# Submission

February 2, 2021

[1]:
```
# Craig Fox
print('Name: Craig Fox')
# TUID: 915781095
print('TUID: 915781095')

import networkx as nx
import matplotlib.pyplot as plt
import csv
import collections
import statistics
```

```
Name: Craig Fox
TUID: 915781095
```

[2]:
```
# Problem 1
print('Problem 1')

G = nx.read_edgelist("com-amazon.ungraph.txt")
```

```
Problem 1
```

[3]:
```
# Size of the largest connected component
lcc = max(nx.connected_components(G), key=len)
print('The number of nodes of the largest connected component is ' +␣
 ↪str(len(lcc)))
```

```
The number of nodes of the largest connected component is 334863
```

[4]:
```
# Number of connected components
print('The number of connected components is ' + str(len(list(nx.
 ↪connected_components(G)))))
```

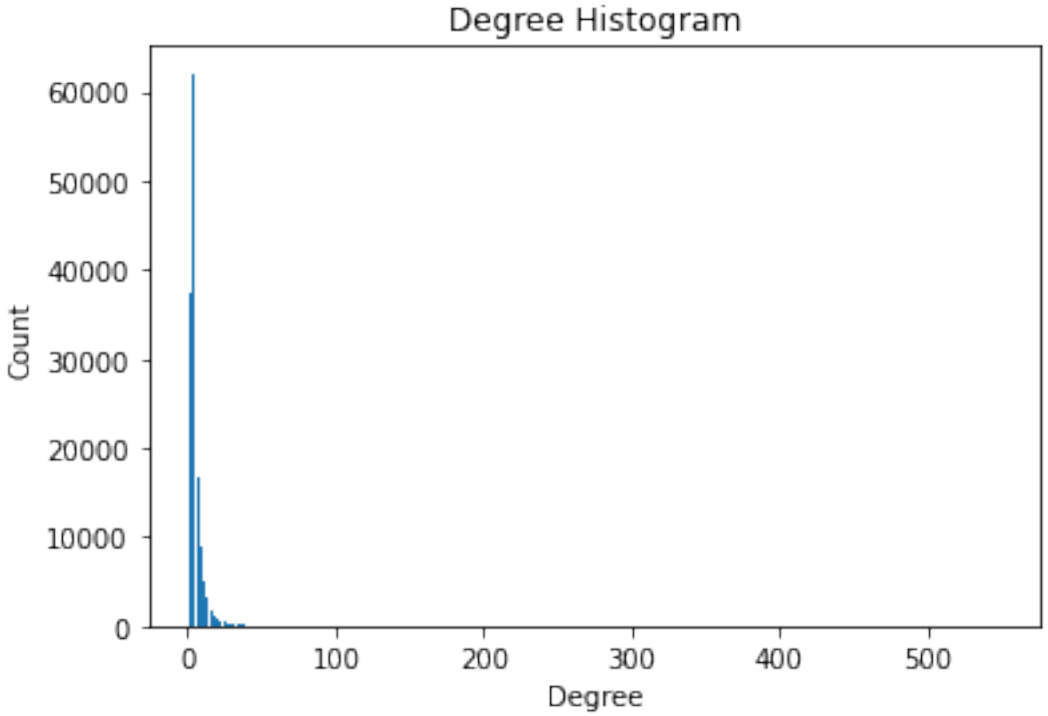
```
The number of connected components is 1
```

[5]:
```
# Degree distribution
degreeCount = collections.Counter(sorted([d for n, d in G.degree()],␣
 ↪reverse=True))
degree, count = zip(*degreeCount.items())
```

```
fig, ax = plt.subplots()
plt.bar(degree, count)

plt.title("Degree Histogram")
plt.ylabel("Count")
plt.xlabel("Degree")

plt.show()
```




[20]:
```
# Path length

# To find the shortest path length for all 334,863 nodes would have taken a␣
 ↪long time, so 1,000 nodes were done to get a estimate
x = []
for r in range(1000):
    m = statistics.mean(nx.shortest_path_length(G, source=(list(G)[r])).
 ↪values())
    x.append(m)
print('The average shortest path length is ' + str(statistics.mean(x))[:5])
```

```
The average shortest path length is 11.60
```

[18]:
```
# Clustering coefficient
print('The average clustering coefficient is ' + str(nx.average_clustering(G))[:
↪5])
```

```
The average clustering coefficient is 0.396
```

[ ]: