Activity A1 : Structural descriptors of complex networks

Hajar Lachheb

1 Introduction

Through this paper, I am presenting the individual work I did. Working on the first activity was a real challenge and a very exciting exercise indeed. First, we had to work on all the networks, then focused on the airport one. In the end, we had to plot the different networks and compare their representations to select the best.

2 Work Description

2.1 Description of what is done in all the work parts

To start with, the software used in my work is Google Colab. Following the use of Google Colab, I used the package NetworkX. The programming language used is Python. The choice of the software and the language can be explained by the fact that I am very confident when it comes to Python and I have used Google Colab a lot so it helped me code better and use the NetworkX Package in a very easy and optimized way.

I also tried working with iGraph but I preferred at the end working with NetworkX.

2.2 First Exercise Method Explanation

For the first exercise, we had to calculate the descriptors corresponding to all the networks. The idea was to have a csv file with all the results corresponding to each descriptor: Number of nodes, Number of edges, Minimum, maximum, and average degree, Average clustering coefficient (average of the clustering coefficient of each node), Assortativity, Average path length (average distance between all pairs of nodes), Diameter (maximum distance between nodes in the network). Before coding this part, I had to convert all the multigraphs to graphs using nx.Graph.

My method was not that smart, because I had to repeat the code three times for every sub-element of the initial folder "A1-networks". Here is the result I get after running the code for each sub-folder.

Figure representing the descriptors results for the folder Model

network_file	num_nodes	num_edges	min_degree	max_degree	avg_degree	avg_clustering	assortativity	avg_path_length	diameter
ER1000k8.net	1000	3956	1	17	7.912	0.008042110177	-0.01683326736	3.569777778	6
SF_1000_g2.7.net	1000	1668	2	24	3.336	0.006650454307	-0.001960898955	5.468826827	12
SF_500_g2.7.net	500	859	2	22	3.436	0.007773809575	-0.0255858811	4.875935872	12
256_4_4_4_13_18_	256	2299	10	25	17.9609375	0.5112995545	0.0006954016351	2.651102941	4
256_4_4_2_15_18_	256	2274	15	23	17.765625	0.7330611456	0.02862392046	2.782107843	5
homorand_N1000_K	1000	2000	4	4	4	0.002		5.64	9
SF_1000_g2.5.net	1000	1905	2	30	3.81	0.00960504089	0.01998152187	4.614932933	10
BA1000.net	1000	3990	4	115	7.98	0.0354487623	-0.05420735055	3.183279279	5
homorand_N1000_K	1000	2994	5	6	5.988	0.0038	0.191902834	4.191299299	6
ws2000.net	2000	6000	3	13	6	0.00333015873	-0.07617895967	4.511093047	7
SF_1000_g3.0.net	1000	1517	2	26	3.034	0.005159194521	-0.008538806598	5.965063063	13
ER5000k8.net	5000	19980	4	17	7.992	0.001388896659	-0.05547627909	4.379741068	6
ws1000.net	1000	3000	3	13	6	0.004364840715	-0.09992885591	4.091303303	6
rb125.net	125	410	4	100	6.56	0.8372950558	-0.1729819599	2.303225806	4

Figure representing the descriptors results for the folder Real

network_file	num_nodes	num_edges	min_degree	max_degree	avg_degree	avg_clustering	assortativity	avg_path_length	diameter
rb25.net	25	66	4	20	5.28	0.9023157895	-0.1635258359	2.033333333	4
graph3+1+3.net	7	8	2	3	2.285714286	0.6666666667	-0.6	2.19047619	4
20x2+5x2.net	50	404	4	22	16.16	0.9715584416	0.9186473754	2.387755102	4
graph4+4.net	8	13	3	4	3.25	0.875	-0.08333333333	1.857142857	3
circle9.net	9	9	2	2	2	0		2.5	4
star.net	9	8	1	8	1.777777778	0	-1	1.777777778	2
wheel.net	9	16	3	8	3.555555556	0.6243386243	-0.3333333333	1.555555556	2
grid-p-6x6.net	36	72	4	4	4	0		3.085714286	6
Airports.csv	36	72	4	4	4	0		3.085714286	6

