. Watson Research Center. As the approaching holidays have brought a predictable halt to life at Watson, I decided to use the break to learn a bit more about my employer. Back then IBM was a synonym of the computer, so I went to Watson's library looking for an introduction into computer science.

Curious about the field's intellectual challenges, I walked away with a book covering an array of problems, from algorithms to Boolean Logic and NP-completeness. One chapter, focusing on the minimal spanning tree problem, particularly piqued my interest. For good reason: I realized that the Kruskal algorithm described in the book mapped into of a well-

### Invasion Percolation and Global Optimization

Albert-László Barabási

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556  
and T.J. Watson Research Center, IBM, P.O. Box 218, Yorktown Heights, New York 10598  
(Received 24 February 1995)

Invasion bond percolation (IBP) is mapped exactly into Prim's algorithm for finding the shortest spanning tree of a weighted random graph. Exploring this mapping, which is valid for arbitrary dimensions and lattices, we introduce a new IBP model that belongs to the same universality class as IBP and generates the minimal energy tree spanning the IBP cluster. [S0031-9007(96)00106-8]

PACS numbers: 47.55.Mh

Flow in a porous medium, a problem with important practical applications, has motivated a large number of theoretical and experimental studies [1]. Aiming to understand the complex interplay between the dynamics of flow processes and randomness characterizing the porous medium, a number of models have been introduced that capture different aspects of various experimental situations. One of the most investigated models in this respect is invasion percolation [2], which describes low flow rate drainage experiments or secondary migration of oil during the formation of underground oil reservoirs [1,3].

When a wetting fluid (e.g., water) is injected slowly into a porous medium saturated with a nonwetting fluid (e.g., oil), capillary forces, inversely proportional to the local pore diameter, are the major driving forces determining the invasion of the fluid. The invasion process is a

this graph is a connected graph of  $n$  vertices and  $n - 1$  bonds. Of the many possible spanning trees one wants to find the one for which the sum of the weights  $p_{ij}$  is the smallest. A well known example is designing a network that connects  $n$  cities with direct city-to-city links (whose length is  $p_{ij}$ ) and shortest possible total length. This is a problem of major interest in the planning of large scale communication networks and is one of the few problems in graph theory that can be considered completely solved. Since for a fully connected graph with  $n$  vertices there are  $n^{n-2}$  spanning trees [5], designing an algorithm that finds the shortest one in nonexponential time steps is a formidable global optimization problem.

An efficient algorithm for finding the shortest spanning tree of an arbitrary connected graph  $G$  was introduced by Prim [6] and involves the following steps: (1) Choose an

Figure 1.1

### 1994-1995: My First Take on Networks

Conceived over the winter break at the end of 1994, my first network paper [1] mapped the minimal spanning tree problem, a well-known algorithm in computer science, into invasion percolation, a much studied problem in statistical physics. It marked the beginning of my long engagement with network science.

known model of statistical physics, called invasion percolation. So exactly two months after Christmas, on Feb 24, 1995, I submitted my first paper on networks to *Physical Review Letters* [1], demonstrating the equivalence of two much studied network problems of physics and computer science (Figure 1.1). While a single-author paper in the prestigious physics journal was undoubtedly a smart career move, its true impact was far more far-reaching: The hidden intellectual flood gates the paper unlocked laid the ground for my subsequent decades-long love affair with networks.

### FAIL 1: THE SECOND PAPER (1995)

The more I learned, the more puzzled I was about how little we knew about real networks. Living in New York City, I imagined the remarkable complexity of the millions of electric, telephone, and Internet cables cramped under Manhattan's pavement. Graph theory envisioned that these networks were wired randomly. That didn't make much sense to me. There must be some organizing principles that govern the numerous networks that we depend on. Finding these principles was a fitting challenge for a statistical physicist trained at the border of order and randomness.

So I devoted the subsequent months to Béla Bollobás' excellent book on random graphs [2] that introduced me to the classical work of Erdős and Rényi [3]. At the same time Stuart Kauffman's visionary writing made me appreciate the importance of networks in biology [4]. Two very different perspectives collided in these books: The theorem-driven dry world of mathematics with the wandering imagination of Stu, that saw no mathematical bounds (Figure 1.2).

Eight months into my postdoctoral position I accepted a faculty position at University of Notre Dame, allowing me to devote the remaining four months at IBM to my second network paper. Entitled *Dynamics of Random Networks: Connectivity and First Order Phase Transitions*, [5] it was my first attempt to probe the implications of altering the topology of

