Math 168 Essay 3

Pratyusha Majumder

October 2020

1 Facebook dataset

The first network I chose was the Facebook dataset, which was collected as part of a study done by Mcauley and Leskovec [1]. I obtained this from the Stanford Network Analysis Project database of datasets. The Facebook dataset has 4039 nodes, 88234 edges, and an average clustering coefficient of 0.6055. It represents Facebook 'friends lists', based on the data collected from a survey participants completed on the Facebook application [1].

First, I created a normal visualization of my network which is shown in Figure 1. It is evident that this representation of the network is not very clear, which then caused me to experiment with with the "spring layout visualization" , as shown in Figure 2. Experimenting with color and shape of the nodes resulted in Figures 3 and 4. The blue nodes in Figure 3 represent the nodes with degree greater than 20 and the yellow nodes represent those with degrees less than or equal to 20.

Figure 5 shows the degree distribution of this network, which is extremely right-skewed but is approximately a Poisson distribution.

2 Butterfly dataset

The Butterfly dataset depicts a similarity network between butterflies, collected as part of Wang, Markert and Everingham's study [2]. The nodes represent the butterflies and the edges represent the visual similarities between them, which are calculated using different species' images. The network has 832 nodes, 86528 edges, and an average clsutering coefficient of 0.595439 [2]. Figure 6 represents the regular visualization, which shows eight clusters connected to a central cluster, and one cluster that is not connected to the central cluster. Using the "spring layout", unlike in the previous case, does not make the visualization much clearer than using the regular visualization, as seen in Figure 7.

I repeated the visualization processes used with the Facebook data set for the Butterfly data set. From Figure 8, one can see that a majority of the nodes have degree greater than 10, in all the groups of nodes. After investigating and decreasing the number of nodes for each visualization, I found that a large percentage of the nodes had a degree equal to 2, showing that no matter which cluster the nodes were in, a majority of them had degree 2. This can be seen from Figure 9 and Figure 10, the degree distribution with the modal value of d=1.

3 Python code

```
#Butterfly- Importing dataset

file = open("butterfly.tsv", "r")
lines = file.read().splitlines()[4:]

sources_b = [int(L.split('\t')[0]) for L in lines]
targets_b = [int(L.split('\t')[1]) for L in lines]
```

```
#Butterfly - Visualizing data
# Make the network
G_b = nx.Graph()
E = len(sources_b)     # Number of edges
for i in range(E):
    G_b.add_edge(sources_b[i], targets_b[i])

# Draw the network.
plt.figure(3,figsize=(12,12))
nx.draw(G_b, with_labels = True)
plt.show()

pos = nx.spring_layout(G_b)   # Uses Fruchterman-Reingold algorithm to find "optimal" node
plt.figure(3,figsize=(12,12))
nx.draw(G_b, pos, with_labels = True)
plt.show()

my_node_colors = ['blue' if G_b.degree[u] > 10 else 'green' for u in G_b]
plt.figure(3,figsize=(12,12))
nx.draw(G_b, pos, node_size = 15, node_color = my_node_colors)
plt.show()

my_node_colors = ['blue' if G_b.degree[u] == 1 else 'green' for u in G_b]
plt.figure(3,figsize=(12,12))
nx.draw(G_b, pos, node_size = 10, node_color = my_node_colors, node_shape = 's')
plt.show()

 # This code snippet was taken from the NetworkX documentation on degree histograms [3]

degree_sequence = sorted([d for n, d in G_b.degree()], reverse=True)
degreeCount = collections.Counter(degree_sequence)
deg, cnt = zip(*degreeCount.items())
fig, ax = plt.subplots()

plt.figure(3,figsize=(12,12))
plt.bar(deg, cnt, width=0.9, color="b")
plt.title("Degree Distribution")
plt.ylabel("Count")
plt.xlabel("Degree")
```

4 References

[1] McAuley, J., and Leskovec, J., Learning to Discover Social Circles in Ego Networks, in *NIPS*, (2012).

[2] Wang, J., Markert, K., and Everingham, M., Learning Models for Object Recognition from Natural Language Descriptions, in *Proceedings of British Machine Vision Conference*, (2009).

[3] Hagberg, A. A., Schult, D. A., and Swart, P. J., Exploring network structure, dynamics, and function using NetworkX, in *Proceedings of the 7th Python in Science Conference (SciPy2008)*, Pasadena (2008).