Figure representing the descriptors results for the folder Toy

network_file	num_nodes	num_edges	min_degree	max_degree	avg_degree	avg_clustering	assortativity	avg_path_length	diameter
dolphins.net	62	159	1	12	5.129032258	0.2589582461	-0.04359402822	3.356953993	8
airports_UW.net	3618	14142	1	250	7.817578773	0.4957489312	0.04622413053	4.439594642	17
zachary_unwh.net	34	78	1	17	4.588235294	0.5706384782	-0.4756130977	2.408199643	5
PGP.net	10680	24316	1	205	4.553558052	0.2659452243	0.2382113717	7.485540051	24

2.3 Second Exercise Method Explanation

For the second exercise, I had to work on a single network which is the Airport one. I am using the numerical descriptors of the nodes such as Degree, Strength, Clustering coefficient, Average path length (average distance to the rest of the nodes), Maximum path length (maximum distance to the rest of the nodes), Betweenness, Eigenvector centrality, and PageRank. We considered that the network is unweighted and we did the same as we did in the previous exercise. In this part, instead of looking manually for the Airport network, we will merge all the networks we have in a tuple, so that we can, later on, navigate through the different networks easily for the next exercise without the need to navigate through each folder and subfolder and look for the adequate file. We get a CSV File then with the results of the descriptor corresponding to all of the aeroports.

Figure representing the descriptors result for the full Airport Network

	Node	Degree	Strenght	Clustering coefficien	Average path length	Maximum path lengt	Betweenness	Eigenvector centrali	PageRank
0	AAA	2	92	1	4.515754561	12	0	8.97E-05	6.47E-05
1	AAE	6	7284	0.7333333333	3.676616915	11	6.39E-05	0.005915562148	0.0002196250765
2	AAL	3	13071	1	3.909342178	11	0	0.003318859326	0.0001834549814
3	AAN	11	3224.5	0.6727272727	3.926755113	11	1.01E-05	0.006981236005	9.87E-05
4	AAQ	10	1690	0.3333333333	3.809010503	11	2.00E-05	0.006565162375	9.26E-05
5	AAR	6	11653	0.9333333333	3.599778883	11	1.70E-08	0.00910678414	0.0001635669682
6	AAT	1	28.5	0	4.681315644	12	0	0.0001246977819	4.22E-05
7	AAU	1	38	0	5.630458817	13	0	8.65E-07	5.06E-05
8	AAY	4	646	0.8333333333	4.516583748	12	6.12E-08	0.0005953009436	7.32E-05
9	ABA	1	164	0	4.871199558	12	0	6.83E-05	4.81E-05
10	ABD	4	2918	0.8333333333	3.954947485	11	3.03E-06	0.002381753466	0.0001116901739
11	ABE	14	21783	0.9230769231	3.644831399	12	2.27E-06	0.02078037394	0.0001911662168
12	ABI	3	2582	1	3.931177446	12	0	0.003548051213	5.88E-05
13	ABJ	28	25044.5	0.3544973545	3.465450525	11	0.000950327909	0.02444291808	0.0006233111675
14	ABK	3	250	1	5.426755113	13	0	7.83E-06	0.0001253324272
15	ABL	4	278	0.6666666667	5.190713101	13	2.75E-06	6.30E-06	0.0001400644832
16	ABM	3	268	0.3333333333	4.463239359	10	0.0002757586063	8.99E-05	7.10E-05
17	ABQ	36	105678	0.3412698413	3.422056385	11	0.004188841409	0.02902630608	0.001067623058
18	ABR	3	2958	0.3333333333	4.045328911	12	0.0005529444291	0.001430211472	0.0001624508565
19	ABS	1	4391	0	4.660862355	12	0	9.17E-05	0.000113697954
20	ABT	2	2210	1	4.089552239	11	0	0.001729300578	7.67E-05
21	ABV	3	3253	0.6666666667	3.706467662	11	8.56E-05	0.003663205566	0.0001334386233
22	ABX	5	3935	0.6	4.212548369	11	5.70E-06	0.0004636435229	0.000147441454
23	ABY	1	1410	0	3.914870094	12	0	0.002168693763	5.05E-05
24	ABZ	22	32727.5	0.2813852814	3.582642344	11	0.003747922599	0.0146878349	0.000610957714