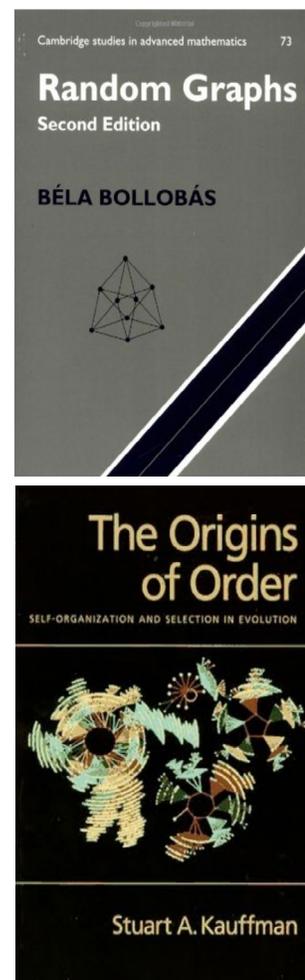
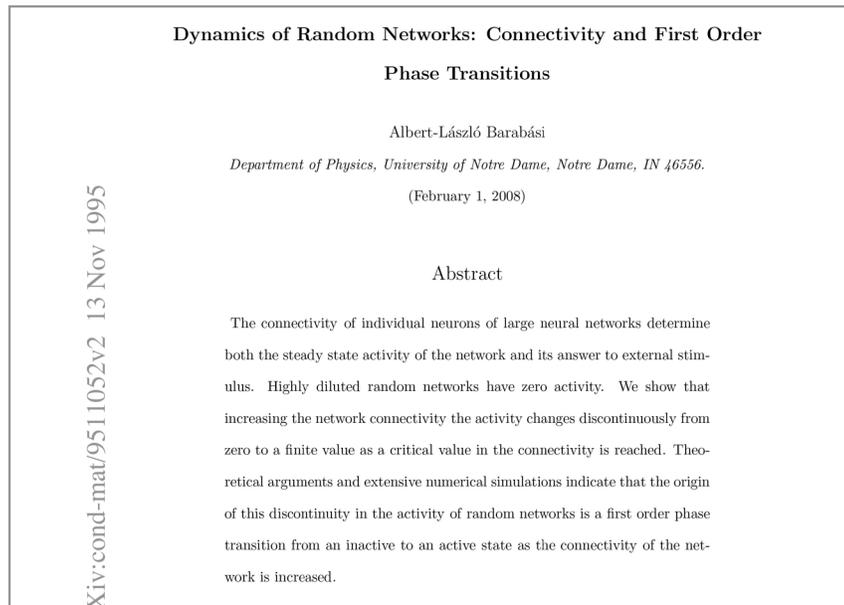


Figure 1.2

### 1995: Order or Randomness?

Two of the three books that inspired my early journey towards network science. I could never track down the first book, whose title (or subtitle) was something like *Fifty Problems in Computer Science*, the one I borrowed in 1994 from the library of IBM's Watson Research Center.



**Figure 1.3**

**1995-1997: The Never Published Network Paper**

My second take on networks, and the first paper in which I explored the role of the network topology. It was posted on the online server Arxiv in November 1995, after it was rejected by four journals. I eventually gave up trying to get it published in a journal.

a network. The paper merged the world of Bollobás and Kaufmann, asking how changes in the network structure affect the dynamics of a Boolean system (Figure 1.3). The underlying observation was simple: If we alter the average degree of a random network, the Boolean system undergoes a dynamic phase transition. Hence we cannot interpret a system's behavior without fully accounting for the structure of the network behind it.

The paper was motivated by a mixture of ideas rooted in cellular networks, the Internet and the World Wide Web. Yet, these topics were largely absent from the physics journals that normally published my work. I struggled, therefore, to find some tangible applications within my own domain. At the end I put the results in the context of neural networks, a much studied problem among physicists. This community, I thought, should be inclined to think positively about networks. I was wrong, of course, and this decision marked the first of a series of failures that trailed my journey towards network science for the next four years.

On Nov 10, 1995 I mailed the finished manuscript to *Science* and returned to Boston for the annual meeting of the Materials Research Society. Philipp Ball, a *Nature* editor with interest in interdisciplinary subjects, was at the meeting, giving me the opportunity to tell him about my fascination with my new subject, networks. So when a few weeks later *Science* rejected the paper without review, I sent it to Philipp, hoping that *Nature* would show more interest. And it did, sending the paper for review.

The referees were much less fascinated, however. One of them put this bluntly writing in the referee report that:

1. It is badly motivated;
2. It is technically very constrained;
3. The speculations (about evolution and Internet), do not materialize.

The referee was right, of course: I failed to explain why we care about networks in the first place. It was all in my head. But barely a year after my Ph.D., relying on a language (english) that I acquired only four years earlier, I could not yet translate my ideas into a story that stuck.

Disappointed, on April 25, 1996, I resubmitted the paper to *Physical Review Letters*. It did not fare much better there either, being rejected after a lengthy review. When on Nov 21, 1997, two years after its first submission, I resubmitted the paper to *Europhysics Letters*, I was already experiencing the second major failure of my network-bound-journey.

## FAIL 2: MAPPING THE WEB (1996)

While struggling to get my second paper published, I became increasingly convinced that to move forward I will need to abandon the graph theoretical path I pursued thus far. I should do instead what physicists are good at: Look at the real world for inspiration. That is, I decided that I need maps. Maps of real networks, to be precise.

Five years after Tim Berners-Lee unleashed the code behind the WWW and two years before Google was founded, the Web just started humming. An odd collection of search engines going by names like JumpStation, RBSE Spider or Webcrawler, hacked together in research labs, were trying to map its link structure. In February, 1996 I sent an email to several researchers running such crawlers (Figure 1.4), hoping to get a sample of their data. A full map would have been ideal. Short of that it would have been sufficient to get the number of links each node had. "I wish to make a simple histogram of the previous data," I wrote, asking for something that we will name only three years later: The degree distribution of the World Wide Web.

No one said no. But no one bothered to answer either. And as I waited for a reply, my second network paper got its final blow, being rejected by *Europhysics Letters* as well.

By that point my journey into networks was quite disappointing. My second network paper has been seen by four journals and three rounds of referees. No one said that it was wrong. The referees' message was simple: Who cares? Then my Plan B to access real data has slowly reached a dead end. Disappointed and under pressure to publish and obtain grants, I gradually replaced networks with a safer line of research on quantum dots.

