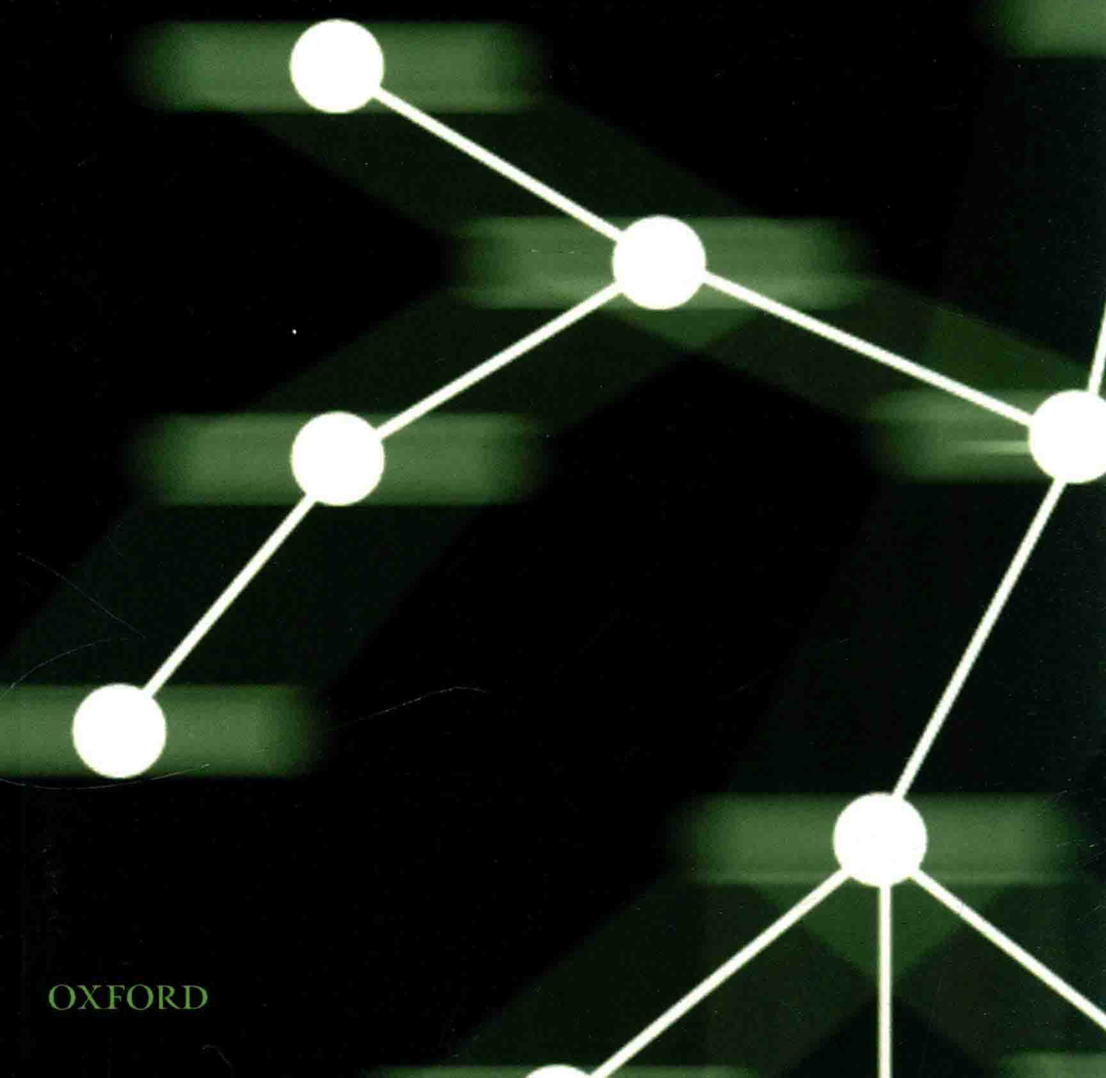
Evolution of Networks

From Biological Nets to the Internet and WWW

S. N. Dorogovtsev | J. F. F. Mendes



OXFORD

Evolution of Networks

From Biological Nets to the Internet and WWW

S. N. Dorogovtsev

University of Aveiro and Ioffe Institute, St Petersburg

J. F. F. Mendes

University of Aveiro

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries

First published 2003
First published in paperback 2013

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

ISBN 978-0-19-851590-6 (hbk.)
ISBN 978-0-19-968671-1 (pbk.)

