

# Index

- $G(n, m)$  model, 107
- $G(n, p)$  model, 106
- $k$ -clan, 183
- $k$ -clique, 182
- $k$ -club, 183
- $k$ -means, 156
- $k$ -plex, 180
  
- actor, *see* node
- Adamic and Adar measure, 324
- adjacency list, 32
- adjacency matrix, 31
- adopter types, 229
- adoption of hybrid seed corn, 229
- affiliation network, 44
- Anderson, Lisa R., 218
- Asch conformity experiment, 216
- Asch, Solomon, 216
- assortativity, 255
  - measuring, 257
    - nominal attributes, 258
    - ordinal attributes, 261
  - significance, 258
- attribute, *see* feature
- average path length, 102, 105
  
- behavior analysis methodology, 320
- betweenness centrality, 78
- BFS, *see* breadth-first search
- big data, 13
- big data paradox, 13
- bipartite graph, 44
- botnet, 171
- breadth-first search, 47
- bridge, 45
- bridge detection, 60
  
- cascade maximization, 225
- causality testing, 321
- centrality, 70
  - betweenness centrality, 78
    - computing, 78
  - closeness centrality, 80
  - degree centrality, 70
    - gregariousness, 70
    - normalized, 71
    - prestige, 70
- eigenvector centrality, 71
- group centrality, 81
  - betweenness, 82
  - closeness, 82

- degree, 81
- Katz centrality, 74
  - divergence in computation, 75
- PageRank, 76
- Christakis, Nicholas A., 245
- citizen journalist, 11
- class, *see* class attribute
- class attribute, 133
- clique identification, 178
- clique percolation method, 180
- closeness centrality, 80
- cluster centroid, 156
- clustering, 155, 156
  - k*-means, 156
  - hierarchical, 191
    - agglomerative, 191
    - divisive, 191
  - partitional, 156
  - spectral, 187
- clustering coefficient, 84, 102
  - global, 85
  - local, 86
- cohesive subgroup, *see* community
- cold start problem, 286
- collaborative filtering, 289, 323
  - memory-based, 289
    - item-based, 292
    - user-based, 290
  - model-based, 289, 293
- collective behavior, 328
- collective behavior modeling, 333
- common neighbors, 324
- community, 171
  - detection, 175
  - emit, *see* explicit
  - etic, *see* implicit
  - evolution, 193
    - explicit, 173
    - implicit, 174
- community detection, 175
  - group-based, 175, 184
  - member-based, 175, 177
    - node degree, 178
    - node reachability, 181
    - node similarity, 183
- community evaluation, 200
- community membership behavior, 317
- commute time, 327
- confounding, 255
- contact network, 237
- content data, 316
- content-based recommendation, 288
- cosine similarity, 92, 183, 291
- covariance, 94, 261
- data point, *see* instance
- data preprocessing, 138
  - aggregation, 138
  - discretization, 138
  - feature extraction, 138
  - feature selection, 138
  - sampling, 138
    - random, 139
    - stratified, 139
  - with/without replacement, 139
- data quality, 137
  - duplicate data, 137
  - missing values, 137
  - noise, 137
  - outliers, 137
- data scientist, 12

- data sparsity, 287
- decision tree learning, 141
- degree, 25, 28
  - distribution, 29
- degree centrality, 70, 264
  - gregariousness, 70
  - prestige, 70
- degree distribution, 29, 102
- dendrogram, 191
- densification, 194
- depth-first search, 46
- DFS, *see* depth-first search
- diameter, 40
  - shrinkage, 196
- diffusion of innovation, 214
- diffusion of innovations, 228
- Dijkstra, Edsger, 49
- Dijkstra's algorithm, 49
- diminishing returns, 318
- directed graph, 28, 33
- distinguishing influence and homophily, 276
  - edge-reversal test, 277
  - randomization test, 278
  - shuffle test, 276
- early adopters, 229
- early majority, 229
- Eckart-Young-Mirsky theorem, 294
- edge, 27
  - betweenness, 191
  - directed, 27
  - incident, 35
  - loop, 28, 32
  - self-link, 32
  - signed, 34
  - traverse, *see* visit
  - undirected, 27
  - visit, 36
- edge list, 32
- edge-reversal test, 277
- eigenvector centrality, 71
- entropy, 143
- epidemics, 214, 236
  - infected, 238
  - recovered, 238
  - removed, 238
  - SI model, 239
  - SIR model, 241
  - SIRS model, 244
  - SIS model, 242
  - susceptible, 238
- Erdős, Paul, 107
- Euclidean distance, 155
- evaluation dilemma, 13
- evolutionary clustering, 198
- external-influence model, 232
- F-measure, 202, 307
- false negative, 201
- false positive, 201
- feature, 133
  - categorical, *see* nominal
  - continuous, 133
  - discrete, 133
  - interval, 134
  - nominal, 134
  - ordinal, 134
  - ratio, 134
- feature selection, 319
- Ford-Fulkerson algorithm, 55
- Frobenius norm, 293
- fully-mixed technique, 237
- giant component, 108, 194