Following is the representation of descriptors of the following airports: PAR, LON, FRA, AMS, CHI, NYC, ATL, BCN, WAW, CHC, DJE, ADA, AGU, TBO, ZVA.

Figure representing the descriptors result - Dataframe

0	ADA	7	10704.0	0.714286	3.632394	11	0.000013	0.010688	0.000205
1	AGU	7	7678.0	0.761905	3.664455	11	0.000006	0.005134	0.000119
2	AMS	192	481335.0	0.142834	2.731343	10	0.040492	0.171452	0.005384
3	ATL	172	1129605.0	0.137835	2.915423	11	0.024896	0.122071	0.008603
4	BCN	80	289105.0	0.328481	3.273079	11	0.001932	0.089146	0.002816
5	CHC	20	64158.5	0.252632	3.565229	10	0.003367	0.004188	0.001615
6	DJE	20	10198.5	0.700000	3.578220	11	0.000146	0.031831	0.000181
7	FRA	237	697513.5	0.116963	2.682145	10	0.065578	0.195546	0.007704
8	LON	242	1464828.0	0.112342	2.635158	10	0.084989	0.200372	0.015606
9	MOW	186	217145.0	0.095844	2.877557	10	0.052211	0.116645	0.005884
10	NYC	179	1524349.5	0.157554	2.708402	11	0.069283	0.160584	0.012471
11	PAR	250	1023424.5	0.089157	2.687673	10	0.093420	0.180287	0.012729
12	TBO	2	234.0	1.000000	4.583195	12	0.000000	0.000123	0.000080
13	WAW	55	86836.5	0.458586	3.243505	11	0.001557	0.075199	0.001120
14	ZVA	1	19.0	0.000000	7.575180	15	0.000000	0.000000	0.000095

2.4 Third Exercise Method Explanation

In this exercise, we simply had to plot the histograms. What is special about my work is that i tried to plot the different possibilities and includes: PDF linlin, CDF linlin, CCDF linlin, PDF loglog, CDF loglog and finally CCDF loglog. The reason why we plotted all of these representations is to be able to choose the best representation following if the figure we get represents a power law and how it is actually presented.
Following is the representation of the networks : ER5000k8, SF1000g2.7, ws1000, airportsUW and PGP.

