Complex Networks
A1. Structural descriptors of complex networks

Aglaia-Elli Galata
Kharitonova Ksenia

Barcelona, Master in Artificial Intelligence

March 16, 2020

Contents

1 Introduction

The objective of this assignment is the calculation of the main structural descriptors for multiple complex networks. Through this part, it will be demonstrated that the networks of the same distribution tend to share common characteristics. Although most of these networks are simple graphs, there are also a few Multigrpahs.

2 Software

The implementation of this assignment was performed on `Jupyter notebook`, using `Python 3.7`. The included packages that were exploited are the following:

- **NetworkX**: All the metrics were calculated with this package. The provided graphs on `Pajek` format were also loaded via this library.
- **Numpy**: Mathematical operations
- **Pandas**: Table usages
- **seaborn, matplotlib**: Plotting

3 Part A. Numerical descriptors of networks

In this section, various numerical descriptors have been calculated over 26 different networks. These descriptors, that are depicted in table 1 are :

Nodes: total number of nodes

Edges: total number of edges

Min: minimum degree of the network

Max: maximum degree of the network

Avg: average degree of the network

Clus: average clustering of the network

Assort: assortativity of the network

Avg_path: average path of the network

Diam: diameter of the network

Name	Nodes	Edges	Min	Max	Avg	Clust	Assort	Avg_path	Diam
SF_1000_g2.5	1000	1905	2	30	3.81	0.0096	0.02	4.6149	10
SF_1000_g2.7	1000	1668	2	24	3.336	0.0067	-0.002	5.4688	12
homorand_N1000_K6_0	1000	2994	5	6	5.988	0.0038	0.1919	4.1913	6
homorand_N1000_K4_0	1000	2000	4	4	4.0	0.002	-	5.64	9
ER5000-kmed8	5000	19980	4	17	7.992	0.0014	-0.0555	4.3797	6
ER1000k8	1000	3956	1	17	7.912	0.008	-0.0168	3.5698	6
SF_1000_g3.0	1000	1517	2	26	3.034	0.0052	-0.0085	5.9651	13
ws2000	2000	6000	3	13	6.0	0.0033	-0.0762	4.5111	7
BA1000	1000	3990	4	115	7.98	0.0354	-0.0542	3.1833	5
256_4_4_4_13_18_p	256	4598	20	50	35.9219	0.5113	0.0007	2.6511	4
SF_500_g2.7	500	859	2	22	3.436	0.0078	-0.0256	4.8759	12
256_4_4_2_15_18_p	256	4548	30	46	35.5312	0.7331	0.0286	2.7821	5
rb125	125	426	4	100	6.816	0.8373	-0.1837	2.3032	4
ws1000	1000	3000	3	13	6.0	0.0044	-0.0999	4.0913	6
zachary_unwh	34	78	1	17	4.5882	0.5706	-0.4756	2.4082	5
dolphins	62	159	1	12	5.129	0.259	-0.0436	3.357	8
airports_UW	3618	14142	1	250	7.8176	0.4957	0.0462	4.4396	17
PGP	10680	24340	1	206	4.5581	0.2659	0.2395	7.4855	24
rb25	25	66	4	20	5.28	0.9023	-0.1635	2.0333	4
wheel	9	16	3	8	3.5556	0.6243	-0.3333	1.5556	2
graph32123	7	8	2	3	2.2857	0.6667	-0.6	2.1905	4
20x226x2	50	404	4	22	16.16	0.9716	0.9186	2.3878	4
graph32425	8	13	3	4	3.25	0.875	-0.0833	1.8571	3
grid-p-6x6	36	72	4	4	4.0	0.0	-	3.0857	6
star	9	8	1	8	1.7778	0.0	-1.0	1.7778	2
circle9	9	9	2	2	2.0	0.0	-	2.5	4

Table 1: Numerical descriptors of the networks

According to the table 1, the networks `homorand_N1000_K6_0`, `grid-p-6x6` and `circle9` have the assortativity metric equal to `nan`. The variance of these graphs is equal to 0.0, so the calculation of the assortativity fails. In other words, all the nodes are connected to an equal number of neighbours (figures 1 and 2).

