Mateo W. Racca

UNIVERSITAT ROVIRA i VIRGILI
Escola Tècnica
Superior d'Enginyeria

COMPLEX NETWORKS

(MESIIA): Structural descriptors of complex networks

Introduction

The first assignment regarding Complex Networks demanded me to delve more into the importance of the networks in our day-to-day life.

The given bibliography is: *Newman, Networks 2019,* the presentations and other materials provided by the professors.

This assignment explores complex networks, focusing on their structural properties and individual node metrics utilizing NetworkX, a comprehensive Python library for network analysis.

The objective is to work in order to dissect both: theoretical models and real-world networks to uncover underlying patterns and dynamics.

Through tasks designed to highlight global network descriptors, node-specific metrics, and degree distributions, I aim to bridge the gap between abstract network theory and tangible insights.

This hands-on approach not only deepens my understanding of complex systems but also showcases the power of computational tools and existent libraries like NetworkX.

All the folders provided for the assignment, as well as the Jupyter notebook used for the assignment and the results are available at https://github.com/raccamateo/complexnet_structural_descriptors.

By cloning the repository and following the instructions provided at the README file the process can be opened, started and replicated in order to evaluate locally on any device/cloud environment.

a) Numerical descriptors of networks

The study of complex networks has become a pivotal part of understanding systems ranging from biological entities to human social structures and technological designs.

The intricate web of connections in these networks defines not only their current state but also their evolutionary dynamics and functional capabilities.

Structural descriptors are quantifiable measures that capture the essence of a network's architecture, providing a window into the behavior and characteristics of complex systems.

For the calculation of structural descriptors, we utilized NetworkX, a Python library specifically designed for the creation, manipulation, and study of complex networks. NetworkX offers an extensive suite of built-in functions that facilitate the computation of various network metrics, making it an ideal choice for this task.

The networks analyzed were sourced from both model-generated and real-world datasets, covering a diverse array of network types. Each network was loaded into NetworkX from a Pajek-format file, ensuring uniformity in data handling. To ensure comparability across networks, all were treated as undirected graphs, simplifying their structure for analysis.

Decisions included network Conversion: multigraphs were converted to simple graphs to eliminate parallel edges and self-loops, focusing on the underlying connectivity structure.

This is crucial for obtaining accurate, comparable measures calculating the following:

- ➔ Number of Nodes and Edges: the size of a network can have implications for its dynamics. In epidemiology, for instance, the number of nodes (individuals) and edges (interactions) can dictate the spread of a disease. In social networks, more nodes and edges might suggest greater complexity and potential for information dissemination.

- ➔ Average Degree: is often interpreted as a network's redundancy. In ecological networks, for example, a higher average degree might suggest greater biodiversity and ecosystem resilience.

➔ Clustering Coefficient: captures the 'small world' phenomenon prevalent in many real-world networks, where nodes tend to create tightly knit groups characterized by a high density of triangles.

➔ Assortativity: networks with a high assortativity coefficient tend to have nodes that preferentially attach to others with a similar degree. This can impact the network's robustness and can influence processes such as opinion formation in social networks.

➔ Average Path Length and Diameter: provide insights into the 'six degrees of separation' concept, highlighting the interconnectedness and efficiency of the network. A low average path length and small diameter are indicative of efficient information transfer, which can be advantageous for communication networks but potentially detrimental for disease spread.

The achieved results for each one of the provided networks can be find on the next table:

