CHAPTER 0

# Personal Introduction

## Introduction

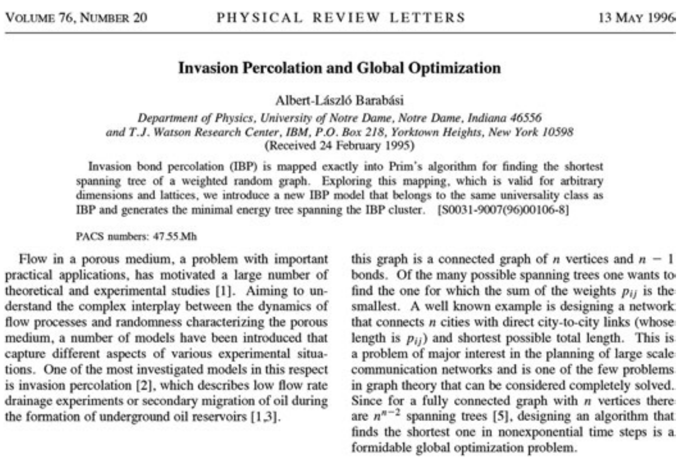
Today, when each year a dozen or so conferences, workshops and schools focus on networks, when over a hundred books and four journals are devoted to the field, when most universities offer network science courses and one can get a PhD in network science on three continents, and when funding agencies have earmarked hundreds of millions of dollars for 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 ours here, 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

Figure 0.1 **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.



VOLUME 76, NUMBER 20 PHYSICAL REVIEW LETTERS 13 MAY 1996

**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].

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.

T. J. Watson Research Center. As the approaching holidays had 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 synonymous with computers, so I went to Watson's library looking for an introduction to 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 a well-known model of statistical physics, called invasion percolation. So exactly two months after Christmas, on February 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 0.1**). While a single-author paper in this prestigious physics journal was undoubtedly a smart career move, its true impact was more far-reaching: the hidden intellectual floodgates 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 pavements. Graph theory envisioned that these networks were wired randomly. That didn't make much sense to me. There must be some organizing principles governing 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], which introduced me to the classical work of Erdős and Rényi [3]. At the same time, Stuart Kaufmann's visionary writing made me appreciate the importance of networks in biology [4]. Two very different perspectives collided in these books: the dry, theorem-driven world of mathematics and the wandering imagination of Stu, which saw no mathematical bounds (**Figure 0.2**).

Eight months into my postdoctoral position I accepted a faculty position at the 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 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 0.3**). 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 (WWW), 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

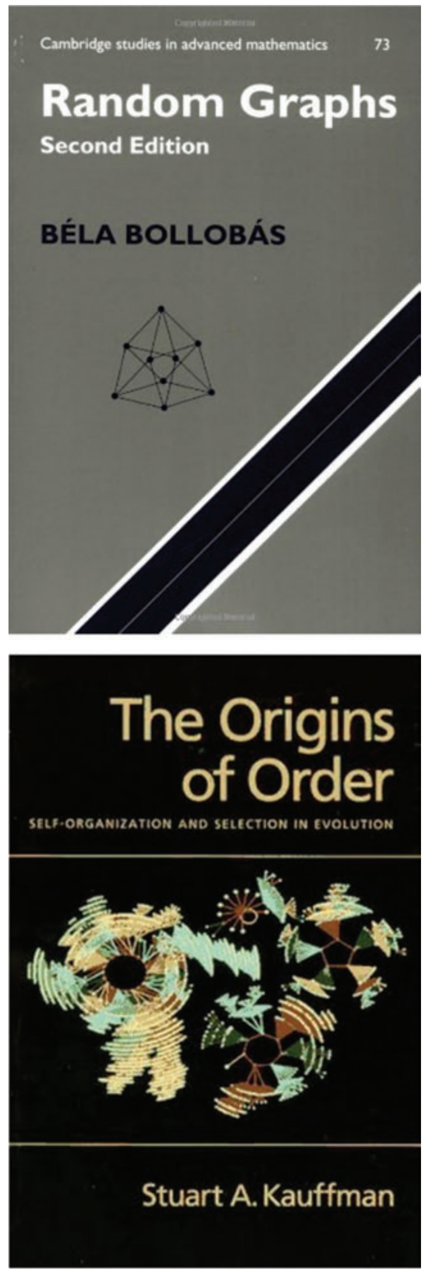
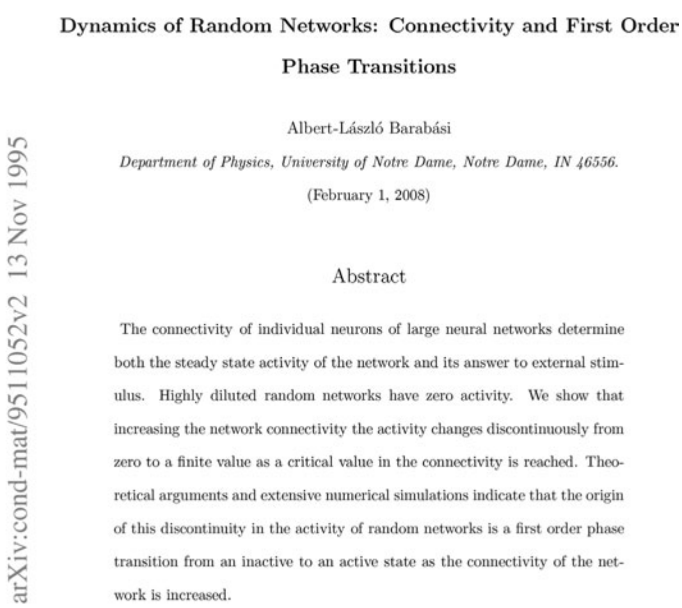



