Exercise 1 question 3

Result of simple and randomized apriori algorithm

Data file name	Sample size(default)	Support	Confidence	Resident set size	Time consumption
T10I4D100K.dat	0.01 (1%), 1009 lines	1%	1%	32186368 bytes	5.8 seconds
T40I10D100K.dat	0.01 (1%), 1043 lines	7%	7%	24600576 bytes	3 seconds
Chess.dat	0.01 (1%), 34 lines	90%	90%	1107095552 bytes	81.3 seconds
Connect.dat	0.01 (1%), 653 lines	90%	90%	2065944576 bytes	321.6 seconds
Mushroom.dat	0.01 (1%), 83 lines	50%	50%	11333632 bytes	0.05 seconds
Pumsb.dat	0.01 (1%), 505 lines	95%	95%	18153472 bytes	0.7 seconds
Pumsb_star.dat	0.01 (1%), 442 lines	45%	45%	211255296 bytes	16.7 seconds

Remarks: In Exercise 1 question 3 and 4, support and confidence were adjusted to reflect the frequent itemsets of each dataset.
Moreover, every single command run 10 times to confirm having similar results.

For dataset T10I4D100K.dat, T40I10D100K.dat, frequent itemsets appear in output when support threshold and confidence are set on below 8%

For dataset Chess.dat, Connect.dat, Pumsb.dat, a huge amount frequent itemsets appear in output even when support threshold and confidence are set on as high as 90% or above.

For dataset Mushroom.dat and Pumsb_star.dat, the size of frequent itemsets reduce gradually when the support and confidence are higher. But frequent itemsets still be output when they are set at 90%.

Result of SON algorithm

Data file name		Support	Confidence	Resident set size	Time consumption
T10I4D100K.dat		1%	1%	144224256 bytes	476 seconds
T40I10D100K.dat		7%	7%	416636928 bytes	182.6 seconds
Chess.dat		90%	90%	88915968 bytes	93.5 seconds
Connect.dat		90%	90%	4055944576 bytes	1021.6 seconds
Mushroom.dat		50%	50%	83619840 bytes	11.7 seconds
Pumsb.dat		95%	95%	382140416 bytes	164.5 seconds
Pumsb_star.dat		45%	45%	3131255296 bytes	504.7 seconds

Since SON algorithm must process the whole dataset, it takes longer time and more memory usage as shown by result. However, for chess.dat, SON only 12 seconds longer while the test of apriori algorithm only process 1% of dataset.

Exercise 1 question 4

Result of running T10I4D100K.dat

Result of simple and random apriori algorithm					
Sample size	Support	Confidence	Resident set size	Time consumption	Rules
1%, 1041 lines	1%	1%	34009088 bytes	5.7 seconds	190
2%, 2073 lines	1%	1%	38502400 bytes	12 seconds	158
5%, 4996 lines	1%	1%	46854144 bytes	27 seconds	87
10%, 10026 lines	1%	1%	52629504 bytes	55.8 seconds	27
40%, 50263 lines	1%	1%	101797888 bytes	197.3 seconds	23
50%, 50263 lines	1%	1%	119083008 bytes	252.5 seconds	21
100%, 100000 lines	1%	1%	198774784 bytes	505.1 seconds	21
Result of SON algorithm					
	1%	1%	144224256 bytes	476 seconds	22

Observation

On time usage, it increases in proportion to the change of sample size while memory usage doesn't. Similar size of associate rule can be retrieved when sample size is above 40%, while 21/23 of results of 40% sample size are same the result of entire dataset, and time usage is approximately 60% lower and memory usage is approximately 50% lower.

However, the memory usage of running the SON algorithm is 27.4% less than that of running the simple and random apriori algorithm on the entire dataset, and SON is running 5.74% faster than the simple and random apriori algorithm. Their results of associate rules is very similar as well, the difference is only 1.