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Evolution of Networks

PREFACE TO THE PAPERBACK EDITION

This book was written eleven years ago. Eleven years is a long time in science, especially in an explosively developing research area. Importantly, not only has the science of networks rapidly progressed, but also many objects of its study, real-world complex networks, have evolved remarkably. Furthermore, new giant social networking services emerged, including Facebook, Twitter, and Instagram, which influence modern society in many ways. In Section 3.2 we indicated that in October 1999, from the point of view of Altavista, the Web consisted of 271×10^6 pages. Nine years later, on 25 July 2008, the official Google blog announced that the Google index 'hit a milestone: 1 trillion (as in 1 000 000 000 000) unique URLs (Uniform Resource Locator) on the web at once!'[1] In February 2013 Google announced that it processed 10^{11} search requests per month, that is about 1 search request per day per human being, and that the size of the Web already exceeded 3×10^{13} pages.[2] The number of neurons in a human brain is much smaller than that, namely of the order of 10^{11}. Progress in information technologies is enormous. The index of Google, keeping track of the WWW, is already over 100 million gigabytes. This big data provides infinite, unseen possibilities. Surprisingly, the Everest of empirical data on a huge variety of networks collected and analysed during these eleven years, new powerful methods, and emerged research directions have not changed the basic principles and approaches described in this book. We understand networks better now, but this deeper understanding is still essentially based on the ideas developed before 2003. As for comprehensive discussion of discoveries based on new empirical data, the readers will find it in the reference books and texts on complex networks published after 2003 (Barrat *et al.* 2008, Caldarelli 2007, Cohen and Havlin 2010, Dorogovtsev 2010, Newman 2010, Palsson 2006, Pastor-Satorras and Vespignani 2004), and in the detailed reviews (Boccaletti *et al.* 2006, Dorogovtsev *et al.* 2008, Newman 2003).

Without introducing major content updates, here we briefly outline several trends and research directions developed after 2003.

Weighted networks. In these extensively explored networks edges differ from each other. The edges are made individual by ascribing a positive number—a weight—to each of them. The weighted networks enable researchers to quantitatively represent processes and flows in numerous real-world networks. In particular, the weight of an edge in transportation networks is naturally defined as traffic through this edge (Barrat *et al.* 2004).

Motifs in complex networks. The motifs are subgraphs that are present in many copies in a network. They play the role of 'building blocks' of complex

[1] See http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html.

[2] How search works: From algorithms to answers, http://www.google.com/insidesearch/howsearchworks/.

networks. Many efforts were made to relate specific motifs and some functions of networks (Milo *et al.* 2002). Newman (2009) proposed a useful generalization of the standard configuration model of uncorrelated networks, which incorporated motifs. This generalization is actually a random graph with a given sequence of motifs attached to its vertices. In the infinite size limit, this network has no finite loops apart from those within the motifs used in this construction.

Communities in complex networks. Increasing research activity on communities in networks (Fortunato 2010, Newman and Girvan 2004) was determined by numerous perspective applications of these studies to information technologies. The problem is how to efficiently detect poorly distinguishable and overlapping communities in large networks. Most efficient algorithms exploited the spectral properties of networks, particularly the structure of eigenvectors (Newman 2006).

k-cores in networks. The k-core of a network is its largest subgraph whose vertices have at least k connections (within this subgraph). In general, there is a set of successively enclosed k-cores in a network, similarly to a Russian nesting doll—'matrioshka'. The resulting onion-like structure (the system of so-called k-shells) was widely used for characterization of the position of individual vertices within a network (Carmi *et al.* 2007). The birth of a k-core is a discontinuous transition, remarkably combining the discontinuity and a critical power-law singularity (Dorogovtsev *et al.* 2006) These specific hybrid transitions, which are related to the limiting metastable states of a first-order phase transition, have attracted much attention in recent years.

