NSF-CBMS Regional Research Conference on
The Combinatorics of Large Sparse Graphs,
held at California State University San Marcos,
June 7–12, 2004

Partially supported by the National Science Foundation

2000 Mathematics Subject Classification. Primary 05Cxx, 68R10, 68W20, 90B10, 90C06, 90C35, 94C15.

For additional information and updates on this book, visit
www.ams.org/bookpages/cbms-107

Library of Congress Cataloging-in-Publication Data
Chung, Fan R. K., 1949–
Complex graphs and networks / Fan Chung, Linyuan Lu.
p. cm. — (CBMS regional conference series in mathematics ; no. 107)
Includes bibliographical references and index.
QA166.C48 2006
511.5—dc22 2006042898

Copying and reprinting. Individual readers of this publication, and nonprofit libraries acting for them, are permitted to make fair use of the material, such as to copy a chapter for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgment of the source is given.

Republication, systematic copying, or multiple reproduction of any material in this publication is permitted only under license from the American Mathematical Society. Requests for such permission should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA. Requests can also be made by e-mail to reprint-permission@ams.org.

© 2006 by the American Mathematical Society. All rights reserved.
The American Mathematical Society retains all rights except those granted to the United States Government.
Printed in the United States of America.

The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at http://www.ams.org/
Contents

Preface vii

Chapter 1. Graph Theory in the Information Age 1
  1.1. Introduction 1
  1.2. Basic definitions 3
  1.3. Degree sequences and the power law 6
  1.4. History of the power law 8
  1.5. Examples of power law graphs 10
  1.6. An outline of the book 17

Chapter 2. Old and New Concentration Inequalities 21
  2.1. The binomial distribution and its asymptotic behavior 21
  2.2. General Chernoff inequalities 25
  2.3. More concentration inequalities 30
  2.4. A concentration inequality with a large error estimate 33
  2.5. Martingales and Azuma’s inequality 35
  2.6. General martingale inequalities 38
  2.7. Supermartingales and Submartingales 41
  2.8. The decision tree and relaxed concentration inequalities 46

Chapter 3. A Generative Model — the Preferential Attachment Scheme 55
  3.1. Basic steps of the preferential attachment scheme 55
  3.2. Analyzing the preferential attachment model 56
  3.3. A useful lemma for rigorous proofs 59
  3.4. The peril of heuristics via an example of balls-and-bins 60
  3.5. Scale-free networks 62
  3.6. The sharp concentration of preferential attachment scheme 64
  3.7. Models for directed graphs 70

Chapter 4. Duplication Models for Biological Networks 75
  4.1. Biological networks 75
  4.2. The duplication model 76
  4.3. Expected degrees of a random graph in the duplication model 77
  4.4. The convergence of the expected degrees 79
  4.5. The generating functions for the expected degrees 83
  4.6. Two concentration results for the duplication model 84
  4.7. Power law distribution of generalized duplication models 89

Chapter 5. Random Graphs with Given Expected Degrees 91
  5.1. The Erdős-Rényi model 91
  5.2. The diameter of $G_{n,p}$ 95
## CONTENTS

5.3. A general random graph model 97
5.4. Size, volume and higher order volumes 97
5.5. Basic properties of $G(w)$ 100
5.6. Neighborhood expansion in random graphs 103
5.7. A random power law graph model 107
5.8. Actual versus expected degree sequence 109

Chapter 6. The Rise of the Giant Component 113
6.1. No giant component if $w < 1$? 114
6.2. Is there a giant component if $\bar{w} > 1$? 115
6.3. No giant component if $\bar{w} < 1$? 116
6.4. Existence and uniqueness of the giant component 117
6.5. A lemma on neighborhood growth 126
6.6. The volume of the giant component 129
6.7. Proving the volume estimate of the giant component 131
6.8. Lower bounds for the volume of the giant component 136
6.9. The complement of the giant component and its size 138
6.10. Comparing theoretical results with the collaboration graph 141

Chapter 7. Average Distance and the Diameter 143
7.1. The small world phenomenon 143
7.2. Preliminaries on the average distance and diameter 144
7.3. A lower bound lemma 146
7.4. An upper bound for the average distance and diameter 147
7.5. Average distance and diameter of random power law graphs 149
7.6. Examples and remarks 158