Number of Nodes	Number of Edges	Minimum Degree	Maximum Degree	Average Degree	Average Clustering Coefficient	Assortativity	Average Path Length	Diameter	Network
25	66	4	20	5.28	0.9023	-0.1635	2.0333	4	toy/rb25.net
9	16	3	8	3.5556	0.6243	-0.3333	1.5556	2	toy/wheel.net
7	8	2	3	2.2857	0.6667	-0.6	2.1905	4	toy/graph3+1+3.net
50	404	4	22	16.16	0.9716	0.9186	2.3878	4	toy/20x2+5x2.net
8	13	3	4	3.25	0.875	-0.0833	1.8571	3	toy/graph4+4.net
36	72	4	4	4.0	0.0		3.0857	6	toy/grid-p-6x6.net
9	8	1	8	1.7778	0.0	-1.0	1.7778	2	toy/star.net
9	9	2	2	2.0	0.0		2.5	4	toy/circle9.net
34	78	1	17	4.5882	0.5706	-0.4756	2.4082	5	real/zachary_unwh.net
62	159	1	12	5.129	0.259	-0.0436	3.357	8	real/dolphins.net
3618	14142	1	250	7.8176	0.4957	0.0462	4.4396	17	real/airports_UW.net
10680	24316	1	205	4.5536	0.2659	0.2382	7.4855	24	real/PGP.net
1000	1905	2	30	3.81	0.0096	0.02	4.6149	10	model/SF_1000_g2.5.net
1000	1668	2	24	3.336	0.0067	-0.002	5.4688	12	model/SF_1000_g2.7.net
1000	2994	5	6	5.988	0.0038	0.1919	4.1913	6	model/homorand_N1000_K6_0.net
1000	2000	4	4	4.0	0.002		5.64	9	model/homorand_N1000_K4_0.net
1000	3956	1	17	7.912	0.008	-0.0168	3.5698	6	model/ER1000k8.net
1000	1517	2	26	3.034	0.0052	-0.0085	5.9651	13	model/SF_1000_g3.0.net
2000	6000	3	13	6.0	0.0033	-0.0762	4.5111	7	model/ws2000.net
1000	3990	4	115	7.98	0.0354	-0.0542	3.1833	5	model/BA1000.net
256	2299	10	25	17.9609	0.5113	0.0007	2.6511	4	model/256_4_4_4_13_18_p.net
5000	19979	4	17	7.9916	0.0014	-0.0555	4.3797	6	model/ER5000k8.net
500	859	2	22	3.436	0.0078	-0.0256	4.8759	12	model/SF_500_g2.7.net
256	2274	15	23	17.7656	0.7331	0.0286	2.7821	5	model/256_4_4_2_15_18_p.net
125	410	4	100	6.56	0.8373	-0.173	2.3032	4	model/rb125.net
1000	3000	3	13	6.0	0.0044	-0.0999	4.0913	6	model/ws1000.net

Insights

Network Size and Connectivity: the vast difference in network sizes, as seen from rb25.net to PGP.net, underscores the diversity of complex networks.
The number of nodes and edges directly impacts the network's potential for complexity and its ability to facilitate diverse interactions.

High-density networks, such as 256_4_4_2_15_18_p.net, potentially indicate a robust system where failure of a single node is less likely to disrupt the network. However, this can also suggest a vulnerability to systematic attacks aimed at high-degree nodes.

Clustering and Network Cohesion: networks with high clustering coefficients, like rb125.net, often signify a strong presence of community structures or functional groups within the network. This can enhance the resilience of the network to external shocks, as communities can maintain functionality even when parts of the network are compromised.
On the other hand, low clustering seen in SF_1000_g2.5.net and ws2000.net may point towards a more tree-like structure, where hierarchical connections dominate, possibly making these networks susceptible to the removal of central nodes.

Assortativity and Network Homogeneity: the variation in assortativity across networks highlights different linking preferences.
Networks with negative assortativity, such as graph3+1+3.net, may possess a core-periphery structure, where highly connected hubs are essential for maintaining network cohesion. This could imply that disrupting these hubs could fragment the network.
Positive assortativity as observed in networks like 20x2+5x2.net suggests a tendency towards homogeneity in node connections, which can foster resilience against random failures but may also lead to echo chambers or reduced innovation in social or informational networks.

Path Length and Network Diameter:
Networks exhibiting small-world properties, like airports_UW.net with its relatively short average path length, demonstrate an efficient information flow across the network, crucial for rapid communication and synchronization across the system.
A large diameter, as seen in PGP.net, while indicative of wide network reach, can also signal inefficiencies in communication or transportation times across the network.
This can be particularly critical in networks where timely response or interaction is vital.