Interdependent and multiplex networks. Many real-world systems are not single networks but rather networks of networks. The significance of these specific networks was realized only recently. In particular, in the interdependent networks, vertices in each network mutually depend on vertices in other networks, so that removal of a fraction of vertices from one of these networks leads to a cascade of back-and-forth damage propagation (Buldyrev *et al.* 2010). This cascade may completely or partially destroy the networks depending on the fraction of initially removed vertices and on the structure of the network. The transition between these two regimes is hybrid, similarly to k-cores. Often the interdependent networks can be reduced to multiplex networks. These have vertices of one type and several different types of edges. In other words, the multiplex networks are graphs with coloured edges. The giant cluster remaining after the cascade of failures has the following property: for every kind of edge and for every two vertices in this cluster, there is an interconnecting path following only edges of this kind within the cluster, which gives a natural generalization of the notion of a connected component.

Evolutionary games on networks. Another issue of great interest is the evolution of individual strategies (cooperation or defection) of the set of players placed on the vertices of a complex network (Szabó and Fáth 2007). In standard evolutionary two-player matrix game models, an individual plays with its neighbours and modifies his or her strategy comparing the resulting pay-off with those of other vertices. The complex structure of a network may change crucially the temporal evolution of this game.

Epidemics in networks. Apart from numerous refined epidemic models (including metapopulation models), many efforts were aimed at a better understanding of the basic SIS epidemic model. The original conclusions of Pastor-Satorras and Vespignani (2000), which we describe in Section 6.9, were based on the so-called heterogeneous mean-field approximation. In this approximation, a network is actually substituted by a fully connected graph with weighted edges (annealed network approximation). One can show that diseases survive longer around hubs in a network, which may result in localized islands of a disease below the original epidemic threshold (Goltsev *et al.* 2012). On the other hand, the SIS model has a final absorbing state in which all the vertices are susceptible. This forces these localized states, which contain a finite number of infective vertices, to decay with time (Lee *et al.* 2013). Note, however, that if highly-connected vertices are sufficiently close to each other, these islands of disease overlap, which leads to a finite fraction of infective vertices and, surprisingly, to a vanishing epidemic threshold in infinite uncorrelated networks with degree distributions decaying slower than an exponential function (Boguñá *et al.* 2013).

Synchronization in networks. Numerous studies were devoted to synchronization in complex networks (Arenas *et al.* 2008). Synchronization transitions differ strongly from phase transitions considered in this book. Specifically, a sharp synchronization may take place in finite systems, even in a set of two coupled oscillators. The major issue was the synchronizability of networks with different architectures.

Optimization driven evolution of networks. The optimality of network design can be regarded as a driving force of network evolution. Despite many efforts, only very simplified optimization based network models were explored. Nonetheless, Papadopoulos *et al.* (2012) found that even a simple model taking into account competition between popularity and similarity in growing networks reproduces surprisingly well the architectures of many real-world networks.

Random walks on networks. After the work of Brin and Page (1998) on the Google PageRank, various problems for random walks and diffusion processes on networks became a topical issue for complex networks. The results of Noh and Rieger (2004) show that the presence of hubs in networks may play a key role in the random walk process. Random walks on evolving networks are particularly interesting (Perra *et al.* 2012). If a network evolves much slower than the random walker moves from vertex to vertex, then the problem is readily reduced to random walks on a static network. The opposite case of a rapidly evolving network and the extremely slow random walker is equivalent to random walks on an annealed network. On the other hand, the non-trivial intermediate situation, in which the time scale of the network reconstruction coincides with hopping times in the random walk process, demands special consideration.

Controllability of networks. Liu *et al.* (2011) applied control theory to the system of coupled linear equations describing the dynamics of states of vertices of directed complex networks. The question was: how many driver nodes can control the dynamics of the entire network? It turned out that homogeneous networks can be easily controlled, while scale-free networks are the most difficult to control.

In his book 'Linked: The New Science of Networks' (2002), László Barabási appealed 'Think networks!' After eleven years, we already observe the common perception of an individuum as a node of a complex system of interconnected evolving networks. We believe that this perception and interpretation of our world in terms of complex networks will be among key factors contributing to the future development of humankind.

Aveiro S.N.D.
June 2013 J.F.F.M.

