UNIVERSITAT ROVIRA I VIRGILI
Dep. of Computer Engineering and Mathematics
March 10, 2021

CN-MAI Assignment 1

Complex Networks
Spring 2021, 5 Credits

Structural Descriptors Of Complex Networks

Name Bartosz Paulewicz
Betty Törnkvist

E-mail bartosz.paulewicz@student.put.poznan.pl
betty.tornkvist@gmail.com

Teacher
Alejandro Arenas Moreno

Contents

1 Introduction

Given three sets of undirected graphs, different numerical descriptors can be calculated for comparison and analysis. The sets are the following:

1. toy (sample networks),
2. model (networks generated from models),
3. real (real networks).

All files are in Pajek format (*.net), a simple text file-format where each line is an element. First up is the vertices(nodes), followed by the edges, see example below,

```
 *Vertices 2
 1 "1" 0.4 0.9 0.5
 2 "2" 0.6 0.9 0.5
*Edges
 1 2 1
```

The first line has information about the number of vertices. Thereafter, each vertex is described by number, name and coordinates. Edges are described by the two nodes which has the edge, followed by the weight of the edge.

2 Method

The task can be divided into three parts:

(a) **Numerical descriptors of networks**
For each network, the following numerical descriptors are calculated:

- Number of nodes - Integer, extracted by reading the first line of the file.
- Number of edges - Integer, extracted by counting the lines after `*Edges`.
- Minimum, maximum and average degree - The number of neighbours for each vertex.
- Average clustering coefficient - The average of the clustering coefficient of each node. The clustering coefficient is the degree of which the vertices clusters together.
- Assortativity - The preference of how vertices connect to similar vertices.
- Average path length - The average distance between all pairs of nodes.
- Diameter - The maximum distance between nodes in the network.

(b) **Numerical descriptors of the nodes of the network real/airports_UW.net**

- Degree - The number of neighbours for each vertex.
- Strength - The sum of the weights connected to each node.
- Clustering coefficient - The degree of which the vertices clusters together.
- Average path length - The average distance to the rest of the nodes.
- Maximum path length - The maximum distance to the rest of the nodes.
- Betweenness - The centrality of a node based on the shortest paths.
- Eigenvector centrality - The influence of a node based on node-scores.
- PageRank - The importance of a node based on incoming connections and importance of those nodes.

(c) **Plots**

- Degree distributions - PDF, probability distribution function.
- Complementary cumulative degree distributions - CCDF, complementary cumulative distribution function.

3 Result

All descriptors are calculated using the Python library igraph [1]. The library contains functions for all the different descriptors. The main numerical descriptors for task *(a)* can be seen in Figure 1, where non-inter values are shown with a precision of four decimals. The result is saved to a .csv file, named `a.csv`. The calculations covers all available networks from the three sets.

Network	Number of nodes	Number of edges	Minimum degree	Maximum degree	Average degree	Average clustering coefficient	Assortativity	Average path length	Diameter
model\256_4_4_2_15_18_p.net	256	4548	30	46	35.5312	0.7331	0.0286	2.7821	5
model\256_4_4_4_13_18_p.net	256	4598	20	50	35.9219	0.5113	0.0007	2.6511	4
model\BA1000.net	1000	3990	4	115	7.98	0.0354	-0.0542	3.1833	5
model\ER1000k8.net	1000	3956	1	17	7.912	0.0081	-0.0168	3.5698	6
model\ER5000k8.net	5000	19980	4	17	7.992	0.0014	-0.0555	4.3797	6
model\homorand_N1000_K4_0.net	1000	2000	4	4	4.0	0.002	NaN	5.6400	9
model\homorand_N1000_K6_0.net	1000	2994	5	6	5.988	0.0038	0.1919	4.1913	6
model\rb125.net	125	426	4	100	6.816	0.8373	-0.1837	2.3032	4
model\SF_1000_g2.5.net	1000	1905	2	30	3.81	0.0096	0.02	4.6149	10
model\SF_1000_g2.7.net	1000	1668	2	24	3.336	0.0067	-0.002	5.4688	12
model\SF_1000_g3.0.net	1000	1517	2	26	3.034	0.0052	-0.0085	5.9651	13
model\SF_500_g2.7.net	500	859	2	22	3.436	0.0078	-0.0256	4.8759	12
model\ws1000.net	1000	3000	3	13	6.0	0.0044	-0.0999	4.0913	6
model\ws2000.net	2000	6000	3	13	6.0	0.0033	-0.0762	4.5111	7
real\airports_UW.net	3618	14142	1	250	7.8176	0.6228	0.0462	4.4396	17
real\dolphins.net	62	159	1	12	5.129	0.3029	-0.0436	3.3570	8
real\PGP.net	10680	24340	1	206	4.5581	0.4403	0.2395	7.4855	24
real\zachary_unwh.net	34	78	1	17	4.5882	0.5879	-0.4756	2.4082	5
toy\20x2+5x2.net	50	404	4	22	16.16	0.9716	0.9186	2.3878	4
toy\circle9.net	9	9	2	2	2.0	0.0	NaN	2.5000	4
toy\graph3+1+3.net	7	8	2	3	2.2857	0.6667	-0.6	2.1905	4
toy\graph4+4.net	8	13	3	4	3.25	0.875	-0.0833	1.8571	3
toy\grid-p-6x6.net	36	72	4	4	4.0	0.0	NaN	3.0857	6
toy\rb25.net	25	66	4	20	5.28	0.9023	-0.1635	2.0333	4
toy\star.net	9	8	1	8	1.7778	0.0	-1.0	1.7778	2
toy\wheel.net	9	16	3	8	3.5556	0.6243	-0.3333	1.5556	2

Figure 1: Main numerical descriptors.

