Contents

Ι	Background		1
1	A whirlwind tour of network science		3
	1.1	Networks as a powerful analogical framework	3
	1.2	Data and theory—the pillars of network science	8
	1.3	Networks are everywhere	9
	1.4	Basic terminology	11
	1.5	Common properties of real networks	15
	1.6	Summary	15
		Exercises	16
2	Network data across fields		17
	2.1	Biology	18
	2.2	Socioeconomic systems	20
	2.3	Other fun networks	22
	2.4	Focal Points: networks used throughout this book	23
	2.5	Summary	25
		Exercises	26
3	Data ethics		27
	3.1	Introduction to data ethics	27
	3.2	Biases in the dataset	28
	3.3	Privacy and surveillance	32
	3.4	Mistakes, misconduct, and how to prevent them	35
	3.5	Summary	37
		Exercises	37
4	Prin	ner	39
	4.1	Coding and computation	39
	4.2	Mathematics	45
	4.3	Statistics	55

	4.4	Summary	61
		Exercises	62
II	Ap	plications, tools, and tasks	63
5	The	life cycle of a network study	65
	5.1	Network questions	66
	5.2	Collecting, constructing, and cleaning network data	68
	5.3	Iterating on the cycle	70
	5.4	Summary	70
		Exercises	71
6	Gath	ering data	73
	6.1	Motives, means and opportunities	74
	6.2	Data gathering across fields	78
	6.3	Summary	81
		Exercises	82
7	Extr	acting networks from data — the "upstream task"	83
	7.1	What is it?	83
	7.2	Why does it matter?	84
	7.3	Summary	86
		Exercises	86
8	Impl	ementation: storing and manipulating network data	89
	8.1	A home for your networks	89
	8.2	Moving in and out of your home	96
	8.3	Software libraries for network analysis	100
	8.4	Summary	101
		Exercises	102
9	Inco	rporating node and edge attributes	105
	9.1	Data surround your network	105
	9.2	Representing attributes	108
	9.3	Patterns and relationships of attributes	110
	9.4	Connecting attributes and the network—record linkage	112
	9.5	Missing attributes	113
	9.6	Summary	115
		Exercises	116
10	Awfu	ll errors and how to amend them	117
	10.1	Errors in data: omission and commission	117
	10.2	Errors in networks	119
	10.3	Sources of network errors	121
	10.4	Fixes	122
	10.5	Thinning spurious data	122

	10.6	Thickening missing data	130
	10.7	Other approaches	133
	10.8	Summary	133
		Exercises	135
11	Explo	re and explain: statistics for network data	137
	11.1	Exploratory analysis	137
	11.2	Network analysis is usually iterative and complementary	140
	11.3	Confirmatory analysis	141
	11.4	Graph models as null models	142
	11.5	The configuration model	142
	11.6	Computational null models (Monte Carlo)	144
	11.7	Case study: the degree distribution	149
	11.8	Reporting considerations	161
	11.9	Summary	162
		Exercises	163
12	Unde	rstanding network structure and organization	165
	12.1	Micro-to-meso-to-macroscale	165
	12.2	Egocentric networks	166
	12.3	Degrees and degree distributions	167
	12.4	Clustering and transitivity	169
	12.5	Mixing patterns and correlations	171
	12.6	Motifs and graphlets	173
	12.7	Communities	175
	12.8	Hierarchy and cross-scale structure	185
	12.9	Centrality measures and ranking	188
	12.10	Distances and connectedness	191
	12.11	Size and density	193
	12.12	Other organizing patterns	193
	12.13	Choosing and designing measures	198
	12.14	Summary	199
		Exercises	202
13	Visua	lizing networks	203
	13.1	Standard network visualization	204
	13.2	Customizing your visualization	208
	13.3	Alternatives to "ball-and-stick" diagrams	214
	13.4	Processing data for visualization	216
	13.5	Emphasizing your network question in your visualization	218
	13.6	Visualization tools	219
	13.7	Summary	220
		Exercises	221
14	Sumn	narizing and comparing networks	223
	14.1	Summarizing networks	223