Implications

Model vs. Real-world Networks: The stark contrasts between model and real-world networks highlight the complexity of real systems, which often evolve under constraints and pressures not captured by idealized models. For instance, the airports_UW.net displays characteristics of both scale-free and small-world networks, reflecting the multifaceted priorities in airport network design, such as efficiency, accessibility, and robustness.

Network Design and Policy Making: Understanding the structural properties of networks has profound implications for designing infrastructure systems, formulating public health strategies, and crafting policies for social governance. For example, enhancing the clustering coefficient in social networks might be desirable to foster community support, whereas increasing connectivity in transportation networks can reduce travel times and costs.

Strategies for Network Resilience: The analysis underscores the importance of network design strategies that enhance resilience. For example, introducing redundancy in highly centralized networks can mitigate the risk of cascading failures, while ensuring diversity in connectivity patterns in assortative networks can prevent isolation of network clusters.

Concluding Remarks

The analysis of complex networks through their structural descriptors reveals the nuanced interplay between network architecture and functionality. It provides a rich tapestry of insights into how networks are organized, how they perform under various conditions, and how they can be optimized or protected. This understanding is crucial not just for theoretical exploration but for practical applications in technology, ecology, sociology, and beyond, where the stability and efficiency of networks can have far-reaching consequences.

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

Part b is focused on the 'airports_UW.net' network, representing a real-world air transportation system.

Software & Libraries:

NetworkX's centrality measures and path analysis functions were utilized to calculate node-specific metrics, leveraging its efficient implementation of these complex algorithms.

NumPy was employed for mathematical operations, especially for handling path length calculations and averages across large sets of data.

Scripts & Decisions:

Efficient Calculation: Given the size of the airports_UW.net network and the computationally intensive nature of metrics like betweenness centrality and average path lengths, we optimized calculations by pre-computing certain metrics and using approximation methods where appropriate. For example, betweenness centrality calculations were conducted using NetworkX's efficient algorithms that avoid redundant computation.

Weighted Analysis for Strength: The network was analyzed as unweighted for most metrics to focus on the structure of connectivity. However, for the strength (weighted degree), edge weights were considered to provide insight into the volume of traffic or connections between airports, reflecting their operational significance.

The achieved results for each one of chosen (filtered) airports can be find on the next table:

Airport	Degree	Strength	Clustering Coefficient	Average Path Length	Maximum Path Length	Betweenness	Eigenvector Centrality	PageRank
ADA	7	10704.0	0.71428571	3.63239359	11	0.00001317	0.01068776	0.00020485
AGU	7	7678.0	0.76190476	3.66445550	11	0.00000576	0.00513412	0.00011933
AMS	192	481335.0	0.14283377	2.73134328	10	0.04049213	0.17145160	0.00538375
ATL	172	1129605.0	0.13783490	2.91542289	11	0.02489618	0.12207139	0.00860287
BCN	80	289105.0	0.32848101	3.27307905	11	0.00193230	0.08914620	0.00281650
CHC	20	64158.5	0.25263158	3.56522941	10	0.00336745	0.00418781	0.00161537
CHI	184	1329505.0	0.13417676	2.80790492	11	0.04444350	0.13810203	0.01017965
DJE	20	10198.5	0.70000000	3.57822001	11	0.00014586	0.03183126	0.00018064
FRA	237	697513.5	0.11696346	2.68214483	10	0.06557771	0.19554603	0.00770430
HOU	144	654154.5	0.16336441	2.98313986	11	0.01745704	0.09628632	0.00516273
LON	242	1464828.0	0.11234183	2.63515755	10	0.08498898	0.20037151	0.01560551
NYC	179	1524349.5	0.15755445	2.70840243	11	0.06928349	0.16058354	0.01247061
PAR	250	1023424.5	0.08915663	2.68767275	10	0.09342038	0.18028732	0.01272929
TBO	2	234.0	1.00000000	4.58319514	12	0.00000000	0.00012303	0.00007989
WAW	55	86836.5	0.45858586	3.24350470	11	0.00155695	0.07519898	0.00111978
ZVA	1	19.0	0.00000000	7.57517966	15	0.00000000	0.00000000	0.00009497

The complete table with all the airports can be retrieved from:
https://github.com/raccamateo/complexnet_structural_descriptors/blob/main/CN_A1_RACCA_Results/airports_node_descriptors.csv

Insights

Degree and Strength: the degree of an airport is indicative of its connectivity.
Paris (PAR), with a high degree of 250, is a central node in the network. It's striking when compared to TBO, which has a degree of only 2, suggesting it's a peripheral airport.
Strength amplifies this story. ATL, with a strength of 112960.5, is not just a hub in terms of routes but also in terms of traffic volume, dwarfing airports like ZVA, which has a strength of 19.0, indicating minimal traffic.