I had no choice, really. Two years into my assistant professorship my startup funds were dwindling and my prospect for tenure looked thin. As much as I believed in networks, all I could show for the past three years was one publication and a string of failures. The transition to the more conventional topic paid off, however: by the end of 2007 I was awarded two research grants, allowing me to hire several students and a postdoctoral researcher.

Figure 1.4  
1996: Begging for Data

One of the emails I sent to computer scientists building web crawlers in the mid 1990s, hoping to convince them to share data on the Web's topology. In hindsight, not a very convincing letter. No wonder no one responded. I had to wait two more years, until Hawoong Jeong joined my research lab and built our own crawler, to get the data that allowed us to discover scale-free networks. Had we gotten the data in 1996, as I originally hoped to, we might have discovered it three years earlier.

Date: Fri, 09 Feb 1996 10:34:17 -0500  
From: Albert-Laszlo Barabasi <alb@nd.edu>  
To: [REDACTED].gov  
Cc: alb@nd.edu  
Subject: Robots  
X-Url: [REDACTED].html

Dear [REDACTED],

I am doing some research on random networks and their statistical mechanical properties. The best available real word example of such networks is the WWW with its almost random links. To try out my approach I need some data that that Robots could provide without much difficulty. I friend of mine (who knows much more about the dangers of writing and operating a poorly working- robot) convinced me that instead of attempting to write my own robot, I should rather check if somebody with an already running robot could either (i) help me with the data I need or (ii) allow me to use his/her robot for this purpose.

I wonder if you are willing to give me a help in this direction? Of course, any help will be carefully acknowledged when the results of this research will be published (this is all-academic, non-profit basic research).

When a robot visits a new site, it finds a number of external links (pointing to other home pages). I need statistics regarding this number. Robots regularly collect this information, since this is how they assemble their database. Thus the only thing I need is to have the robot write this info into a file in a structured form, that would allow me to extract this information. Maybe some of the robots do save the obtained data in a format that would allow me to simply collect these numbers.

For example, if the robot visits the home page <http://www.new.homepage.edu/bbb.html> it finds that there are for example four links there, pointing to the addresses:

<http://www.aps.org/xxx.html>  
<http://www.my.best.friend/home.html>  
<http://www.my.hobby/joke.html>  
<http://www.my.preffered.newspaper/news.html>

So the type of list I could most use is this one (or something equivalent):

<http://www.new.homepage.edu/bbb.html>  
HAS LINKS TO:

<http://www.aps.org/xxx.html>  
<http://www.my.best.friend/home.html>  
<http://www.my.hobby/joke.html>  
<http://www.my.preffered.newspaper/news.html>

Moreover, to start with it would be enough less information, for example a just listing the number of links he found:

4

After visiting a fair number of home pages the table would look like this:

4  
2  
3  
2  
0  
19  
10  
1  
0  
1

How many datapoints do I need? Well, I wish to make a simple histogram of the previous data, thus I need enough data to obtain a smooth histogram. This histogram will be the starting point of my investigation.

I hope you are willing to help me to obtain this information. If you are not running your robot currently, but are willing to lend me your code so that I can run from my computer to collect this data (I have an IBM RISC 6000 that I could use for this purpose), that is also a solution. Again, I do not plan to use the robot for any other purpose than collecting this (and similar) statistics on the connectivity of the web. If you are interested in more details regarding the nature of the scientific questions I am investigating, I am happy to provide it to you.

laszlo

Albert-Laszlo Barabasi  
Assistant Professor

## REBOOT (1998)

In 1997 I lived in Chicago, commuting every second day to Notre Dame. To kill the boredom of the two-hour drive, I started to listen to books on tape. One day I picked up from the library Asimov's *Foundation*, a book that I devoured as a child (Figure 1.5). As I slipped into the magical world of the *Second Foundation*, I was captivated by Harry Seldon's ability to forecast the fate of humanity hundreds of years into the future. It was the best of science fiction: fascinating, out of reach, but still plausible in some abstract dimension.

The monolithic corn fields that surround Route 90 that connects Notre Dame to Chicago allowed my mind to contemplate a whole range of quixotic questions: What would it take to turn Asimov's fiction into reality? Could one indeed formulate a set of equations that could predict the future of a system as complex as the society? Is there anything I could do to help achieve this? As my research on quantum dots blossomed, Asimov kept pulling my mind back to the questions that never stopped fascinating me despite the many setbacks I experienced earlier: networks and complex systems.

By early 1998 I was ready to try it again. I started by sketching out a new network-related research project and in March I invited Réka Albert to lunch at *Sorins*, the most elegant restaurant Notre Dame had to offer. Réka, a year and a half into her graduate studies, was on to a stellar career. Her paper on granular media have just made the cover of *Nature* and the preliminary results of her ongoing projects were just as promising. Hence my purpose with the lunch defied all wisdom: I wanted to persuade her to give up the research she has been so successful at. I wanted her to explore networks instead.

As I asked my best student to join me on my network crusade, I could offer little encouragement. I had to tell her that my second paper on the subject has been rejected by four journals and that I could never get it published [5]. Networks had no community, no journal and no funding. I had to be honest, confessing to her that no one seemed to care about the subject. She was therefore risking a sudden end to the success story she has so far experienced.

Yet, I also told her that to succeed we must take risks. And that in my view networks were worth the gamble.

