EVOLUTION AND STRUCTURE OF THE INTERNET

A Statistical Physics Approach

This book describes the application of statistical physics and complex systems theory to the study of the evolution and structure of the Internet.

Using a statistical physics approach, the Internet is viewed as a growing system that evolves in time through the addition and removal of nodes and links. This perspective permits us to outline the dynamical theory required for a description of the macroscopic evolution of the Internet. The presence of such a theoretical framework appears to be a revolutionary and promising path towards our understanding of the Internet and the various processes taking place on this network, including, for example, the spread of computer viruses or resilience to random or intentional damages.

The presentation focuses on statistical regularities observed in the large-scale structure of the network, the so-called "global Internet" as well as on the importance of dynamics in the formulation of adequate models. Using this approach it is possible to provide a unified picture of results obtained on the Internet in the context of different scientific communities. This makes use of methods and concepts that have proven to be extremely useful in the analysis of more classical statistical physics systems, such as percolation theory, mean-field methods, and cellular automata simulations.

This book will be of interest to graduate students and researchers in statistical physics, computer science, and mathematics studying the structure and evolution of the internet.

ROMUALDO PASTOR-SATORRAS received his Ph.D. at the University of Barcelona. He has been a research fellow at Yale University and at the Massachusetts Institute of Technology in Cambridge, MA. He spent two years as a research fellow at the International Center for Theoretical Physics (UNESCO) and then moved back to Spain in 2000 as Assistant Professor at the University of Barcelona. Since 2001, Pastor-Satorras has been a research scientist and lecturer at the Universitat Politècnica de Catalunya. He is the author of more than 40 research papers in different areas of non-equilibrium statistical physics, condensed matter theory, and complex systems analysis.

ALESSANDRO VESPIGNANI obtained his Ph.D. at the University of Rome "La Sapienza." After holding research positions at the Universities of Yale and Leiden, he joined the condensed matter research group at the International Center for Theoretical Physics (UNESCO) in Trieste where he has carried out research and teaching activities for more than five years. He has authored more than 100 scientific papers on the statistical physics of non-equilibrium phenomena, critical phase transitions, and complex and disordered systems. At present he is Senior Research Scientist of the French National Council of Scientific Research (CNRS), at the Université de Paris-Sud, France.

EVOLUTION AND STRUCTURE OF THE INTERNET

A Statistical Physics Approach

ROMUALDO PASTOR-SATORRAS
Universitat Politécnica de Catalunya
Barcelona, Spain

ALESSANDRO VESPIGNANI
Laboratoire de Physique Théorique, Université de Paris-Sud
Orsay, France

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2004

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 11/14 pt. *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library

Library of Congress Cataloging in Publication data

Pastor-Satorras, R. (Romualdo), 1967–
Evolution and structure of the Internet: a statistical physics approach / Romualdo Pastor-Satorras,
Alessandro Vespignani.
p. cm.
Includes bibliographical references and index.
ISBN 0 521 82698 5
1. Internet. 2. Statistical physics. I. Vespignani, Alessandro, 1965– II. Title.
TK5105.875.I57P385 2004
004.67′8–dc21 2003047257

ISBN 0 521 82698 5 hardback

Contents

Cambridge University Press
978-0-521-82698-3 - Evolution and Structure of the Internet: A Statistical Physics Approach
Romualdo Pastor-Satorras and Alessandro Vespignani
Frontmatter
More information

Preface

For the majority of people the word "Internet" means access to an e-mail account and the ability to mine data through any one of the most popular public web search engines. The Internet, however, is much more than that. In simple terms, it is a physical system that can be defined as a collection of independently administered computer networks, each one of them (providers, academic and governmental institutions, private companies, etc.) having its own administration, rules, and policies. There is no central authority overseeing the growth of this networks-of-networks, where new connection lines (links) and computers (nodes) are being added on a daily basis. Therefore, while conceived by human design, the Internet can be considered as a prominent example of a self-organized system that combines human associative capabilities and technical skills.

The exponential growth of this network has led many researchers to realize that a scientific understanding of the Internet is necessarily related to the mathematical and physical characterization of its structure. Drawing a map of the Internet's physical architecture is the natural starting point for this enterprise, and various research projects have been devoted to collecting data on Internet nodes and their physical connections. The result of this effort has been the construction of graph-like representations of large portions of the Internet. The statistical analysis of these maps has highlighted, to the surprise of many, a very complex and heterogeneous topology with statistical fluctuations extending over many scale lengths. These are the typical signatures of emergent phenomena, as we observe in nature in many complex systems which are subject to dynamical evolution.

