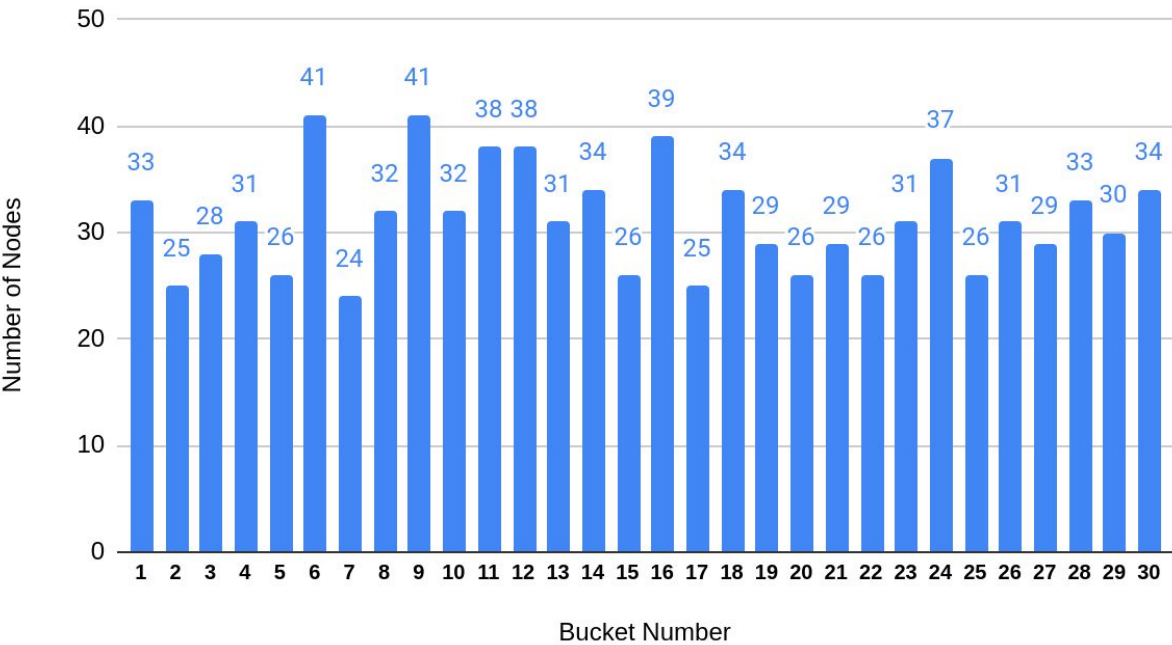Dhanushka S.M.V.
E/16/083

Hash Tables - Report

Initially consider the sample-text1.txt
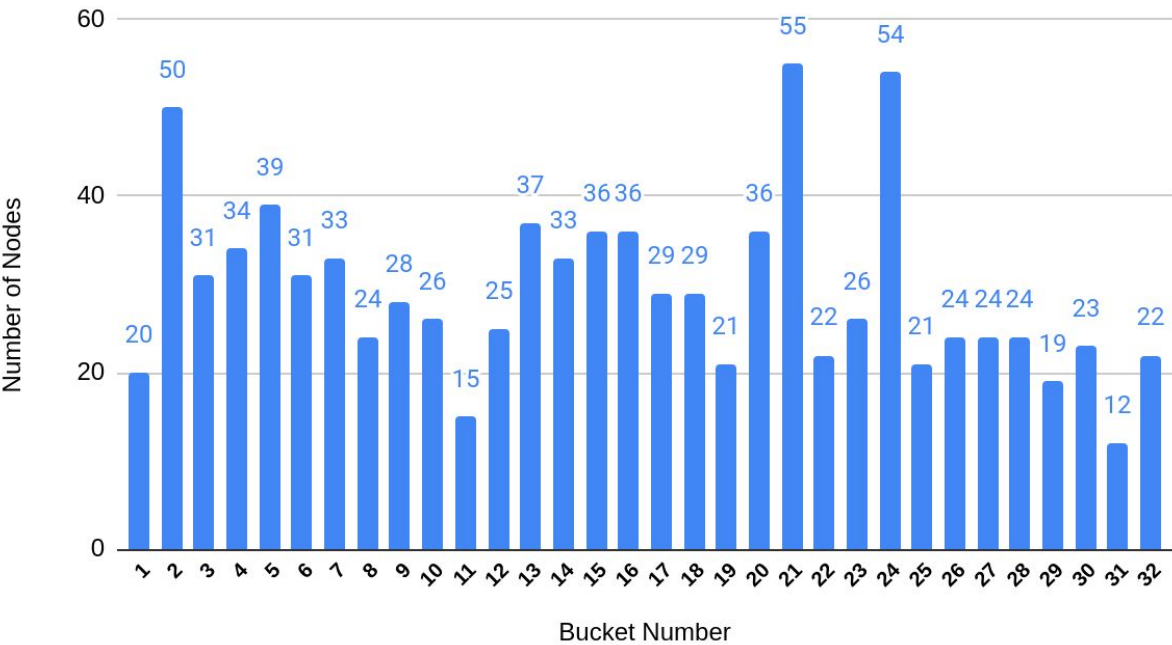
1) Hash function 1

```java
public int hashing(String key) {
    int hash = 0;
