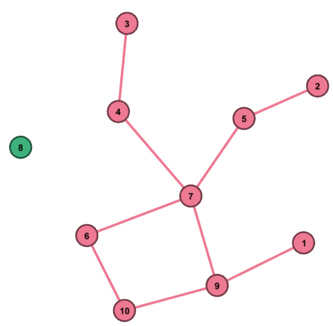
Homework 1

Jackson Paull

September 2023

Problem 1

Graph 1 (p=0.166)



Graph 1 has **one large connected component and one isolated node**. The connected component is centered around node 7, with several branches. Interestingly no triangles formed, meaning that the clustering coefficient is 0.00 for all nodes. That said, one quadrilateral did form. It has **no cycles**, an **average degree of 3.65**.

Unreachable Pairs of Nodes:

- (8, 1)
- (8, 2)
- (8, 3)
- (8, 4)
- (8, 5)
- (8, 6)
- (8, 7)
- (8, 9)
- (8, 10)

Network Diameter	4
Clustering Coefficient	0
Average Distance	2.33

**Work for these statistics on next page*

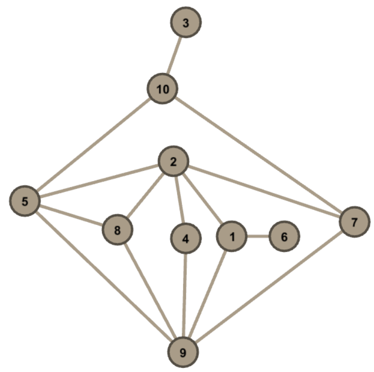
Path Length Between Nodes:	2	3	4	5	6	7	9	10
1	4	4	3	3	3	2	1	2
2		4	3	1	3	2	3	4
3			1	3	3	2	3	4
4				2	2	1	2	3
5					2	1	2	3
6						1	2	1
7							1	2
9								1

$$\text{Diameter} = \max_{ij} \{ l_{ij} \} = 4$$

$$\text{Avg Path length} = \frac{2}{N(N-1)} \sum_{ij} l_{ij} = \frac{2}{9(8)} \cdot 84 = 2.33$$

Clustering Coeff $= 0$ (no triangles)

Graph 2 (p=0.33)



Graph 2 features only a single component, thus there are no unreachable nodes. Also, there is an emergence of triangles, so the clustering coeff is actually defined

Network Diameter	5
Clustering Coefficient	0.12
Average Distance	2.067

**Work for these statistics on next page*

Path Length Between Nodes:	2	3	4	5	6	7	8	9	10
1	1	4	2	2	1	2	2	1	3
2		3	1	1	2	1	1	2	2
3			4	2	5	2	3	3	1
4				2	3	2	2	1	3
5					3	2	1	1	1
6						3	3	2	4
7							2	1	1
8								1	2
9									2

$$\text{Diameter} = \max_{ij}\{l_{ij}\} = 5$$

$$\text{Avg Path Length} = \frac{2}{10\cdot 9}\cdot[1\,66] = \quad 6$$

Clustering Coeff: 2 triangles 2-5-8, 5-8-9 ⇒ Nodes 2, 5, 8, 9

$$2:\ \frac{2\cdot 1}{5\cdot 4} = 0.1$$

$$5:\ \frac{2\cdot 2}{4\cdot 3} = 0.33 \qquad \Rightarrow CC = \frac{1}{10}[1.2] = 0.12$$

$$8:\ \frac{2\cdot 2}{3\cdot 2} = 0.67$$

$$9:\ \frac{2\cdot 1}{5\cdot 4} = 0.1$$

c)

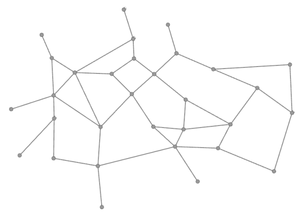


Figure 1: Asphalt

Average Degree	3.907
Average Path Length	3.613
Network Diameter	8
Clustering Coefficient	0.12

