Morgan Baccus
CptS 315
Homework 1

Analytical Part

Q1.

a) 4
b) 4/6 = .66 or 66%
c) 4/6 = .66 or 66%

Q2.

a) For n = 20 and pair (7, 8)
position = (i – 1) * (n – (i / 2)) + j
= (7 – 1) * (20 – (7/2)) + *8 - 7)
= 100

b) Since only 10% of the total pairs will have a non-zero count, we will prefer the tabular method as it will consume less memory compared to the triangular matrix.

Q3.

a)

Item	**Support**
1	4
2	6
3	8
4	8
5	6
6	4

Pair	**Support**
(1, 2)	2
(1, 3)	3
(2, 3)	3
(2, 4)	4
(3, 4)	4
(3, 5)	3
(4, 5)	3
(4, 6)	3
(5, 6)	2
(1, 5)	1
(2, 6)	1
(1, 4)	2
(2, 5)	2
(3, 6)	2

b)

Pair	**Support**
(1, 2)	2
(1, 3)	3
(2, 3)	6
(2, 4)	8
(3, 4)	1
(3, 5)	4
(4, 5)	9
(4, 6)	2
(5, 6)	8
(1, 5)	5
(2, 6)	1
(1, 4)	4
(2, 5)	10
(3, 6)	7

c) Buckets 1, 2, 4, and 8 are frequent

d) {1,2}, {1,4}, {2,4}, {2,6}, {3,4}, {3,5}, {4,6}, {5,6}

Q4.

This paper goes into details of how to detect partial copies and why that is more difficult than finding full copies. The fingerprint technique used to do this is called winnowing. Many of the applications that use winnowing record what the fingerprint is, as well as the locations within the documents that are being compared. There are many uses for such algorithms including plagiarism. To take winnowing a step further, an algorithm called MOSS can determine the source of the fingerprint and determine if copying would be acceptable or not. For example, math equations could be fingerprints that are labeled as a non-plagiarism sources, but an article on Wikipedia would be labeled as a plagiarism source. This reduces the number of false positives reported by the algorithm. The paper goes into further detail about how winnowing works and its different uses. In conclusion, winnowing is a very beneficial local fingerprinting algorithm with many different applications.