REFERENCES IN THE PAPERBACK EDITION PREFACE

A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou (2008), Synchronization in complex networks, *Phys. Rep.* **469**, 93.

A.-L. Barabási (2002), *Linked: The New Science of Networks* (Perseus, Cambridge MA).

A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani (2004), The architecture of complex weighted networks, *PNAS* **101**, 3747.

A. Barrat, M. Barthélemy, and A. Vespignani (2008), *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge).

S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang (2006), Complex networks: structure and dynamics, *Phys. Rep.* **424**, 175.

M. Boguñá, C. Castellano, R. Pastor-Satorras (2013), Nature of the epidemic threshold for the Susceptible-Infected-Susceptible dynamics in networks, *Phys. Rev. Lett.* **111**, 068701.

S. Brin and L. Page (1998), The anatomy of a large-scale hypertextual web search engine, *Proc. of the Seventh Int. World Wide Web Conf.* p. 107.

S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, and S. Havlin (2010), Catastrophic cascade of failures in interdependent networks, *Nature* **464**, 1025.

G. Caldarelli (2007), *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford Finance, Oxford University Press, Oxford).

S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir (2007), A model of Internet topology using k-shell decomposition, *PNAS* **104**, 11150.

R. Cohen and S. Havlin (2010), *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge).

S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes (2006), k-core organization of complex networks, *Phys. Rev. Lett.* **96**, 040601.

S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes (2008), Critical phenomena in complex networks, *Rev. Mod. Phys.* **80**, 1275.

S.N. Dorogovtsev (2010), *Lectures on Complex Networks* (Oxford University Press, Oxford); the book companion site: https://sites.google.com/site/sergeydorogovtsev/lectures_on_complex_networks.

S. Fortunato (2010), Community detection in graphs, *Phys. Rep.* **486**, 75.

A.V. Goltsev, S.N. Dorogovtsev, J.G. Oliveira, J.F.F. Mendes (2012), Localization and spreading of diseases in complex networks, *Phys. Rev. Lett.* **109**, 128702.

H.K. Lee, P.-S. Shim, and J.D. Noh (2013), Epidemic threshold of Susceptible-Infected-Susceptible model on complex networks, *Phys. Rev. E* **87**, 062812.

Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási (2011), Controllability of complex networks, *Nature* **473**, 167.

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon (2002), Network motifs: Simple building blocks of complex networks, *Science* **298**, 824.

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

M.E.J. Newman and M. Girvan (2004), Finding and evaluating community structure in networks, *Phys. Rev. E* **69**, 026113.

M.E.J. Newman (2006), Finding community structure in networks using the eigenvectors of matrices, *Phys. Rev. E* **74**, 036104.

M.E.J. Newman (2009), Random graphs with clustering, *Phys. Rev. Lett.* **103**, 058701.

M.E.J. Newman (2010), *Networks: An Introduction* (Oxford University Press, Oxford).

J.D. Noh and H. Rieger (2004), Random walks on complex networks, *Phys. Rev. Lett.* **92**, 118701.

B.Ø. Palsson (2006), *Systems Biology: Properties of Reconstructed Networks* (Cambridge University Press, Cambridge).

F. Papadopoulos, M. Kitsak, M.Á. Serrano, M. Boguñá, and D. Krioukov (2012), Popularity versus similarity in growing networks, *Nature* **489**, 537.

R. Pastor-Satorras and A. Vespignani (2000), Epidemic spreading in scale-free networks, *Phys. Rev. Lett.* **86**, 3200.

R. Pastor-Satorras and A. Vespignani (2004), *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge).

N. Perra, A. Baronchelli, D. Mocanu, B. Gonçcalves, R. Pastor-Satorras, and A. Vespignani (2012), Random walks and search in time-varying networks, *Phys. Rev. Lett.* **109**, 238701.

G. Szabó and G. Fáth (2007), Evolutionary games on graphs, *Phys. Rep.* **446**, 97.

PREFACE

This book is about the growth and structure of random networks. The book is written by physicists and presents the point of view of a physicist, but is addressed to all researchers involved in this subject and students.

