COMP9313 Assignment

ID: Z5085901
Name: Jingjie JIN

Question 1. MapReduce

```
class Pair
    node1, node2
    int compareTo (Pair p)
        int ret =  this.node1.compareTo(p.node1)
        if (ret == 0)
            ret = this.node2.compareTo(p.node2)
        return ret

class Mapper
    method Map (key, adjacency list L)
        node = L.split()[0]
        adjList = L.split().remove(node)
        initialize a list pairList to store each Pair pair containing node and adjacency
        foreach (adj in adjList)
            if (node.toInt < adj.toInt)
                pair = Pair (node, adj)
                pairList.add(pair)
            else
                pair = Pair (adj, node)
                pairList.add(pair)
        foreach (pair_i in pairList)
            foreach (pair_j in pairList)
                if (i != j)
                    Emit (pair_i, pair_j)

class Partitioner
    method int getPartition(key, value, int numPartitions)
        return key.first.hashCode() % Integer.MAX_VALUE % numPartitions

class PairGroupingComparator

class Reducer
    method Reduce (pair, edgesList ([node1, adj1], [node1, adj2], …, [node2, adj1]
…))
        Emit (pair, edgesList)
```

Question 2. LSH

(i). Set of 2-shingles:

S(A) = {“the sky”, “sky is”, “is blue”, “blue the”, “the sun”, “sun is”, “is bright”}

S(B) = {“the sun”, “sun in”, “in the”, “the sky”, “sky is”, “is bright”}

Jaccard similarity:

Sim(A, B) = | S(A) intersect S(B) | / | S(A) union S(B) |

= |4| / |7+6-4|

= 4 / 9

= 0.444

(ii).

Row	2-shingles	A	B	h1= 5n-1 mod 9	h2 = 2n+1 mod 9
0	the sky	1	1	8	1
1	sky is	1	1	4	3
2	is blue	1	0	0	5
3	blue the	1	0	5	7
4	the sun	1	1	1	0
5	sun is	1	0	6	2
6	is bright	1	1	2	4
7	sun in	0	1	7	6
8	in the	0	1	3	8

Initialize:

	Sig(A)	Sig(B)
h1	∞	∞
h2	∞	∞

Row 0:

	Sig(A)	Sig(B)
h1	8	8
h2	1	1

Row 1:

	Sig(A)	Sig(B)
h1	4	4
h2	1	1

Row 2:

	Sig(A)	Sig(B)
h1	0	4
h2	1	1

Row 3:

	Sig(A)	Sig(B)
h1	0	4
h2	1	1

Row 4:

	Sig(A)	Sig(B)
h1	0	1
h2	0	0

Row 5:

	Sig(A)	Sig(B)
h1	0	1
h2	0	0

Row 6:

	Sig(A)	Sig(B)
h1	0	1
h2	0	0

Row 7:

	Sig(A)	Sig(B)
h1	0	1
h2	0	0

Row 8: (the resulting signatures for A and B)

	Sig(A)	Sig(B)
h1	0	1
h2	0	0