At the end of the lunch I gave Réka a densely typed document, my early vision of network science. I estimated that it will take us about six months to quantify the network topology and another six months to understand the impact of the topology on network dynamics. Then we could move on to the real problem, and explore the joint evolution of network topology and dynamics.

I was completely off the mark, of course: I could not foresee the fantastic richness the topology had to offer. But that was besides the point back then. What mattered is that in her quiet and gracious manner, Réka agreed

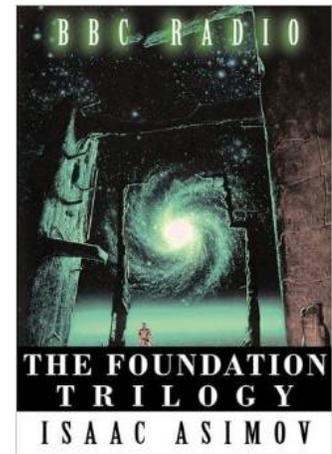


Figure 1.5  
1997: Reboot

Isaac Asimov's science fiction trilogy, that inspired my return to networks.

to join me on this risky network-bound journey.

### FAIL 3: SMALL WORLDS (1998)

I still find puzzling how disjoint were the communities that were thinking about networks before 1999. On one end there was a small but active social network community, whose roots went back to the 1940s. Indeed, much of what we know today about the small world problem is in a little-known paper written around 1960 by the social scientist Ithiel de Sola Pool and the mathematician Manfred Kochen. While their work remained unpublished until 1978 [6], its preprint was widely circulated in the social network community, inspiring Milgram's 1967 small-world experiment [7]. And it was Milgram's work that a quarter century later inspired the playwright John Guare to invent the *six degrees of separation* phrase.

While Pool and Kochen relied on the same models that the graph theorists Erdős and Rényi explored in parallel, no sociology paper showed even the faintest evidence that they were aware of the massive mathematical literature emerging on random graphs. On the other end there was the extensive random graph literature inspired by Erdős and Rényi's pioneering work. Yet, no one in graph theory had any awareness of the social network community, nor they made any reference to small worlds.

This disciplinary gap was reflected in the different questions the two communities asked: The graph theorists worried about phase transitions, subgraphs and giant components, while social scientists were fascinated by small worlds, weak ties and communities. While for social scientists networks with a hundred nodes were beyond comprehension, mathematicians got excited only in the  $N \rightarrow \infty$  limit.

When the Watts and Strogatz paper about small world networks was published in *Nature* in 1998 [8], it first brought back memories of my failed attempt to publish my second network paper in the same journal three years earlier. The roots of my failure became painfully obvious as I read their paper: I had a massive framing problem. Both papers used the random network paradigm. Yet, I asked questions of interest to physicists, while directing the paper to neuroscientists. In contrast, the questions asked by Duncan and Steve were deeply rooted in sociology, six degrees offering a brilliant narrative for their manuscript.

At the same time the small world model appeared to be a dead end for the questions Réka and I were pursuing at that time. As physicists we cared about patterns that could *not* be produced by randomness. Hence we were searching for phenomena that went beyond both regular lattices, the over-explored bread-and-butter of solid state physics, and the purely random network model of Erdős and Rényi. The Watts-Strogatz model interpolated between a regular and a random network, precisely the two limits we sought to avoid. So we set the paper aside, seeing it as a distraction from the path we embarked on. I pulled it out again only months later, when the small world framing offered some unexpected help in our journey.

## MAPPING THE WEB (1998)

When Hawoong Jeong joined my group as a postdoctoral researcher in 1998, Réka and I were already deeply immersed in networks. A graduate of Korea's prestigious Seoul National University, Hawoong's knowledge of computers were prodigious. One night, in the fall of 1998, I dropped by his office to chat about his progress on quantum dots, his main project at that time. Somehow we slipped into networks, prompting me to tell him about my failures to access real data on the topology of the WWW. I asked if he knows how to build a robot, the colloquial term for a Web crawler. He never built one, responded Hawoong, but he was willing to give it a try. And try he did: A few weeks later Hawoong's robot was busily crawling the Web, reviving my failed Plan B to explore the structure of the World Wide Web.

We decided to use the data collected by Hawoong to continue where I left off in 1996 (Figure 1.4), measuring the degree distribution of the WWW. We were motivated by a simple question: Has the WWW reached its percolation threshold? Erdős and Rényi predicted that under a critical link density a network is fragmented into many isolated clusters. Yet once the density reaches a critical threshold a giant component, something that we would perceive as a network, emerges.

Could the WWW be still broken into many disconnected components? Or is it already one big network, as everyone perceived it back then? These were intriguing questions, no matter the outcome. To answer it we needed the Web's degree distribution, which was now provided by Hawoong's robot. The data granted us our first real surprise: We did not see the Poisson distribution that random network theory predicted. A power law greeted us instead.

Hawoong's data was a shocking departure from everything I learned during my four year journey into networks. There was no trace in the literature of a network with a power law degree distribution. In fact, no one seemed to care much about the degree distribution up to that point: Both the random graph and the social network literature took the Poisson form for granted. The power law observed by us predicted that the Web has hubs, nodes with a huge number of links, outliers forbidden in a random universe. None of the existing models could account for them.