Clustering Coefficient: clustering coefficients range from 1.0 for TBO, meaning every possible triangular connection is fulfilled, to 0.0 for ZVA, indicating no interconnectivity between neighboring nodes. CHI's clustering coefficient (0.1341) suggests a low probability that two airports connected to Chicago are also connected to each other, while WAW, with a clustering coefficient of 0.4585, reveals a regional tight-knit structure.

Average and Maximum Path Length: average path lengths vary, with ZVA at a stark 7.57, hinting at its isolation within the network's structure.
AMS boasts a path length of 2.7313, indicating its efficiency in connecting to any other airport with minimal steps.
Maximum path lengths paint a similar picture; ZVA has the highest at 15, suggestive of its remote connectivity within the network.

Betweenness Centrality: betweenness centrality captures an airport's influence over the flow of traffic through the network.
LON scores high at 0.0849, making it a crucial node for routing.
ATL's betweenness centrality is 0.0249, reinforcing its role as a significant transit point in the network, which, if disrupted, could significantly impact the network's efficiency.

Eigenvector Centrality and PageRank: eigenvector centrality gives us insight into the influence of an airport within the context of the network's overall connectivity. FRA, with an eigenvector centrality of 0.1955, is not just a well-connected airport but is also linked to other well-connected airports.

PageRank corroborates these findings, with PAR scoring 0.0127, indicating its structural importance and resilience within the network due to its robust connections.

Concluding Insights

Delving into the numerical data, we can infer that:

- ➔ Major Hubs (PAR, LON, FRA, AMS): These airports are not just transit hubs; they are integral to the network's robustness. Their high degree and strength point towards their role in maintaining the integrity and efficiency of the network. They serve as key nodes that keep the network tightly integrated and functional.
- ➔ Peripheral Airports (TBO, ZVA): These represent the other extreme, being less connected and less central. They are indicative of the network's reach but also highlight the potential risks of network disintegration if major hubs are compromised.
- ➔ Intermediate Nodes (CHI, ATL, BCN): They exhibit a mix of centrality and clustering that highlights their role as important but not dominant nodes within the network. They serve as regional hubs, connecting less central airports to the core of the network.

Other key inferred approaches: from a design perspective, the data reveals the need for strategic investment in both the major and peripheral airports to enhance the network's resilience and efficiency.

Ensuring robustness in major hubs and improving connectivity for peripheral airports could be crucial for the network's overall health.

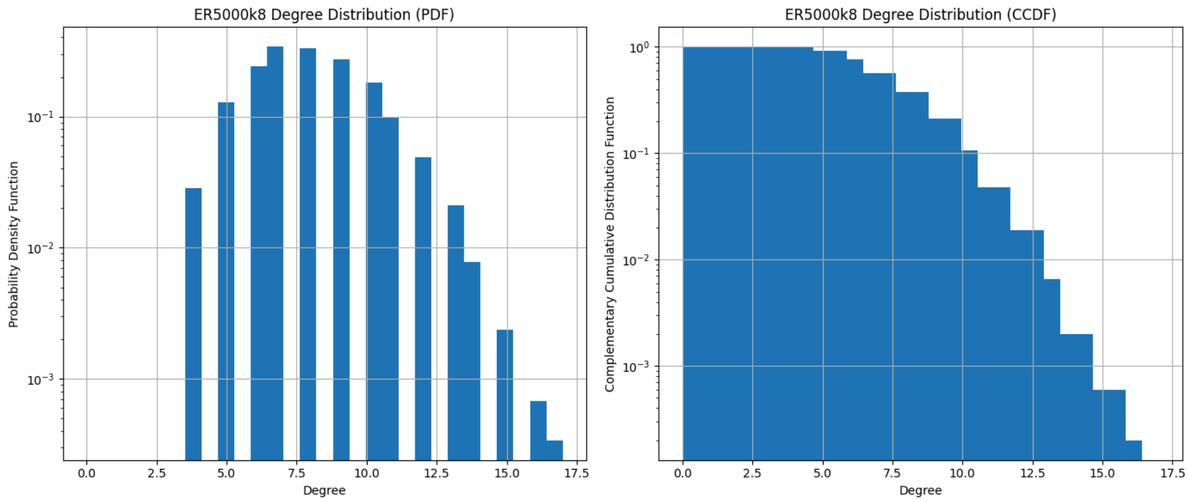
Another key could be for airport managers and policymakers, these metrics can inform critical decisions regarding capacity planning, route scheduling, infrastructure investment, and emergency planning. Recognizing the importance of each airport's role could lead to more effective management and development strategies.

