PageRank Algorithm Implementation

- **Technology/Concept:**
  - **Graph Processing**: The PageRank algorithm is a graph-based computation that is commonly implemented in big data frameworks like **Apache Spark (GraphX)** or **Apache Giraph**, which are optimized for handling large-scale graphs and distributed computations.
  - **Distributed Computing**: If the dataset (graph) is large, tools like Spark can help distribute the PageRank computation across multiple nodes.

Steps:

1. Read the text file and create an RDD.
2. Pass this RDD into a parsing function and the edges node wise
3. Count the number of linkages
4. Create ranks
5. Iterating until the distance get conversed, assuming that the comparing distance(delta) is 0.1

Output:

```
# Passing the text through the parser function and grouping it as per the nodes as key and values being the connections tha
txt_lnk = text.map(lambda x: passed(x)).distinct().groupByKey().mapValues(lambda x: list(x)).cache()
txt_lnk.collect()
```

```
[('1', ['1', '2']), ('2', ['1', '3']), ('3', ['2'])]
```

```

```

```
# Counting the number of nodes.
X = txt_lnk.count()
print(X)
```

```
3
```

```
# Generating ranks for each node
rank = txt_lnk.map(lambda node: (node[0],1.0/N))
print(rank.collect())
```

```
[('1', 0.3333333333333333), ('2', 0.3333333333333333), ('3', 0.3333333333333333)]
```

```
In [28]:
old_rank = rank.collect()
while True:
    print("Old Rank - {0}".format(old_rank))
    ranks = txt_lnk.join(rank).flatMap(lambda x : [(i, float(x[1][1])/len(x[1][0])) for i in x[1][0]])
    ranks = ranks.reduceByKey(lambda x,y: x+y)
    new_rank = ranks.sortByKey().collect()
    print("New_rank - {0}".format(new_rank))
    if check_conv(old_rank, new_rank):
        print("Page ranking is completed.")
        break
    old_rank = new_rank
```

```
Old Rank - [('1', 0.3333333333333333), ('2', 0.3333333333333333), ('3', 0.3333333333333333)]
New_rank - [('1', 0.3333333333333333), ('2', 0.5), ('3', 0.16666666666666666)]
Old Rank - [('1', 0.3333333333333333), ('2', 0.5), ('3', 0.16666666666666666)]
New_rank - [('1', 0.3333333333333333), ('2', 0.5), ('3', 0.16666666666666666)]
Page ranking is completed.
```