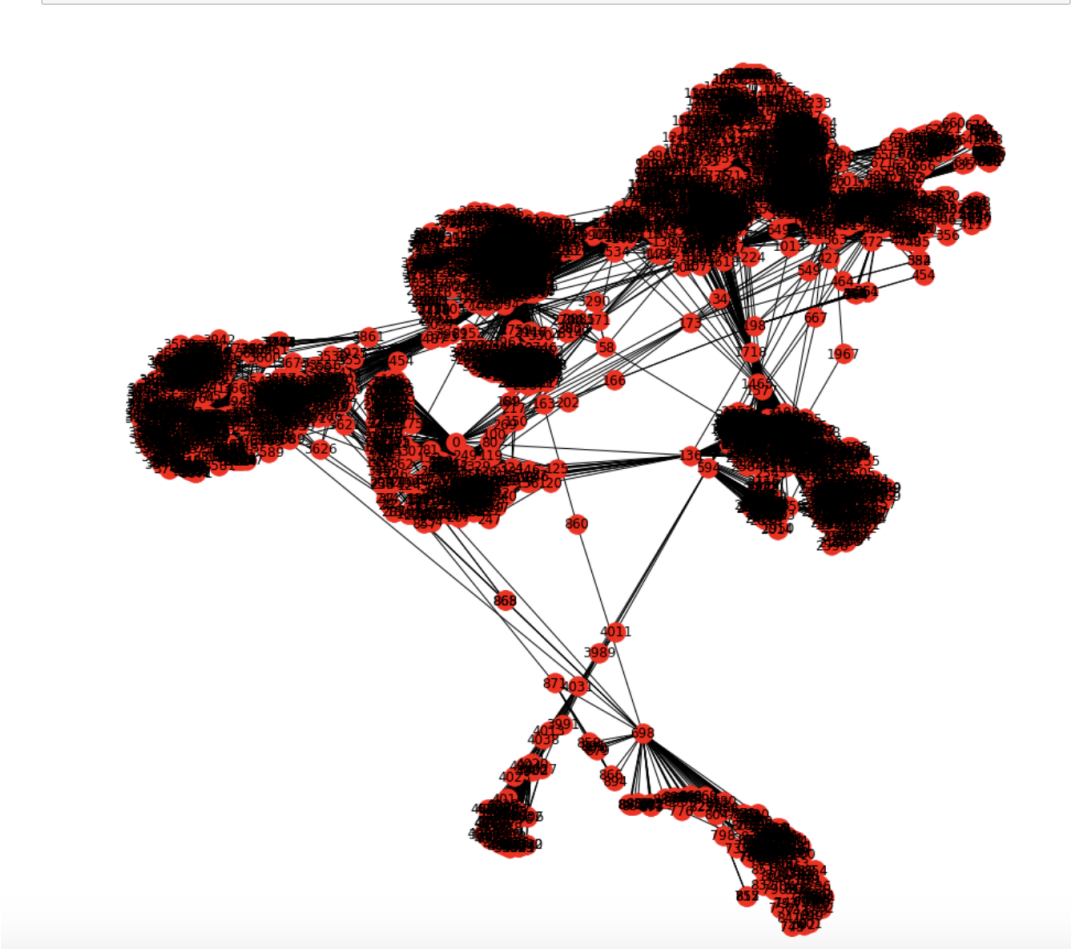

Figure 1: Facebook social network representation.