Chapter 8. Eigenvalues of the Adjacency Matrix of $G(w)$ 161
8.1. The spectral radius of a graph 161
8.2. The Perron-Frobenius Theorem and several useful facts 162
8.3. Two lower bounds for the spectral radius 163
8.4. An eigenvalue upper bound for $G(w)$ 164
8.5. Eigenvalue theorems for $G(w)$ 165
8.6. Examples and counterexamples 169
8.7. The spectrum of the adjacency matrix of power law graphs 170

Chapter 9. The Semi-Circle Law for $G(w)$ 173
9.1. Random matrices and Wigner’s semi-circle law 173
9.2. Three spectra of a graph 174
9.3. The Laplacian of a graph 175
9.4. The Laplacian of a random graph in $G(w)$ 176
9.5. A bound for random graphs with large minimum degree 177
9.6. The semi-circle law for Laplacian eigenvalues of graphs 179
9.7. An upper bound on the spectral norm of the Laplacian 180
9.8. Implications of Laplacian eigenvalues for $G(w)$ 185
9.9. An example of eigenvalues of a random power law graph 187

Chapter 10. Coupling On-line and Off-line Analyses of Random Graphs 189
10.1. On-line versus off-line 189
10.2. Comparing random graphs 190
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.3</td>
<td>Edge-independent and almost edge-independent random graphs</td>
<td>194</td>
</tr>
<tr>
<td>10.4</td>
<td>A growth-deletion model for random power law graphs</td>
<td>198</td>
</tr>
<tr>
<td>10.5</td>
<td>Coupling on-line and off-line random graph models</td>
<td>200</td>
</tr>
<tr>
<td>10.6</td>
<td>Concentration results for the growth-deletion model</td>
<td>205</td>
</tr>
<tr>
<td>10.7</td>
<td>The proofs of the main theorems</td>
<td>215</td>
</tr>
</tbody>
</table>

Chapter 11. The Configuration Model for Power Law Graphs

11.1. Models for random graphs with given degree sequences | 223 |
11.2. The evolution of random power law graphs | 224 |
11.3. A criterion for the giant component in the configuration model | 225 |
11.4. The sizes of connected components in certain ranges for $\beta$ | 225 |
11.5. The distribution of connected components for $\beta > 4$ | 229 |
11.6. On the size of the second largest component | 232 |
11.7. Various properties of a random graph of the configuration model | 236 |
11.8. Comparisons with realistic massive graphs | 237 |

Chapter 12. The Small World Phenomenon in Hybrid Graphs

12.1. Modeling the small world phenomenon | 241 |
12.2. Local graphs with many short paths between local edges | 242 |
12.3. The hybrid power law model | 244 |
12.4. The diameter of the hybrid model | 248 |
12.5. Local graphs and local flows | 250 |
12.6. Extracting the local graph | 251 |
12.7. Communities and examples | 253 |

Bibliography | 255 |

Index | 261 |
Preface

In many ways, working on graph theory problems over the years has always seemed like fun and games. Recently, through examples of large sparse graphs in realistic networks, research in graph theory has been forging ahead into an exciting new dimension. Graph theory has emerged as a primary tool for detecting numerous hidden structures in various information networks, including Internet graphs, social networks, biological networks, or more generally, any graph representing relations in massive data sets.

How will we explain from first principles the universal and ubiquitous coherence in the structure of these realistic but complex networks? In order to analyze these large sparse graphs we will need to use all the tools at our disposal, including combinatorial, probabilistic and spectral methods. Time and again, we have been pushed beyond the limit of the existing techniques and have had to create new and better tools to be able to analyze these networks. The examples of these networks have led us to focus on new, general and powerful ways to look at graph theory. In the other direction, we hope that these new perspectives on graph theory contribute to a sound scientific foundation for our understanding of the discrete networks that permeate this information age.

This book is based on ten lectures given at the CBMS Workshop on the Combinatorics of Large Sparse Graphs in June 2004 at the California State University at San Marcos. Various portions of the twelve chapters here are based on several papers coauthored with many collaborators. Indeed, to deal with the numerous leads in such an emerging area it is crucial to have partners to sound out the right approaches, to separate what can be rigorously proved and under what conditions from what cannot be proved, to face seemingly overwhelming obstacles and yet still gather enough energy to overcome one more challenge. Special thanks are due to our coauthors, including Bill Aiello, Reid Andersen, David Galas, Greg Dewey, Shirin Handjani, Doug Jungreis, and Van Vu.