Reflection

Throughout the implementation and running experiments, it is very difficult to know if I get a correct result or not. I must test each dataset with different support threshold and confidence and find out how 'frequent' is the dataset. For some data file, if you set the support and confidence low, it will take a very long time to run the program. Moreover, I need to check time usage, memory usage, the rules number and similarity of frequent itemsets of different sample size to assess the efficiency and accuracy of the algorithm, which is very time-consuming. When program run so long on some datasets, I may suspect

there are bugs in my code, however, I still needed to wait, since I didn't know it's due to my code functionality or the size of dataset.

Moreover, multiprocessing library for Python doesn't work, and I can't find another way to do before the deadline. Therefore, I can only split the data and process each data in apriori algorithm and then finish the second mapreduce of SON algorithm, which means my code takes longer time to finish.

Outcome of 1% sample size

({'862'}) - ({'494'})
({'494'}) - ({'862'})
({'217'}) - ({'346'})
({'346'}) - ({'217'})
({'888'}) - ({'208'})
({'208'}) - ({'888'})
({'290'}) - ({'888'})
({'888'}) - ({'290'})
({'290'}) - ({'208'})
({'208'}) - ({'290'})
({'458'}) - ({'888'})
({'888'}) - ({'458'})
({'458'}) - ({'208'})
({'208'}) - ({'458'})
({'888'}) - ({'969'})
({'969'}) - ({'888'})
({'208'}) - ({'969'})
({'969'}) - ({'208'})
({'290'}) - ({'969'})
({'969'}) - ({'290'})
({'458'}) - ({'969'})
({'969'}) - ({'458'})
({'766'}) - ({'572'})
({'572'}) - ({'766'})
({'368'}) - ({'829'})
({'829'}) - ({'368'})
({'789'}) - ({'829'})
({'829'}) - ({'789'})
({'509'}) - ({'335'})
({'335'}) - ({'509'})
({'947'}) - ({'661'})
({'661'}) - ({'947'})
({'217'}) - ({'661'})
({'661'}) - ({'217'})
({'947'}) - ({'546'})
({'546'}) - ({'947'})

({'661'}) - ({'546'})
({'546'}) - ({'661'})
({'217'}) - ({'546'})
({'546'}) - ({'217'})
({'923'}) - ({'947'})
({'947'}) - ({'923'})
({'923'}) - ({'661'})
({'661'}) - ({'923'})
({'923'}) - ({'217'})
({'217'}) - ({'923'})
({'923'}) - ({'546'})
({'546'}) - ({'923'})
({'758'}) - ({'471'})
({'471'}) - ({'758'})
({'947'}) - ({'217'})
({'217'}) - ({'947'})
({'510'}) - ({'487'})
({'487'}) - ({'510'})
({'510'}) - ({'217'})
({'217'}) - ({'510'})
({'75'}) - ({'684'})
({'684'}) - ({'75'})
({'57'}) - ({'937'})
({'937'}) - ({'57'})
({'529'}) - ({'598'})
({'598'}) - ({'529'})
({'598'}) - ({'782'})
({'782'}) - ({'598'})
({'704'}) - ({'39'})
({'39'}) - ({'704'})
({'825'}) - ({'39'})
({'39'}) - ({'825'})
({'825'}) - ({'704'})
({'704'}) - ({'825'})
({'458'}) - ({'969', '888'})
({'888'}) - ({'969', '458'})
({'969'}) - ({'888', '458'})
({'888'}) - ({'969', '290'})
({'969'}) - ({'888', '290'})
({'290'}) - ({'969', '888'})
({'888'}) - ({'969', '208'})
({'208'}) - ({'969', '888'})
({'969'}) - ({'208', '888'})
({'888'}) - ({'208', '290'})
({'208'}) - ({'888', '290'})
({'290'}) - ({'208', '888'})
({'923'}) - ({'217', '947'})

