`Project Summary`

The summary of tested networks and their respective properties are given below,

Dataset 1:

- **File name** : *el-Gnutella08_undirected.txt*
- **Description** : Gnutella peer-to-peer file sharing network from August 2002.
- **Nodes** : hosts in the Gnutella network.
- **Edges** : represent connections between the Gnutella hosts at that instant.
- **Type** : Undirected

Dataset 2:

- **File name** : *el-email-Eu-directed.txt*
- **Description** : This is a private email network of a large European research institution
- **Nodes** : members of the research institution
- **Edges** : There is an edge (u, v) in the network if person u sent person v at least one email
- **Type** : Directed

Dataset 3:

- **File name** : *el-facebook_undirected.txt*
- **Description** : This dataset consists of 'circles' (or 'friends lists') from Facebook
- **Nodes** : Users
- **Edges** : represents friendship between two users
- **Type** : Undirected

	Gnutella	email-EU-core	Facebook
Number of connected Components	2	20	1
Maximum Degree	97	347	1045
Average Degree	6.59483	33.24577	43.69101
Average clustering coefficient	0.01087	0.39935	0.60555
Degree distribution	*Please look into the output folder*		
Exponent	1.25934	1.09363	1.05000
Average path length	4.64299	2.58693	3.69251
Diameter	9	7	8

Out of 3 networks in sample one is directed and other two are undirected. From their properties the following points can be concluded,

1. The higher value of **maximum degree** and **average degree** shows that connectivity in *facebook* network is far more than that of *gnutella* or *email-eu*.

2. The clustering coefficient is the measure of the extent to which one's friends are also friends of each other. All of out test cases has a high **clustering coefficient**. Which proves that they are not random graphs.

3. All of the examples have a heavily tailed **degree distribution.**

4. Its also observed that the degree distribution of the test networks follows power law so they are scale free network.

5. Both *gnutella* and the *email-eu* graph has more than one connected component so path length between any two nodes from different component is infinite. So **diameter** and **average path length** doesn't make sense. But here the they are calculated on their largest connected component.

6. Low value of **diameter** denotes that these networks can be viewed as small world.

Resources:

Dataset : https://snap.stanford.edu/data/index.html

Algorithm insights : http://networksciencebook.com/

Networkx tutorial : https://networkx.github.io/documentation/stable/tutorial.html