In [1]:
```
#Alan Uthuppan
```

In [2]:
```
import networkx as nx
import matplotlib.pyplot as plt
import statistics
import csv
import collections
```

In [4]:
```
#Part a

g = nx.read_edgelist("com-amazon.ungraph.txt.gz")
ans = max(nx.connected_components(g), key=len)
print("Largest connected component size: " + str(len(ans)))
```

```
Largest connected component size: 334863
```

In [7]:
```
#Part b

print("# of connected components: " + str(len(list(nx.connected_components(g)))
```

```
# of connected components: 1
```

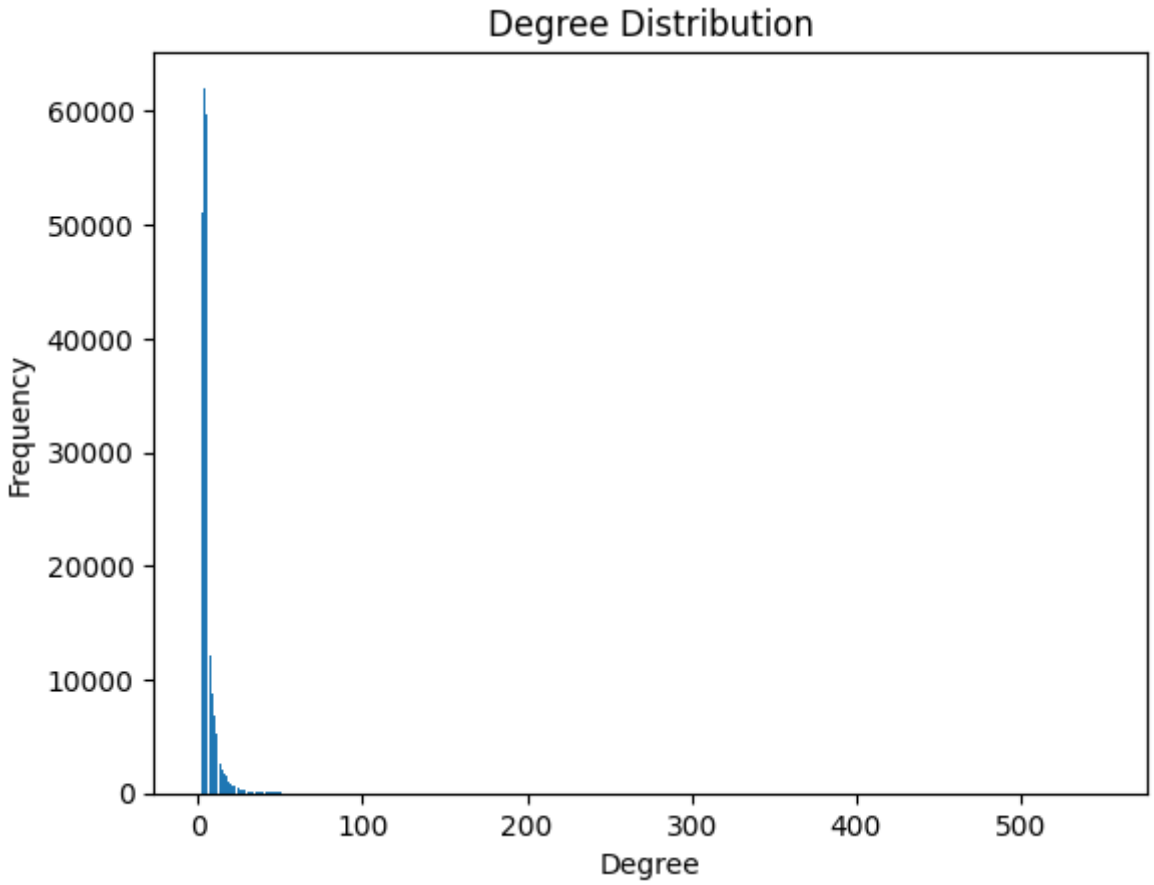
In [13]:
```
#Part c

distribution = collections.Counter(sorted([d for n, d in g.degree()], reverse=T
degree, freq = zip(*distribution.items())

fig, ax = plt.subplots()
plt.bar(degree, freq)

plt.title("Degree Distribution")
plt.ylabel("Frequency")
plt.xlabel("Degree")

plt.show()
```



In [8]:
```
#Part d

#Will calculate average shortest path for 500 nodes

res = []
for i in range(500):
    avg = statistics.mean(nx.shortest_path_length(g, source=(list(g)[i])).value
    res.append(avg)

print("Average shortest path length for sample of 500 nodes: " + str(statistics
```

```
Average shortest path length for sample of 500 nodes: 11.56
```

In [5]:
```
#Part e

print("Average clustering coefficient: " + str(nx.average_clustering(g))[:5])
```

```
Average clustering coefficient: 0.396
```