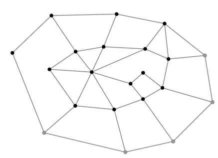


Figure 2: Wood

Average Degree	3.524
Average Path Length	2.567
Network Diameter	5
Clustering Coefficient	0.162

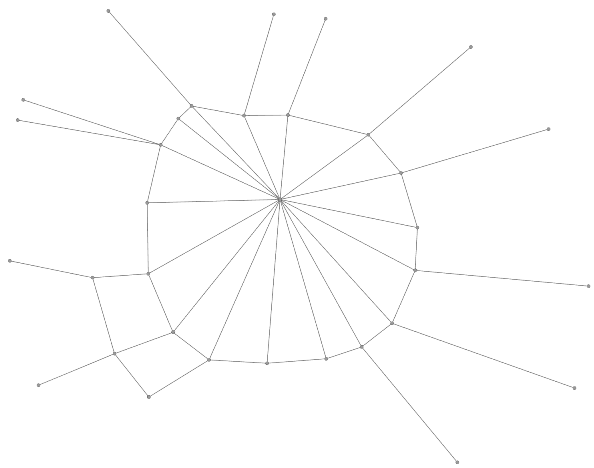


Figure 3: Glass

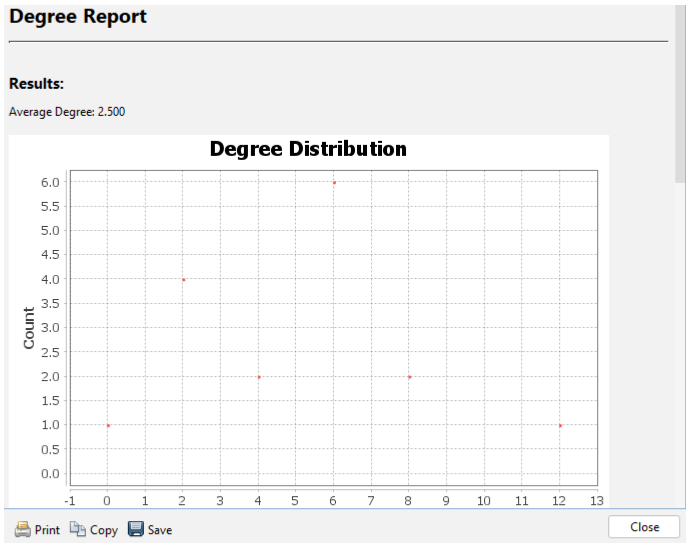
Average Degree	3.273
Average Path Length	2.752
Network Diameter	5
Clustering Coefficient	0.335

Asphalt's high average degree implies that the breakage is distributed more uniformly. In addition, since there is no "central" node, the average path length and especially diameter is particularly high. This might be caused by the fact that asphalt cracks form from expansion and contraction resulting from ambient changes in temperature; since the changes in temperature affect all parts of the asphalt equally, the cracks form uniformly leading to these results.

Wood develops continually around its core, creating a shape with a "center" and several "rings" around it, and occasional hops connecting them. This structure leads to a low diameter and average path length because any node can, at worst, be connected though the center to the target node. Additionally, the center node is a part of a lot of triangles, leading to a higher clustering coefficient.

Glass has a structure similar to wood, explaining its similarities. They both stem from a central node, but where wood is created from continually adding on to the center, glass propagates from a single shattering point. In practice, most of the analysis from wood applies to glass as well. The central node being a part of a lot of triangles leads to a high clustering coefficient. The central node leads to lower average path length and network diameter. In contrast to wood, glass propagates further, leading to a slightly higher average degree.