According to a surviving email I sent to Hawoong, I started writing my third network paper on March 30, 1999, my 32nd birthday. It was tempting to focus on the true discovery, which was simple: the WWW represents a new type of network, a previously unrecognized form of organization. I sensed, however, that this would be a mistake. By then I was convinced that the failure of my second network paper had little to do with its science, but was a framing problem. Focusing on the inherent, scientific value of the observation was too dry, unlikely to excite the editors of *Nature*. So I decided to use a Trojan horse instead: I hid the discovery under six degrees. We entitled the paper the *Diameter of the World Wide Web*, and trumped up the fact that six is really nineteen on the Web, and mailed it off it to *Nature* [9].

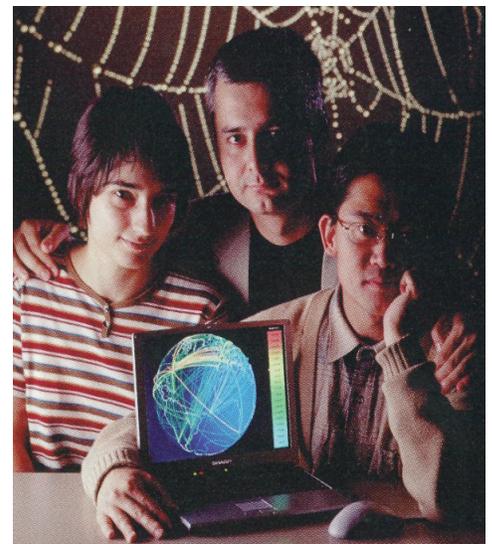


Figure 1.6  
1999: The 19 Degrees Team

A photo taken for Business 2.0 magazine in 2000, showing Réka Albert, Hawoong Jeong and the author, soon after our publication of the paper on the topology of the WWW.

To: ~~AL~~ REKA ALBERT  
 From: A.L. Barabási

Réka: Próbáld meg beprogramozni a következőket  
 ezt modellt; ami esetleg megoldja a véletlenszerű  
 végt.

MODEL 1:  $t=0$ :  $(M)$  vertices,  $(M)$  edges.  
 at time  $t$ : always add a new vertex, which  
 has  $m$  edges (non-directed) coming out of it.  
 The ends of the edges will be randomly  
 connected to the existing vertices in the system.

Thus:  $t=1$ : we have  $(M+1)$  vertices, and  $(M)$  edges,  
 coming out of the new vertex.  
 $t=2$ :  $(M+2)$  vertices;  $2M$  edges  
 $t \geq 2$ :  $(M+t)$  vertices;  $tM$  edges

Thus a new vertex always has  $m$  edges  
 whose ends are connected randomly to the  
 already present vertices in the system.

The average connectivity after time  $t$ :  $\langle k \rangle = \frac{tM}{tM} = 1$

Determine  $P(k)$  at different times  $t_1, t_2, \dots, t_n$   
 (where  $t_1 \dots$  are large),  
 and see if it is exponential or power law!

Model 2: ~~same~~ same as model 1, with  
 the following difference:  
 when you add a new vertex with  $m$   
 edges, you do not connect ~~edges~~ with equal  
 probability to the existing vertices, but:  
 $P_i$ : probability that the new vertex will be  
 connected to vertex  $i$  with  $k_i$  edges is  
 $P_i = \frac{k_i^2}{\sum k_i}$ ; where:  $(\sum k_i)$  is the total  
 number of vertices in the  
 system at time  $t$ .  
 (total degree, proportional)

This model will create flocking: the more connected  
 vertices will grow faster than those less connected,  
 hopefully giving a power law scaling...

Atta Again: determine  $P(k)$  after time  $t_1, t_2, \dots$

Kérlek Farad el a dorabát a következő fax  
 száma: 351 26082622, analóg kontárolt  
 még működik (att) hátul már nem vopok it).

P.S. Most várom, hogy megadalmírd. Model 2 power law at  
 all ( $P(k) \sim k^{-2}$ ), míg a model 1:  $P(k) \sim e^{-k}$ .

Amint le-hat-tam vésztini az a következő:  
 az adott vételek-nél a  $k_i$ -je: Model 1:  $k_i(t) \sim k_i(t) \frac{(k_i(t))}{(M)}$   
 (hogis hogyan né az adott vételek-konnectivitásán (közben)) Model 2:  $k_i(t) \sim t^m$

Figure 1.7  
 1999: Scale-free Fax

The fax sent to Réka Albert on June 14, 1999 from Porto, Portugal. In a mixture of Hungarian and English, the fax describes the algorithm called today the Barabási-Albert model, that explains the origins of scale-free networks, and outlines the continuum theory to calculate the degree exponent (CHAPTER 5).

THE DISCOVERY (1999)

Shortly after the submission I started a two-week long trip to Spain and Portugal, whose last leg was a workshop at the University of Porto. As I drove across the Iberian peninsula, I could not help carrying with me the question: Why hubs? Why power-laws?

To understand what makes the WWW so special, we needed to learn more about other networks. Therefore, before boarding my flight to Europe, I embarked on an aggressive pursuit of additional network maps. The first map came from Jay Brockman, a computer science professor at Notre Dame, who gave us the wiring diagram of a computer chip manufactured by IBM. Duncan Watts sent us the map of the power grid, and Brett Tjaden shared the Hollywood actor database. I left them all with Réka, to analyze them while I was traveling.

On June 14, 1999 I was already in Porto when Réka sent a message detailing some ongoing activities. At the end of her email, more like an afterthought, she added a line: "I looked at the connectivity distribution too, and in almost all systems (IBM, actors, power grid), the tail of the distribution follows a power law."