vii

	14.2	Comparing networks	227
	14.3	Clustering networks	229
	14.4	Summary	232
		Exercises	232
15	Dyna	mics and dynamic networks	235
	15.1	Dynamic networks and dynamics on networks	235
	15.2	Representations	236
	15.3	Quantifying dynamic networks	239
	15.4	Null models	242
	15.5	Visualization	244
	15.6	Further considerations	247
	15.7	Summary	248
		Exercises	249
16	Mach	ine learning	251
	16.1	Common network machine learning tasks	251
	16.2	Supervised learning	254
	16.3	Unsupervised, self-supervised, and representation learning	259
	16.4	Overfitting, bias-variance tradeoff, and regularization	262
	16.5	Model selection	266
	16.6	Data hygiene and evaluation	267
	16.7	Graph embedding	270
	16.8	Challenges and practical considerations	274
	16.9	Summary	276
		Exercises	277
In	terlu	de — Good practices for scientific computing	279
17	Resea	arch record-keeping	283
	17.1	Establishing a research record	284
	17.2	Backups and backup practices	286
	17.3	Summary	287
18	Data	provenance	289
	18.1	Why should we care?	289
	18.2	Best practices for data provenance	289
	18.3	Backups	292
	18.4	Summary	292
19	Repr	oducible and reliable code	293
	19.1	Coding for readability	293
	19.2	Coding for record-keeping	297
	19.3	Summary	299
20	Help	ful tools	301

Computational notebooks	301
Pipelines	302
Working with remote computers	303
UNIX—I know this system	306
Version control	309
Backups	310
Selecting tools for yourself	311
Summary	312
	Computational notebooks Pipelines Working with remote computers UNIX—I know this system Version control Backups Selecting tools for yourself Summary

II	[Fi	indamentals	315
21	Netw	orks demand network thinking: the friendship paradox	317
	21.1	Is the friendship paradox a paradox?	317
	21.2	Paradox under the extremes	319
	21.3	The random neighbors' degree distribution	320
	21.4	Do random graphs show the paradox?	321
	21.5	Generalized friendship paradox	324
	21.6	Summary	324
		Exercises	325
22	Network models		327
	22.1	Mechanistic and statistical models	327
	22.2	Randomness in models	328
	22.3	Erdős–Rényi model	329
	22.4	Configuration model	333
	22.5	It's a small world!	340
	22.6	Power-law networks—what's all the hubbub?	344
	22.7	Summary	349
		Exercises	349
23	Stati	stical models and inference	351
	23.1	Statistical models we've seen before	351
	23.2	Stochastic block models	352
	23.3	Witness me: the edge observer model	358
	23.4	Other modeling approaches	365
	23.5	Ensembles	373
	23.6	Summary	374
		Exercises	375
24	Unce	rtainty quantification and error analysis	377
	24.1	Computational and mathematical approaches	377
	24.2	Missing data and its effects	378
	24.3	Community structure	380
	24.4	Uncertain networks as probabilistic graphs	385
	24.5	Size estimation	387

CONTENTS

	24.6	Other approaches	394
	24.7	Summary	394
		Exercises	395
25	Ghos	t in the matrix: spectral methods for networks	397
	25.1	Networks as matrices	397
	25.2	Spectral properties reflect network properties	406
	25.3	Some spectral applications	406
	25.4	Centrality	407
	25.5	Partitioning	408
	25.6	Community detection	417
	25.7	Spectral clustering	421
	25.8	Summary	426
		Exercises	428
26	Embe	edding and machine learning	429
	26.1	Embeddings as representations	429
	26.2	Language models and word2vec	431
	26.3	From writing to walking: embedding networks	434
	26.4	Embedding as matrix factorization	438
	26.5	Graph neural networks	441
	26.6	Summary	444
		Exercises	445
27	Big d	ata and scalability	447
	27.1	Do you <i>really</i> have big data?	447
	27.2	When networks become large	447
	27.3	What do we mean by "scalability"?	450
	27.4	Compressing, distributing, and streaming graphs	450
	27.5	Approximations and local methods	454
	27.6	Updating schemes for network statistics	459
	27.7	Making graphs	466
	27.8	Summary	468
		Exercises	469
Co	nclusi	on	471
Bil	oliogra	aphy	476
Inc	lex		513

х