Network Science

Networks are everywhere, from the Internet, to social networks, and the genetic networks that determine our biological existence. Illustrated throughout in full color, this pioneering textbook, spanning a wide range of topics from physics to computer science, engineering, economics and social sciences, introduces network science to an interdisciplinary audience.

From the origins of the six degrees of separation to explaining why networks are robust to failures and fragile to attacks, the author explores how viruses like Ebola and H1N1 spread, and why it is that our friends have more friends than we do. Using numerous real-world examples, this innovative text includes clear delineation between undergraduate- and graduate-level material. The mathematical formulas and derivations are included within Advanced Topics sections, enabling use at a range of levels. Extensive online resources, including films and software for network analysis, make this a multifaceted companion for anyone with an interest in network science.

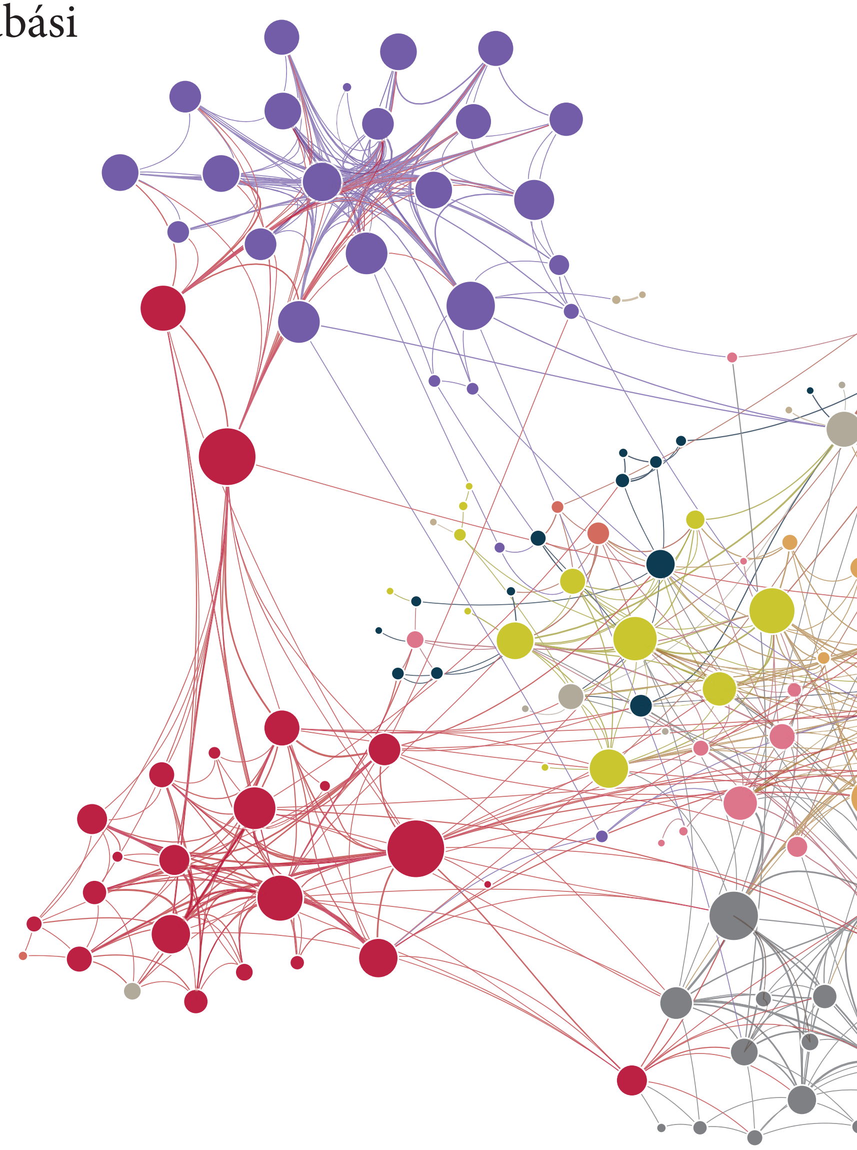
Albert-László Barabási is Robert Gray Dodge Professor of Network Science and Director of the Center for Complex Network Research at Northeastern University, with appointments at the Harvard Medical School and the Central European University in Budapest. His work in network science has led to the discovery of scale-free networks and elucidated many key network properties, from robustness to control.

