

List of Figures

2-1 Garfield's historiographs are a pioneering visualization of citation link structures	13
2-2 Garfield's cluster maps for visualizing co-citation based clusters of documents	14
2-3 Typical document citation structures for emerging, stable, and declining disciplines	15
2-4 Citation structure of a discipline under the oscillating model	15
2-5 Clustering of documents based on citation structure	17
2-6 Example of chaining in co-citation single-linkage clustering	18
2-7 Chen's visualization of author influence network	26
2-8 Chen's StarWalker VRML application for navigation of influence network	27
2-9 Visualizations of document collections from Pacific Northwest Laboratory: Galaxies scatter plot and ThemeScapes document landscape	28
2-10 Narcissus visualizes Web links directly, rather than employing measures of page similarity derived from link information	29
2-11 InXight's hyperbolic lens visualization for Focus+Context	30
2-12 Munzner's 3-dimensional generalization of the hyperbolic lens visualization	31
2-13 The cone tree for visualizing hierarchies, developed at Xerox PARC	32
2-14 Butterfly user interface for searching citation links, developed at Xerox PARC	32
3-1 Full and reduced citation adjacency matrix from citation or hyperlink graph	36
3-2 Co-citation count	37
Citation matrix for SCI (Science Citation Index) "Microtubules" data set	40

3-3 Co-citation counts and cited document correlation for SCI “Microtubules” data set	41
3-5 Correlation matrix column minimum values for SCI “Microtubules” data set	41
3-6 Co-citation counts, correlations, and corresponding multiplicative and additive inverse distances for SCI “Microtubules” data set	42
3-7 Inter-cluster distances for single-linkage, average-linkage, and complete-linkage ...	45
3-8 Dendrogram tree for visualizing clustering hierarchy	46
3-9 Interpreting clusters from dendrogram for given clustering threshold distance	47
3-10 Distance matrices for Figures 3-11 through 3-13, in which mean of Gaussian distribution defines cluster separations	48
3-11 Dendrograms for 2 very well separated clusters (Δ mean = 80)	48
3-12 Dendrograms for 2 well separated clusters (Δ mean = 30)	49
3-13 Dendrograms for 2 poorly separated clusters (Δ mean = 20)	50
3-14 Dendrograms for 2 clusters with chain of points between them	51
3-15 Dendrograms for “count inverse” distances for “Microtubules” data set	52
3-16 Dendrograms for “count linear decrease” distances for “Microtubules” data set ...	53
3-17 Dendrograms for “correlation inverse” distances for “Microtubules” data set	53
3-18 Dendrograms for “correlation linear decrease” distances for “Microtubules” data set	54
3-19 Hierarchical clustering for uniform random points	55
3-20 Hierarchical clustering for fractal random points	55
3-21 Single-linkage chaining for co-citation similarities	57
3-22 Stronger clustering criterion for complete-linkage with co-citation similarities	58
3-23 Higher-order co-citation (association mining itemset) is an even stronger association than complete-linkage cluster	58

3-24 Itemset lattice for a set of 4 documents, visualized with Hasse diagram	61
3-25 Clustering versus frequent itemsets for “Wavelets (1-100)” data set	63
3-26 Clustering versus frequent itemsets for “Wavelets (1-150)” data set	66
3-27 Augmenting dendrogram with frequent itemsets and text for information retrieval	68
3-28 Itemset-matching clustering metric is the average portion occupied by an itemset within the minimal cluster containing it	71
3-29 General method for computing itemset-matching clustering metric	71
3-30 Pairwise document similarities by summing supports over all itemsets containing a given document pair	74
3-31 Document similarities by summing support features over all itemsets of a given cardinality containing a given document pair	75
3-32 Quadratic nonlinearity applied to itemset-support features in computing hybrid pairwise/higher-order similarities	77
3-33 Clustering versus frequent itemsets for cubic transformation of itemset-support features	82
3-34 Clustering versus frequent itemsets for 4 th power transformation of itemset-support features	83
3-35 Clustering versus frequent itemsets for 5 th power transformation of itemset-support features	84
3-36 Document similarities by summing 4 th power supports over itemsets of multiple cardinalities	85
3-37 Exponential nonlinearity applied to itemset-support features	86
3-38 General method for computing itemset- matching clustering metric	88
3-39 Cardinality-4 itemset supports for (a) the data set in Table A-1 and (b) the data set in Table A-8	93
3-40 Cardinality-4 itemset supports for the data set in Table A-2	94