Problem 2

a) 

b) The average distance between the largest strongly connected component is 2.5. The network diameter is 5, so the average path length is roughly half that. The average degree is 2.5, so a small world network would expect to see $\frac{ln(n)}{ln(k)} = \frac{ln(16)}{ln(2.5)} = 3.0$ So we can observe some small world effects, but the degree distribution is too spread out to support this.

c) The most important nodes by betweenness and degree are Medici, Guadagni, Albizzi, Salviati, and Ridolfi. Typically, a higher degree implies higher betweenness. However, Salviati has a betweenness of 26.0 and degree 4, while Ridolfi has betweenness 20.6 and degree 6. In this case, despite Salviati's lower degree, it holds a more crucial position in the network as the only node connected to Pazzi hence all shortests path to Pazzi must pass through Salviati. This is not the case for Ridolfi. It has a high degree, yet it can be circumvented hence it will lie on fewer shorter paths and have a lower betweenness.

Problem 3

a)

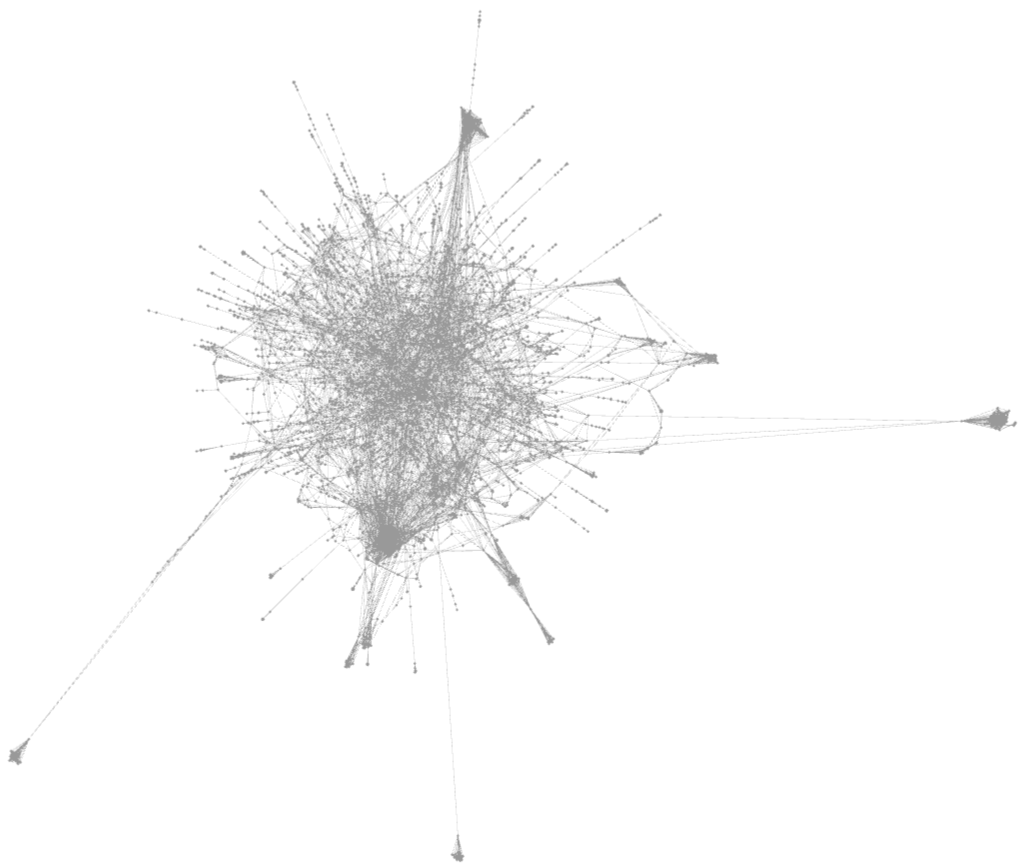


Figure 4: arXiv_lcc.gml

Average Degree	6.459
Average Path Length	6.04
Network Diameter	17
Clustering Coefficient	0.665

b)