When looking at networks from the point of view of complex systems, the focus is placed on the microscopic processes that rule the appearance and disappearance of nodes and links. Then, since the system is composed of very many interacting units, a detailed evaluation of the dynamics of each unit is avoided in favor of the understanding of the cooperative phenomena originated by their dynamical interactions and the statistical laws governing the system. Such a methodology is akin to the statistical physics approach that has been successfully applied to link the microscopic dynamics and interactions of atoms and matter to the statistical regularities of macroscopic physical systems.

Following this path, an intense research activity has been devoted in recent times to apply the statistical physics approach to the study of complex growing networks in general, and the Internet in particular. In the statistical physics approach the Internet is viewed as a growing system that evolves in time by adding and removing nodes and links. This perspective is somehow opposite to the traditional static graph modeling and allows the identification of some basic models that, while still missing many details, appear to outline the dynamical theory required for the description of the macroscopic Internet's evolution. The presence of such a theoretical framework appears as a revolutionary and promising path in our understanding of the Internet and other complex technological and natural networks.

The introduction of the statistical physics approach into the field of network studies has also provided new techniques and methods with which to approach problems related to network topology, such as resilience to damage and diffusion or searching processes. In this case, well-established techniques in statistical physics, such as percolation theory, mean-field methods, cellular automata simulations, etc., can be used to gain a deeper understanding of the Internet's properties.

The purpose of this book is to provide a unified picture of the results obtained about the Internet in the context of different scientific communities by privileging the use of methods and concepts that have proven to be extremely useful in the analysis of more classical statistical physics systems. We shall therefore make a strong emphasis on the statistical regularities observed in the large-scale structure of the network, the so-called global Internet, and the importance of the dynamics in the formulation of adequate models. In doing this, we have made a special effort to bridge the language gap that might occur among different communities by devoting the two initial chapters to an outline of the Internet's history and an elementary description of its functioning. This will allow us to build up a basic Internet glossary and outline the main elements that make the Internet work. We also provide an appendix summarizing the main concepts of graph theory, which are used in the topological description of Internet maps.

The road map of the book can be schematized in two main parts. The first six chapters are essentially devoted to the physical Internet. In these chapters we review the various experimental projects dealing with data collection, focusing on the various mapping strategies and the level of description achieved with different tools. Following this, we present the statistical analysis of the most recent data available, discussing in detail the main topological features characterizing the Internet's large-scale topology. The ensuing chapter contains an overview of models which propose to represent the Internet. Here we emphasize the "physicist" point of view by introducing the reader to the modern field of growing network models. Finally, we report in Chapter 6 an analysis of the Internet's resilience to damage by casting the problem in the general framework of phase transitions and percolation phenomena.

The second part consisting of Chapters 7, 8, and 9 is instead focused on the virtual networks hosted by the Internet, such as the World Wide Web, peer-to-peer systems, and other social communities, and to dynamical phenomena that occur on them, such as search processes and epidemic spreading. Finally, Chapter 10 is a short discussion of important features that are likely going to represent the main challenges for a full understanding of the Internet in the near future.

The systematic study of the large-scale properties of the Internet and its view as a complex evolving network, while a relatively recent field, has generated quite a large number of works and a vast literature on the subject. We have made every effort to account and mention all the works relevant for a proper understanding of each chapter. It is, however, quite impossible to discuss in detail all the contributions to the field and we have therefore made some choices based on our perception of what is more relevant to the focus of the present book. We apologize to all the colleagues who feel that their specific contributions have been overlooked. We hope that our effort will result in a comprehensive and useful presentation of the subject to everybody working in the field, and, more specially, to any researcher or student who intends to enter it. In this sense, by conveying the idea that the Internet is a paradigmatic example of complex system, we believe that the book can be of interest to computer scientists, physicists, and mathematicians alike.

Many people have contributed to the preparation of this book, specially by shaping our own understanding of the subject. Most of what we know about the Internet is the result of past and present scientific collaborations with M. Boguñá, G. Caldarelli, Y. Moreno, R. Percacci, R. V. Solé, and A. Vázquez. Many of the subtle technicalities of the Internet working has been explained to us by I. Alvarez-Hamelin and S. Visintin to whom goes our deepest gratitude. Our views and knowledge of complex networks have been refined through invaluable scientific discussions with L. Adamic, I. Alvarez-Hamelin, A.-L. Barabási, A. Barrat, G. Bianconi, C. Castellano, A. Flammini, A. Krzywicki, S. Havlin, M. Latapy, A. Lloyd, R. May, F. Menczer, A. Maritan, M. E. J. Newman, R. Percacci, L. Pietronero, N. Schabanel, R. V. Solé and F. van Wijland. The first drafts of the book were read and criticized by I. Alvarez-Hamelin, A.-L. Barabási, A. Barrat, M. Barthelemy, G. Caldarelli, L. Fabbian, A. Flammini, D. Iaschi, C. Magnien, F. Menczer, M.-C. Miguel, Y. Moreno, T. Pernice, A. Vázquez, and M. Vergassola. The book would look very different, and much worse, without their thorough revisions and comments. Many of the figures and plots reported in the book would not be there without the kind help and raw data retrieval of G. Caldarelli, G. Bianconi, S. Bornholdt, H. Ebel, H. Jeong, S. Leonardi, E. Ravasz and A. Vázquez. Particularly, we want to express our warmest gratitude to A.-L. Barabási for the encouragement and the warm hospitality in the University of Notre Dame, where parts of the book were written. We are also grateful to the friends at the International Center for Theoretical Physics in Trieste, where the book writing was initiated. We wish to thank the constant editorial guidance of Simon Capelin and the continuous help and support of the outstanding editorial staff at Cambridge University Press. We also thank the various institutions which have made possible our work through their generous support at the various stages of this project: the FET Open project "COSIN" IST-2001-33555, the Ministerio de Ciencia y Tecnología (Spain) through its program "Ramón y Cajal" and grant BFM 2001–2154, the Action Specifique CNRS/STIC "*Dynamo*" and the BQR "*Graphes du Web*" of Orsay University.

Romualdo Pastor-Satorras
Alessandro Vespignani

Cambridge University Press
978-0-521-82698-3 - Evolution and Structure of the Internet: A Statistical Physics Approach
Romualdo Pastor-Satorras and Alessandro Vespignani
Frontmatter
More information

Abbreviations

APNIC	Asia Pacific Network Information Centre
ARIN	American Registry for Internet Numbers
ARPA	Advanced Research Project Agency
ARPANET	ARPA Network
AS	Autonomous System
BBN	Bolt, Beranek, and Newman Corporation
BGP	Border Gateway Protocol
BITNET	Because It's Time Network
CAIDA	Cooperative Association for Internet Data Analysis
CERN	Counseil Europeen pour la Recherche Nucleaire
CPU	Central Processing Unit
CSNET	Computer Science Research Network
DARPA	Defense Advanced Research Project Agency
DC	Disconnected Components
DNS	Domain Name System
EGP	Exterior Gateway Protocol
FTP	File Transmission Protocol
GIN	Giant In-Component
GOUT	Giant Out-Component
GSCC	Giant Strongly Connected Component
GWCC	Giant Weakly Connected Component
HEPNET	High Energy Physics Research Network
HOT	Heuristically Optimized Trade-off
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
ICMP	Internet Control Message Protocol

IESG	Internet Steering Group
IETF	Internet Engineering Task Force
IGP	Interior Gateway Protocol
IMP	Interface Message Processors
INWG	Internetworking Working Group
IP	Internet Protocol
IPMA	Internet Performance Measurement and Analysis Project
IR	Internet Router
IRR	Internet Routing Registry
ISO	International Organization for Standardization
ISP	Internet Service Provider
LACNIC	Latin American and Caribbean IP address Regional Registry
LAN	Local Area Network
MAN	Metropolitan Area Network
MFENET	Magnetic Fusion Energy Research Nerwork
MILNET	Military Network
MIT	Massachusetts Institute of Technology
NASA	National Aeronautics and Space Administration
NCP	Network Control Protocol
NLANR	National Laboratory for Applied Network Research
NSF	National Science Foundation
NSFNET	NSF Network
OSI	Open System Interconnection
P2P	Peer-to-Peer system
RAM	Random Access Memory
RFC	Request for Comments
RIPE	Réseaux IP Européens
RIP	Routing Information Protocol
RTT	round-trip-time
SIR	Susceptible-Infected-Removed model
SIS	Susceptible-Infected-Susceptible model
SMTP	Simple Mail Transfer Protocol
SPAN	Space Physics Analysis Network
SRI	Stanford Research Institute
TCP	Transmission Control Protocol
TTL	time-to-live
UCLA	University of California at Los Angeles

UDP	User Datagram Protocol
URL	Uniform Resource Locator
USENET	Unix User Network
UUCP	Unix User Control Protocol
WAN	Wide Area Network
WWW	World Wide Web