Gilbert, Edgar, 106  
 Girvan-Newman algorithm, 191  
 global clustering coefficient, 85  
 Granger causality, 321  
 Granger, Clive W. J., 321  
 graph  
    $k$ -connected, 188  
   bipartite graph, 44  
   bridge, 45  
     detection, 60  
   complete graph, 42  
   component, 38  
   connected, 38  
   connectivity, 35  
     circuit, *see* tour  
     cycle, 36  
     path, 36  
     tour, 36  
     trail, 36  
     walk, 36  
   cut, 185  
   densification, 194  
   density, 190  
   diameter, 40  
   directed, 28  
   directed graph, 33  
   edge, 25, 27  
   empty graph, 33  
   forest, 40  
   minimum cut, 185  
   mixed graph, 33  
   multigraph, 34  
   node, 25, 26  
   null graph, 33  
   partition, 44  
   planar graph, 43  
   regular graph, 44  
   representation  
     adjacency list, 32  
     adjacency matrix, 31  
     edge list, 32  
   shortest path, 39  
     Dijkstra's algorithm, 49  
   signed graph, 34  
   simple graph, 34  
   strongly connected, 38  
   subgraph, 30  
     minimum spanning tree, 41  
     spanning tree, 41  
     Steiner tree, 42  
   traversal, 45  
     breadth-first search, 47  
     depth-first search, 46  
   tree, 40  
   undirected graph, 33  
   weakly connected, 38  
   weighted graph, 34  
 graph Laplacian, 187  
 graph traversal, 45  
   breadth-first search, 47  
   depth-first search, 46  
 group, *see* community  
 group centrality, 81  
   betweenness, 82  
   closeness, 82  
   degree, 81  
  
 herd behavior, 214  
 hitting time, 327  
 Holt, Charles A., 218  
 homophily, 255, 274  
   measuring, 274  
   modeling, 274  
  
 ICM, *see* independent cascade  
   model

- in-degree, 28
- independent cascade model, 222
- individual behavior, 315
  - user-community behavior, 316
  - user-entity behavior, 316
  - user-user behavior, 316
- individual behavior modeling, 322
- influence, 255, 264
  - measuring, 264
    - observation-based, 264
    - prediction-based, 264
  - modeling, 269
- influence flow, 266
- influence modeling, 269
  - explicit networks, 269
  - implicit networks, 271
- information cascade, 214, 221, 323
- information pathways, 78
- information provenance, 249
- innovation characteristics, 228
- innovators, 229
- instance, 133
  - labeled, 133
  - unlabeled, 133
- internal-influence model, 232
- intervention, 214, 220, 227, 235, 245
  
- Jaccard similarity, 92, 183, 324
  
- Katz centrality, 74
  - divergence in computation, 75
- Katz model, 230
- Katz, Elihu, 230
- KDD process, *see* knowledge discovery in databases
  
- Kendall's tau, 308
- knowledge discovery in databases, 131
- Kronecker delta function, 258
  
- labeled data, 140
- laggards, 229
- Laplacian, 187
- largest connected component, 108
- late majority, 229
- levels of measurement, 133
- LIM, *see* linear influence model
- linear influence model, 269, 334
- linear threshold model, 269, 322
- link data, 316
- link prediction, 299, 323
  - node neighborhood-based methods, 324
  - path-based methods, 326
- local clustering coefficient, 86
- logistic function, 276
- logistic growth function, 240
- logistic regression, 333
- logit function, 152
- loop, 28, 32
- low-rank matrix approximation, 293
- LTM, *see* linear threshold model
  
- MAE, *see* mean absolute error
- Mahalanobis distance, 155
- matrix factorization, 301
- maximum bipartite matching, 57
- mean absolute error, 305
- measurement, *see* feature
- measuring homophily, 274
- Milgram, Stanley, 105
- minimum spanning tree, 41

- Prim's algorithm, 52
- mixed graph, 33
- mixed-influence model, 232
- modeling homophily, 274
- modularity, 189, 259
- multi-step flow model, 230
- mutual information, 203
  