[1] https://igraph.org/python/

The numerical descriptors of the airport network in *(b)* can be seen in Figure 2. Non-integer values are shown with a precision of eight decimals. The result is saved to a .csv file, named `b.csv`. The descriptors are calculated for following airports (nodes):

PAR, LON, FRA, AMS, MOW, CHI, NYC, ATL, BCN, WAW, CHC, DJE, ADA, AGU, TBO, ZVA

Airport	Degree	Strength	Clustering coefficient	Average path length	Maximum path length	Betweenness	Eigenvector centrality	PageRank
ADA	7	10704.0	0.71428571	4.63239359	12	86.13209062	0.05336155	0.00018651
AGU	7	7678.0	0.76190476	4.6644555	12	37.68371295	0.02559141	0.00019367
AMS	192	481335.0	0.14283377	3.73134328	11	264799.7485435	0.85577995	0.00401098
ATL	172	1129605.0	0.1378349	3.91542289	12	162809.45400881	0.60864673	0.00381899
BCN	80	289105.0	0.32848101	4.27307905	12	12636.35135909	0.44507124	0.00160535
CHC	20	64158.5	0.25263158	4.56522941	11	22021.5322141	0.02089875	0.00084125
CHI	184	1329505.0	0.13417676	3.80790492	12	290639.84924374	0.68866292	0.00406115
DJE	20	10198.5	0.7	4.57822001	12	953.83921258	0.15892481	0.00041071
FRA	237	697513.5	0.11696346	3.68214483	11	428847.8032452	0.97602259	0.00496923
LON	242	1464828.0	0.11234183	3.63515755	11	555788.52171987	1.0	0.00526564
MOW	186	217145.0	0.09584423	3.87755666	11	341434.48646073	0.5823272	0.00511005
NYC	179	1524349.5	0.15755445	3.70840243	12	453081.88267624	0.80100266	0.00389207
PAR	250	1023424.5	0.08915663	3.68767275	11	610925.94208172	0.89986318	0.00614411
TBO	2	234.0	1.0	5.58319514	13	0.0	0.00061431	0.00012092
WAW	55	86836.5	0.45858586	4.2435047	12	10181.73242408	0.37541514	0.00115741
ZVA	1	19.0	NaN	8.57517966	16	0.0	0.00000001	0.00013213

Figure 2: Main numerical descriptors of selected nodes in the network real/airports_UW.net

The plots for task *(c)*, PDF and CCDF, where generated for the following networks:

- model/ER5000k8.net (linear)
- model/SF_1000_g2.7.net (linear)
- model/ws1000.net (log-scale)
- real/airports_UW.net (log-scale)

The results can be seen in Figure 3.

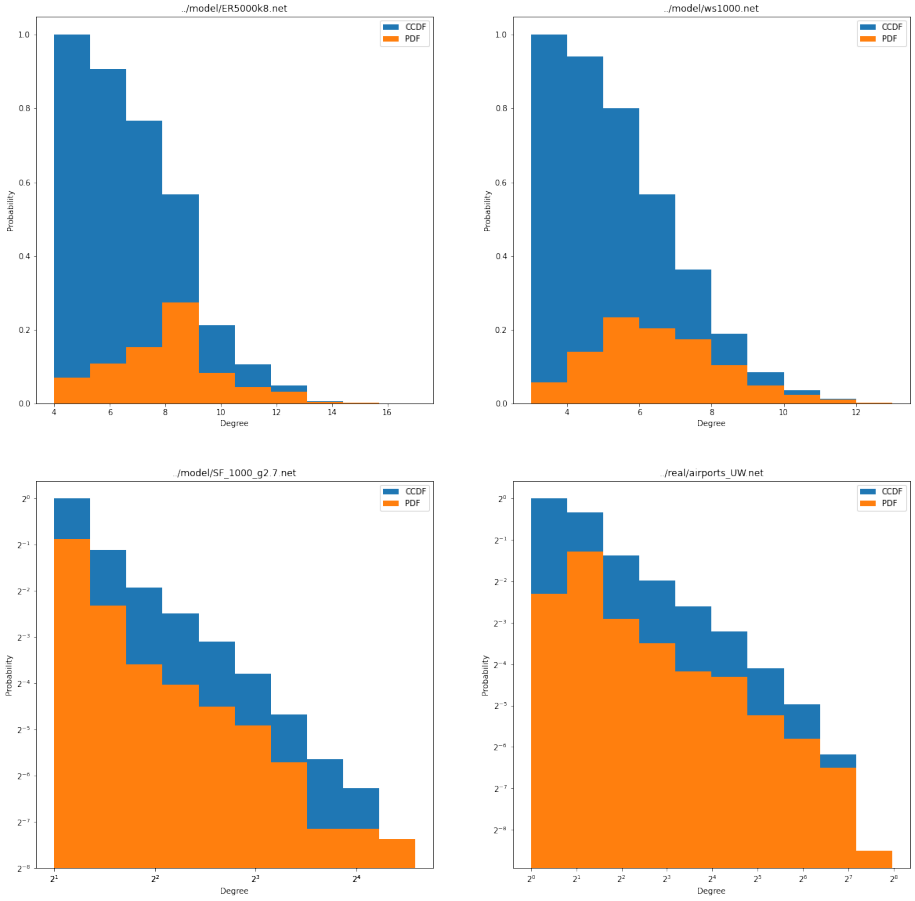

Figure 3: Normalised histograms for PDF and CCDF. Top plots in linear form, and lower plots in log-scale.