Réka's sentence hit me like a thunderbolt. I could not pay attention to the talks any longer. My mind was spinning about its implications. If networks as different as the Web and Hollywood display the same power-law degree distribution, then the property we saw earlier on the WWW is universal! Hence some common law or mechanism must be responsible for its

emergence! And if it applies to systems as different as actors, a computer chip or the Web, the explanation needs to be fundamental and simple.

I needed a quiet spot to think about this. So I left the workshop to withdraw to *Casa Diocesana*, the seminary that housed us during the conference.

I did not get too far, however. During the fifteen-minute walk from the University to the seminary I found the explanation. I made a few frantic calculations in my room to formulate the idea in mathematical terms and wrote a fax to Réka, asking her to perform a few numerical simulations to verify my quick conclusions (Figure 1.7).

A few hours later her answer arrived by email. To my great astonishment, it worked. A simple model, relying on only two ingredients, *growth* and *preferential attachment*, could explain the power laws we spotted on the Web and Hollywood.

### THE RUSH (1999)

After Portugal I had only seven days at Notre Dame before my month-long vacation in Transylvania. Yet, I could not see myself sitting for a full month on the discovery. Which meant that I had two more days in Portugal and seven in the US to write the paper.

I wanted to get started right away. My girlfriend reminded me, however, my promise that there would be no work during the last two days of the trip. We planned to take a vacation in Lisbon. So I postponed the writing until the eight hour flight from Lisbon to New York, joining her to explore the city instead. My brain could not free itself of networks, however: The paper was slowly taking shape in my mind as my foot mapped out the narrow streets of Santa Cruz.

Once the plane took off, I started typing frantically. I just finished the manuscript's introduction when the flight attendant, handing a Coke to the passenger next to me, poured the content of the glass on my keyboard. With that she ruined both my brand new laptop and my dream to finish the first draft of the paper on the plane.

I could not give up, however. So I ended up writing the manuscript on a pad the remorseful flight attendant gave me following the accident. And by the end of the week the paper was submitted to *Science*.

I was paranoid, however. At that point we had two manuscripts under submission. The first reported the discovery of scale-free networks in the context of the WWW, which was under review at *Nature*. The second, just submitted to *Science*, showed that the scale-free property was universal and proposed a theory to explain its origin. To be sure, *Nature* and *Science* were the most prestigious journals out there. They are equally infamous, however, for their huge rejection rate. With less than 10% of the submissions pub-

lished, the likelihood that both papers would be rejected was over 81%. More disturbing, the odds that *both* papers would be accepted was less than 1%.

My previous network paper languished for two years before I gave up on it. What if these two manuscripts suffer a similar fate? And what if in the meantime someone else makes the same discovery? The phenomena was so robust and obvious that an independent discovery was rather plausible. I needed a backup plan.

There is an expectation in the physics community that short publications, like a *Nature*, *Science* or a *Physical Review Letters* paper, should be followed by a “long paper,” offering a detailed exposition of the results. So in those seven days Réka, Hawoong and I wrote the long paper as well. I called Gene Stanley, the Editor in Chief of *Physica A*, telling him that I will send him our hottest paper, if he is willing to make a quick decision on it. I doubt that he believed me on the “hot” issue. He promised to act on it quickly, nevertheless.

It took only a few days to learn that my paranoia was well founded. I barely arrived to Csíkszereda, my hometown in Transylvania, when the rejection from *Science* arrived. Disappointed but convinced that the paper was important, I did something that I have never done before: I called the editor who rejected the paper in a desperate attempt to change his mind.

To my astonishment, I succeeded, and he sent the paper for review. And a few months later the 1% scenario came true: *Nature*, *Science* and *Physica A* all accepted our papers [9,10,11]. It was an unexpected but delightful payoff of my five years of setbacks with network science.

### LEAP OF FAITH (1999)

With four students and one postdoc, my research group was modest back then. All but Réka were working on surfaces and quantum dots. A few days after the *Science* paper was accepted I called a group meeting to make an announcement, one that promised to be shocking to several group members. I told them that I was quitting materials science. The reason was simple: I did not want to divide my time and attention any longer between my passion and the topic that paid the bills. So with three years left on my tenure clock, I decided to switch fields, from quantum dots to networks. I offered each group member a choice: Join me in this new journey, or leave.

Two students bailed the ship. The rest followed me on this new, untested voyage.

### FAIL 4: FUNDING (1999)

If you enter an established field and do good work, it is only a matter of time until you get funded. Entering, however, a nonexistent field, does create some puzzling difficulties.

The good news was that I just got awarded a new grant by the National

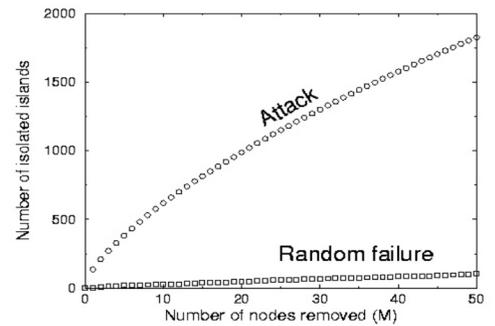


Figure 1.8  
1999: Funding Robustness

A figure from the proposal submitted to DARPA on November 1, 1999, illustrating the impact of the network topology on a network’s error and attack tolerance. The original caption foreshadows our error tolerance paper published a year later in *Nature* [12]:

“The effect of attacks or accidental failures on the connectivity of power networks. We constructed a power network with 40,000 nodes such that the probability that a node has  $k$  links follows  $P(k) \sim k^{-3}$ . Attacks typically target the most connected nodes in the system. To investigate their effect, we removed from the network the  $M$  nodes with the largest number of links. The upper curve shows the number of isolated island as a function of  $M$ , indicating for example, that the removal of only 10 of the largest nodes broke the system into 500 islands, that do not communicate with each other. Random failures uniformly affect any node in the system, and since most nodes have only a few links, failures rarely have a large effect. Indeed, removing randomly  $M$  nodes breaks the system only into a few clusters, leaving the overall connectivity practically unchanged.”

