Name: _____ USC ID: _____

INF 553 – Spring 2018

Quiz 3: PCY algorithm (10 points), **15 minutes**

Consider the PCY algorithm. Support the minimum support threshold = 3 and the hash function h(i,j)= (i+j)%5.

1,2,3
2,3
1,2,4
3,4
1,2,3,4

1. [5 points] Show the content of count table and the **frequent-buckets** table produced by PCY in the first pass. For the frequent-buckets table, show the actual counts in the bucket and also the bitmap generated from the table.

1	3
2	4
3	4
4	3

0	$(1,4)^2$, $(2,3)^3$	5	1
1	$(2,4)^1$	1	0
2	$(3,4)^2$	2	0
3	$(1,2)^3$	3	1
4	$(1,3)^2$	2	0

2. [3 points] Show the candidate itemsets that the frequent-buckets table helps remove from consideration in the second pass.

   Removed itemsets: (2,4), (3,4), (1,3)

3. [2 points] Show all the frequent 2-itemsets discovered by PCY.

   (1,4), (2,3), (1,2)