Where was physics 50 years ago, and where is it now? At first sight, the role of physics is decreasing; other natural sciences are developing more rapidly. However, physics has penetrated into all sciences. A natural step for a physicist is to jump from the traditional topics of physics to new intriguing problems. Actually, our book describes a flight from physics to the new interdisciplinary field of networks. This escape is, however, still dependent on physics.

For many years the term 'random graphs' usually meant to mathematicians static, 'equilibrium' networks with a Poisson-type distribution of connections. Mathematicians have made truly great advances in the description of such networks.

Only recently have we realized that we reside in a world of networks. The Internet and World Wide Web (WWW) are changing our lives. Our physical existence is based on various biological networks. The extent of the development of communications networks is a good indicator of the level of development in a country. 'Network' turns out to be a central notion in our time, and the explosion of interest in networks is already a social and cultural phenomenon.

Graph theory has made great progress. However, the most important natural and artificial networks have a specific architecture based on a fat-tailed distribution of the number of connections of vertices that differs crucially from the 'classical random graphs' studied by mathematicians. As a rule, these networks are not static but evolving objects. Their state is far from equilibrium and their structure cannot be understood without insight into the principles of their evolution. Only in the last few years have physicists started extensive empirical and theoretical research into networks organized in such a way. Earlier, physicists' interest was rather in neural and Boolean networks where the arrangement of connections was secondary.

We think that the physics approach is the most advantageous for understanding the evolution of networks. Actually, what we physicists are now doing on this active topic is a direct generalization of the usual physics of growth, percolation phenomena, diffusion, self-organized criticality, mesoscopic systems, etc.

Our aim is to understand networks: that is, to understand the basic principles of their structural organization and evolution. We believe that this understanding is necessary to find the best solutions to the problems of real networks.

We decided to present a concise informative book which could be used even by students without a deep knowledge of mathematics and statistical physics and which would be a good source of reference material. Therefore we have tried to introduce the main ideas and concepts in as simple a manner as possible, with minimal mathematics. Special attention is given to real networks, both natural and artificial. We discuss in detail the collected empirical data and numerous real applications of existing theories. The urgent problems of communication networks are highlighted and discussed.

For a description of network evolution, we prefer to use a simpler continuum approach. We feel that it is more important to be understood than to be perfectly rigorous. Also, we follow the hierarchy of values in Western science: an experiment and empirical data are more valuable than an estimate; an estimate is more valuable than an approximate calculation; an approximate calculation is more valuable than a rigorous result. More cumbersome calculations and supplementary materials are placed in appendices. We hope that all of the results and statements that we discuss can be easily found in the text and understood without undertaking detailed calculations. Therefore, we ask our brave readers to skim over difficult pages without hesitation and not to pay any attention to footnotes. However, as this is a monograph written by theoretical physicists, we try to keep a 'physical level' of strictness in our explanations and definitions. Although, we try to avoid superfluous words, we are not afraid to repeat important statements at a different level. We hope that the book will also be useful to mathematicians, as a source of interesting new objects and ideas.

We thank our friends and colleagues for their help. Foremost among these are our collaborators in this field: Alexander V. Goltsev and Alexander N. Samukhin from the Ioffe Institute in St Petersburg. We did not reprint figures with empirical data from original papers but made sketches of data. We are grateful to Albert-László Barabási, Stefan Bornholdt, Jonathan Doye, Jennifer Dunne, Lee Giles, Ramesh Govindan, Byungnam Kahng, Ravi Kumar, Neo Martinez, Sergei Maslov, Mark Newman, Sidney Redner, Ricard Solé, Alessandro Vespignani, and their coauthors for permission to use data from their original figures for derivative reproduction. We are much indebted to John Bulger, Ester Richards, Chris Fowler, David Duckitt, Goutam Tripathy, and Neville Hankins, the copy editor at Oxford University Press for correcting the English of our book. Our computers did not crash only thanks to Miguel Dias Costa and João Viana Lopes. When this book was written, one of us (SND) was on leave from his native Ioffe Institute, and he acknowledges the Centre of Physics of Porto for their support and hospitality.

Porto S.N.D.
May 2002 J.F.F.M.