Science Foundation (NSF) to explore quantum dots. The bad news was that I had no interest in pursuing that topic any longer. Of course, I could have just let the funds pour in and pretend that I was carrying on with the subject. I was not comfortable with that option, however. So I called the NSF program manager, asking if I could use the funds to pursue networks instead.

He said no. I either do what I promised, or I need to return the money to NSF. This was a Catch 22: I had money to pursue an incremental project that I completely lost interest in. But I had no funds to pursue the problem that was potentially transformative.

At the end I followed my dream and returned the funds. That, however, left my group in a precarious situation: We desperately needed a new grant but no funding agency considered networks a legitimate research subject. Until I noticed a call by the Defense Advanced Research Projects Agency (DARPA) for technologies that would “allow the networks of the future to be resistant to attacks and continue to provide network services.”

In hindsight it is obvious that the call was intended for networking experts, an active branch of computer science. I was convinced, however, that no one can build fault tolerant networks without first understanding the network topology and its inherent vulnerabilities. The scale-free property we just discovered, with its accompanying hubs, must impact any technology DARPA will develop. With the insights we just reported in *Nature* and *Science* we can nail this, I decided, and immersed myself into proposal writing.

I felt, however, that we need a ‘smoking gun’: something that will convince the program manager of the key role the network topology plays in robustness. Therefore Réka started to randomly remove nodes from a scale-free network, mimicking component failures. She compared the impact of these failures to an attack, when only the hubs were removed. The difference was dramatic: Scale-free networks proved to be surprisingly resistant to random failures but shockingly sensitive to attacks. We quickly incorporated this discovery into the proposal, convinced that we now demonstrated beyond doubt the key role the network topology plays in fault tolerance (Figure 1.8).

After submitting the proposal by the November 1st deadline, I suggested to Réka and Hawoong that the questions we formulated in the proposal were too exciting to wait for the DARPA funding to act on them. We must pursue the problem right away. So we expanded our findings and sent the manuscript to *Science*, only to be rejected without review again.

I called the editor once more, to be told that in his view the paper offered little advances over our previous *Science* paper. I was flabbergasted, but failed to convince him otherwise. So we resubmitted it to *Nature*.

A few months later DARPA rejected our proposal. *Nature*, however, accepted the paper and put it on its cover (Figure 1.9) [12].



Figure 1.9  
2000: Achilles' Heel

The cover of the July 27, 2000 issue of *Nature*, highlighting our paper *Attack and error tolerance of complex networks* inspired by the (unsuccessful) DARPA proposal [12].

## FAIL 5: “COMICALLY WRONG”

Each time there is a major scientific discovery, some researchers will feel that their mission in life is to restore the balance of the Universe by doing everything in their power to erase the topic from the face of the earth. If network science were to live up to its transformative power, it had to have its own nemesis.

“Comically wrong” was the phrase John Doyle, a control theorist from Caltech and a self-proclaimed expert on networks, used in one of his numerous interviews that dotted his decade-long crusade against network science. The small world property was surprising at first, but it was easy to derive and had decades of research backing it. The scale-free property was a different story altogether, raising a host of questions that originally we could not answer. If it was so universal, how did we miss it for decades? Isn’t growth and preferential attachment just one of the many potential explanations? We knew about power laws since Pareto, why is this any different? Béla Bollobás was even more blunt, telling me during our first meeting in the Buda Castle that absent an exact mathematical proof, the scale-free property “does not exist.”

These were legitimate questions with consequences: Only researchers with a strong mathematical training could build on the scale-free concept. The resulting vacuum of understanding gave room for confusion and misinformation. This void was filled by the bombastic statements John Doyle made each time a journalist waved a mike in front of him.

Then slowly the tide turned. José Mendes, Sergey Dorogovtsev and Sid Redner used a rate equation approach to put the continuum theory of scale-free networks on a firm mathematical footing [13,14]. Béla Bollobás and several colleagues in a landmark paper offered the exact proof of the scale-free property [15]. Shlomo Havlin and his students connected robustness to percolation theory [16] and Bollobás and Riordan waved in with an exact proof [17]. A string of discoveries started to document how deeply the scale-free property alters a network’s behavior, like Romualdo Pastor-Satorras and Alessandro Vespignani’s classic result on the disappearance of the epidemic threshold [18]. With that the community started to appreciate the central role the degree distribution plays in networks. As we will see throughout this textbook, virtually all network characteristics, from six degrees to robustness and community structure, must be interpreted in its context. With hundreds of researchers attracted to networks by the many fundamental questions the field posed, gradually network science has taken shape.