We are particularly grateful to Ross Richardson and Reid Andersen for many beautiful illustrations in the book and to the students in Math261 spring 2004 at UCSD for taking valuable lecture notes. In the course of writing, we have greatly benefitted from discussions with Alan Frieze, Joe Buhler and Herb Wilf. Most of all, we are indebted to Steve Butler and Ron Graham for their thoughtful readings and invaluable comments without which this book would not have so swiftly converged.

Fan Chung and Lincoln Lu, May 2006
This page intentionally left blank
Bibliography


[52] P. Erdős and T. Gallai, Gráfok előírt fokú pontokkal (Graphs with points of prescribed degrees, in Hungarian), Mat. Lapok 11 (1961), 264–274.
[76] M. Henzinger, private communication.
This page intentionally left blank
Index

\( \beta, \) 7
\( d, \) 98, 104
\( \bar{d}, \) 98, 104
\( \bar{d}_x, \) 98
\( D(G), \) 186
\( D(X,Y), \) 186
\( \partial(S), \) 103
e\( (S,T), \) 102
||G||, 161
\( \Gamma(S), \) 103
\( \Gamma^k(S), \) 103
\( G(w), \) 97
\( G_{n,p}, \) 92
\( h_G, \) 174
\( L, \) 174
\( \mathcal{L}, \) 174
\( \rho, \) 97
\( \sigma\)-field, 35
\( w, \) 98
\( w_{max}, \) 98
\( \bar{w}, \) 98
\( w_k, \) 98
\( \text{vol}(S), \) 97
\( \text{vol}_k(S), \) 98
\( \text{vol}(G), \) 98
\( \text{vol}_k(G), \) 98
\( \text{Vol}(S), \) 98
\( \text{Vol}_k(S), \) 98
\( \zeta(s), \) 108

Abello, J., 9, 10, 75, 238, 255
adjacency matrix, 161
admissible, 144
admissible conditions, 47
Aiello, W., xii, 3, 10, 17, 19, 75, 91, 255
Albert, R., 3, 9, 10, 16, 62, 75, 91, 143, 255
Alderson, D., 3, 258
Algorithm\((k,l),\) 243
Alon, N., 46, 255
Andersen, R., xii, 19, 251, 255
Auerbach, F., 8, 255
autonomous system, 10
average distance, 5, 143
Azuma’s inequality, 36, 227
balls-and-bins, 60
Barabási, A.-L., 3, 9, 10, 16, 62, 75, 91, 143,
170, 255, 257
Bender, E. A., 223, 255
Bernoulli trials, 21
BGP, 10
Bhalla, U. S., 9, 10, 255
bicyclic, 94
Biggs, N. L., 1, 255
binomial distribution, 22
biological networks, 16, 75
Bollobás, B., 94, 96, 199, 255
Bouman, P., 17, 75, 257
Boyles, S., 250, 255
Broder, A., 9, 143, 256
Broido, A., 11, 256
Buchbbaum, A., 9, 10, 75, 238, 255
Buhler, J., xii
Butler, S., xii
Burtin, J. D., 96, 256
call graph, 13
Canfield, E. R., 223, 255
Cavalcanti, A., 17, 75, 257
Central Limit Theorem, 22
Cheeger constant, 174
Cheeger inequality, 174, 185
Chen, F.-C., 17, 75, 257
Chernoff inequalities, 25
Chernoff, H., 25, 256
Chung, F., 3, 10, 17, 19, 54, 61, 62, 75, 91,
163, 176, 251, 255, 256
Claffy, K. C., 11, 256
clustering effect, 241
collaboration graphs, 13, 159
collaboration multigraph, 15
collaboration graph of the second kind, 14
conductance, 185
configuration model, 18, 223
connected, 5
Cooper, C., 199, 256
cycles, 93
De Castro, R., 75, 257
degree, 5
degree distribution, 7
degree sequence, 6
DeMoivre-Laplace Limit Theorem, 22