Cambridge University Press
978-1-107-07626-6 - Network Science
Albert-László Barabási and Márton Pósfai
Frontmatter
More information

Network Science

Albert-László Barabási

with
Márton Pósfai
Data analysis and simulations



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107076266

First published 2016

Printed in the United Kingdom by Bell and Bain Ltd

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

ISBN 978-1-107-07626-6 Hardback

Contents

Cambridge University Press
978-1-107-07626-6 - Network Science
Albert-László Barabási and Márton Pósfai
Frontmatter
More information

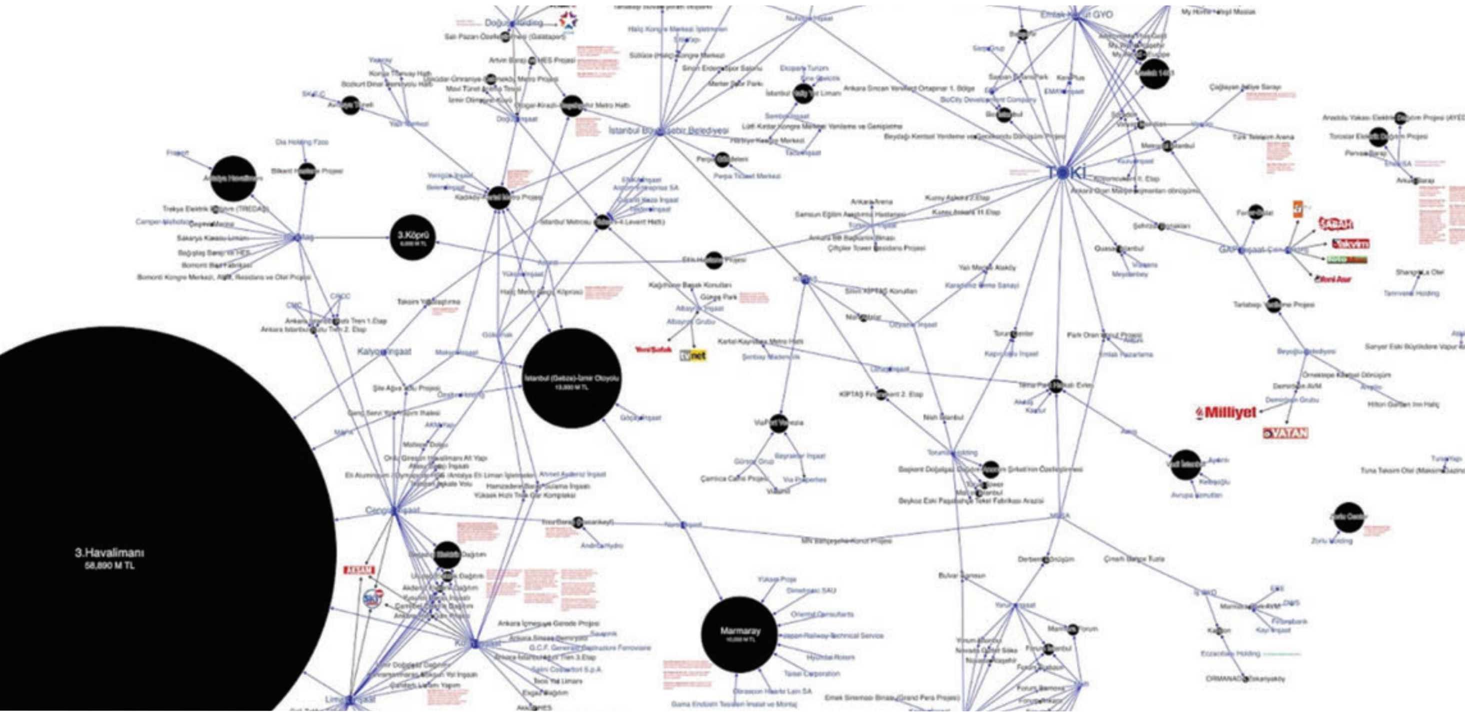


Figure 0 **Art & Networks: Networks of Dispossession**
In 2013 a popular protest movement shook Turkey, prompting thousands of activists and protesters to decamp at Gezi Park. The protests were accompanied by online campaigns, using Twitter or the WWW to mobilize supporters. A central component of this campaign was *Networks of Dispossession*, generated by a coalition of artists, lawyers, activists and journalists that mapped the complex financial relationships behind Istanbul's political and business elite. First exhibited at the Istanbul Biennial in 2013, the map reproduced here shows "dispossession" projects as black circles. The size of each circle represents the monetary value of the project. Corporations and media outlets, shown in blue, are linked directly to their projects. Work-related crimes are noted in red and supporters of Turkey's Olympic bid are shown in purple, while the sponsors of the Istanbul Biennial are in turquoise. The map was developed by Yaşar Adanalı, Burak Arıkan, Özgül Şen, Zeyno Üstün, Özlem Zıngıland and anonymous participants using Graph Commons (http://graphcommons.com/).

Preface

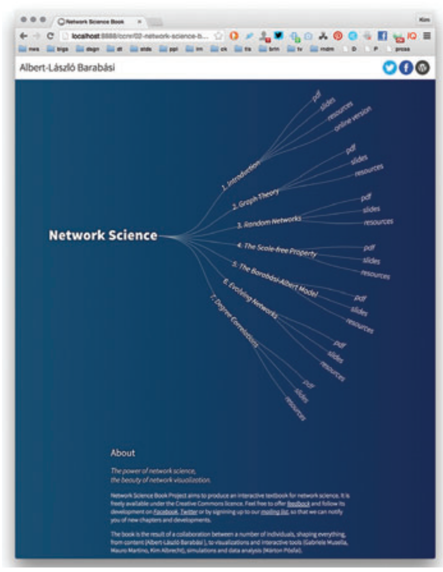
Teaching Network Science

The perspective offered by networks is indispensable for those who wish to understand today's interlinked world. This textbook is the best avenue I have found to share this perspective, offering anyone the opportunity to become a bit of a network scientist. Many of the choices I made in selecting the topics and in presenting the material were guided by the desire to offer a quantitative yet easy-to-follow introduction to the field. At the same time I have tried to pass on the many insights networks offer about the many complex systems that surround us. To resolve these often-conflicting desires, I have paired the technical advances with historical notes and boxes that illustrate the roots and the applications of the key discoveries.

This Preface has two puposes. On the one hand, by describing the class that motivated this text, it offers some practical tips on how to best use the textbook. On the other hand, and equally importantly, it acknowledges the long list of individuals who helped move this textbook forward.

Online Compendium

Network science is rich in content and knowledge that is best appreciated online. Therefore, throughout the chapters we



Online Resource 1
barabasi.com/NetworkScienceBook
The website offers online access to the textbook, the videos, software and interactive tools mentioned in the chapters, the slides I use to teach network science and the datasets analyzed in the book.

encounter numerous **Online resources** that point to pertinent online material – videos, software, interactive tools, datasets and data sources. These resources are available on the http://barabasi.com/NetworkScienceBook website.

The website also contains the PowerPoint slides that I use to teach network science, mirroring the content of this textbook. Anyone teaching networks should feel free to use these slides and modify them as they see fit to offer the best classroom experience. There is no need to ask the author for permission to use these slides in educational settings.

Given the empirical roots of network science, the book has a strong emphasis on the analysis of real networks. We have therefore assembled ten network maps that are frequently used in the literature to test various network characteristics. They were chosen to represent the diversity of the networks explored in network science, describing social, biological, technological and informational systems. The Online Compendium offers access to these datasets, which are used throughout the book to illustrate the tools of network science.

Finally, for those teaching the book in different languages, the website also mirrors the ongoing translation projects.

Network Science Class

I have taught network science in two different settings. The first is a full-semester class that attracts graduate and advanced undergraduate students with a physics, computer science and engineering background. The second is a three-week, two-credit class for students with an economics and social science background. The textbook builds on both teaching experiences. In the full-semester class I cover the full text, integrating into the lectures the proofs and derivations contained in the Advanced topics. In the shorter class I only cover the content of the main sections, omitting the Advanced topics and the chapter on degree correlation (**Chapter 7**).

In both settings key components of the class are assignments and the research project described below.

Homework Problems

For the longer class we assign as homework a subset of the problems listed at the end of each chapter, testing the technical proficiency of the students with the material, and their problem-solving ability. Two rounds of homework cover the material as we progress with the class.

Wiki Assignment

We ask each student to select a concept or a term related to network science and write a Wikipedia page on it (**Figure 1**). What makes this assignment somewhat challenging is that the topic must not already be covered by Wikipedia, yet must be sufficiently notable to be covered. The Wiki assignment tests the students' ability to synthesize and distill material in an easy-to-understand encyclopedic style, potentially turning them into regular Wikipedia contributors. At the same time, the assignment enriches Wikipedia with network science content, offering a service to the whole community. Those teaching network science in other languages should consider contributing to Wikipedia in their native language.

Social Network Analysis

As a warmup to network analysis, students are asked to analyze the social network of the class. This requires a bit of preparation and the help of a teaching assistant. In the very first class the instructor hands out the class list and asks everyone to check that they are on the list or add their name if they are missing. The teaching assistant takes the final list, and during the class prints an accurate class list for each student. At the end of the class each student is asked to mark everyone they knew before coming to the class. To help students match the faces with the names, each student is asked to briefly introduce themselves – also offering a chance for the instructor to learn more about the students in the class. These lists are then compiled to generate a social network of the class, enriching the nodes with gender and the name of the program the students are engaged in. The anonymized version of this network is returned to the class halfway through the course, the assignment being to analyze its properties using the network science tools the students have acquired up to that point. This allows them to explore a relatively small network that they are invested in and understand. The assignment offers a preparation for the more extensive network analysis they will perform for their final research project. This homework is assigned after

WIKI ASSIGNMENT

1. Select a keyword related to network science and check that it is not already covered in Wikipedia. "Related" is defined widely – you can select a technical concept (degree distribution), a network-related concept (terrorist networks), an application of networks (networks in finance), you can write about a network scientist, or anything else that you can convincingly relate to networks.

2. You are not expected to generate original material. Instead you need to identify 2-5 sources on the subject (research papers, books, etc.) and write a succinct, self-contained encyclopedic-style summary with references, graphs, tables, images, photos, as required to best cover the material. Observe Wikipedia's copyright and notability guidelines.

3. Upload your page on Wikipedia and send us the link. You will need to sign up for an account in Wikipedia, as anonymous editors cannot add new pages. Make sure that the page is not deleted by the Wikipedia administrators, which happens when the concept is not well documented or referenced, or is not written in an encyclopedic style.

4. The grade will reflect how understandable, pertinent, self-contained and accurate the content of your page is.

Figure 1 **Wikipedia Assignment Guidelines**

PRELIMINARY PROJECT

Present 5 slides in no more than five minutes:

- Introduce your network, discussing its nodes and links.
- Tell us how you will collect the data and estimate size of the network (N, L). Make sure that $N > 100$.
- Tell us what questions you are planning to ask. We understand that they may change as you advance with your project and the class.
- Tell us why you care about your network.

Figure 2 **Preliminary Project Guidelines**

the hands-on class on software, so that the students are already familiar with the online tools available for network analysis.

Final Research Project

The final project is the most rewarding part of the class, offering students the opportunity to combine and utilize all the knowledge they have acquired. Students are asked to select a network of interest to them, map it out and analyze it. Some procedural details enrich this assignment.

(a) The project is carried out in pairs. If the class composition allows, the students are asked to form professionally heterogenous pairs: undergraduate students are asked to pair up with graduate students, or students from different programs are asked to work together, like a physics student with a biology student. This forces the students to collaborate outside their expertise level and comfort zone, a common ingredient of interdisciplinary research. The instructor does not do the pairing, but students are encouraged to find their own partners.

(b) A few weeks into the course one class is devoted to *preliminary project presentations*. Each group is asked to offer a five-minute presentation with no more than five slides, offering a preview of the dataset they selected (**Figure 2**). Students are advised to collect their own data – simply downloading a dataset already prepared for network analysis is not acceptable. Indeed, one of the goals of the project is to experience the choices and compromises one must make in network mapping. Manual mapping is allowed, like looking up the ingredients of recipes in a cookbook or the interaction of characters in a novel or historical text. Digital mapping is encouraged, like scrapping data from a website or a database that is not explicitly organized as a network map, but the students must reinterpret and clean the data to make it amenable for network analysis. For example, one can systematically scrap data from Wikipedia to identify relationships between writers and scientists or concepts.

(c) It is important to always emphasize that the purpose of the final project is to test a student's ability to analyze a network. Consequently, students must stay focused on exploring the network aspect of the data, and avoid being carried away by other tempting questions their dataset poses that would take them away from this goal.

(d) The course ends with the final project presentations. Depending on the size of the class, we devote one or two classes to this (**Figure 3**).

The choice of Wikipedia keywords, partner selection for the research project and choice of topic for the final project require repeated feedback from the instructor, making sure that all students are on track. To achieve this, the last ten minutes of each class is devoted to asking everyone: Have you chosen a network that you wish to analyze? What are your nodes and your links? Do you know how to get the data? Do you have a partner for your final project? What is your Wiki word? Did you check if it is already covered by Wikipedia? Did you collect literature pertaining to it? The answers range from "Not yet," to firm or vague ideas the students are entertaining. Providing public feedback about the appropriateness and the feasibility of their plans helps those who are behind to crystallize their ideas, and to identify potential partners with common interests. After a few classes typically everyone finds a partner and identifies a research project and a Wikipedia keyword, at which point this end-of-class ritual finishes.

Software

We devote one class to various network analysis and visualization software, like Gephi, Cytoscape or NetworkX. In the longer course we devote another class to other numerical procedures, like fitting, log-binning or network visualization. We ask students to bring their laptops to these classes, so that they can try out these tools immediately.

Movie Night

We devote one night, typically outside class time, to a movie night, where we screen the documentary *Connected* by Annamaria Talas. The one-hour documentary features many contributors to network science, and offers a compelling narrative of the field's importance. Movie night is advertised university wide, offering a chance to reach out to the wider community.

FINAL PROJECT

Each group has 10 minutes to present their final project. Time limit is strictly enforced. On the first slide, give your title, name and program.

Tell us about your data and the data collection method. Show an entry of the data source to offer a sense of where you started from.

Measure: N, L, and their time dependence if you have a time-dependent network; degree distribution, average path length, clustering coefficient $C(k)$, the weight distribution $P(w)$ if you have a weighted network. Visualize communities; discuss network robustness and spreading, degree correlations, whichever is appropriate for your project.

It is not sufficient to simply measure things - you need to discuss the insights you gained, always asking:

- What was your expectation?
- What is the proper random reference?
- How do the results compare to your expectation?
- What did you learn from each quantity?

Grading criteria:

- Use of network tools (completeness/correctness);
- Ability to extract information/insights from your data using the network tools;
- Overall quality of the project/presentation.

No need to write a report - email us the presentation as a pdf file.

Figure 3 **Final Project Guidelines**

COMPLEX NETWORKS: SYLLABUS

Week 1
- **Class 1** *Ch. 1: Introduction*
- **Class 2** *Ch. 2: Graph Theory*

Week 2
- **Class 1** *Ch. 3: Random Networks*
- **Class 2** *Ch. 3: Random Networks*

Week 3
- **Class 1** *Ch. 4: The Scale-Free Property*
- **Class 2** *Ch. 4: The Scale-Free Property*

Hand-out Assignment 1 (Problems for Chapters 1-5)

Week 4
- **Class 1** *Ch. 5: The Barabási-Albert model*
- **Class 2** *Ch. 5: The Barabási-Albert model*

Week 5
- **Class 1** *Preliminary Project Presentations*
- **Class 2** *Hands-on Class: Graph representation, binning, fitting*

Week 6
- **Class 1** *Hands-on Class: Gephi and Python*

Collect Assignment 1;
Hand-out Assignment 2: Class Network Analysis
- **Class 2** *Guest Speaker*

Week 7
- Class 1 *Ch. 6: Evolving Networks*
- Class 2 *Ch. 6: Evolving Networks*

Week 8
- **Class 1** *Guest Speaker*

Collect Assignment 2
- **Class 2** *Ch. 7: Degree Correlations*

Hand out Assignment 3 (Problems for Chapters 6-10)

Week 9
- **Class 1** *Ch. 8: Network Robustness*

Hand out Assignment 4: Wikipedia Page
- **Class 2** *Ch. 8: Network Robustness*

Week 10
- **Class 1** *Ch. 9: Communities*
- **Class 2** *Ch. 9: Communities*

Movie Night: *Connected*, by Annamaria Talas

Week 11
- **Class 1** *Ch. 10: Spreading Phenomena*
- **Class 2** *Ch. 10: Spreading Phenomena*

Week 12
- **Class 1** *Guest Speaker*
- **Class 2** *Ch. 10: Spreading Phenomena*

Collect Assignment 4

Week 13
- **Class 1** *Guest Speaker*
- **Class 2** *Open-Door class (Research Project Discussions)*

Collect Assignment 3

Week 14
- **Exam Week** *Final Project Presentations (10 min per group)*

Figure 4 **The Syllabus**
The week-by-week schedule of the four-credit network science class, which meets twice a week.

CAMBRIDGE

Cambridge University Press
978-1-107-07626-6 - Network Science
Albert-László Barabási and Márton Pósfai
Frontmatter
More information

Guest Speakers

In the full-semester class we invite researchers from the area to give research seminars about their work pertaining to networks. This offers the students a sense of what cutting-edge research looks like in this area. This is typically (but not always) done toward the end of the class, by which point most theoretical tools have been covered and the students are focusing on their final project. Such talks, advertised and open to the local research community, often inspire additional perspectives and ideas for the final project.

Figure 5 details the grading system used for the one-semester class. To aid in planning, **Figure 4** offers the schedule of the full-semester class I co-taught before this book went to print.

GRADE DISTRIBUTION

(1) Assignment 1 (Homework 1): 15%
(2) Assignment 2 (Homework 2): 15%
(3) Assignment 3 (Class Network): 15%
(4) Assignment 4 (Wikipedia): 15%
(5) Preliminary Project Presentation: No grading, only feedback.
(6) Final Project: 40%

Figure 5 **Grading**
The grading system used in the one-semester class.

Acknowledgments

Writing a book, any book, is an exercise in lonely endurance. This project was no different, dominating all my free time between 2011 and 2015. It was mostly time spent alone, working in one of the many coffeehouses I frequent in Boston and Budapest, or wherever in the world the morning found me. Despite this, the book is far from being a lonely achievement: during these four years a number of individuals donated their time and expertise to help move the project forward, offering me the opportunity to discuss the subject with colleagues, friends and lab members. I also shared the chapters on the Internet for everyone to use, receiving valuable feedback from many individuals. In this section I wish to acknowledge the professional network that stepped in to help at various stages of this long journey.

Formulas, Graphs, Simulations

A textbook must ensure that everything works as promised: that one can derive the key formulas, and that the measures described in the text – when applied to real data – work as the theory predicts. There is only one way to achieve this: one must check and repeat each calculation, measurement and simulation. This was a heroic job, most of it done by *Márton Pósfai* (**Figure 6**), who joined the project when he was a visiting

Figure 6 **The Math Team**
Márton Pósfai was responsible for the calculations, simulations and measurements in the textbook.

student in my lab in Boston and stayed with it throughout his PhD work in Budapest. He checked all derivations, if necessary helped re-derive key formulas, performed all the simulations and measurements and prepared the book's figures and tables. Many figures and tables amounted to small research projects, their outcome forcing us to de-emphasize some quantities because they did not work as promised, or helping us appreciate and understand the importance of others. His deep understanding of the network science literature and his careful work offered many subtle insights that enriched the book. There is no way I could have achieved this depth and reliability if it wasn't for Márton's tireless dedication to the project.

The Design

The ambition to create a book that had a clear aesthetic and visual appeal was planted by *Mauro Martino*, a data visualization expert in my lab. He created the first face of the chapters and many visual elements designed by him stayed with us until the end. After Mauro moved on to lead a team of designers at IBM Research, *Gabriele Musella* took over the design. He standardized the color palette and designed the basic elements of the info-graphics appearing throughout the book, also redrawing most images. He worked with us until the fall of 2014, when he too had to return to London to take up his dream job. At that time the design was taken over by *Nicole Samay*, who tirelessly and gently retouched the whole book as we neared the finish line (**Figure 7**). The website for the book was designed by *Kim Albrecht*, who is currently collaborating with Mauro to design the online experience that trails the book.

An important component of the visual design are the images included at the beginning of each chapter, illustrating the

Figure 7 **The Design Team**
Mauro Martino, Gabriele Musella and Nicole Samay developed the look and feel of the chapters and the figures, offering the book an elegant and consistent style. Kim Albrecht designed the online experience that trails the book.

interplay between networks and art. In selecting these images I have benefited from advice and discussions with several artists and designers, academics and practicing artists alike. Many thanks go to *Isabel Meirelles* and *Dietmar Offenhuber* from the Art and Design Department at Northeastern, *Mathew Ritchie* from Columbia University and *Meredith Tromble* from the San Francisco Art Institute, for helping me navigate the boundaries of art, data and network science.

The Daily Drill: Typing, Editing

I remain an old-fashioned writer, who writes with a pencil rather than a computer. I am lost, therefore, without editors and typists, who integrate my hand-written notes, corrections and recommendations into each chapter. *Sabrina Rabello* and *Galen Wilkerson* helped get this project started, yet the bulk of editing fell on the shoulders of three individuals (**Figure 8**): *Payam Parsinejad* worked with me during the first year of the project; after he had to refocus on his research, *Amal Al-Husseini*, a former student from my network science class, joined us and stayed until the very end; equally defining was the help of *Sarah Morrison*, my former assistant, who joined the project after she moved to Lucca, Italy. Her timely and accurate editing were essential in finishing the book.

Each chapter, before being released on our webpage, has undergone a final check by *Philipp Hoevel* (**Figure 9**), who

Figure 8 **The Editorial Team**
Payam Parsinejad, Amal Al-Hussieni and Sarah Morrison worked on the book on a daily basis, editing and correcting it.

Figure 9 **Accuracy and Rights**
Philipp Hoevel acted as our first reader and last editor. The rights were obtained and managed by Brett Common.

Figure 10 **Homework**
Roberta Sinatra conceived and compiled the homework at the end of each chapter in the textbook.

joined the project while visiting my lab and continued to work with us even after he returned to Berlin to run his own lab. Philipp methodically reviewed everything, from the science to notations, becoming our first reader and final filter.

Brett Common has worked tirelessly to secure all the permissions for the visual materials used throughout the textbook. This was a major project on its own, whose magnitude and difficulty was hard to anticipate.

Homework

The homework at the end of each chapter was conceived and curated by *Roberta Sinatra* (**Figure 10**). As a research faculty affiliated with my lab, Roberta co-taught the network science class with me in the fall of 2014, helping also to catch and correct many typos and misunderstandings that surfaced while teaching the material.