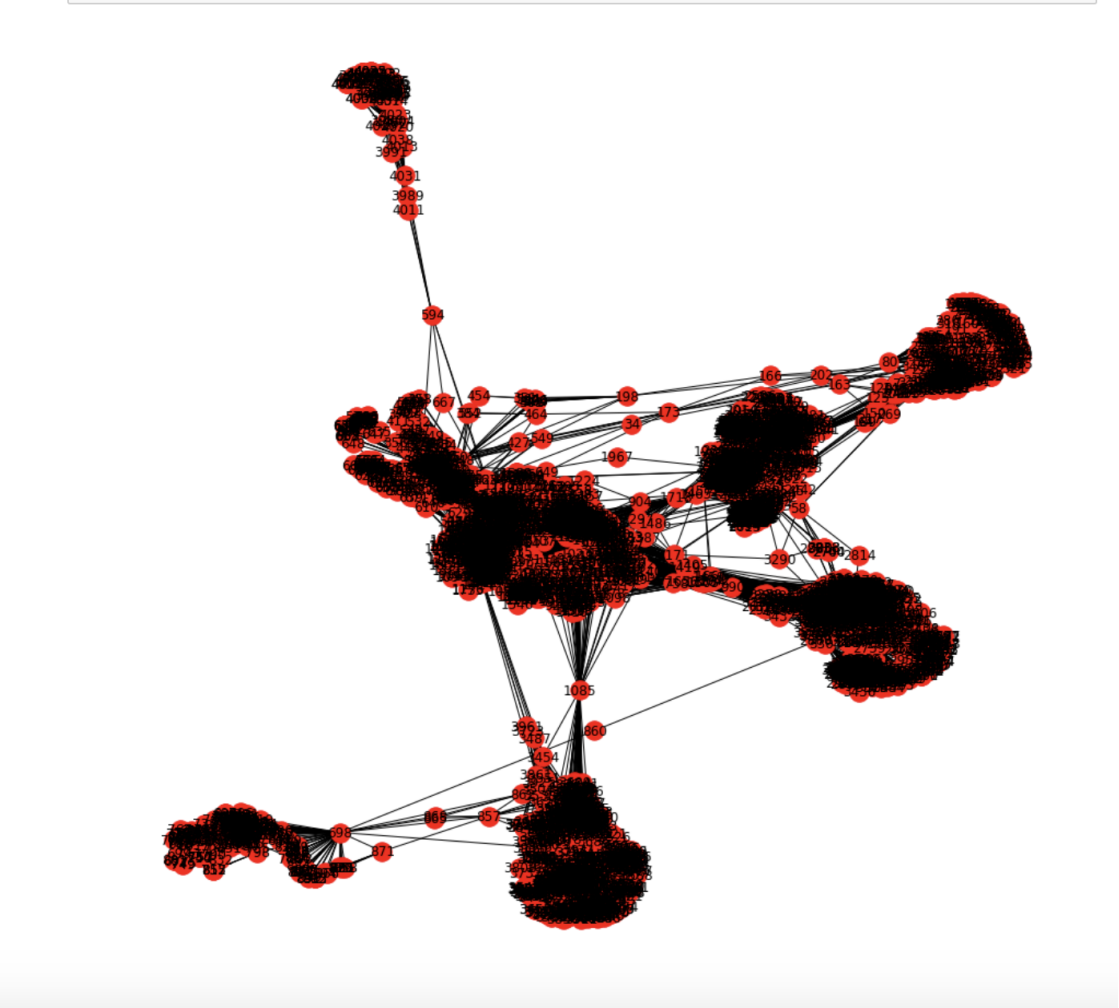

Figure 2: Facebook social network representation using the "spring layout".

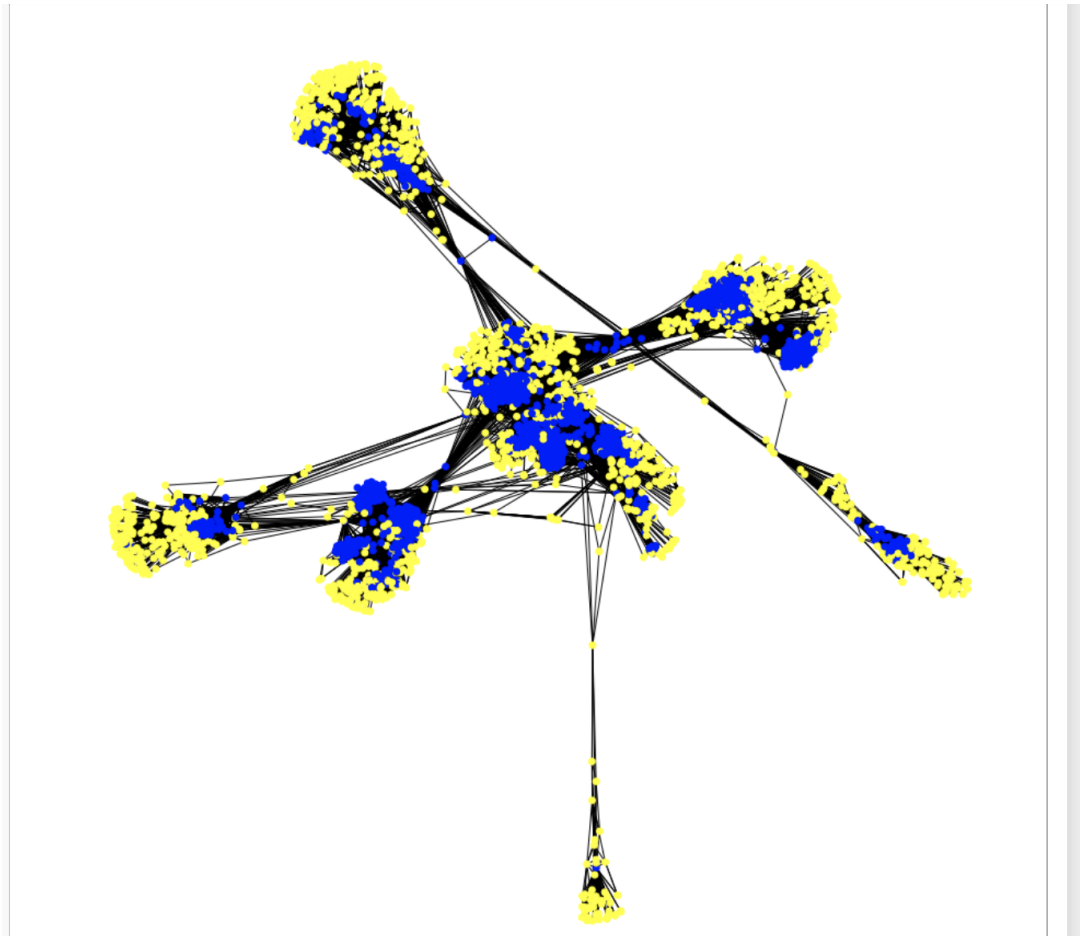

Figure 3: Facebook social network representation: nodes with degrees greater than 20 in blue , nodes with degrees less than or equal to 20 in yellow.