- naive bayes classifier, 144
- nearest neighbor classifier, 146
- neighborhood, 28
- network flow, 53
  - augmenting path, 56
  - capacity, 54
  - capacity constraint, 54
  - flow, 54
  - flow conservation constraint, 54
  - flow network, 54
  - Ford-Fulkerson algorithm, 55
  - residual capacity, 56
  - residual network, 56
  - source and sink, 54
- network measure, 69
- network segmentation, 194
- NMI, *see* normalized mutual information
- node, 26
  - adjacent, 35
  - connected, 38
  - degree, 25, 28
  - in-degree, 28
  - neighborhood, 28
  - out-degree, 28
  - reachable, 38
- noise removal fallacy, 13
- normalized cut, 185
  
- normalized mean absolute error, 305
- normalized mutual information, 203
- nuke attack, 287
  
- observation, *see* instance
- obtaining sufficient samples, 13
- out-degree, 28
- overfitting, 301
  
- PageRank, 76, 264
- parametric estimation, 273
- Pearson correlation coefficient, 94, 263, 290
- Perron-Frobenius theorem, 72
- Poisson distribution, 112
- Power Iteration Method, 98
- power-law distribution, 30
- powerlaw distribution, 103, 272, 276, 333
- precision, 201, 307
- preferential attachment measure, 325
- preferential attachment model, 333
- prestige, 70
- prominence, *see* prestige
- purity, 203
- push attack, 287
  
- quasi-clique, 191
  
- random graph, 106
  - average path length, 114
  - clustering coefficient, 112
  - degree distribution, 111
  - evolution, 108
  - phase transition, 109

random walk, 36  
 randomization test, 278, 321  
 rank correlation, 268  
 Rapoport, Anatol, 106  
 ratio cut, 185  
 raw data, 131  
 recall, 201, 307  
 reciprocity, 87  
 recommendation to groups, 297  
     least misery, 297  
     maximizing average satisfaction, 297  
     most pleasure, 298  
 recommender systems, 285  
 regression, 322  
 regular equivalence, 94, 184  
 regular graph, 44  
     ring lattice, 115  
 regular ring lattice, 115  
 regularization term, 302  
 relationships, *see* edge  
 relaxing cliques, 180  
     *k*-plex, 180  
 RMSE, *see* root mean squared error  
 Rogers, Everett M., 228  
 root mean squared error, 305  
 rooted PageRank, 327  
 Rényi, Alfred, 107  
  
 scale-free network, 105  
 self-link, *see* loop, 32  
 sender-centric model, 222  
 sentiment analysis, 164  
 shortest path, 39  
     Dijkstra's algorithm, 49  
 shuffle test, 276  
 SI model, 239  
  
 similarity, 91  
     cosine similarity, 92  
     Jaccard similarity, 92, 94  
     regular equivalence, 94  
     structural equivalence, 92  
 SimRank, 328  
 singleton, 194  
 singular value decomposition, 152, 293  
 SIR model, 241  
 SIRS model, 244  
 SIS model, 242  
 six degrees of separation, 105  
 small world model, 333  
 small-world, 105  
 small-world model, 115  
     average path length, 118  
     clustering coefficient, 118  
     degree distribution, 117  
 Social Atom, 12  
 social balance, 89  
 social correlation, 276  
 social media, 11  
 social media mining, 12  
 Social Molecule, 12  
 social network, 25  
 social similarity, 255  
 social status, 35, 90  
 social tie, *see* edge  
 sociomatrix, *see* adjacency matrix  
 Solomonoff, Ray, 106  
 sparse matrix, 32  
 Spearman's rank correlation coefficient, 268, 308  
 spectral clustering, 187  
 star, 194  
 Stevens, Stanley Smith, 133  
 Strogatz, Steven, 115

structural equivalence, 92  
 submodular function, 225  
 supervised learning, 140
 

- classification, 141
  - decision tree learning, 141
  - naive bayes classifier, 144
  - nearest neighbor classifier, 146
  - with network information, 147
- evaluation
  - $k$ -fold cross validation, 154
  - accuracy, 154
  - leave-one-out, 154
- evaluations, 153
- regression, 141, 150
  - linear, 151
  - logistic, 152

SVD, *see* singular value decomposition

tabular data, 133  
 tag clouds, 205  
 target data, 131  
 term frequency-inverse document frequency, 135  
 test data, 140

TF-IDF, *see* term frequency-inverse document frequency  
 theory of scales, 133  
 training data, *see* labeled data  
 transitivity, 83
 

- clustering coefficient, 84
- global, 85
- local, 86

Trotter, Wilfred, 216

true negative, 201

true positive, 201

undirected graph, 33

unsupervised learning, 155
 

- evaluation, 158
  - cohesiveness, 159
  - separateness, 159
  - silhouette index, 160

user migration, 329

user-item matrix, 289

vector-space model, 135

vectorization, 135

vertex, *see* node

Watts, Duncan J., 115

Zachary's karate club, 173