({'217'}) - ({'947', '923'})
({'947'}) - ({'217', '923'})
({'923'}) - ({'546', '217'})
({'217'}) - ({'546', '923'})
({'546'}) - ({'217', '923'})
({'923'}) - ({'546', '661'})
({'661'}) - ({'546', '923'})
({'546'}) - ({'661', '923'})
({'217'}) - ({'546', '661'})
({'661'}) - ({'546', '217'})
({'546'}) - ({'661', '217'})
({'923'}) - ({'546', '947'})
({'947'}) - ({'546', '923'})
({'546'}) - ({'947', '923'})
({'947'}) - ({'546', '217'})
({'217'}) - ({'546', '947'})
({'546'}) - ({'217', '947'})
({'947'}) - ({'546', '661'})
({'661'}) - ({'546', '947'})
({'546'}) - ({'661', '947'})
({'923'}) - ({'661', '217'})
({'217'}) - ({'661', '923'})
({'661'}) - ({'217', '923'})
({'923'}) - ({'661', '947'})
({'947'}) - ({'661', '923'})
({'661'}) - ({'947', '923'})
({'947'}) - ({'661', '217'})
({'217'}) - ({'661', '947'})
({'661'}) - ({'217', '947'})
({'825'}) - ({'39', '704'})
({'704'}) - ({'39', '825'})
({'39'}) - ({'704', '825'})
({'947', '923'}) - ({'661', '217'})
({'217', '923'}) - ({'661', '947'})
({'217', '947'}) - ({'661', '923'})
({'661', '923'}) - ({'217', '947'})
({'661', '947'}) - ({'217', '923'})
({'661', '217'}) - ({'947', '923'})
({'923'}) - ({'661', '217', '947'})
({'947'}) - ({'661', '217', '923'})
({'217'}) - ({'661', '947', '923'})
({'661'}) - ({'217', '947', '923'})
({'546', '947'}) - ({'661', '217'})
({'217', '947'}) - ({'546', '661'})
({'546', '217'}) - ({'661', '947'})
({'661', '947'}) - ({'546', '217'})
({'546', '661'}) - ({'217', '947'})

({'661', '217'}) - ({'546', '947'})
({'546'}) - ({'661', '217', '947'})
({'947'}) - ({'546', '661', '217'})
({'217'}) - ({'546', '661', '947'})
({'661'}) - ({'546', '217', '947'})
({'947', '923'}) - ({'546', '661'})
({'546', '923'}) - ({'661', '947'})
({'546', '947'}) - ({'661', '923'})
({'661', '923'}) - ({'546', '947'})
({'661', '947'}) - ({'546', '923'})
({'546', '661'}) - ({'947', '923'})
({'923'}) - ({'546', '661', '947'})
({'947'}) - ({'546', '661', '923'})
({'546'}) - ({'661', '947', '923'})
({'661'}) - ({'546', '947', '923'})
({'546', '923'}) - ({'661', '217'})
({'217', '923'}) - ({'546', '661'})
({'546', '217'}) - ({'661', '923'})
({'661', '923'}) - ({'546', '217'})
({'546', '661'}) - ({'217', '923'})
({'661', '217'}) - ({'546', '923'})
({'546'}) - ({'661', '217', '923'})
({'923'}) - ({'546', '661', '217'})
({'217'}) - ({'546', '661', '923'})
({'661'}) - ({'546', '217', '923'})
({'947', '923'}) - ({'546', '217'})
({'546', '923'}) - ({'217', '947'})
({'546', '947'}) - ({'217', '923'})
({'217', '923'}) - ({'546', '947'})
({'217', '947'}) - ({'546', '923'})
({'546', '217'}) - ({'947', '923'})
({'923'}) - ({'546', '217', '947'})
({'947'}) - ({'546', '217', '923'})
({'546'}) - ({'217', '947', '923'})
({'217'}) - ({'546', '947', '923'})
({'661', '947', '923'}) - ({'546', '217'})
({'546', '661', '923'}) - ({'217', '947'})
({'546', '661', '947'}) - ({'217', '923'})
({'546', '947', '923'}) - ({'661', '217'})
({'661', '217', '923'}) - ({'546', '947'})
({'661', '217', '947'}) - ({'546', '923'})
({'217', '947', '923'}) - ({'546', '661'})
({'546', '661', '217'}) - ({'947', '923'})
({'546', '217', '923'}) - ({'661', '947'})
({'546', '217', '947'}) - ({'661', '923'})
({'661', '923'}) - ({'546', '217', '947'})
({'661', '947'}) - ({'546', '217', '923'})