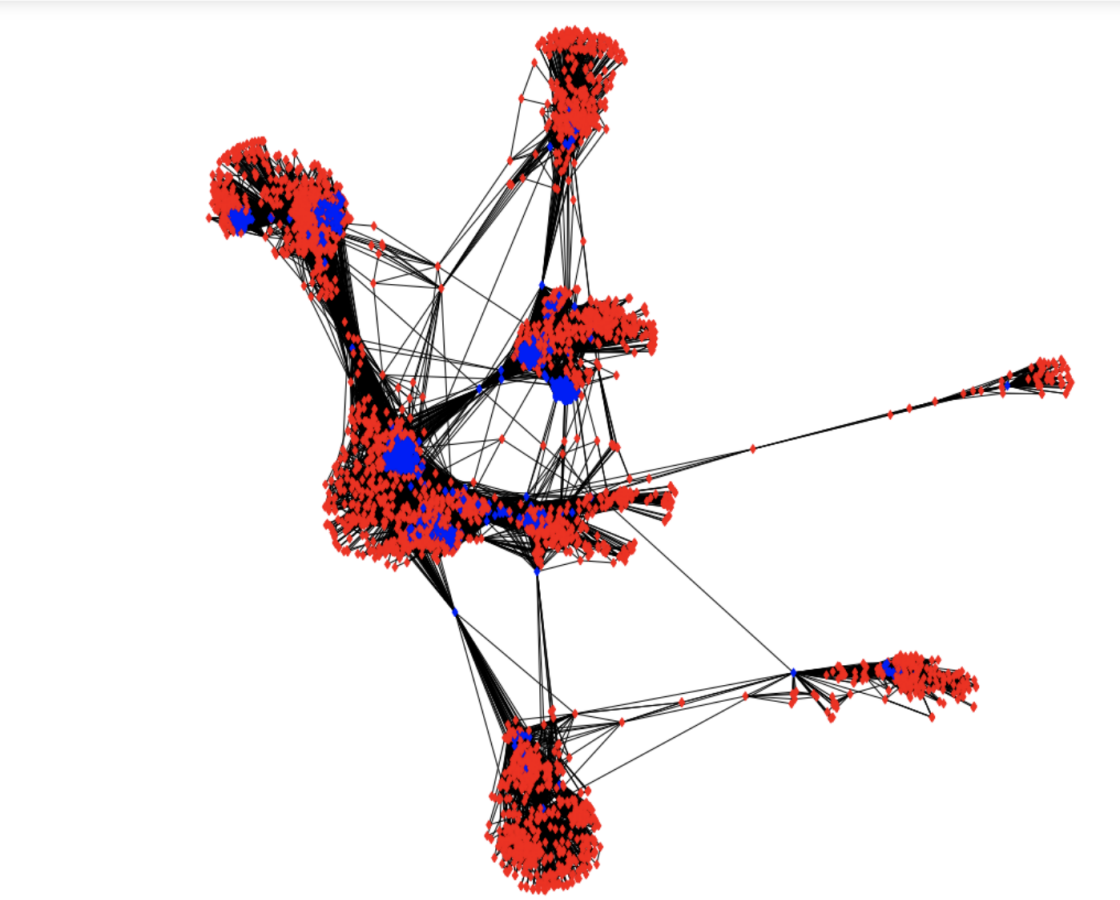

Figure 4: Facebook social network representation: nodes with degrees greater than 50 in blue, nodes with degrees less than or equal to 50 in red, and all diamond-shaped nodes.