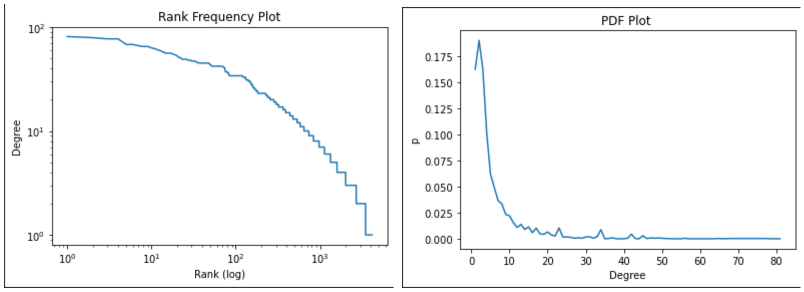
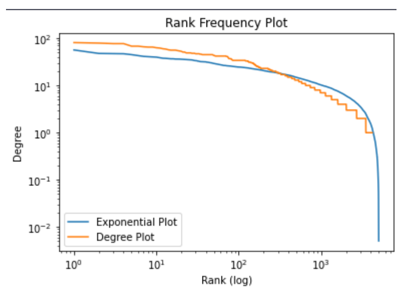


Figure 5: arXiv_lcc.gml Rank Frequency and PDF plots

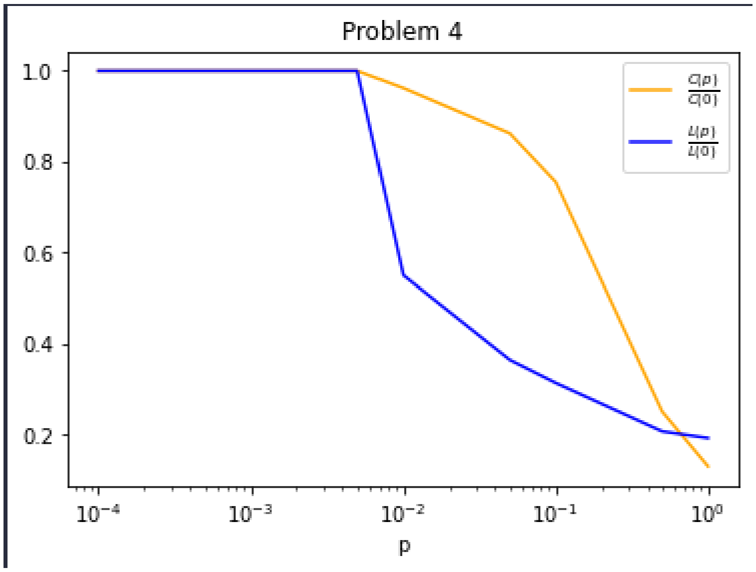
c)



We can see that the exponential numbers follow fairly closely, but at low ranks there is a higher degree and high ranks have lower degree. Additionally, low ranks drop almost linearly, but never reach 0 since degree is discrete. In contrast, the exponential numbers are able to take on decimal values and therefore reach values much closer to 0.

This network seems to be best classified as a power law as we can observe a roughly linear relationship between rank and degree in log log scale. However the network seems to have an inflated number of middling degree nodes, whereas a more typical power law network would have a sharper drop off in degree (and therefore a stronger linear relationship)

Problem 4



Here we can observe the small world effect in action, where the path length decreases quickly, while clustering coefficient remains high for probabilities around 0.1. In this regime, the path length is low while the clustering coefficient is high, meaning that the graph is relatively easily traversable, while still being as densely connected as possible.

Problem 5

a)