4-1	Original and nonlinearly transformed itemset supports for 3 different values of <i>minsup</i>	100
4-2	Cardinality-4 itemset-support distribution versus right-sided Laplacian for SCI “Wavelets (1-100)” data set	103
4-3	Cardinality-3 itemset-support distribution versus right-sided Laplacian for SCI “Microtubules” data set	105
4-4	Itemset-support distributions for SCI “Wavelets” data set, cardinalities with 100, 150, and 200 citing documents	105
4-5	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Adaptive Optics” data set	107
4-6	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Adaptive Optics” data set (bibliographic coupling)	108
4-7	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Collagen” data set	109
4-8	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Genetic Algorithms and Neural Networks” data set	110
4-9	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Quantum Gravity and Strings” data set	111
4-10	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Quantum Gravity and Strings” data set (bibliographic coupling)	112
4-11	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Wavelets (1 -100)” data set	113
4-12	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Wavelets (1 -500)” data set	114
4-13	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Wavelets and Brownian” data set	115
4-14	Association mining itemset supports and their distributions for cardinalities 2,3,4 for SCI “Wavelets and Brownian” data set (bibliographic coupling)	116

4-15 Citation matrix with transaction and item weights for “Wavelets and Brownian” data set	121
4-16 Distance matrices for “Wavelets and Brownian” data set: (a) standard pairwise, (b) transaction weighting, (c) item weighting, (d) hybrid from cardinality-3 itemsets, and (e) hybrid from cardinality-4 itemsets	122
4-17 Standard pairwise similarities versus frequent itemsets for complete-linkage clustering of “Wavelets and Brownian” data set	124
4-18 Transaction weighting similarities, itemset-matching metric for complete-linkage clustering of “Wavelets and Brownian” data set	125
4-19 Item weighting similarities, itemset-matching metric for complete-linkage clustering of “Wavelets and Brownian” data set	125
4-20 Cardinality-3 similarities, itemset-matching metric for complete-linkage clustering of “Wavelets and Brownian” data set	126
4-21 Cardinality-4 similarities, itemset-matching metric for complete-linkage clustering of “Wavelets and Brownian” data set	126
5-1 Cut $(S, V - S)$ of graph for computing the minimum spanning tree	133
5-2 Multidimensional scaling for synthetic distances with Euclidean norm	136
5-3 Multidimensional scaling fails for real-life SCI documents	136
5-4 Attractive, repulsive, and total forces versus distance	140
5-5 Iterations of spring algorithm for placing vertices of minimum spanning tree	141
5-6 Minimum spanning tree placement for pairwise distances computed via co-citation count versus citation correlation, for data set “Wavelets 1999 (1-100)”	142
5-7 Minimum spanning tree placement for pairwise distances computed via co-citation count versus citation correlation, for data set “Wavelets 1999 (1-150)”	142
5-8 Minimum spanning tree placement for pairwise distances computed via co-citation count versus citation correlation, for data set “Wavelets 1999 (1-200)”	143
5-9 Minimum spanning tree visualization serves as network of document influences ..	144