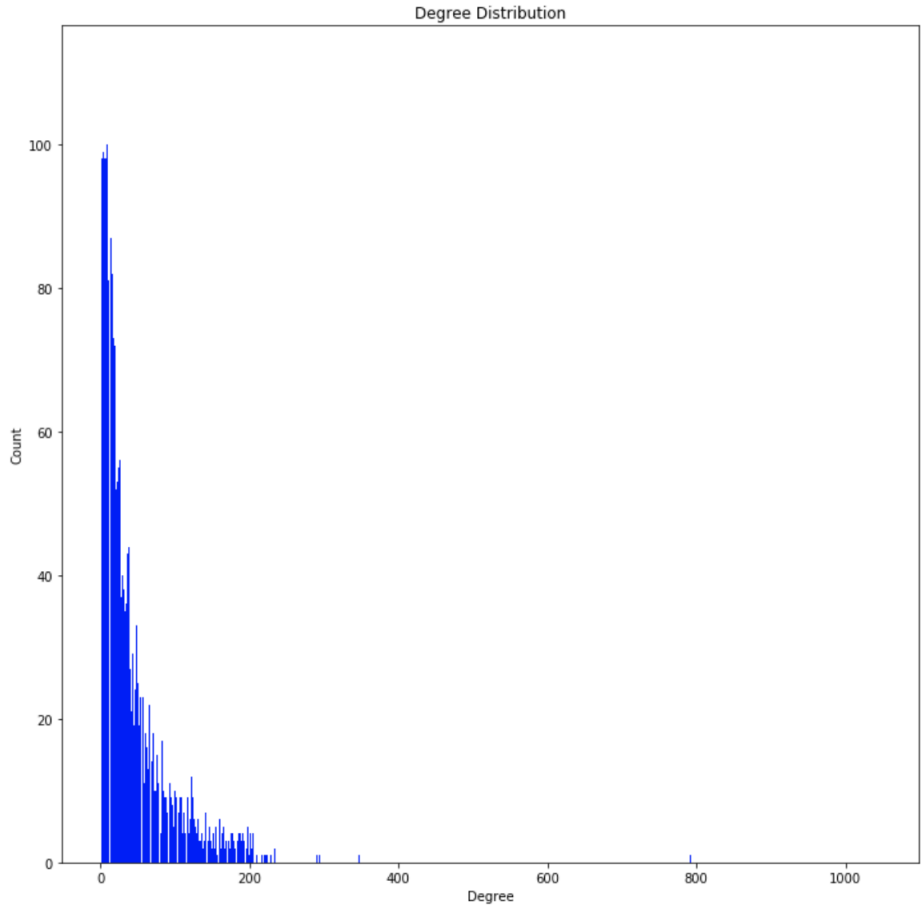

Figure 5: Facebook social network degree distribution.

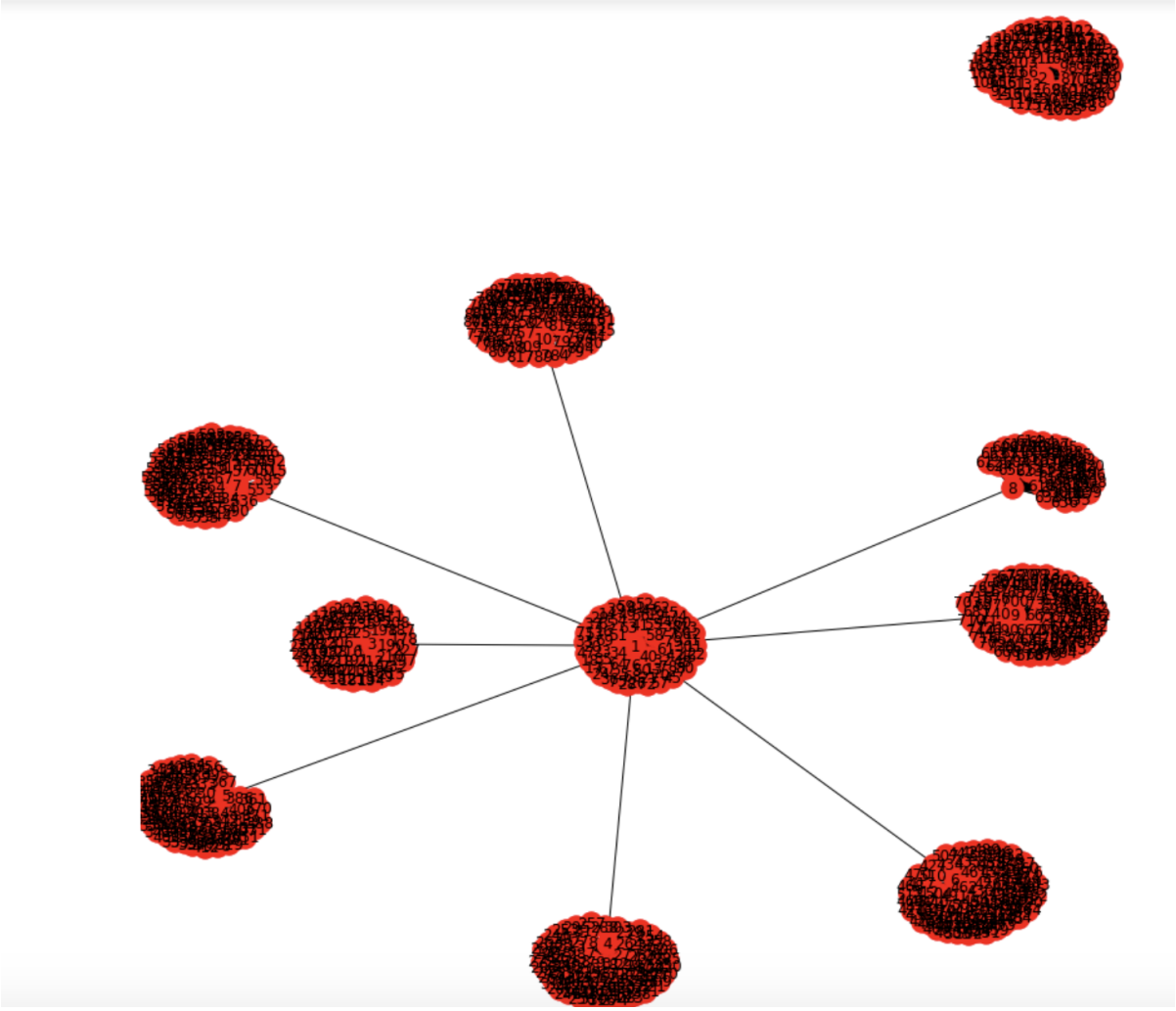

Figure 6: Butterfly similarity network.

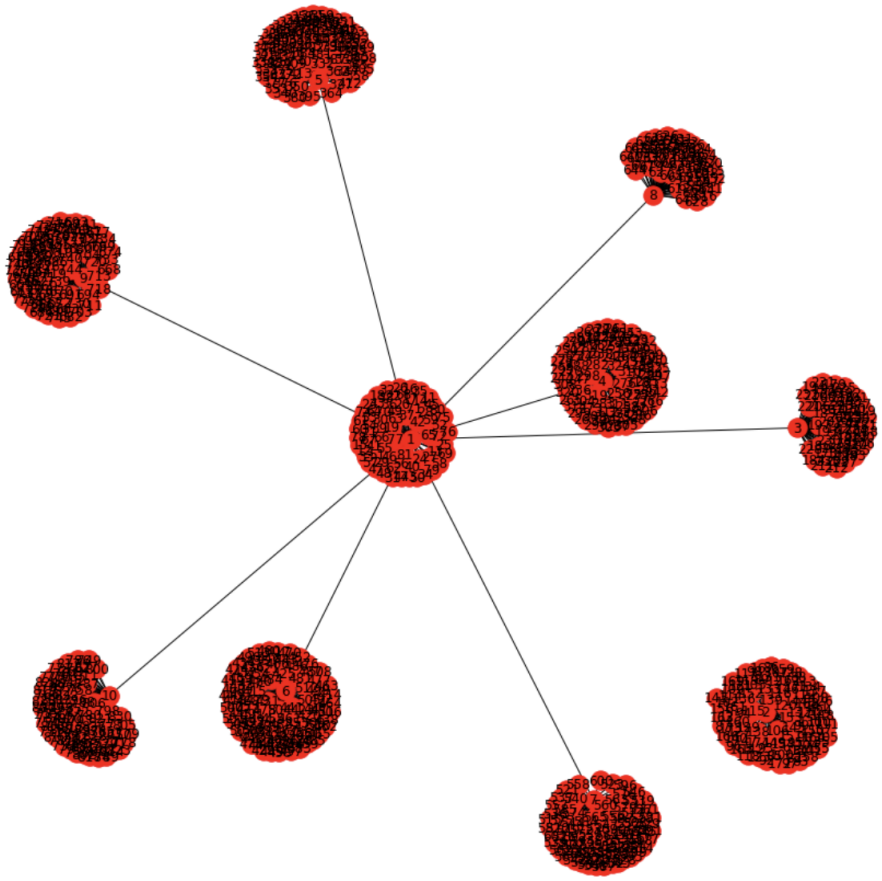

Figure 7: Butterfly similarity network with "spring layout".