261
Derényi, I., 170, 257
Dewey, G., xii, 19, 256
diameter, 5, 95, 143, 186
directed graph, 5
discrepancy, 186
distance, 5, 143
Dorogovtsev, S. N., 3, 17, 75, 256
double jump, 94
Doyle, J. C., 3, 258
Drinea, E., 61, 256
duplication model, 76
duplication step, 76
duplication-deletion model, 90
edge, 4
double boundary, 103
double-duplication, 77
double-independent, 194
Edwards, J. S., 10, 258
eigenvalues, 161, 175
Einstein, A., 3, 4
Enachescu, M., 61, 256
Erdős number, 13
Erdős, P., 6, 7, 13, 14, 91, 92, 95, 114, 115, 129, 151, 248, 256
Erdős-Rényi model, 17
Estoup, J. B., 8, 257
Euler, 1
exceptional probability, 51
Exoo, G., 250, 255
expected second-order average degree, 98
expected average degree, 98
exponent, 8
exponents of power law graphs, 75
exponents of random power law graphs, 108
Fabriciak, A., 241, 257
Faloutsos, M., 9, 10, 75, 170, 257
Faloutsos, P., 9, 10, 75, 170, 257
Faloutsos, C., 9, 10, 75, 170, 257
Farkas, I. J., 170, 257
Feller, W., 22, 35, 257
Fernholz, D., 237, 257
filter, 36
FKG inequality, 146
Flake, G. W., 253, 257
fractional packing, 251
frequency, 7
Friedman, R., 17, 75, 257
Frieze, A., xii, 61, 199, 256
full duplication model, 89
Füredi, Z., 173, 257
Gabai, X., 8, 257
Galan, D. J., xii, 19, 257
Gallai, T., 6, 7, 256
Garay, M. R., 257
Garg, N., 251, 257
giant component, 99, 113, 114
Gkantsidis, C., 236, 257
Godsil, C., 174, 257
Goh, K.-I., 170, 257
Graham, R. L., xii, 23, 256, 257
graph, 4
graph metrics, 3
graphical, 7
Grossman, J., 13, 75, 141, 143, 159, 257
growth-deletion model, 198
Gu, Z., 17, 75, 257
Hakimi, S., 7, 257
Handjani, S., xii, 61, 62, 256
Handshaking Theorem, 7
Havel, V., 7, 257
Hayes, B., 257
Henzinger, M., 242, 257
Hollywood graph, 16
Horn, R. A., 257
Hughes, A., 17, 75, 257
hybrid graph, 242
indegree, 6
interlacing theorem, 163
Internet graph, 10
Ion, P., 75, 257
irreducible matrix, 162
Itai, A., 250, 257
Iyengar, R., 9, 10, 255
Janson, S., 36, 94, 138, 257
Jeong, H., 9, 62, 91, 143, 255
Johnson, C. R., 257
Johnson, D. S., 257
Jungreis, D., xii, 61, 62, 256
Kahng, B., 170, 257
Kevin Bacon number, 16
Kim, D., 170, 257
Kim, J. H., 46, 51, 255, 257
Kirchhoff, G., 117, 174, 258
Klee, V., 96, 258
Kleinberg, J., 70, 143, 241, 258
Knuth, D. E., 23, 94, 257
Komlós, J., 173, 257
Koutoupias, E., 241, 257
Könemann, J. K., 251, 257
Krapivsky, P. L., 61, 258
Kumar, R., 9, 10, 75, 143, 256, 258
Laplacian, 174, 175
Larman, D., 96, 258
Li, L., 3, 258
Li, W.-H., 17, 75, 257
linear, 1
Lipschitz, 47
Lipschitz condition, 36
near-e-Lipschitz, 51
Lloyd, E. K., 1, 255
INDEX

local connectivity, 242
loops, 4
Lotka’s law, 8
Lotka, A. J., 10, 258
Lovász, L., 250, 258
Lu, L., 3, 10, 17, 19, 54, 75, 91, 163, 176, 251, 255, 256
Luczak, T., 36, 94, 96, 97, 138, 257, 258