({'947', '923'}) - ({'546', '661', '217'})
({'546', '661'}) - ({'217', '947', '923'})
({'546', '923'}) - ({'661', '217', '947'})
({'546', '947'}) - ({'661', '217', '923'})
({'661', '217'}) - ({'546', '947', '923'})
({'217', '923'}) - ({'546', '661', '947'})
({'217', '947'}) - ({'546', '661', '923'})
({'546', '217'}) - ({'661', '947', '923'})
({'661'}) - ({'217', '546', '947', '923'})
({'923'}) - ({'661', '217', '546', '947'})
({'947'}) - ({'661', '217', '546', '923'})
({'546'}) - ({'661', '217', '947', '923'})
({'217'}) - ({'661', '546', '947', '923'})

Result: 190 rules

Outcome of 2% sample size

({'829'}) - ({'789'})
({'789'}) - ({'829'})
({'829'}) - ({'368'})
({'368'}) - ({'829'})
({'682'}) - ({'368'})
({'368'}) - ({'682'})
({'541'}) - ({'72'})
({'72'}) - ({'541'})
({'346'}) - ({'217'})
({'217'}) - ({'346'})
({'782'}) - ({'598'})
({'598'}) - ({'782'})
({'598'}) - ({'529'})
({'529'}) - ({'598'})
({'795'}) - ({'571'})
({'571'}) - ({'795'})
({'623'}) - ({'571'})
({'571'}) - ({'623'})
({'57'}) - ({'937'})
({'937'}) - ({'57'})
({'290'}) - ({'208'})
({'208'}) - ({'290'})
({'290'}) - ({'458'})
({'458'}) - ({'290'})
({'458'}) - ({'208'})
({'208'}) - ({'458'})
({'888'}) - ({'290'})
({'290'}) - ({'888'})
({'888'}) - ({'208'})

({'208'}) - ({'888'})
({'888'}) - ({'458'})
({'458'}) - ({'888'})
({'969'}) - ({'290'})
({'290'}) - ({'969'})
({'969'}) - ({'208'})
({'208'}) - ({'969'})
({'969'}) - ({'458'})
({'458'}) - ({'969'})
({'969'}) - ({'888'})
({'888'}) - ({'969'})
({'676'}) - ({'6'})
({'6'}) - ({'676'})
({'390'}) - ({'227'})
({'227'}) - ({'390'})
({'227'}) - ({'722'})
({'722'}) - ({'227'})
({'390'}) - ({'722'})
({'722'}) - ({'390'})
({'862'}) - ({'392'})
({'392'}) - ({'862'})
({'969'}) - ({'458', '888'})
({'888'}) - ({'458', '969'})
({'458'}) - ({'888', '969'})
({'969'}) - ({'208', '888'})
({'888'}) - ({'208', '969'})
({'208'}) - ({'888', '969'})
({'969'}) - ({'290', '888'})
({'888'}) - ({'290', '969'})
({'290'}) - ({'888', '969'})
({'969'}) - ({'208', '458'})
({'458'}) - ({'208', '969'})
({'208'}) - ({'458', '969'})
({'888'}) - ({'208', '458'})
({'458'}) - ({'208', '888'})
({'208'}) - ({'458', '888'})
({'290'}) - ({'458', '969'})
({'969'}) - ({'458', '290'})
({'458'}) - ({'290', '969'})
({'290'}) - ({'458', '888'})
({'888'}) - ({'458', '290'})
({'458'}) - ({'290', '888'})
({'969'}) - ({'208', '290'})
({'290'}) - ({'208', '969'})
({'208'}) - ({'290', '969'})
({'888'}) - ({'208', '290'})
({'290'}) - ({'208', '888'})