CONTENTS

0

MODERN ARCHITECTURE OF RANDOM GRAPHS

The first natural questions to ask about a network are:

- What does it look like?
- What is its structure and its topology?
- Is it large or small?
- Why does it have the features it has?
- How did it emerge and develop?
- What can we do with it?

In fact, this book is devoted to a discussion of just these issues. Note that we are not very interested in the internal states of the vertices or edges of networks which are of primary importance in neural and Boolean nets. We restrict ourselves to problems related to the topological structure of random networks, to the evolution of this structure, and to the direct consequences of the particular structural organization of nets.

For many years, the structure of networks with random connections was an object of immense interest for researchers in various sciences, namely mathematics (graph theory), computer science, communications, biology, sociology, economics, etc. Physics may be practically omitted from this incomplete list if we forget neural networks. These sciences provided separate views of distinct networks. A general insight was absent.

In the late 1990s the study of the evolution and structure of networks became a new field of physics. Now we can speak about the statistical physics of networks. What happened?

Here we must explain the strange sounding title of this introduction. How can graphs be modern or traditional? The point is that a few years ago the common interest moved from graphs with a rapidly decreasing distribution of connections (a Poisson degree distribution) to those with a fat-tailed degree distribution; that is, those with many highly connected vertices. The difference between these two architectures is so striking that the transition to the study of the latter actually leads to a revolution in network science.

This transition has been induced by empirical observations of power-law distributions of connections in many real networks, above all, the WWW (Albert, Jeong, and Barabási 1999, Huberman and Adamic 1999) and the Internet (Faloutsos, Faloutsos, and Faloutsos 1999, Govindan and Tangmunarunkit 2000). Mathematical graph theory turned out to be rather a long way from real needs since it focused mainly on 'too simple' static random graphs with Poisson distributions of connections (Erdős and Rényi 1959, 1960), where hubs are not essential. In these random graphs of graph theory, edges are distributed at random between a fixed number of vertices. In our book, we call such simple nets *classical random graphs*. Moreover, mathematicians did not really study evolving random networks.

By the middle of the 1990s, the impact of large, growing communications nets with complex architectures, the Internet and the WWW, on our civilization became incredible. However, understanding of their global organization and functioning was absent.

On the other hand, knowledge of the principles of the evolution and structuration of networks was of vital practical importance. For example, the effective working of search engines is hardly possible without this knowledge. We should note that the first concepts of functioning and practical organization of large communication networks were elaborated by one of the 'parents' of the Internet, Paul Baran (1964). Actually, many present studies develop his initial outstanding ideas and use his terminology. What is the optimal design of communication networks? How can one afford their stability and safety? These and many other vital problems were first studied by Baran at a practical level.

Communication networks are well documented. The data on their structure can be obtained using special programs—*robots*. The most difficult aspect of these empirical studies is that really very large nets are necessary for good statistics. Indeed, the results of such observations are usually various statistical distributions. Finite-size effects cut off their tails and narrow the field of observation. One should note that a number of effects in networks cannot be explained without accounting for their finite sizes. In this sense, most real networks are *mesoscopic objects*. However, for the observation of the fat tails of degree distributions, these finite sizes are only a complicating factor.

Several years ago large artificial nets had approached sizes that allowed statistically reliable data to be obtained. At present, the largest of them, the WWW, contains about 10^9 vertices (the documents of the WWW, that is pages) connected by about 10^{10} edges (hyperlinks). The first empirical study of the WWW (Albert, Jeong, and Barabási 1999) showed that (1) *it is a surprisingly compact network*: the average length of the shortest directed path between two of its randomly chosen pages is only about 19 steps ('clicks'), that is of the order of the logarithm of the WWW size and (2) *the distribution of the numbers of connections of its vertices has an unusual fat-tailed form.*

At present, the first result seems quite obvious. Indeed, this value of the average shortest-path length is typical of networks with random connections. This smallness of the average shortest-path length is usually referred to as the *small-world effect*. Even the introduction of a single shortcut between widely separated sites in a finite lattice essentially reduces the average shortest-path length. A low concentration of shortcuts between randomly chosen lattice sites produces the shortest-path lengths typical of classical random graphs (Watts and Strogatz 1998). In random networks where all edges are actually such shortcuts, the av-