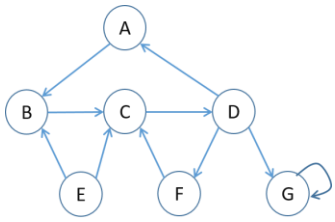
CS4225/CS5425 BIG DATA SYSTEMS FOR DATA SCIENCE

Tutorial 8: Large Graph Processing II

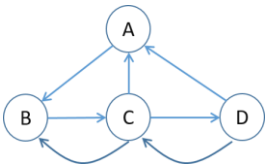
1. Given the following graph,

1) how many dead ends are there in the graph? For each dead end (if any), please indicate the set of vertices forming the dead end.

2) how many spider traps are there in the graph? For each spider trap (if any), please indicate the set of vertices forming the spider trap.



2. Set up the PageRank equations for the below graph, assuming β = 0.8 (jump probability = 1−β). Denote the PageRank of node *x* by *r(x)*.



3. Suppose you have a large graph, and you will implement breadth first traversal on the graph. Each vertex has three attributes: 1) *id*, the vertex ID, 2) *isVisited*, indicating the vertex has been visited or not, and 3) *vList*, the list of neighbour vertices. Show the pseudo code on how you would use Pregel to perform a breadth first traversal on the graph, starting with a pre-defined vertex V_0. You must follow the pseudo code of the below format.

```
compute (vertex v) {

/* your pseudo code*/

/* you can use two APIs given in Pregel: 1) getSuperStep(): Retrieves the current superstep;
```

```
2) voteToHalt(): After this is called, the compute() code will no longer be called for this vertex.*/

}
```