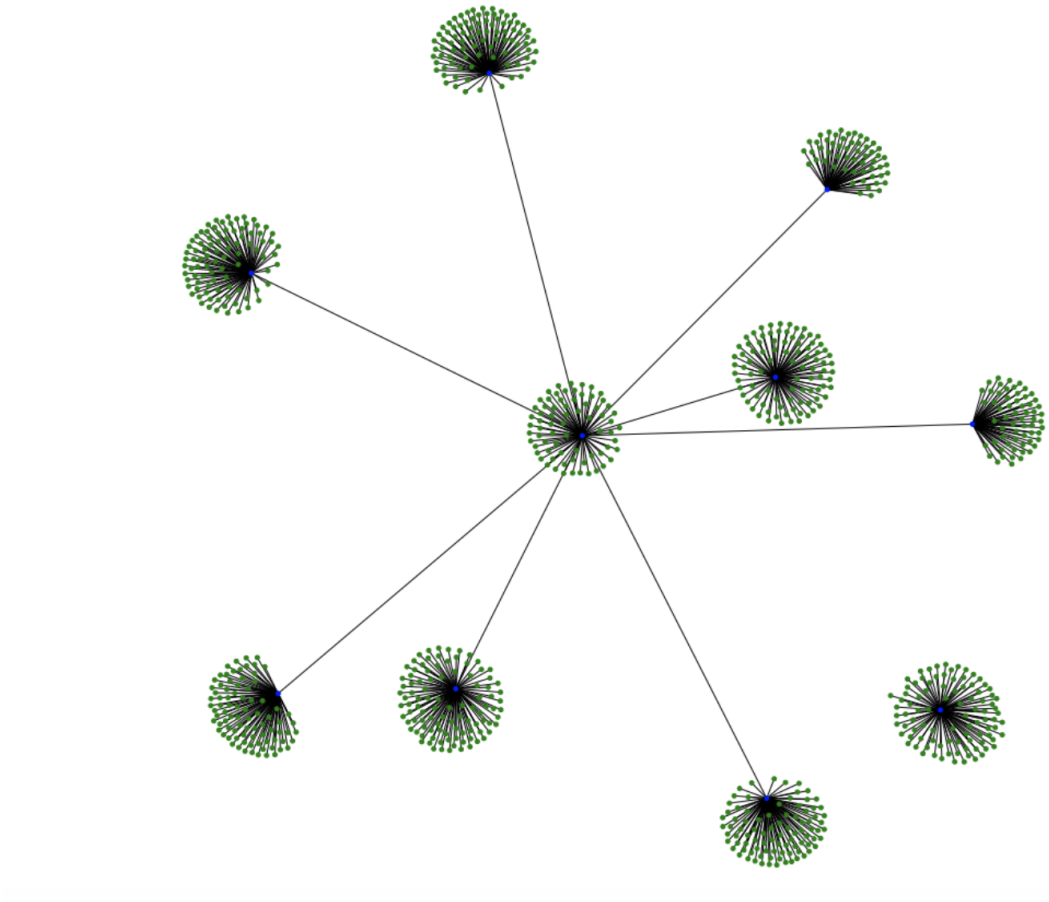

Figure 8: Butterfly similarity network: nodes with degrees greater than 10 in blue, nodes with degrees less than or equal to 10 in green.