    for (int i = key.length() - 1; i >= 0; i--) {
      hash = (hash + (key.charAt(i) * (int) (Math.pow(128, i)))) % table.length;
    }
    return (hash % table.length);
  }
```

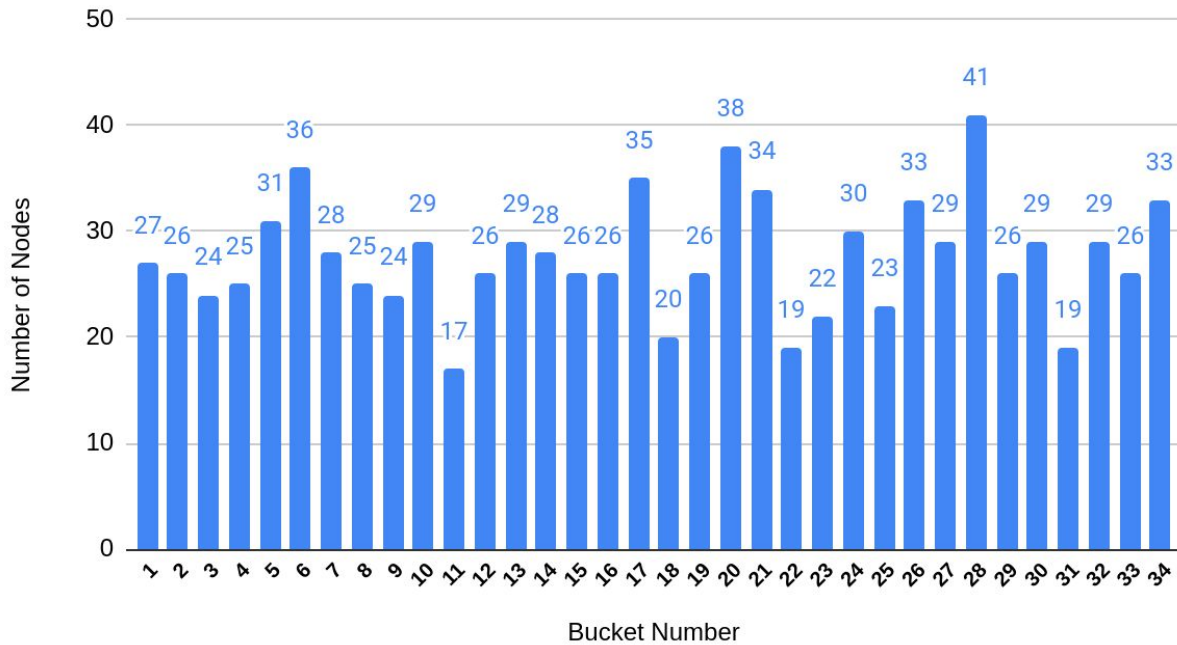# Distribution of nodes when bucket size is 30 and using hash method 1