Figure representing the best representation for the network SF1000g2.7

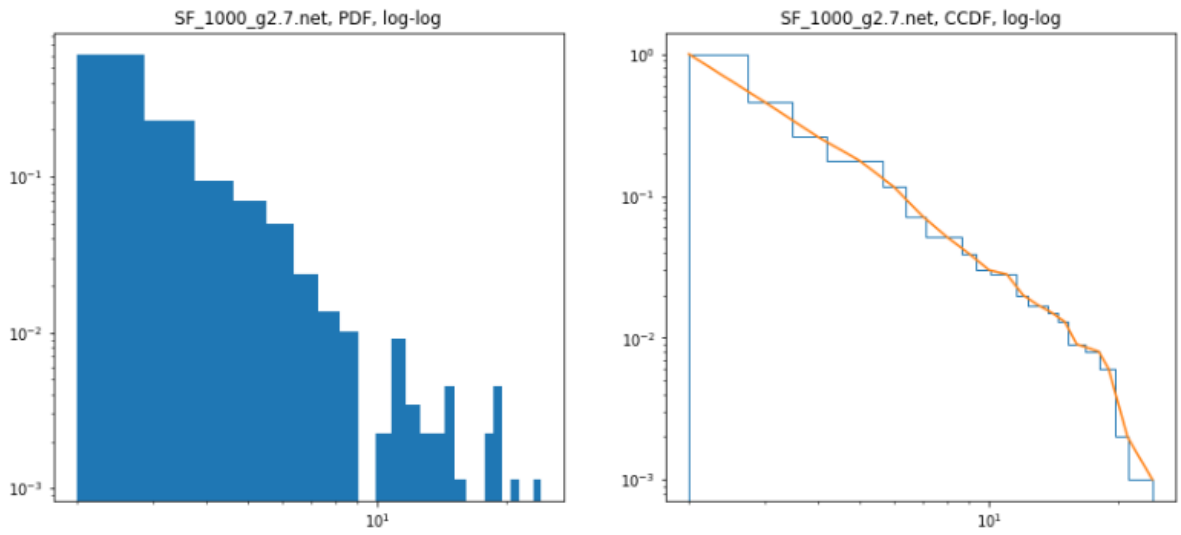

Figure representing the best representation for the network ws1000

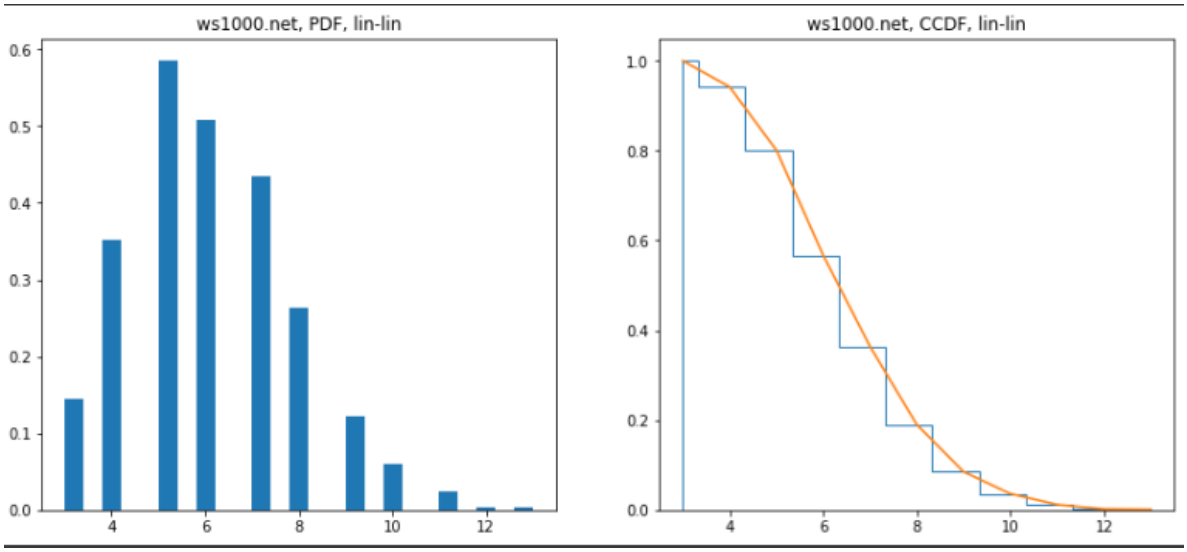