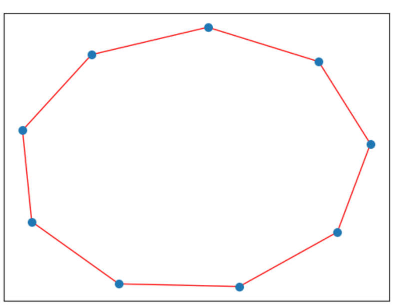

Figure 1: Circle9 network

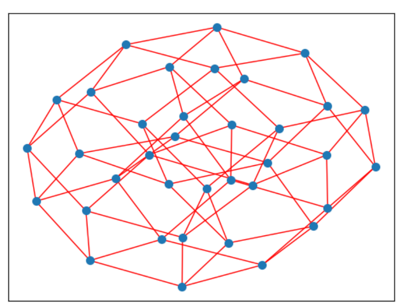

Figure 2: grid-p-6x6 network

4 Part B. Numerical descriptors of the nodes of the network real/airports_UW

In this section, we are analyzing 16 different airports of the airport network. For each one of them, we calculate multiple numerical descriptors (table 2)
Given this small sample of nodes, we can assume that the hubs of the network are the airports of *Paris, France, London* and *new York*. The last one has the biggest strength even if its degree is not the highest. Thus, it can be assumed that the edges - roots around New York are more weighted than the one in Paris.

Node	Degree	Strength	Clust	Avg_path	Max_path	Betweenness	Eigenvect_centr	Pagerank
ADA	7	10704.0	0.7143	3.6334	11	0.0	0.0107	0.0002
AGU	7	7678.0	0.7619	3.6655	11	0.0	0.0051	0.0001
AMS	192	481335.0	0.1428	2.7321	10	0.0405	0.1715	0.0054
ATL	172	1129605.0	0.1378	2.9162	11	0.0249	0.1221	0.0086
BCN	80	289105.0	0.3285	3.274	11	0.0019	0.0891	0.0028
CAN	80	204848.0	0.1712	3.3166	10	0.0076	0.0153	0.0028
CHC	20	64158.5	0.2526	3.5662	10	0.0034	0.0042	0.0016
CHI	184	1329505.0	0.1342	2.8087	11	0.0444	0.1381	0.0102
DJE	20	10198.5	0.7	3.5792	11	0.0001	0.0318	0.0002
FRA	237	697513.5	0.117	2.6829	10	0.0656	0.1955	0.0077
LON	242	1464828.0	0.1123	2.6359	10	0.085	0.2004	0.0156
MOW	186	217145.0	0.0958	2.8784	10	0.0522	0.1166	0.0059
NYC	179	1524349.5	0.1576	2.7092	11	0.0693	0.1606	0.0125
PAR	250	1023424.5	0.0892	2.6884	10	0.0934	0.1803	0.0127
TBO	2	234.0	1.0	4.5845	12	0.0	0.0001	0.0001
ZVA	1	19.0	0.0	7.5773	15	0.0	0.0	0.0001

Table 2: Airports: numerical descriptors

5 Part C. Plots of PDF and CCDF

In this section, it is analyzed the probability distribution of four different networks. The histogram forms demonstrate the way that the degrees of each node are spreading along.

5.1 ER1000k8

This is an homogeneous random network, since its probability distribution function follows a Gaussian like distribution. This conclusion can be extracted from both the figure 3 and table 3. The bins of the network contain a similar number of nodes. Moreover, the average degree of this network according to the calculations of section 3 corresponds to the bin with the maximum probability (7.9).