Maximum: 41
Minimum: 24
Avg: 31.3
Deviation: 4.828049149889632

Distribution of nodes when bucket size is 32 and using hash method 1

Maximum: 55
Minimum: 12
Avg: 29.34375
Deviation: 9.947076753373324

Distribution of nodes when bucket size is 34 and using hash method 1
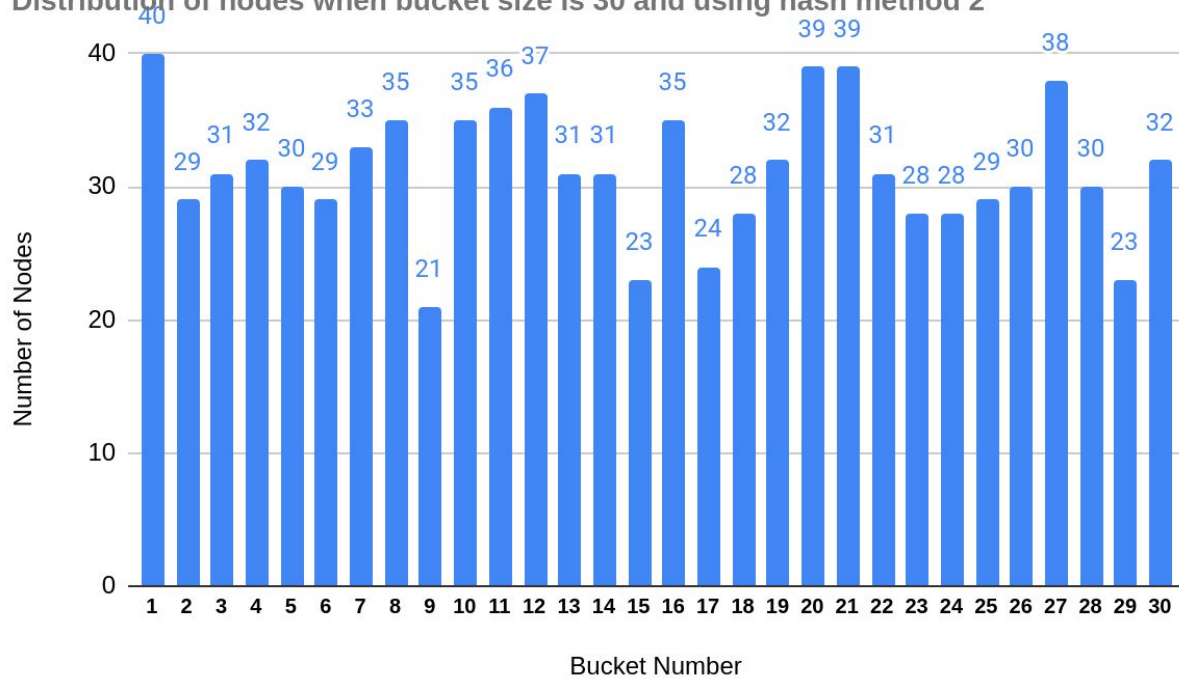

Maximum: 41
Minimum: 17
Avg: 27.617647
Deviation: 5.316536839413769

2) Hash function 2

```java
public int hashing(String key) {
    int hash = 0;
    for (int i = 0; i < key.length(); i++) {
        hash = (31 * hash + key.charAt(i)) % table.length;
    }
    return (hash % table.length);
}
```

Java String function combines successive characters by multiplying the current hash by 31 and then adding on the new character