## SUMMARY

One could easily view the events recounted above as a string of successes: In the next decade our 1999 *Science* paper on scale-free networks became the most cited paper of physical sciences. The 2000 *Nature* paper on error and attack tolerance not only made the journal’s cover [12], but had a profound impact on our understanding of network robustness. Réka

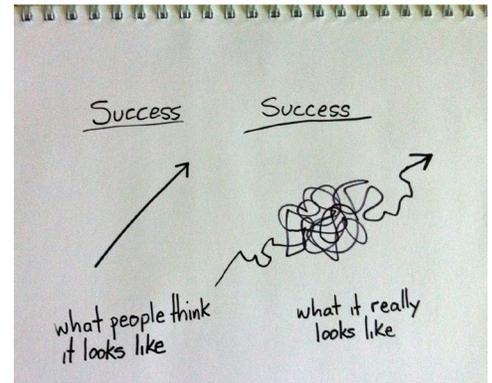
and I spent the subsequent year writing a review on networks, which formalized the field's intellectual foundations, eventually becoming the most cited paper of *Review of Modern Physics* [19]. Then the *National Research Council* report of 2005, published by the US National Academies, coined the term *network science* and persuaded the US government to spend hundreds of millions of dollars to support this new research field as a new and separate discipline. Eventually the most venerable scientific publishers, like Cambridge University Press and Oxford University Press, and the top engineering society, IEEE, have launched journals to cover the field's advances. By all measures a new discipline was born, supported by a vibrant interdisciplinary community.

Science rarely follows a straight path to success (Figure 1.10). New ideas require years of gestation. One could see the theory of scale-free networks as an exception, a spark that took only ten days from the idea to the paper's submission. Yet, had it not been preceded by five years of apparently fruitless work on the problem, the spark could have never started a fire.

Network science offers a reminder of the important role collaboration and mentorship plays in science. Before Réka and Hawoong joined me in this journey, all I could produce was a string of ideas and failures. Without Hawoong's skill to map out the World Wide Web we would have never discovered the scale-free property. Réka's ability to roll the math was essential to develop the theory behind the scale-free model. Our subsequent research on biological networks would have never happened if Zoltán Oltvai, a medical doctor and researcher at Northwestern University, had not convinced us to apply networks to cell biology, patiently guiding us through the maze of proteins and metabolites [20,21,22]. These were not the work of one individual, but truly joint discoveries.

Today many fields consider network science their own. Mathematicians rightly claim ownership and priority through graph theory; the exploration of social networks by sociologists goes back decades; physics lent the universality concept and infused many analytical tools that are now unavoidable in the study of networks; biology invested hundreds of millions of dollars into mapping subcellular networks; computer science offered an algorithmic perspective, allowing us to explore very large networks; engineering invested considerable efforts into the exploration of infrastructural networks. It is remarkable how these many disparate pieces managed to fit together, giving birth to a new discipline.

This textbook is a testimony of the amazing progress that this community has achieved on this fascinating journey. The continued success of network science depends on our ability to maintain its multi-disciplinary nature, allowing each scientist to infuse its unique perspective. This clash of ideas and viewpoints remains the field's strength and its intellectual engine.



**Figure 1.10**  
**Different Paths to Success**

A cartoonist's take on success that vividly captures the convoluted path, punctuated by failures and dead ends, that characterized my early research in networks.



# BIBLIOGRAPHY

[1] A.-L. Barabási. Invasion Percolation and Global Optimization. *Physical Review Letters*, 76:3750, 1996.

[2] B. Bollobás. *Random Graphs*. Cambridge University Press, 2001.

[3] P. Erdős and A. Rényi. On random graphs, I. *Publicationes Mathematicae (Debrecen)*, 6:290-297, 1959.

[4] S. Kauffman. *Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.

[5] A.-L. Barabási. Dynamics of Random Networks: Connectivity and First Order Phase Transitions. <http://arxiv.org/abs/cond-mat/9511052>, 1995.

[6] I. de Sola Pool, M. Kochen. Contacts and influence. *Social Networks*, 1:5-51, 1978.

[7] S. Milgram. The Small World Problem. *Psychology Today*, 2: 60-67, 1967.

[8] D.J. Watts and S.H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393: 409–10, 1998.

[9] H. Jeong, R. Albert and A.L. Barabási. Internet: Diameter of the world-wide web. *Nature*, 401:130-131, 1999.

[10] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509-512, 1999.

[11] A.-L. Barabási, H. Jeong, R. Albert. Mean-field theory for scale free random networks. *Physica A*, 272:173-187, 1999.

[12] R. Albert, H. Jeong, A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406: 378–482, 2000.

- [13] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin. Structure of growing networks with preferential linking, *Physical Review Letters*, 85:4633, 2000.
- [14] P. L. Krapivsky and S. Redner. Statistics of changes in lead node in connectivity-driven networks, *Physical Review Letters*, 89:258703, 2002.
- [15] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18:279-290, 2001.
- [16] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin. Resilience of the Internet to random breakdowns. *Physical Review Letters*, 85:4626, 2000.
- [17] B. Bollobás and O. Riordan. Robustness and Vulnerability of Scale-Free Random Graphs. *Internet Mathematics*, 1, 2003.
- [18] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86:3200–3203, 2001.
- [19] R. Albert and A.-L. Barabási. Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74: 47, 2002.
- [20] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.-L. Barabási . The large-scale organization of metabolic networks. *Nature*, 407:651–655, 2000.
- [21] H. Jeong, S. P. Mason, A.-L. Barabási, and Z.N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411:41-42, 2001.
- [22] A.-L. Barabási, and Z.N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101-113, 2004.