Figure representing the best representation for the network ER5000k8

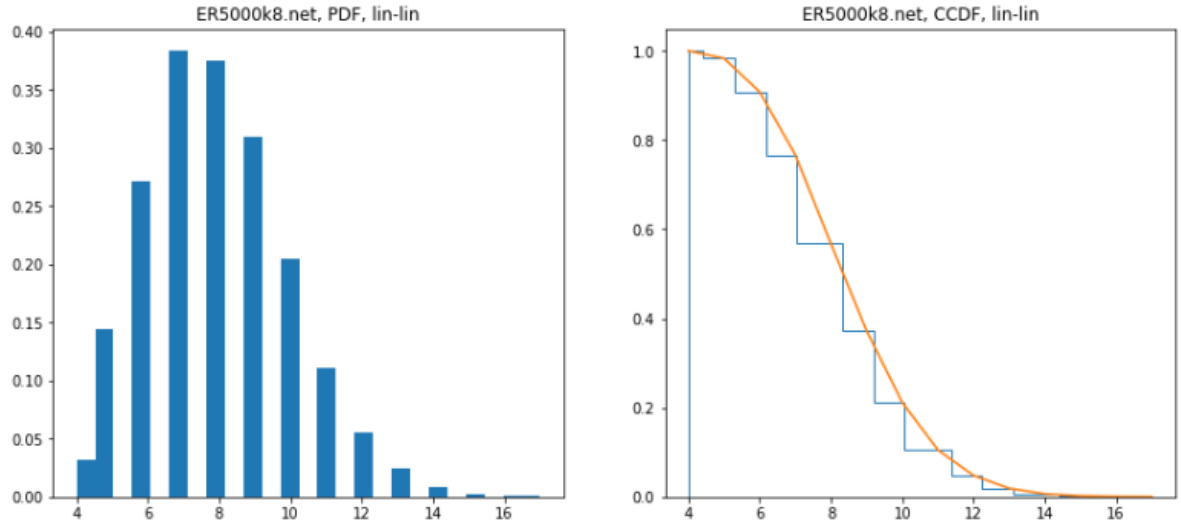

Figure representing the best representation for the network PGP

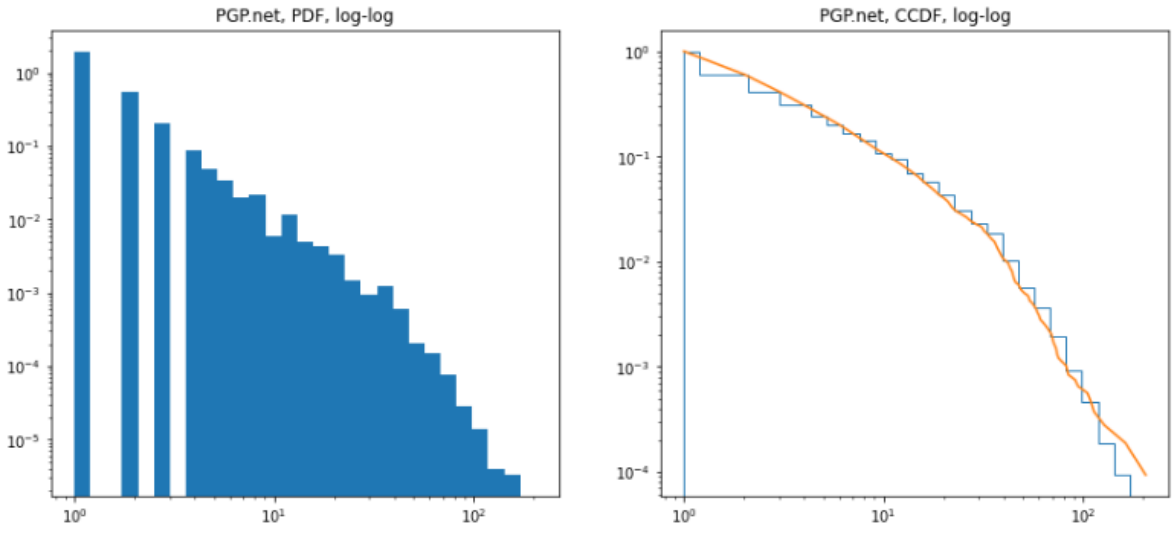

Figure representing the best representation for the network airportsUW