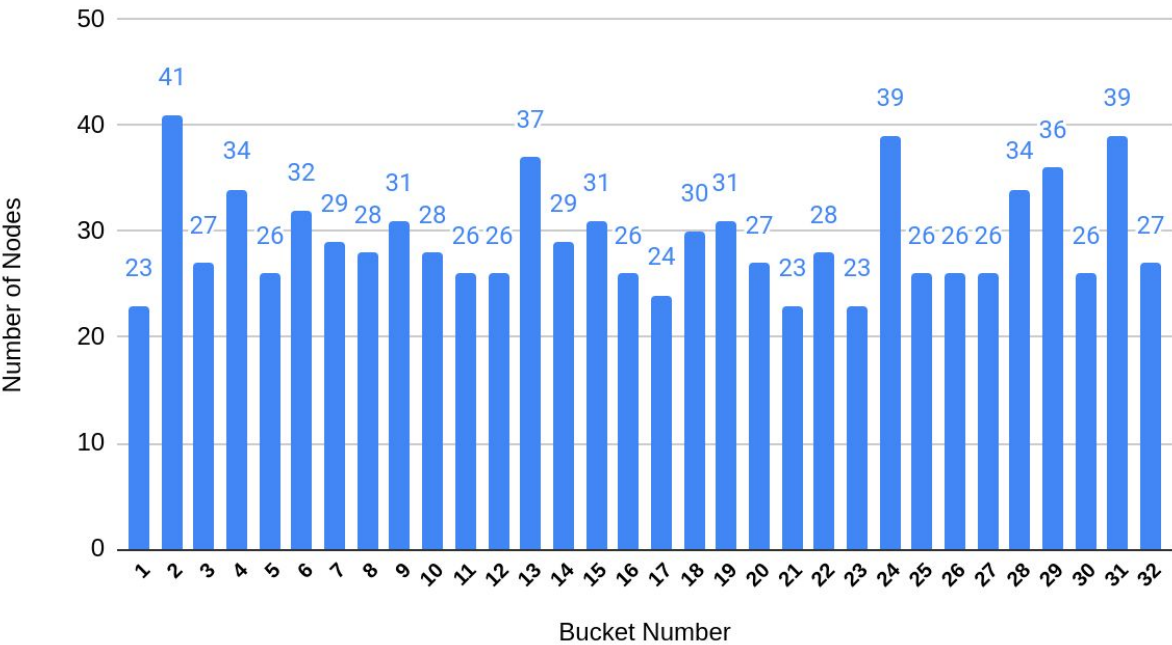**Distribution of nodes when bucket size is 30 and using hash method 2**
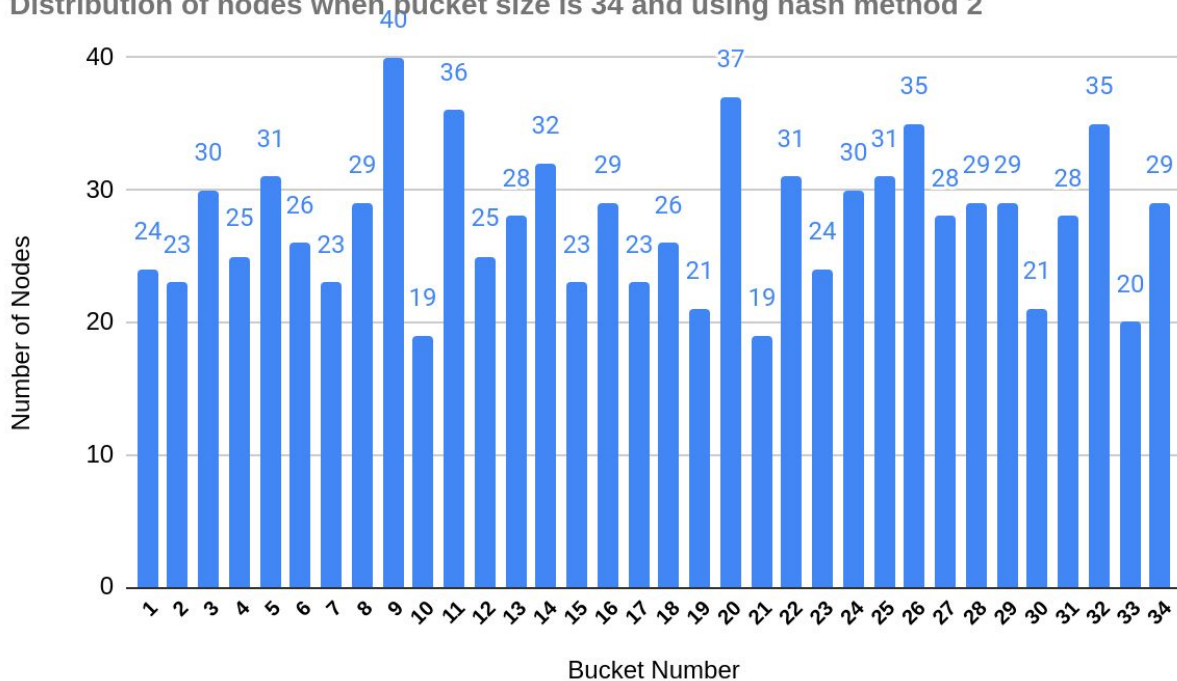

