Index

$G(n, m)$ model, 85
$G(n, p)$ model, 84
$k$-clan, 152
$k$-clique, 151
$k$-club, 152
$k$-means, 128
$k$-plex, 149
actor, see node
Adamic and Adar measure, 280
adjacency list, 19
adjacency matrix, 18
adopter types, 193
adoption of hybrid seed corn, 193
affiliation network, 30
Anderson, Lisa R., 183
Asch conformity experiment, 182
Asch, Solomon, 182
assortativity, 217
measuring, 218
nominal attributes, 219
ordinal attributes, 222
significance, 220
attribute, see feature
average path length, 81, 84
behavior analysis methodology, 276
betweenness centrality, 59
BFS, see breadth-first search
big data, 2
big data paradox, 2
bipartite graph, 29
botnet, 141
breadth-first search, 34
bridge, 31
bridge detection, 45
cascade maximization, 189
causality testing, 277
centrality, 51
betweenness centrality, 59
computing, 60
closeness centrality, 61
degree centrality, 52
gregariousness, 52
normalized, 52
prestige, 52
eigenvector centrality, 53
group centrality, 63
betweenness, 63
closeness, 64
degree, 63
Katz centrality, 56
divergence in computation, 57
PageRank, 58
citizen journalist, 1
class, see class attribute
class attribute, 107
clique identification, 148
clique percolation method, 149
closeness centrality, 61
cluster centroid, 127
clustering, 127, 128
$k$-means, 128
hierarchical, 160
agglomerative, 160
divisive, 160
Index

316

clustering (cont.)
  partitional, 128
  spectral, 155
clustering coefficient, 65, 81
  global, 66
  local, 67
cohesive subgroup, see community
cold-start problem, 246
collaborative filtering, 248, 278
  memory-based, 248
  item-based, 251
  user-based, 250
  model-based, 249, 252
collective behavior, 283
collective behavior modeling, 288
common neighbors, 279
community, 141
  detection, 144
  emit, see explicit
  etic, see implicit
evolution, 161
  explicit, 143
  implicit, 143
community detection, 144
  group-based, 145, 153
  member-based, 145
  node degree, 147
  node reachability, 150
  node similarity, 152
community evaluation, 168
community membership behavior, 273
commute time, 282
confounding, 217
contact network, 201
content data, 272
content-based recommendation, 247
cosine similarity, 72, 152, 250
covariance, 74, 222

data point, see instance
data preprocessing, 111
  aggregation, 111
  discretization, 111
feature extraction, 112
feature selection, 112
sampling, 112
  random, 112

Index

stratified, 112
  with/without replacement, 112
data quality, 110
  duplicate data, 111
  missing values, 111
  noise, 110
  outliers, 110
data scientist, 2
data sparsity, 246
decision tree learning, 115
degree, 13, 15
degree distribution, 16
degree centrality, 52, 225
gregariousness, 52
prestige, 52
degree distribution, 16, 81
dendrogram, 160
densification, 163
depth-first search, 32
DFS, see depth-first search
diameter, 26
  shrinkage, 164
diffusion of innovations, 180, 193
Dijkstra, Edsger, 35
Dijkstra’s algorithm, 35
diminishing returns, 274
directed graph, 15, 20
distinguishing influence and homophily, 236
deugraph, 37
  randomization test, 238
  shuffle test, 236
early adopters, 194
early majority, 194
Eckart-Young-Mirsky theorem, 253
degree, 14
  betweenness, 160
  directed, 15
  incident, 22
  loop, 15, 19
  self-link, 19
  signed, 22
  traverse, see visit
undirected, 15
visit, 23
edge list, 20
deugraph, 237
Index

eigenvector centrality, 53
eigenvector centrality, 53
entropy, 116
epidemics, 180, 200
infected, 202
recovered, 202
removed, 202
SI model, 202
SIR model, 204
SIRS model, 207
SIS model, 206
susceptible, 202
Erdős, Paul, 85
Euclidean distance, 127
evaluation dilemma, 3
evolutionary clustering, 166
external-influence model, 197

F-measure, 170, 265
false negative, 169
false positive, 169
feature, 107
categorical, see nominal
continuous, 107
discrete, 107
interval, 108
nominal, 107
ordinal, 108
ratio, 108
feature selection, 274
Ford-Fulkerson algorithm, 41
Frobenius norm, 253
fully mixed technique, 201
giant component, 86, 162
Gilbert, Edgar, 85
Girvan-Newman algorithm, 160
global clustering coefficient, 66
Granger causality, 277
Granger, Clive W. J., 277
graph
\( k \)-connected, 157
bipartite graph, 29
bridge, 31
detection, 45
complete graph, 28
component, 25
connected, 25
connectivity, 22
circuit, see tour
cycle, 23
path, 23
tour, 23
trail, 23
walk, 23
cut, 153
densification, 163
density, 159
diameter, 26
directed graph, 15, 20
element, 13, 14
empty graph, 20
forest, 27
minimum cut, 153
mixed graph, 20
multigraph, 21
node, 13, 14
null graph, 20
partition, 29
planar graph, 29
regular graph, 31
representation
adjacency list, 19
adjacency matrix, 18
edge list, 20
shortest path, 26
Dijsktra's algorithm, 35
signed graph, 22
simple graph, 21
strongly connected, 25
subgraph, 18
minimum spanning tree, 27
spanning tree, 27
Steiner tree, 28
traversal, 32
breadth-first search, 34
depth-first search, 32
tree, 27
undirected graph, 20
weakly connected, 25
weighted graph, 21
graph Laplacian, 155
graph traversal, 32
breadth-first search, 34
depth-first search, 32
Index

