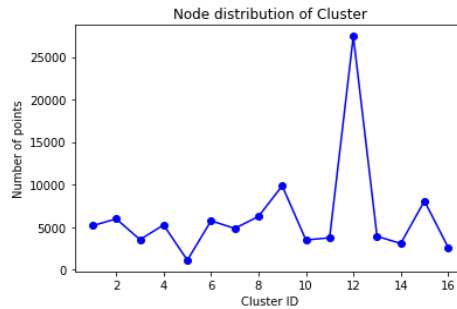


Visualization

Number of nodes in individual cluster/Number of reviews per cluster

```
In [5]: x = []
y = []
for cluster in data["clusters"]:
    x.append(cluster['id'])
    y.append(cluster['size'])

plt.plot(x, y, 'bo-')
plt.xlabel('Cluster ID')
plt.ylabel('Number of points')
plt.title('Node distribution of Cluster')
plt.show()
```



The above diagrams tells clearly that the reviews are mostly clustered in a single cluster.

Form a graph with nodes as cluster id(virtual) and the top 10 frequent words as nodes and edges as their occurrence in cluster nodes and printing other details along with the visualization of the graph.

```
In [10]: print(nx.info(G))
```

```
Name:
Type: Graph
Number of nodes: 40
Number of edges: 160
Average degree: 8.0000
```

```
In [11]: print('Components', nx.number_connected_components(G))
print('Density', nx.density(G))
print('Degree', G.degree())
```

```
Components 1
Density 0.20512820512820512
Degree [(1, 10), ('room', 1), ('burger', 1), ('would', 16), ('get', 16), ('one', 16), ('good', 14), ('like', 16), ('place', 12), ('time', 16), ('us', 6), (2, 10), ('food', 11), ('great', 8), ('service', 6), (3, 10), ('really', 1), (4, 10), ('back', 8), ('said', 2), ('told', 3), (5, 10), (6, 10), ('hair', 1), ('company', 1), ('cut', 1), (7, 10), ('breakfast', 1), (8, 10), (9, 10), (10, 10), ('cheese', 1), (11, 10), (12, 10), (13, 10), (14, 10), (15, 10), (16, 10), ('car', 1), ('call', 1)]
```

```
nx.draw(G, node_size=5)
plt.gcf().set_size_inches(5, 5)
plt.show()
```