Maximum: 40
Minimum: 21
Avg: 31.3
Deviation: 4.723352474011441

Distribution of nodes when bucket size is 32 and using hash method 2

Maximum: 41
Minimum: 23
Avg: 29.34375
Deviation: 4.790024628068211

Distribution of nodes when bucket size is 34 and using hash method 2

Maximum: 40
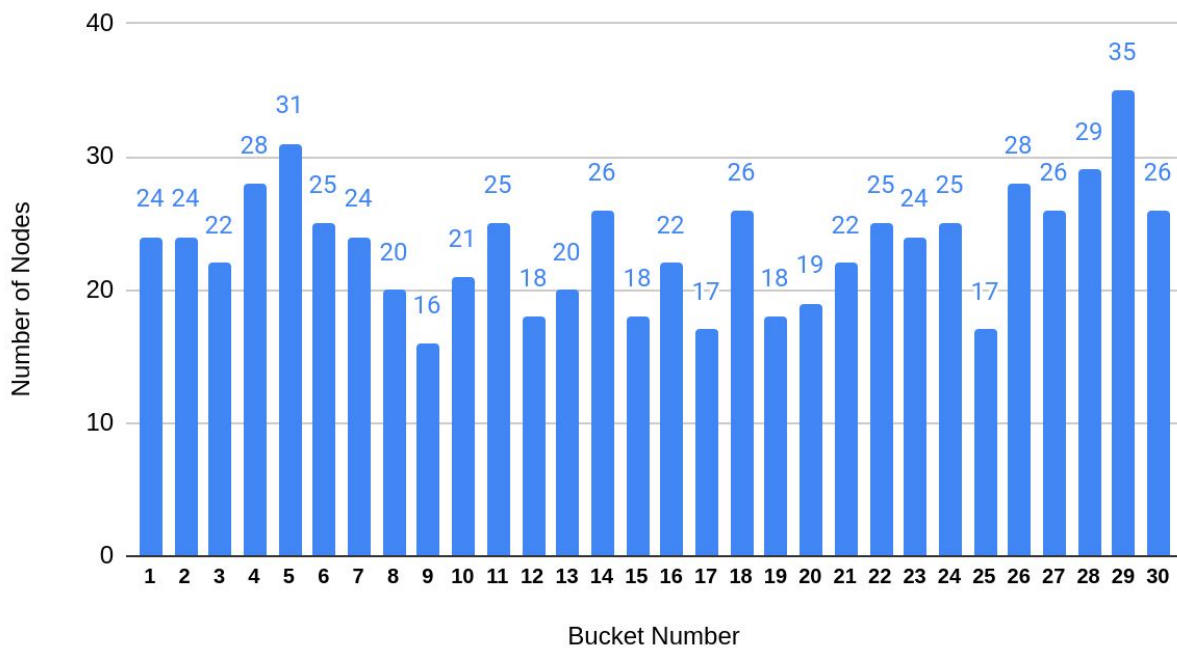Minimum: 19
Avg: 27.617647
Deviation: 5.12499404534715

For both hash functions, for different numbers of buckets it shows different distributions. But all of them have a distribution which is very much close to uniform distribution.

