Mapper Task 1:

1. In the code we are reading the data sample content from command line argument
2. We checked whether the line is valid or not
3. If its valid we split the line to get the two variables srcNode , destNode and passed it to the reducer.

Format given to reducer: srcNode destNode

Reducer Task 1:

1. Opens the v file
2. The reducer takes the content passed from the mapper and splits them into lines. Each line is then split into srcNode and destNode
3. We used a variable called currNode which stores the state of the current source node.
4. Based on the current node the destNode is appended into the corresponding adjacency list which has the destination nodes or it gets printed as output.

Format of Output: srcNode [List of destNodes]

1. A contribution of 1 will get added for every source node in the V file.

Format in V file : srcNode,1

Mapper task 2:

1. Here we are reading the content of v file along with the embedding json file. The output of the task 1 reducer is passed to the Mapper.
2. The mapper reads each line and split them into srcNode, to\_Nodes . The to\_Node is evaluated to a list.
3. For every destination node in the to\_Nodes we are finding the total contribution due to single srcNode which is the product of the srcNode contribution and similarity(got this using embedding json file), it is passed to the reducer.

Format of output: destNode ,srcNode, final contribution

Reducer Task 2:

1. V1 file is opened here
2. The output passed in mapper task 2 is taken as the input here, which gets split into lines
3. Each lines is then split into destination node , source node , page contribution
4. For every destination node all the page contributions are combined using the formula 0.15+ 0.85(Sum of all the page contributions) and gets printed into the v1 file

Format in v1 file: destNode,Final Contribution