

Appendix C

Clustering Metrics for Transaction and Item Weighting

This appendix gives detailed results for the experiments conducted in Chapter 4, Section 4.3. In particular, it gives clustering metrics for each test case, along with the corresponding choice of experimental inputs. The results are presented Tables C-1 through C-5, corresponding to the 5 data sets employed.

Table C-1: Clustering metrics for transaction and item weighting, “Collagen” data set. Red and black text marks standard-versus-hybrid distance comparisons.

Linkage	Distances	Frequent itemsets	Clustering metric
complete	pairwise	$o_3, k = 1$	0.6000
complete	transaction weight	$o_3, k = 1$	0.7500
complete	item weight	$o_3, k = 1$	0.4286
average	pairwise	$o_3, k = 1$	0.6000
average	transaction weight	$o_3, k = 1$	0.7500
average	item weight	$o_3, k = 1$	0.1250
single	pairwise	$o_3, k = 1$	0.3750
single	transaction weight	$o_3, k = 1$	0.7500
single	item weight	$o_3, k = 1$	0.1875
complete	pairwise	$o_3, k = 5$	0.5713
complete	transaction weight	$o_3, k = 5$	0.6394
complete	item weight	$o_3, k = 5$	0.6714
average	pairwise	$o_3, k = 5$	0.7100

average	transaction weight	$o_3, k = 5$	0.7400
average	item weight	$o_3, k = 5$	0.5850
single	pairwise	$o_3, k = 5$	0.5450
single	transaction weight	$o_3, k = 5$	0.6210
single	item weight	$o_3, k = 5$	0.1755
complete	pairwise	$o_3, k = 10$	0.5990
complete	transaction weight	$o_3, k = 10$	0.6197
complete	item weight	$o_3, k = 10$	0.5321
average	pairwise	$o_3, k = 10$	0.6779
average	transaction weight	$o_3, k = 10$	0.6829
average	item weight	$o_3, k = 10$	0.3550
single	pairwise	$o_3, k = 10$	0.4600
single	transaction weight	$o_3, k = 10$	0.5962
single	item weight	$o_3, k = 10$	0.1764
complete	pairwise	$o_4, k = 1$	0.8000
complete	transaction weight	$o_4, k = 1$	1.0000
complete	item weight	$o_4, k = 1$	0.5714
average	pairwise	$o_4, k = 1$	0.8000
average	transaction weight	$o_4, k = 1$	1.0000
average	item weight	$o_4, k = 1$	0.1667
single	pairwise	$o_4, k = 1$	0.5000
single	transaction weight	$o_4, k = 1$	1.0000
single	item weight	$o_4, k = 1$	0.2500
complete	pairwise	$o_4, k = 5$	0.5867
complete	transaction weight	$o_4, k = 5$	0.6000
complete	item weight	$o_4, k = 5$	0.4286
average	pairwise	$o_4, k = 5$	0.6629
average	transaction weight	$o_4, k = 5$	0.7029
average	item weight	$o_4, k = 5$	0.4095
single	pairwise	$o_4, k = 5$	0.3000
single	transaction weight	$o_4, k = 5$	0.5810
single	item weight	$o_4, k = 5$	0.2043
complete	pairwise	$o_4, k = 10$	0.5511
complete	transaction weight	$o_4, k = 10$	0.5400
complete	item weight	$o_4, k = 10$	0.4476
average	pairwise	$o_4, k = 10$	0.6424

average	transaction weight	$o_4, k = 10$	0.6524
average	item weight	$o_4, k = 10$	0.4619
single	pairwise	$o_4, k = 10$	0.3000
single	transaction weight	$o_4, k = 10$	0.5317
single	item weight	$o_4, k = 10$	0.1983

Table C-2: Clustering metrics for transaction and item weighting, “Quantum Gravity and Strings” data set. Red and black text marks standard-vs-hybrid distance comparisons.