If we compare hash function 1 and hash function 2, it is obvious that for a given number of buckets the deviation is relatively high for hash function 1. **Therefore hash function 2 is the best when compared to hash function 1.**

Consider sample-text2.txt

1) Hash function 1

Distribution of nodes when bucket size is 30 and using hash method 1
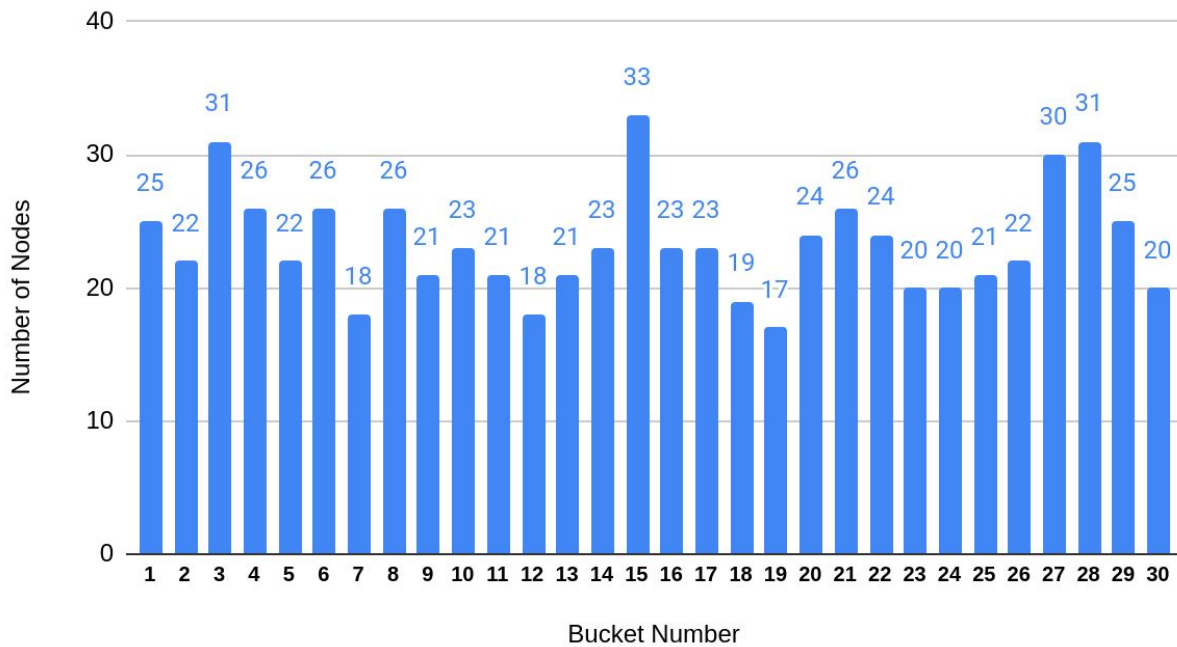

Maximum: 35
Minimum: 16
Avg: 23.366667
Deviation: 4.358772919892421

2) Hash function 2

Distribution of nodes when bucket size is 30 and using hash method 2

Maximum: 33
Minimum: 17
Avg: 23.366667
Deviation: 3.8728415107240703

Since different text files have different distributions of words, even for the same hash function it will give different results. Also if we compare the above 2 graphs with the previously obtained corresponding graphs, it is obvious that the distributions are different.