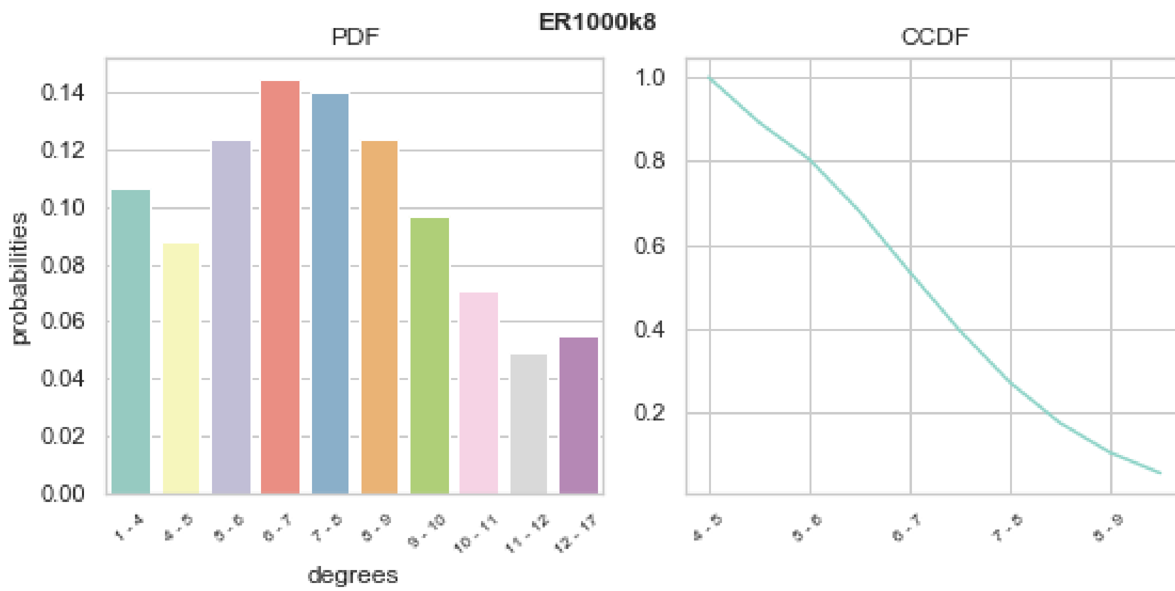

Figure 3: Probability Distributions of ER1000k8 network

Degrees	Count	PDF	CCDF
1 - 3	45	0.045	1.0
3 - 4	62	0.062	0.9551
4 - 5	88	0.088	0.893
5 - 6	124	0.124	0.805
6 - 7	145	0.145	0.681
7 - 8	140	0.14	0.536
8 - 9	124	0.124	0.396
9 - 10	97	0.097	0.272
10 - 11	71	0.071	0.175
11 - 12	49	0.049	0.1041
12 - 13	23	0.023	0.055
13 - 17	32	0.032	0.032

Table 3: 1_{st} column: Node degrees, 2_{nd} column: number of nodes in this interval, PDF: probability distribution function, CCDF: complementary cumulative distribution function

5.2 SF_1000_g2.7

This network follows a power law distribution. The majority of the nodes are concentrated to the first bin, where the degree is from 2 to 3. However, a percentage lower than 10% has degrees from 6 to 24, which opposed to a homogenious normal distribution. These results are verified on figure 3 and table 3.