Linkage	Distances	Frequent itemsets	Clustering metric
complete	pairwise	$o_3, k = 1$	1.0000
complete	transaction weight	$o_3, k = 1$	1.0000
complete	item weight	$o_3, k = 1$	1.0000
average	pairwise	$o_3, k = 1$	1.0000
average	transaction weight	$o_3, k = 1$	1.0000
average	item weight	$o_3, k = 1$	1.0000
single	pairwise	$o_3, k = 1$	1.0000
single	transaction weight	$o_3, k = 1$	1.0000
single	item weight	$o_3, k = 1$	1.0000
complete	pairwise	$o_3, k = 5$	0.9000
complete	transaction weight	$o_3, k = 5$	0.8545
complete	item weight	$o_3, k = 5$	0.9500
average	pairwise	$o_3, k = 5$	0.9000
average	transaction weight	$o_3, k = 5$	0.9500
average	item weight	$o_3, k = 5$	0.9500
single	pairwise	$o_3, k = 5$	0.9000
single	transaction weight	$o_3, k = 5$	0.9500
single	item weight	$o_3, k = 5$	0.8700
complete	pairwise	$o_3, k = 10$	0.7123
complete	transaction weight	$o_3, k = 10$	0.6474
complete	item weight	$o_3, k = 10$	0.7825
average	pairwise	$o_3, k = 10$	0.7425
average	transaction weight	$o_3, k = 10$	0.7475
average	item weight	$o_3, k = 10$	0.7575
single	pairwise	$o_3, k = 10$	0.6857
single	transaction weight	$o_3, k = 10$	0.7475
single	item weight	$o_3, k = 10$	0.6650
complete	pairwise	$o_4, k = 1$	0.6667
complete	transaction weight	$o_4, k = 1$	0.5000
complete	item weight	$o_4, k = 1$	0.8000

average	pairwise	$o_4, k = 1$	0.8000
average	transaction weight	$o_4, k = 1$	0.6667
average	item weight	$o_4, k = 1$	0.6667
single	pairwise	$o_4, k = 1$	0.5714
single	transaction weight	$o_4, k = 1$	0.6667
single	item weight	$o_4, k = 1$	0.1667
complete	pairwise	$o_4, k = 5$	0.7394
complete	transaction weight	$o_4, k = 5$	0.5870
complete	item weight	$o_4, k = 5$	0.7933
average	pairwise	$o_4, k = 5$	0.7933
average	transaction weight	$o_4, k = 5$	0.7667
average	item weight	$o_4, k = 5$	0.7476
single	pairwise	$o_4, k = 5$	0.7286
single	transaction weight	$o_4, k = 5$	0.8267
single	item weight	$o_4, k = 5$	0.5200
complete	pairwise	$o_4, k = 10$	0.7030
complete	transaction weight	$o_4, k = 10$	0.5435
complete	item weight	$o_4, k = 10$	0.7433
average	pairwise	$o_4, k = 10$	0.7433
average	transaction weight	$o_4, k = 10$	0.7167
average	item weight	$o_4, k = 10$	0.6690
single	pairwise	$o_4, k = 10$	0.6500
single	transaction weight	$o_4, k = 10$	0.7467
single	item weight	$o_4, k = 10$	0.3300

Table C-3: Clustering metrics for transaction and item weighting, “Wavelets (1-500)” data set. Red and black text marks standard-vs-hybrid distance comparisons.

Linkage	Distances	Frequent itemsets	Clustering metric
complete	pairwise	$o_3, k = 1$	1.0000
complete	transaction weight	$o_3, k = 1$	1.0000
complete	item weight	$o_3, k = 1$	1.0000
average	pairwise	$o_3, k = 1$	1.0000
average	transaction weight	$o_3, k = 1$	1.0000
average	item weight	$o_3, k = 1$	1.0000
single	pairwise	$o_3, k = 1$	1.0000
single	transaction weight	$o_3, k = 1$	1.0000
single	item weight	$o_3, k = 1$	0.2143
complete	pairwise	$o_3, k = 5$	0.5722
complete	transaction weight	$o_3, k = 5$	0.5722
complete	item weight	$o_3, k = 5$	0.5722
average	pairwise	$o_3, k = 5$	0.5972
average	transaction weight	$o_3, k = 5$	0.6333
average	item weight	$o_3, k = 5$	0.3281
single	pairwise	$o_3, k = 5$	0.4069
single	transaction weight	$o_3, k = 5$	0.3622
single	item weight	$o_3, k = 5$	0.1767
complete	pairwise	$o_3, k = 10$	0.5222
complete	transaction weight	$o_3, k = 10$	0.4528
complete	item weight	$o_3, k = 10$	0.4528
average	pairwise	$o_3, k = 10$	0.5336
average	transaction weight	$o_3, k = 10$	0.5059
average	item weight	$o_3, k = 10$	0.2222
single	pairwise	$o_3, k = 10$	0.3609
single	transaction weight	$o_3, k = 10$	0.2628
single	item weight	$o_3, k = 10$	0.1403
complete	pairwise	$o_4, k = 1$	1.0000
complete	transaction weight	$o_4, k = 1$	1.0000
complete	item weight	$o_4, k = 1$	1.0000