Figure 0.2 **1995: Order or Randomness?**
Two of the three books that inspired my early journey toward network science. I could never track down the first book, whose title (or subtitle) was something like *Fifty Problems in Computer Science*, the one I borrowed in 1994 from the library of IBM's Watson Research Center.


**Dynamics of Random Networks: Connectivity and First Order Phase Transitions**

Albert-László Barabási

*Department of Physics, University of Notre Dame, Notre Dame, IN 46556.*

(February 1, 2008)

arXiv:cond-mat/9511052v2 13 Nov 1995

Abstract

The connectivity of individual neurons of large neural networks determine both the steady state activity of the network and its answer to external stimulus. Highly diluted random networks have zero activity. We show that increasing the network connectivity the activity changes discontinuously from zero to a finite value as a critical value in the connectivity is reached. Theoretical arguments and extensive numerical simulations indicate that the origin of this discontinuity in the activity of random networks is a first order phase transition from an inactive to an active state as the connectivity of the network is increased.

Figure 0.3 **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.

trailed my journey toward network science for the next four years.

On November 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 an 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 the 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 PhD, relying on a language (English) that I had acquired only four years earlier, I could not yet translate my ideas into a story that sticked.

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 November 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 would need to abandon the graph-theoretical path I had pursued thus far. I should instead do what physicists are good at: look at the real world for inspiration. That is, I decided that I needed 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 0.4**), 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 would name only three years later: the degree distribution of the WWW.

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 had 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 had 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 prospects of tenure looked thin. As much as I believed in networks, all I could show for the past three years was one publication and

Figure 0.4 **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

a string of failures. The transition to the more conventional topic paid off, however: by the end of 1997 I was awarded two research grants, allowing me to hire several students and a postdoctoral researcher.

## Reboot (1998)

In 1997 I was living 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 had devoured as a child (**Figure 0.5**). 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 cornfields that surround Route 90, connecting 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 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 had 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 had 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 had 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 had 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

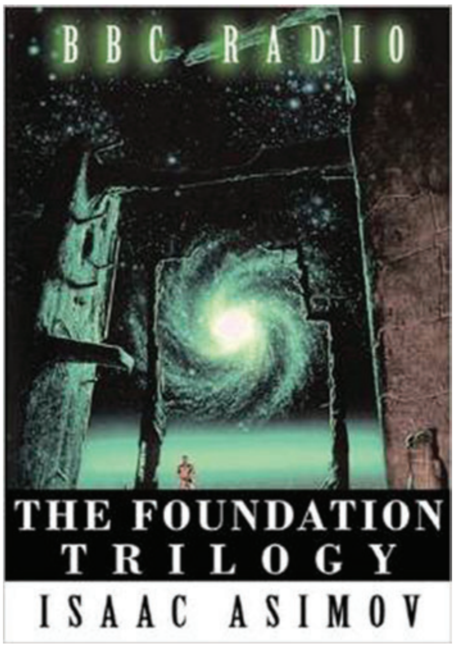



Figure 0.5 **1997: Reboot**
Isaac Asimov's science fiction trilogy that inspired my return to networks.

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 had 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 would 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, exploring 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 was that in her quiet and gracious manner, Réka agreed to join me on this risky network-bound journey.

## Fail 3: Small Worlds (1998)

I still find it puzzling how disjoint the communities were in thinking about networks prior to 1999. On the one hand 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 contained 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 Stanley Milgram's 1967 small-world experiment [7]. And it was Milgram's work that a quarter of a 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 hand 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 did they make 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; the 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 $N \to \infty$ 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. 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 had embarked on. I pulled it out again only months later, when the small-world framing offered some unexpected help on 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 was 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 knew how to build a robot, the colloquial term for a Web crawler. He responded that he had never built one, but 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 WWW.

We decided to use the data collected by Hawoong to continue where I left off in 1996 (**Figure 0.4**), measuring the degree distribution of the WWW. We were motivated by a simple question: Had 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 still be broken into many disconnected components? Or was it already one big network, as everyone perceived it back then? These were intriguing questions, no matter what the outcome. To answer them we needed the Web's degree distribution, which was now being 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 had 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 these.

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,