({'208'}) - ({'290', '888'})
({'458'}) - ({'208', '290'})
({'290'}) - ({'208', '458'})
({'208'}) - ({'290', '458'})
({'390'}) - ({'722', '227'})
({'227'}) - ({'722', '390'})
({'722'}) - ({'227', '390'})
({'208', '888'}) - ({'290', '458'})
({'208', '458'}) - ({'290', '888'})
({'458', '888'}) - ({'208', '290'})
({'208', '290'}) - ({'458', '888'})
({'290', '888'}) - ({'208', '458'})
({'290', '458'}) - ({'208', '888'})
({'208'}) - ({'290', '888', '458'})
({'888'}) - ({'208', '290', '458'})
({'458'}) - ({'208', '290', '888'})
({'290'}) - ({'208', '458', '888'})
({'208', '969'}) - ({'290', '458'})
({'458', '969'}) - ({'208', '290'})
({'208', '458'}) - ({'290', '969'})
({'290', '969'}) - ({'208', '458'})
({'208', '290'}) - ({'458', '969'})
({'290', '458'}) - ({'208', '969'})
({'208'}) - ({'290', '458', '969'})
({'969'}) - ({'208', '290', '458'})
({'458'}) - ({'208', '290', '969'})
({'290'}) - ({'208', '458', '969'})
({'208', '969'}) - ({'290', '888'})
({'888', '969'}) - ({'208', '290'})
({'208', '888'}) - ({'290', '969'})
({'290', '969'}) - ({'208', '888'})
({'208', '290'}) - ({'888', '969'})
({'290', '888'}) - ({'208', '969'})
({'208'}) - ({'290', '888', '969'})
({'969'}) - ({'208', '290', '888'})
({'888'}) - ({'208', '290', '969'})
({'290'}) - ({'208', '888', '969'})
({'208', '969'}) - ({'458', '888'})
({'888', '969'}) - ({'208', '458'})
({'208', '888'}) - ({'458', '969'})
({'458', '969'}) - ({'208', '888'})
({'208', '458'}) - ({'888', '969'})
({'458', '888'}) - ({'208', '969'})
({'208'}) - ({'458', '888', '969'})
({'969'}) - ({'208', '458', '888'})
({'888'}) - ({'208', '458', '969'})
({'458'}) - ({'208', '888', '969'})

({'888', '969'}) - ({'458', '290'})
({'290', '969'}) - ({'458', '888'})
({'290', '888'}) - ({'458', '969'})
({'458', '969'}) - ({'290', '888'})
({'458', '888'}) - ({'290', '969'})
({'458', '290'}) - ({'888', '969'})
({'969'}) - ({'458', '888', '290'})
({'888'}) - ({'458', '969', '290'})
({'290'}) - ({'458', '888', '969'})
({'458'}) - ({'290', '888', '969'})
({'290', '458', '969'}) - ({'208', '888'})
({'208', '290', '458'}) - ({'888', '969'})
({'208', '458', '969'}) - ({'290', '888'})
({'208', '290', '969'}) - ({'458', '888'})
({'290', '888', '458'}) - ({'208', '969'})
({'458', '888', '969'}) - ({'208', '290'})
({'290', '888', '969'}) - ({'208', '458'})
({'208', '458', '888'}) - ({'290', '969'})
({'208', '290', '888'}) - ({'458', '969'})
({'208', '888', '969'}) - ({'290', '458'})
({'290', '458'}) - ({'208', '888', '969'})
({'458', '969'}) - ({'208', '290', '888'})
({'290', '969'}) - ({'208', '458', '888'})
({'208', '458'}) - ({'290', '888', '969'})
({'208', '969'}) - ({'290', '888', '458'})
({'208', '290'}) - ({'458', '888', '969'})
({'458', '888'}) - ({'208', '290', '969'})
({'290', '888'}) - ({'208', '458', '969'})
({'888', '969'}) - ({'208', '290', '458'})
({'208', '888'}) - ({'290', '458', '969'})
({'458'}) - ({'290', '888', '208', '969'})
({'969'}) - ({'290', '458', '888', '208'})
({'290'}) - ({'458', '888', '208', '969'})
({'208'}) - ({'290', '458', '888', '969'})
({'888'}) - ({'290', '458', '208', '969'})

Result: 158 rules

Outcome of 5% sample size

({'390'}) - ({'722'})
({'722'}) - ({'390'})
({'368'}) - ({'494'})
({'494'}) - ({'368'})
({'704'}) - ({'39'})
({'39'}) - ({'704'})
({'704'}) - ({'825'})