average	pairwise	$o_4, k = 1$	1.0000
average	transaction weight	$o_4, k = 1$	1.0000
average	item weight	$o_4, k = 1$	0.1379
single	pairwise	$o_4, k = 1$	0.2857
single	transaction weight	$o_4, k = 1$	0.2353
single	item weight	$o_4, k = 1$	0.1053
complete	pairwise	$o_4, k = 5$	0.2593
complete	transaction weight	$o_4, k = 5$	0.2593
complete	item weight	$o_4, k = 5$	0.2593
average	pairwise	$o_4, k = 5$	0.3296
average	transaction weight	$o_4, k = 5$	0.4000
average	item weight	$o_4, k = 5$	0.1459
single	pairwise	$o_4, k = 5$	0.2684
single	transaction weight	$o_4, k = 5$	0.2253
single	item weight	$o_4, k = 5$	0.1290
complete	pairwise	$o_4, k = 10$	0.1667
complete	transaction weight	$o_4, k = 10$	0.2259
complete	item weight	$o_4, k = 10$	0.1667
average	pairwise	$o_4, k = 10$	0.2407
average	transaction weight	$o_4, k = 10$	0.3556
average	item weight	$o_4, k = 10$	0.1289
single	pairwise	$o_4, k = 10$	0.2277
single	transaction weight	$o_4, k = 10$	0.2034
single	item weight	$o_4, k = 10$	0.1280

Table C-4: Clustering metrics for transaction and item weighting, “Wavelets and Brownian” data set. Red and black text marks standard-versus-hybrid distance comparisons.

Linkage	Distances	Frequent itemsets	Clustering metric
complete	pairwise	$o_3, k = 1$	1.0000
complete	transaction weight	$o_3, k = 1$	1.0000
complete	item weight	$o_3, k = 1$	0.5000
average	pairwise	$o_3, k = 1$	1.0000
average	transaction weight	$o_3, k = 1$	0.6000
average	item weight	$o_3, k = 1$	0.5000
single	pairwise	$o_3, k = 1$	0.6000
single	transaction weight	$o_3, k = 1$	0.6000
single	item weight	$o_3, k = 1$	0.4286
complete	pairwise	$o_3, k = 5$	0.5600
average	transaction weight	$o_3, k = 5$	0.6145
average	item weight	$o_3, k = 5$	0.5857
average	pairwise	$o_3, k = 5$	0.6857
average	transaction weight	$o_3, k = 5$	0.6757
average	item weight	$o_3, k = 5$	0.5857
single	pairwise	$o_3, k = 5$	0.5657
single	transaction weight	$o_3, k = 5$	0.6257
single	item weight	$o_3, k = 5$	0.5714
complete	pairwise	$o_3, k = 10$	0.5800
complete	transaction weight	$o_3, k = 10$	0.5773
complete	item weight	$o_3, k = 10$	0.6129
average	pairwise	$o_3, k = 10$	0.6479
average	transaction weight	$o_3, k = 10$	0.6229
average	item weight	$o_3, k = 10$	0.6129
single	pairwise	$o_3, k = 10$	0.5314
single	transaction weight	$o_3, k = 10$	0.6229
single	item weight	$o_3, k = 10$	0.5914
complete	pairwise	$o_4, k = 1$	0.6667
complete	transaction weight	$o_4, k = 1$	0.8000
complete	item weight	$o_4, k = 1$	0.6667