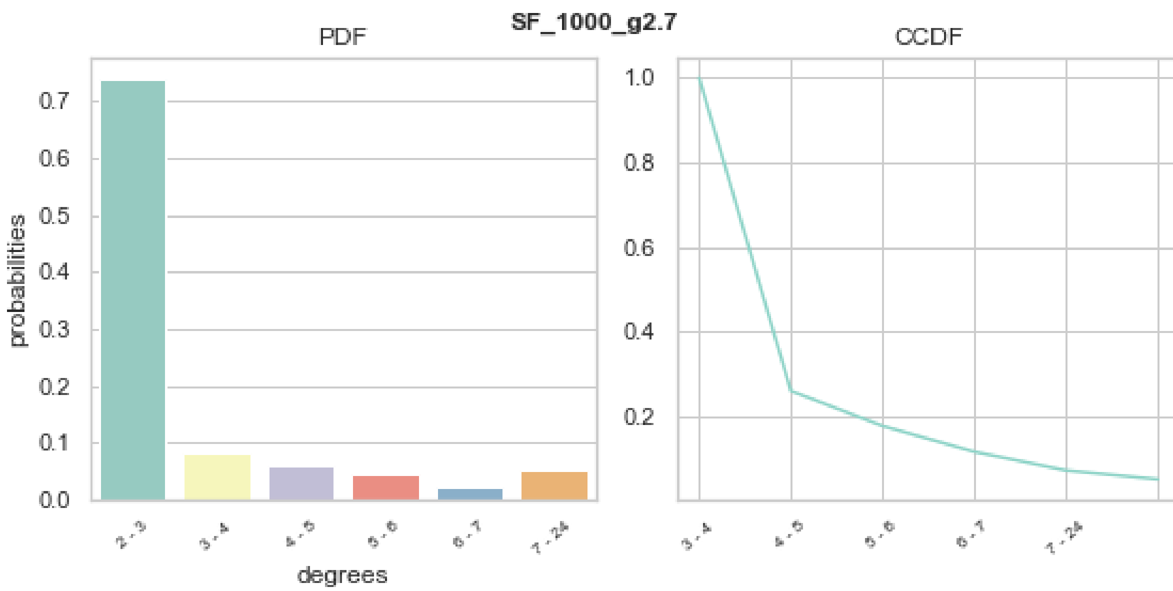

Figure 4: Probability Distributions of SF_1000_g2 network

Degrees	Count	PDF	CCDF
2 - 3	740	0.74	1.0
3 - 4	83	0.083	0.26
4 - 5	61	0.061	0.177
5 - 6	44	0.044	0.116
6 - 7	21	0.021	0.0721
7 - 9	21	0.021	0.0514
9 - 24	30	0.03	0.03

Table 4: 1_{st} column: Node degrees, 2_{nd} column: number of nodes in this interval, PDF: probability distribution function, CCDF: complementary cumulative distribution function

5.3 ws1000

This network follows a linear distribution. There is linear analogy on the probability of nodes contained in each bin. These results are shown on figure 5 and on table 5.

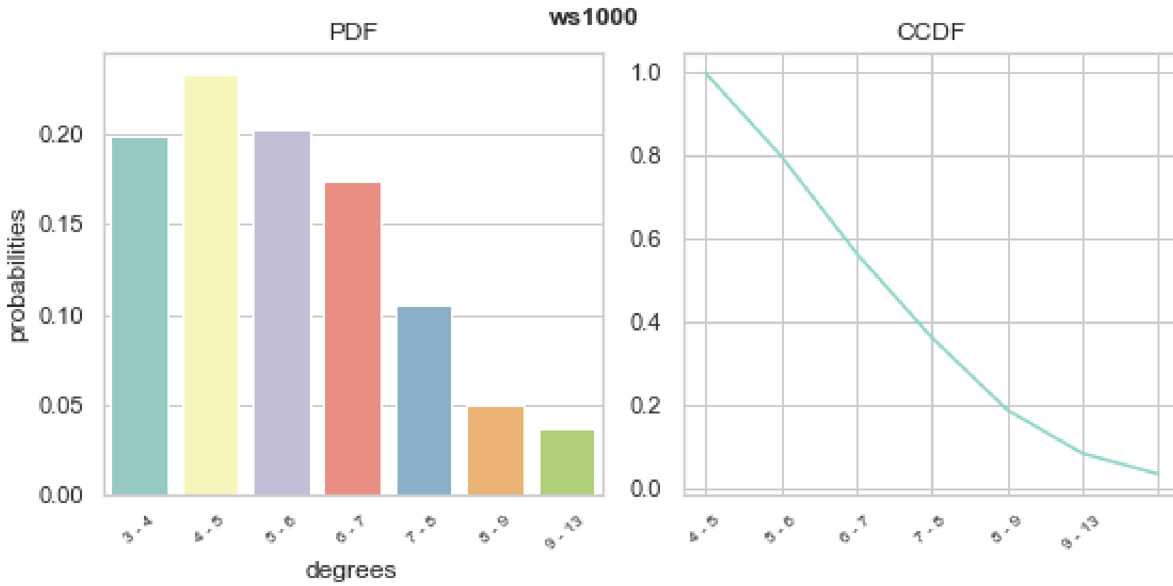

Figure 5: Probability Distributions of ws1000 network

Degrees	Count	PDF	CCDF
3 - 4	199	0.199	1.0
4 - 5	234	0.234	0.801
5 - 6	203	0.203	0.567
6 - 7	174	0.174	0.364
7 - 8	105	0.105	0.19
8 - 9	49	0.049	0.085
9 - 13	36	0.036	0.036