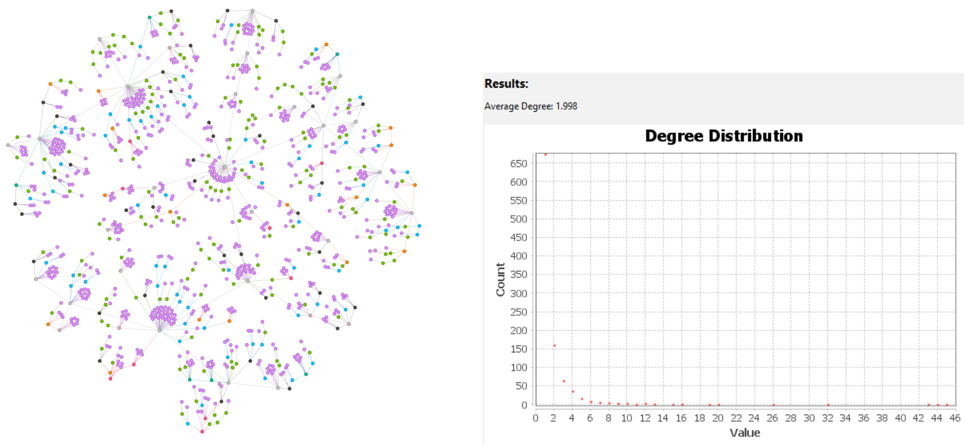


Figure 6: Scale Free Network and Degree Distribution

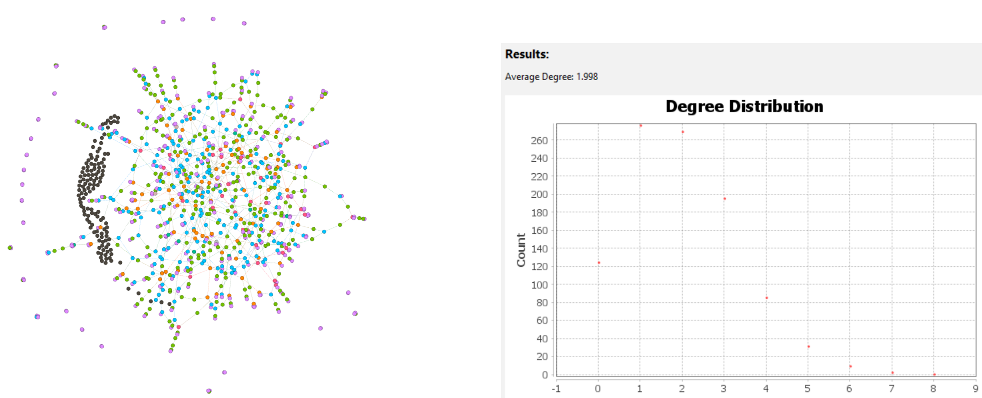
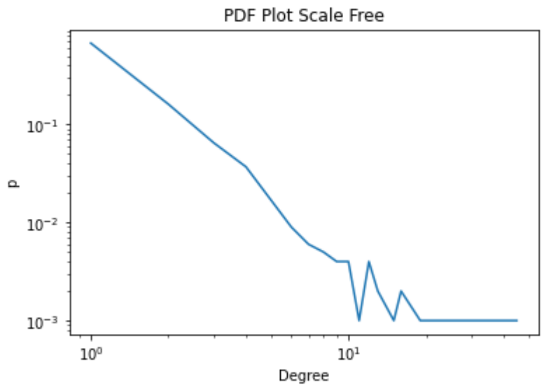
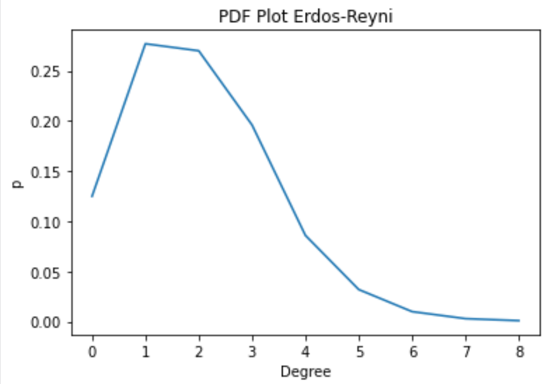


Figure 7: Erdos-Reyni Network and Degree Distribution

b)





c)

Network	Clustering Coeff
Erdos Reyni	0.00136
Scale Free	0.0000

d)

Network	Clustering Coeff
Erdos Reyni	0.00139
Scale Free	0.0000

The scale free network had a CC of 0 to begin with, so removing any nodes or edges could not increase this beyond 0. Interestingly enough, removing the top 2% of nodes increased the CC of the Erdos-Reyni network. Despite having high degree, the removed nodes were not a part of many triangles, meaning that their removal increased the ratio of triangles to possible triangles in the network, leading to an increase in normalized CC.