5-10	Minimum spanning tree network of influence for wavelets documents cited in 1999	145
5-11	Documents at the lower levels of the single-linkage dendrogram tend to be near the center of the minimum spanning tree	147
5-12	Comparison of clustering, frequent itemsets, and minimum spanning tree for data set “Wavelets 1999 (1-100)” with pairwise distances	148
5-13	Comparison of clustering, frequent itemsets, and minimum spanning tree for data set “Wavelets 1999 (1-100)” with distances from order-4 co-citations	149
5-14	Comparison of clustering, frequent itemsets, and minimum spanning tree for data set “Wavelets 1999 (1-100),” distances from co-citations of orders 2, 3, 4..	150
5-15	Comparison of clustering, frequent itemsets, and minimum spanning tree for data set “Wavelets 1999 (1-150)” with pairwise distances	151
5-16	Comparison of clustering, frequent itemsets, and minimum spanning tree for data set “Wavelets 1999 (1-150)” with distances from higher-order co-citations of cardinalities 2, 3, 4	152
5-17	Comparison of clustering, frequent itemsets, and minimum spanning tree for data set “Wavelets 1999 (1-150)” with distances from higher-order co-citations of cardinalities 3, 4	153
5-18	Number of connected components of frequent itemsets forms basis for itemset-matching minimum spanning tree metric	154
5-19	Metric for minimum spanning tree is inverse of average number of connected components per itemset	155
5-20	Minimum spanning tree itemset-degree metrics for “Adaptive Optics” data set ..	165
5-21	Minimum spanning tree itemset-degree metrics for “Adaptive Optics” data set, with bibliographic coupling	166
5-22	Minimum spanning tree itemset-degree metrics for “Collagen” data set	166
5-23	Minimum spanning tree itemset-degree metrics for “Genetic Algorithms and Neural Networks” data set	166
5-24	Minimum spanning tree itemset-degree metrics for “Quantum Gravity and Strings” data set	167

5-25	Minimum spanning tree itemset-degree metrics for “Quantum Gravity and Strings” data set, with bibliographic coupling	167
5-26	Minimum spanning tree itemset-degree metrics for “Wavelets (1-100)” data set.	167
5-27	Minimum spanning tree itemset-degree metrics for “Wavelets (1-500)” data set.	168
5-28	Minimum spanning tree itemset-degree metrics for “Wavelets and Brownian” data set	168
5-29	Minimum spanning tree itemset-degree metrics for “Wavelets and Brownian” data set, with bibliographic coupling	168
5-30	Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Adaptive Optics” data set	169
5-31	Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Adaptive Optics” data set, with bibliographic coupling	170
5-32	Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Collagen” data set	170
5-33	Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Genetic Algorithms and Neural Networks” data set	171
5-34	Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Quantum Gravity and Strings” data set.	171
5-35	Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Quantum Gravity and Strings” data set, with bibliographic coupling	172
5-36	Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Wavelets (1-100)” data set	172
5-37	Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Wavelets (1-500)” data set	173
5-38	Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Wavelets and Brownian” data set	173

5-39 Standard versus hybrid distances in minimum spanning tree visualization with most frequent cardinality-4 itemset, for “Wavelets and Brownian” data set, with bibliographic coupling	174
5-40 Haar, Daubechies, and nearly-symmetric Daubechies wavelets	180
5-41 Translated and scaled wavelets	181
5-42 Fast pyramid algorithm for computing the discrete wavelet transform	182
5-43 Wavelet approximation to signal at various resolutions	183
5-44 One smooth (upper left) and 3 detail basis functions for 2-dimensional extension of Haar wavelet	185
5-45 First stage of 2-dimensional pyramid algorithm	186
5-46 Organization of wavelet coefficients for 2 levels of the 2-dimensional transform.	186
5-47 Image of box with diagonals	187
5-48 3-Level wavelet transform of box image	187
5-49 Outputs of various levels of wavelet low-pass spatial filter for minimum spanning tree density visualization	188
5-50 Minimum spanning tree embedded on its density landscape surface	189
5-51 Interpretation of landscape surface as true clustering mechanism via local attractors of optimization algorithms	190
5-52 Application of spatial thresholds to document landscape to generate crisp clusters	191
6-1 Extension of minimum spanning tree visualization to one higher spatial dimension	197