average	pairwise	$o_4, k = 1$	1.0000
average	transaction weight	$o_4, k = 1$	0.8000
average	item weight	$o_4, k = 1$	0.6667
single	pairwise	$o_4, k = 1$	0.8000
single	transaction weight	$o_4, k = 1$	0.8000
single	item weight	$o_4, k = 1$	0.5714
complete	pairwise	$o_4, k = 5$	0.6667
complete	transaction weight	$o_4, k = 5$	0.7733
complete	item weight	$o_4, k = 5$	0.7333
average	pairwise	$o_4, k = 5$	0.7600
average	transaction weight	$o_4, k = 5$	0.8133
average	item weight	$o_4, k = 5$	0.7333
single	pairwise	$o_4, k = 5$	0.7543
single	transaction weight	$o_4, k = 5$	0.8133
single	item weight	$o_4, k = 5$	0.6571
complete	pairwise	$o_4, k = 10$	0.6810
complete	transaction weight	$o_4, k = 10$	0.7143
complete	item weight	$o_4, k = 10$	0.6689
average	pairwise	$o_4, k = 10$	0.6933
average	transaction weight	$o_4, k = 10$	0.7200
average	item weight	$o_4, k = 10$	0.6467
single	pairwise	$o_4, k = 10$	0.6514
single	transaction weight	$o_4, k = 10$	0.6621
single	item weight	$o_4, k = 10$	0.5956

Table C-5: Clustering metrics with bibliographic coupling for transaction and item weighting, “Wavelets and Brownian” data set. Red and black text marks standard-vs-hybrid distance comparisons.

Linkage	Distances	Frequent itemsets	Clustering metric
complete	pairwise	$o_3, k = 1$	0.0508
complete	transaction weight	$o_3, k = 1$	0.0508
complete	item weight	$o_3, k = 1$	1.0000
average	pairwise	$o_3, k = 1$	0.2500
average	transaction weight	$o_3, k = 1$	0.6000
average	item weight	$o_3, k = 1$	1.0000
single	pairwise	$o_3, k = 1$	0.2308
single	transaction weight	$o_3, k = 1$	0.6000
single	item weight	$o_3, k = 1$	0.4286
complete	pairwise	$o_3, k = 5$	0.3305
complete	transaction weight	$o_3, k = 5$	0.3903
complete	item weight	$o_3, k = 5$	0.4305
average	pairwise	$o_3, k = 5$	0.5500
average	transaction weight	$o_3, k = 5$	0.6900
average	item weight	$o_3, k = 5$	0.6286
single	pairwise	$o_3, k = 5$	0.5923
single	transaction weight	$o_3, k = 5$	0.6900
single	item weight	$o_3, k = 5$	0.2819
complete	pairwise	$o_3, k = 10$	0.1907
complete	transaction weight	$o_3, k = 10$	0.4853
complete	item weight	$o_3, k = 10$	0.2407
average	pairwise	$o_3, k = 10$	0.3125
average	transaction weight	$o_3, k = 10$	0.6575
average	item weight	$o_3, k = 10$	0.4015
single	pairwise	$o_3, k = 10$	0.4115
single	transaction weight	$o_3, k = 10$	0.6575
single	item weight	$o_3, k = 10$	0.3221
complete	pairwise	$o_4, k = 1$	0.0678
complete	transaction weight	$o_4, k = 1$	1.0000
complete	item weight	$o_4, k = 1$	0.0678

average	pairwise	$o_4, k = 1$	0.1000
average	transaction weight	$o_4, k = 1$	1.0000
average	item weight	$o_4, k = 1$	0.2500
single	pairwise	$o_4, k = 1$	0.3077
single	transaction weight	$o_4, k = 1$	1.0000
single	item weight	$o_4, k = 1$	0.4000
complete	pairwise	$o_4, k = 5$	0.0678
complete	transaction weight	$o_4, k = 5$	0.4007
complete	item weight	$o_4, k = 5$	0.0678
average	pairwise	$o_4, k = 5$	0.2400
average	transaction weight	$o_4, k = 5$	0.6933
average	item weight	$o_4, k = 5$	0.3103
single	pairwise	$o_4, k = 5$	0.3077
single	transaction weight	$o_4, k = 5$	0.7219
single	item weight	$o_4, k = 5$	0.4015
complete	pairwise	$o_4, k = 10$	0.0678
complete	transaction weight	$o_4, k = 10$	0.3807
complete	item weight	$o_4, k = 10$	0.0678
average	pairwise	$o_4, k = 10$	0.1700
average	transaction weight	$o_4, k = 10$	0.6900
average	item weight	$o_4, k = 10$	0.2655
single	pairwise	$o_4, k = 10$	0.3077
single	transaction weight	$o_4, k = 10$	0.7114
single	item weight	$o_4, k = 10$	0.3995