Table 5: 1_{st} column: Node degrees, 2_{nd} column: number of nodes in this interval, PDF: probability distribution function, CCDF: complementary cumulative distribution function

5.4 airports_UW

Finally, the airport network follows also a power law distribution. The majority of the nodes are concentrated to the first bins. Its interesting that the hubs of the network have a maximum degree equal to 250, while the average degree for the entire network that was calculated in 3 is just 7.8. Furthermore, the analysis of 5 shows the aforementioned hubs of the network i.e. Paris, New York, London which are contradicted the ones in Madagascar and in Tabora. The distributions and the detailed table are depicted on figure 6 and table 6.

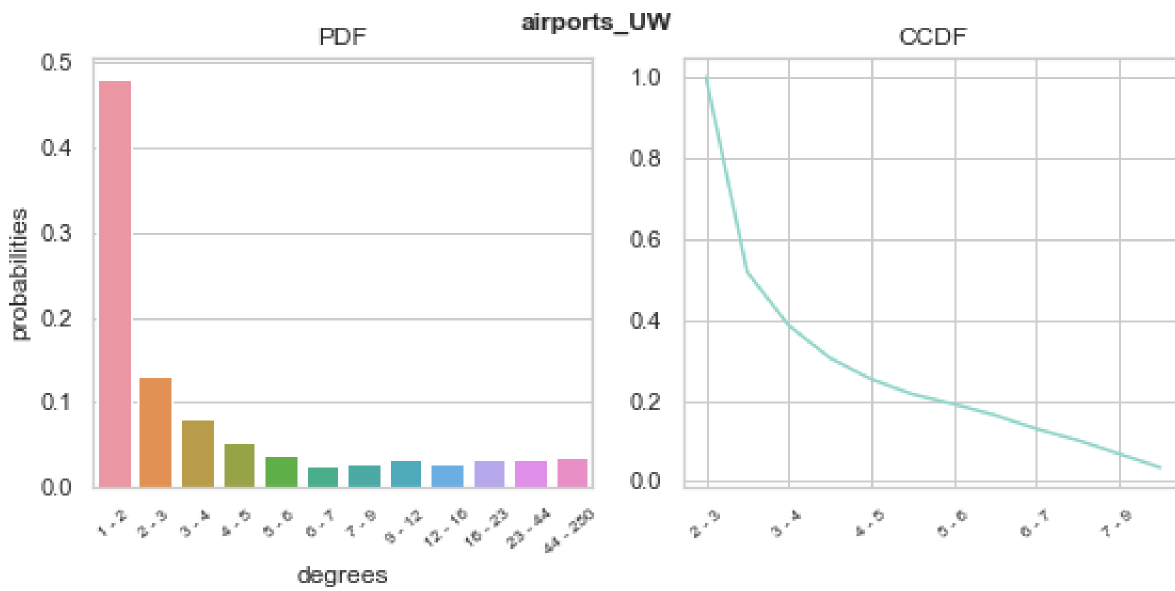

Figure 6: Probability Distributions of airports_UW network

Degrees	Count	PDF	CCDF
1 - 2	1744	0.48203	1.00001
2 - 3	476	0.13156	0.51798
3 - 4	292	0.08071	0.38642
4 - 5	191	0.05279	0.30571
5 - 6	134	0.03704	0.25292
6 - 7	89	0.0246	0.21588
7 - 8	54	0.01493	0.19128
8 - 10	100	0.02764	0.17635
10 - 11	37	0.01023	0.14871
11 - 14	79	0.02184	0.13848
14 - 17	52	0.01437	0.11664
17 - 20	83	0.02294	0.10227
20 - 26	70	0.01935	0.07933
26 - 38	76	0.02101	0.05998
38 - 63	67	0.01852	0.03897
63 - 250	74	0.02045	0.02045

Table 6: 1_{st} column: Node degrees, 2_{nd} column: number of nodes in this interval, PDF: probability distribution function, CCDF: complementary cumulative distribution function

References

[1] Mark E. J. Newman. (2003). *Statistical Mechanics (cond-mat.stat-mech); Disordered Systems and Neural Network. 10.1137/S003614450342480*