Science Input

Throughout the project I have received comments, recommendations, advice, clarifications and key material from numerous scientists and students. It is impossible to recall them all, but I will try.

Chaoming Song helped estimate the degree exponent of scale-free networks and helped me uncover the literature pertaining to cascading failures. The mathematician *Endre Csóka* helped clarify the subtle details of the Bollobás model. I benefited from a great discussion with *Raissa D'Souza* on optimization models, with *Ginestra Bianconi* on the fitness model and with *Erzsébet Ravasz Reagan* on the Ravasz algorithm. *Alex Vespignani* was a great resource on spreading processes and degree correlations. *Marian Boguña* snapped the picture for the Karate Club Trophy. *Huawei Shen* calculated the future citations of research papers. *Gergely Palla* and *Tamás Vicsek* helped me understand the CFinder algorithm and *Martin Rosvall* pointed us to some key material on the InfoMap algorithm. *Gergely Palla, Sune Lehmann* and *Santo Fortunato* offered critical comments on the community detection chapter. *Yong-Yeol Ahn* helped me develop an early version of the material on spreading phenomena and *Kate Coronges* helped improve the clarity of the first four chapters. *Ramis Movassagh, Hiroki Sayama* and *Sid Redner* provided careful feedback on several chapters.

Publishing

Simon Capelin, my longtime editor at Cambridge University Press, had been encouraging this project even before I was ready to write it. He also had the patience to see the book to its completion, through many missed deadlines. *Róisín Munnelly* helped move the book through production.

Institutions

This book would not have been possible if several institutions had not offered inspiring environments and a supporting infrastructure. First and foremost I need to thank the leadership of *Northeastern University*: President *Joseph Aoun*, Provost *Steve Director*, my deans *Murray Gibson* and *Larry Finkelstein* and my department chair *Paul Champion*, who were true champions of network science, turning it into a major cross-disciplinary topic within Northeastern. Their relentless support has led to the hiring of several superb faculty focusing on networks, spanning all domains of inquiry, from physics and mathematics to social, political, computer and health sciences, turning Northeastern into the leading institution in this area. They have also urged and supported the creation of a network science PhD program and helped found the *Network Science Institute* led by *Alessandro Vespignani*.

My appointment at *Harvard Medical School*, through the *Network Medicine Division* at *Brigham and Women's Hospital* and the *Center for Cancer Systems Biology* at *Dana Farber Cancer Institute (DFCI)*, offered a window on the applications of network science in cell biology and medicine. Many thanks to *Marc Vidal* from DFCI and *Joe Loscalzo* from Brigham who, as colleagues and mentors, have defined my work in this area – an experience that found its way into this book as well.

My visiting appointment at *Central European University (CEU)*, and the network science class I teach there in the summer, have exposed me to a student body with an economics and social science background, an experience that has shaped this textbook. *Balázs Vedres* had the vision to bring network science to CEU, *George Soros* convinced me to get involved with the university and President *John Shattuck* and Provosts *Farkas Katalin* and *Liviu Matei*, with their relentless support, have smoothed the path toward CEU's superb program in this area, giving birth to CEU's PhD program in network science.

Thanks to the place where it all began: when I was a young Assistant Professor, the *University of Notre Dame* offered me the support and the serene environment to think about something different. And big thanks to *Suzanne Aleva*, who followed my lab from Notre Dame to Northeastern, and worked tirelessly for over a decade to foster an environment where I can focus, uninterrupted, on science.

Finally, I am truly indebted to my children, Dániel, Izabella and Lénárd, and my wife, Janet, who accepted that I devote countless hours to this book. Without their understanding and patience I would never have finished *Network Science*.

Cambridge University Press
978-1-107-07626-6 - Network Science
Albert-László Barabási and Márton Pósfai
Frontmatter
More information

Figure 0.0 **Art & Networks: Chiharu Shiota**
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 embraces the inherent tension between order and disorder, the very topic of network science.