318

Index

KDD process, see knowledge discovery in databases
Kendall’s tau, 266
knowledge discovery in databases, 105
Kronecker delta function, 219

labeled data, 113
laggards, 194
Laplacian, 155
largest connected component, 86
late majority, 194
levels of measurement, 107
LIM, see linear influence model
linear influence model, 230, 288
linear threshold model, 230, 277
link data, 272
link prediction, 258, 278
node neighborhood-based methods, 279
path-based methods, 281
local clustering coefficient, 67
logistic function, 236
logistic growth function, 203
logistic regression, 287
logit function, 125
loop, 15, 19
low-rank matrix approximation, 253
LTM, see linear threshold model

MAE, see mean absolute error
Mahalanobis distance, 127
matrix factorization, 259
maximum bipartite matching, 44
mean absolute error, 263
measurement, see feature measuring homophily, 234
Milgram, Stanley, 84
minimum spanning tree, 27
Prim’s algorithm, 38
mixed graph, 20
mixed-influence model, 197
modeling homophily, 235
modularity, 157, 220
multistep flow model, 195
mutual information, 171
naive Bayes classifier, 117
nearest neighbor classifier, 119
neighborhood, 15
network flow, 38
  augmenting path, 42
capacity, 40
capacity constraint, 40
  flow, 40
flow conservation constraint, 40
flow network, 40
Ford-Fulkerson algorithm, 41
residual capacity, 41
residual network, 41
  source and sink, 40
network measure, 51
network segmentation, 162
NMI, see normalized mutual information
node, 14
  adjacent, 22
  connected, 25
  degree, 13, 15
  in-degree, 16
  neighborhood, 15
  out-degree, 16
  reachable, 25
noise removal fallacy, 3
normalized cut, 154
normalized mean absolute error, 264
normalized mutual information, 171
nuke attack, 247

observation, see instance
obtaining sufficient samples, 3
out-degree, 16
overfitting, 260
PageRank, 58, 225
parametric estimation, 233
Pearson correlation coefficient, 74, 224, 249
Perron-Frobenius theorem, 54
Poisson distribution, 90
power iteration method, 77
power-law distribution, 17, 81, 232, 288, 297
precision, 168, 265
preferential attachment measure, 280
preferential attachment model, 288
prestige, 52
prominence, see prestige
purity, 170
push attack, 247
quasi-clique, 159
random graph, 84
  average path length, 91
  clustering coefficient, 90
  degree distribution, 89
  evolution, 86
  phase transition, 87
  random walk, 23
  randomization test, 238, 276
rank correlation, 229
Rapoport, Anatol, 85
ratio cut, 154
raw data, 105
recall, 168, 265
reciprocity, 68
recommender systems, 245
  recommendation to groups, 256
  least misery, 256
  most pleasure, 257
regression, 277
  regular equivalence, 74, 153
  regular graph, 31
  ring lattice, 93
  regular ring lattice, 93
  regularization term, 260
  relationships, see edge
  relaxing cliques, 149
  k-plex, 149
RMSE, see root mean squared error
Rogers, Everett M., 193
root mean squared error, 264
rooted PageRank, 282
Rényi, Alfréd, 85
scale-free network, 82
self-link, see loop, 19
sender-centric model, 187
sentiment analysis, 135
shortest path, 26
Dijskra’s algorithm, 35
shuffle test, 236
SI model, 202
similarity, 71
cosine similarity, 72
Jaccard similarity, 72, 74
regular equivalence, 74
structural equivalence, 72
SimRank, 282
singleton, 163
singular value decomposition, 124, 252
SIR model, 204
SIRS model, 207
SIS model, 206
six degrees of separation, 84
small-world model, 84, 93, 288
average path length, 96
clustering coefficient, 95
degree distribution, 95
social atom, 1
social balance, 70
social correlation, 236
social media, 1
social media mining, 2
social molecule, 2
social network, 13
social similarity, 217
social status, 22, 71
social tie, see edge
sociomatrix, see adjacency matrix
Solomonoff, Ray, 85
sparse matrix, 19
Spearman’s rank correlation coefficient, 229, 266
spectral clustering, 155
star, 162
Stevens, Stanley Smith, 107
Strogatz, Steven, 93
structural equivalence, 72
submodular function, 190
supervised learning, 113
classification, 114
decision tree learning, 115
naive Bayes classifier, 117
nearest neighbor classifier, 119
with network information, 119, 120
evaluation, 126
k-fold cross validation, 126
accuracy, 126
leave-one-out, 126
regression, 114, 122
linear, 123
logistic, 124, 125
SVD, see singular value decomposition
tabular data, 106
tag clouds, 173
target data, 105
term frequency-inverse document frequency, 109
test data, 114
TF-IDF, see term frequency-inverse document frequency
theory of scales, 107
training data, see labeled data
transitivity, 65
clustering coefficient, 65
global, 66
local, 67
Trotter, Wilfred, 181
true negative, 169
true positive, 169
undirected graph, 20
unsupervised learning, 127
evaluation, 130
cohesiveness, 130
separateness, 131
silhouette index, 131
user migration, 284
user-item matrix, 248
vector-space model, 108
vectorization, 108
vertex, see node
Watts, Duncan J., 93
Zachary’s karate club, 142