({'825'}) - ({'704'})
({'825'}) - ({'39'})
({'39'}) - ({'825'})
({'283'}) - ({'33'})
({'33'}) - ({'283'})
({'217'}) - ({'33'})
({'33'}) - ({'217'})
({'346'}) - ({'33'})
({'33'}) - ({'346'})
({'283'}) - ({'346'})
({'346'}) - ({'283'})
({'515'}) - ({'33'})
({'33'}) - ({'515'})
({'515'}) - ({'283'})
({'283'}) - ({'515'})
({'515'}) - ({'217'})
({'217'}) - ({'515'})
({'515'}) - ({'346'})
({'346'}) - ({'515'})
({'283'}) - ({'217'})
({'217'}) - ({'283'})
({'217'}) - ({'346'})
({'346'}) - ({'217'})
({'722'}) - ({'227'})
({'227'}) - ({'722'})
({'494'}) - ({'682'})
({'682'}) - ({'494'})
({'368'}) - ({'829'})
({'829'}) - ({'368'})
({'923'}) - ({'947'})
({'947'}) - ({'923'})
({'368'}) - ({'682'})
({'682'}) - ({'368'})
({'598'}) - ({'529'})
({'529'}) - ({'598'})
({'862'}) - ({'392'})
({'392'}) - ({'862'})
({'390'}) - ({'227'})
({'227'}) - ({'390'})
({'515'}) - ({'346', '217'})
({'217'}) - ({'346', '515'})
({'346'}) - ({'217', '515'})
({'515'}) - ({'217', '283'})
({'283'}) - ({'217', '515'})
({'217'}) - ({'283', '515'})
({'515'}) - ({'33', '346'})
({'346'}) - ({'33', '515'})

({'33'}) - ({'346', '515'})
({'515'}) - ({'33', '217'})
({'217'}) - ({'33', '515'})
({'33'}) - ({'217', '515'})
({'217'}) - ({'33', '346'})
({'346'}) - ({'33', '217'})
({'33'}) - ({'346', '217'})
({'515'}) - ({'33', '283'})
({'283'}) - ({'33', '515'})
({'33'}) - ({'283', '515'})
({'283'}) - ({'33', '217'})
({'217'}) - ({'33', '283'})
({'33'}) - ({'217', '283'})
({'283', '515'}) - ({'33', '217'})
({'33', '515'}) - ({'217', '283'})
({'33', '283'}) - ({'217', '515'})
({'217', '515'}) - ({'33', '283'})
({'217', '283'}) - ({'33', '515'})
({'33', '217'}) - ({'283', '515'})
({'515'}) - ({'33', '217', '283'})
({'283'}) - ({'33', '217', '515'})
({'33'}) - ({'217', '283', '515'})
({'217'}) - ({'33', '283', '515'})
({'33', '515'}) - ({'346', '217'})
({'217', '515'}) - ({'33', '346'})
({'33', '217'}) - ({'346', '515'})
({'346', '515'}) - ({'33', '217'})
({'33', '346'}) - ({'217', '515'})
({'346', '217'}) - ({'33', '515'})
({'33'}) - ({'346', '217', '515'})
({'515'}) - ({'33', '346', '217'})
({'217'}) - ({'33', '346', '515'})
({'346'}) - ({'33', '217', '515'})

Result: 87 rules

Outcome of 10% of sample size

({'390'}) - ({'227'})
({'227'}) - ({'390'})
({'390'}) - ({'722'})
({'722'}) - ({'390'})
({'682'}) - ({'368'})
({'368'}) - ({'682'})
({'368'}) - ({'829'})
({'829'}) - ({'368'})
({'217'}) - ({'283'})

({'283'}) - ({'217'})
({'346'}) - ({'283'})
({'283'}) - ({'346'})
({'346'}) - ({'217'})
({'217'}) - ({'346'})
({'39'}) - ({'704'})
({'704'}) - ({'39'})
({'825'}) - ({'704'})
({'704'}) - ({'825'})
({'825'}) - ({'39'})
({'39'}) - ({'825'})
({'789'}) - ({'829'})
({'829'}) - ({'789'})
({'862'}) - ({'392'})
({'392'}) - ({'862'})
({'825'}) - ({'704', '39'})
({'39'}) - ({'704', '825'})
({'704'}) - ({'39', '825'})

