Name: _____ USC ID: _____

INF 553 – Spring 2018

Quiz 4: Jaccard & Minhash signatures (10 points), 15 minutes

Consider three sets: S1, S2, and S3 whose characteristic matrix is shown in columns 1-4 of the table below.

Row #	S1	S2	S3	f1(x)=(x+3)%5	f2(x)=(3x+1)%5
0	0	1	1	3	1
1	1	1	0	4	4
2	0	0	1	0	2
3	1	1	0	1	0
4	0	1	1	2	3

1. [3 points] Suppose we generate signatures of the sets using hash functions f1(x) and f2(x) shown in the table above. Recall that x is the old row #. Fill in the blanks with the new row numbers generated by each hash function.

2. [3 points] Fill in the signature matrix below with minHash values generated by f1 and f2 for the three sets.

	S1	S2	S3
f1	1	1	0
f2	0	0	1

3. [4 points] Fill in the table below with the actual similarity of pairs (i.e., their Jaccard similarity) and their similarity estimated using the signatures.

Pair	Actual Sim.	Estimated Sim.
(S1, S2)		
(S1, S3)		
(S2, S3)		

Pair	Actual Sim	Estimated Sim
(S1,S2)	1/2	1
(S1,S3)	0	0
(S2,S3)	2/5	0

Actual Sim worth 2 points
Estimated Sim worth 2 points