Maghoul, F., 9, 143, 256
Mahmoud, H., 9, 63, 258, 259
Mandelbrot, B. B., 8, 258
Manning, C. D., 8, 258
Martinez, N. D., 9, 259
martingale, 36, 47
matrix-tree theorem, 174
maximum degree, 109
McDiarmid, C., 27, 258
Mendes, J. F. F., 3, 17, 75, 256
metabolic networks, 17
Mihail, M., 170, 236, 257, 258
Milgram, S., 143, 258
minimum degree, 109
Mitzenmacher, M., 3, 8, 61, 199, 256, 258
mixed duplication model, 89
Molloy, M., 225, 258
monotone decreasing, 146
monotone property, 92, 190
Motwani, R., 33, 258
multi-commodity flow, 187

neighborhood expansion, 103
neighborhood growth, 126
Neumann-Lara, V., 250, 258
Newman, M. E. J., 3, 258
normal distribution, 22
NP-complete, 185

off-line model, 17
Oliveira, R., 61, 258
on-line model, 17, 189
outdegree, 6

Palsson, B. O., 10, 258
Papadimitriou, C. H., 170, 241, 257, 258
Pareto, V., 8, 10, 258
Patashnik, O., 23, 257
path, 5
Perl, Y., 250, 257
Perron, O., 162, 258
Perron-Frobenius Theorem, 162
Pittel, B., 94, 257, 258
Plummer, M., 250, 258
Pólya’s urn problem, 60
power law, 2, 3, 7
power law distribution, 171
preferential attachment model, 189
preferential attachment scheme, 8, 55
probability space, 17
protein-protein networks, 17
Raghavan, P., 9, 10, 33, 75, 143, 256, 258
Rajagopalan, S., 9, 10, 75, 143, 256, 258
Ramachandran, V., 237, 257
random graph model G(w), 18
random graphs, 18
random power law graphs, 107, 109
Redner, S., 61, 258
Reed, B., 225, 258
regular, 5
Rényi, A., 91, 92, 95, 114, 115, 129, 151, 248, 256
Richardson, R., xii, 19
Riemann zeta-function, 173
Riordan, O., 199, 255
routing, 187
Royle, G., 174, 257
Rucinski, A., 36, 138, 257
Sárközy, A., 15
Saberi, A., 236, 257
scale-free network, 3, 62
Schilling, C. H., 10, 258
Schütze, H., 8, 258
semi-circle law, 173
Shiloach, Y., 250, 257
Simon, H. A., 8, 10, 55, 259
simple graph, 4
six degrees of separation, 143
small world phenomenon, 143, 241
Smythe, R., 9, 63, 258, 259
sparse, 1
specially admissible, 145
spectral radius, 161
Spencer, J. H., 46, 61, 255, 258
standard normal distribution, 22
Stata, R., 9, 143, 256
Stirling’s formula, 23, 34
Strogatz, S. H., 9, 10, 16, 143, 241, 259
strongly sparse, 144
Stubbs, L., 17, 75, 259
submartingale, 41, 47
supermartingale, 41, 47
Szmyński, J., 9, 63, 258

Tarjan, R. E., 253, 257
Tomkins, A., 9, 10, 75, 143, 256, 258
trees, 92
Tsioutsios, K., 253, 257
unicyclic, 93

Vera, J., 199, 256
vertex boundary, 103
vertex-duplication, 77
Vicsek, T., 170, 257
volume, 97
Vu, V., xii, 19, 46, 51, 163, 176, 256, 257

walk, 5
Watts, D. J., 3, 10, 16, 143, 241, 259
Westbrook, J., 9, 10, 75, 238, 255
Wiener, J., 9, 143, 256
Wierman, J. C., 94, 258
Wiener, E. P., 173, 259
Wilf, H. S., xii, 6, 259
Williams, R. J., 9, 259
Willinger, W., 3, 258
Willis, J. C., 8, 259
Wilson, R. J., 1, 255
Wormald, N. C., 46, 259

Yule, G. U., 8, 10, 259

Zipf's law, 8, 55
Zipf, G. K., 10, 259
Through examples of large complex graphs in realistic networks, research in graph theory has been forging ahead into exciting new directions. Graph theory has emerged as a primary tool for detecting numerous hidden structures in various information networks, including Internet graphs, social networks, biological networks, or, more generally, any graph representing relations in massive data sets.

How will we explain from first principles the universal and ubiquitous coherence in the structure of these realistic but complex networks? In order to analyze these large sparse graphs, we use combinatorial, probabilistic, and spectral methods, as well as new and improved tools to analyze these networks. The examples of these networks have led us to focus on new, general, and powerful ways to look at graph theory. The book, based on lectures given at the CBMS Workshop on the Combinatorics of Large Sparse Graphs, presents new perspectives in graph theory and helps to contribute to a sound scientific foundation for our understanding of discrete networks that permeate this information age.