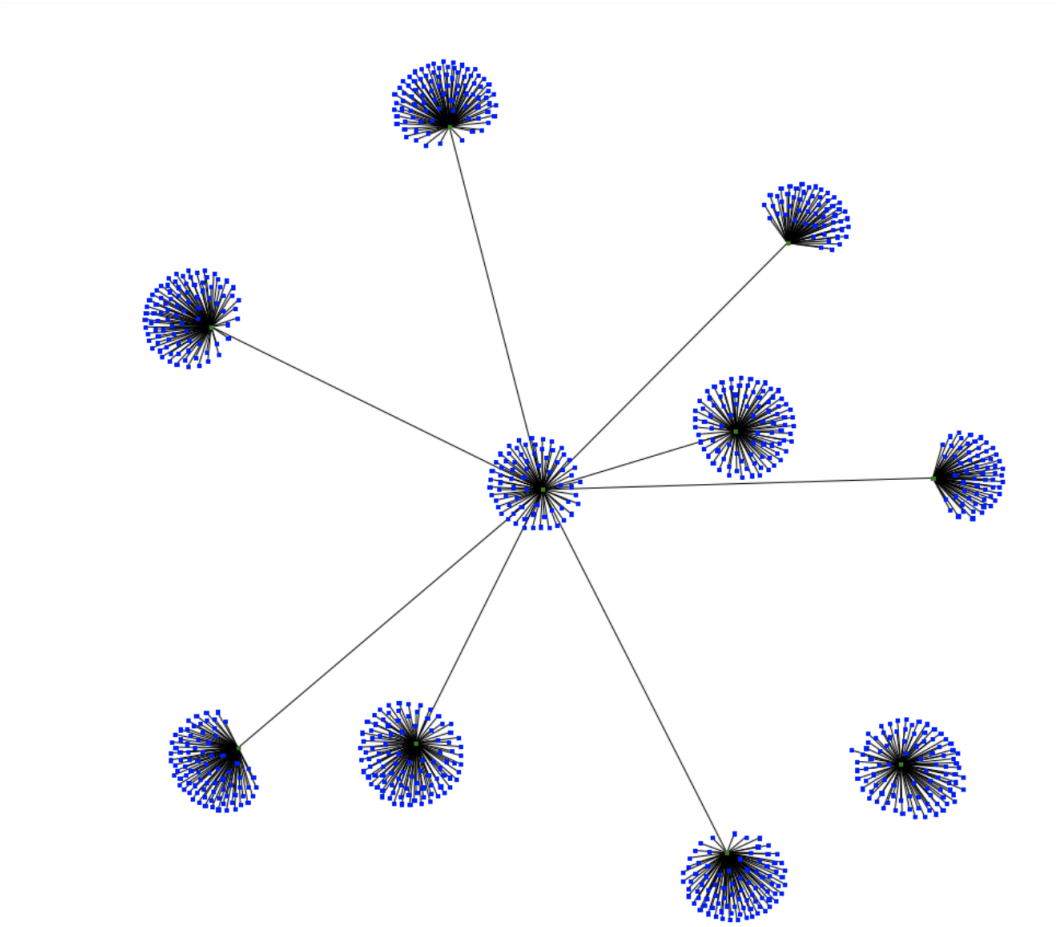

Figure 9: Butterfly similarity network: nodes with degrees equal to 1 in blue, nodes with degrees greater than or less than 1 in green, and all square nodes.

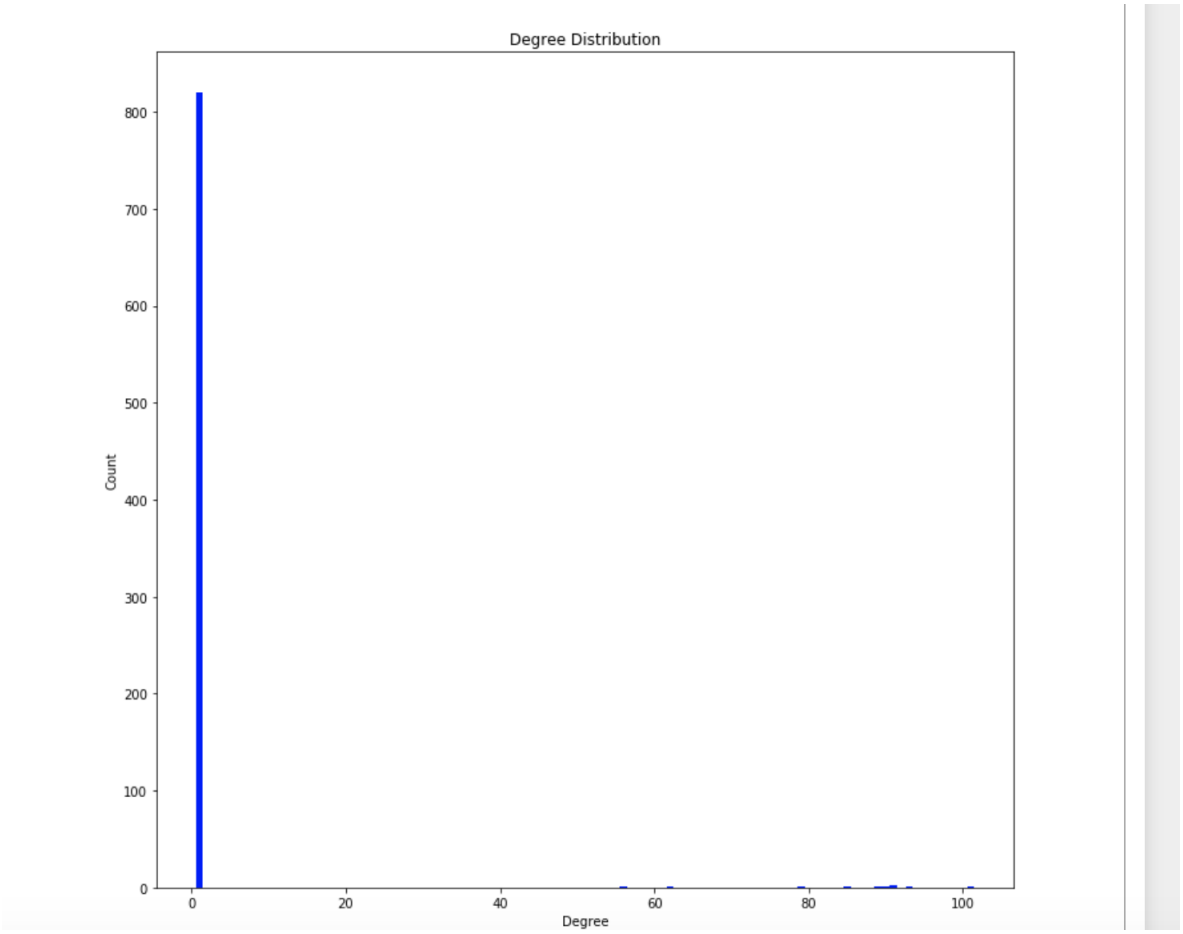

Figure 10: Butterfly similarity network degree distribution.