Result: 27 rules

Outcome of 40% sample size

({'722'}) - ({'227'})
({'227'}) - ({'722'})
({'722'}) - ({'390'})
({'390'}) - ({'722'})
({'390'}) - ({'227'})
({'227'}) - ({'390'})
({'368'}) - ({'682'})
({'682'}) - ({'368'})
({'829'}) - ({'368'})
({'368'}) - ({'829'})
({'829'}) - ({'789'})
({'789'}) - ({'829'})
({'217'}) - ({'346'})
({'346'}) - ({'217'})
({'825'}) - ({'39'})
({'39'}) - ({'825'})
({'825'}) - ({'704'})
({'704'}) - ({'825'})
({'39'}) - ({'704'})
({'704'}) - ({'39'})
({'825'}) - ({'704', '39'})
({'39'}) - ({'704', '825'})
({'704'}) - ({'39', '825'})

Result: 23 rules

Outcome of 50% sample size

({'722'}) - ({'390'})
({'390'}) - ({'722'})
({'39'}) - ({'704'})
({'704'}) - ({'39'})
({'682'}) - ({'368'})
({'368'}) - ({'682'})
({'39'}) - ({'825'})
({'825'}) - ({'39'})
({'346'}) - ({'217'})
({'217'}) - ({'346'})
({'789'}) - ({'829'})
({'829'}) - ({'789'})
({'704'}) - ({'825'})
({'825'}) - ({'704'})
({'368'}) - ({'829'})
({'829'}) - ({'368'})
({'227'}) - ({'390'})
({'390'}) - ({'227'})
({'39'}) - ({'825', '704'})
({'704'}) - ({'825', '39'})
({'825'}) - ({'704', '39'})

Result: 21 rules

Outcome of 100% sample size

({'704'}) - ({'39'})
({'39'}) - ({'704'})
({'346'}) - ({'217'})
({'217'}) - ({'346'})
({'227'}) - ({'390'})
({'390'}) - ({'227'})
({'390'}) - ({'722'})
({'722'}) - ({'390'})
({'829'}) - ({'368'})
({'368'}) - ({'829'})
({'368'}) - ({'682'})
({'682'}) - ({'368'})
({'825'}) - ({'39'})
({'39'}) - ({'825'})
({'829'}) - ({'789'})

({'789'}) - ({'829'})
({'825'}) - ({'704'})
({'704'}) - ({'825'})
({'825'}) - ({'39', '704'})
({'704'}) - ({'39', '825'})
({'39'}) - ({'704', '825'})

Result: 21 rules

Outcome of SON algo in q4:

[[frozenset({'368'}), frozenset({'682'})],
[frozenset({'39'}), frozenset({'825'})],
[frozenset({'39'}), frozenset({'704'})],
[frozenset({'368'}), frozenset({'829'})],
[frozenset({'704'}), frozenset({'825', '39'})],
[frozenset({'682'}), frozenset({'368'})],
[frozenset({'825'}), frozenset({'39'})],
[frozenset({'704'}), frozenset({'39'})],
[frozenset({'829'}), frozenset({'368'})],
[frozenset({'390'}), frozenset({'722'})],
[frozenset({'390'}), frozenset({'227'})],
[frozenset({'789'}), frozenset({'829'})],
[frozenset({'825'}), frozenset({'704'})],
[frozenset({'346'}), frozenset({'217'})],
[frozenset({'825'}), frozenset({'704', '39'})],
[frozenset({'722'}), frozenset({'390'})],
[frozenset({'227'}), frozenset({'390'})],
[frozenset({'829'}), frozenset({'789'})],
[frozenset({'704'}), frozenset({'825'})],
[frozenset({'217'}), frozenset({'346'})],
[frozenset({'39'}), frozenset({'704', '825'})]]

22 rules