In summary, the data retrieved provides a quantitative backbone for our understanding of the network's structure. The analysis extends beyond just identifying key airports; it offers a roadmap for enhancing the network's robustness and efficiency by informing targeted investments and strategic initiatives.

c) Plot the histograms of the degree distributions (PDF, probability distribution function) and the complementary cumulative degree distributions (CCDF) for the following networks:

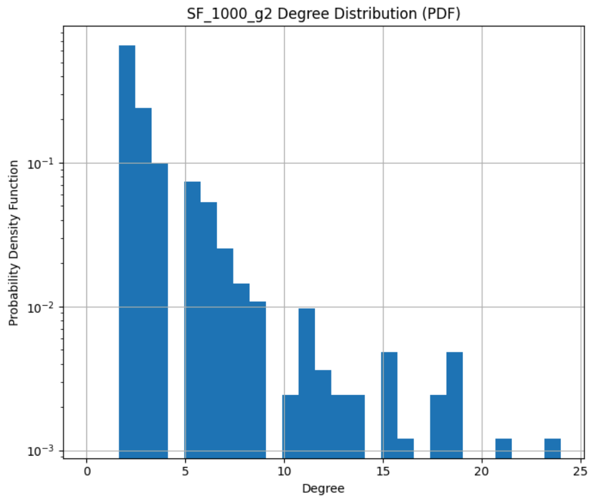
- model/ER5000k8.net
- model/SF_1000_g2.7.net
- model/ws2000.net
- real/airports_UW.net
- real/PGP.net

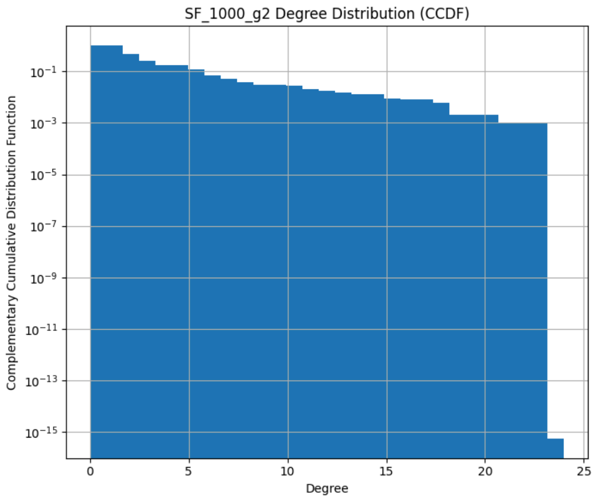
ER5000k8 Network



PDF: the ER5000k8 network's PDF presents a bell-shaped distribution, typical of Erdős-Rényi random networks. The peak of the distribution suggests most nodes have a similar degree, which is characteristic of the random nature of connections and the absence of a preferential attachment mechanism.
CCDF: the CCDF's swift decay indicates that there are almost no nodes with a degree significantly higher than the average, confirming the lack of hubs within the network. This random structure implies that the network is generally robust against random attacks, as there are no critical nodes whose removal would disproportionately disrupt the network.

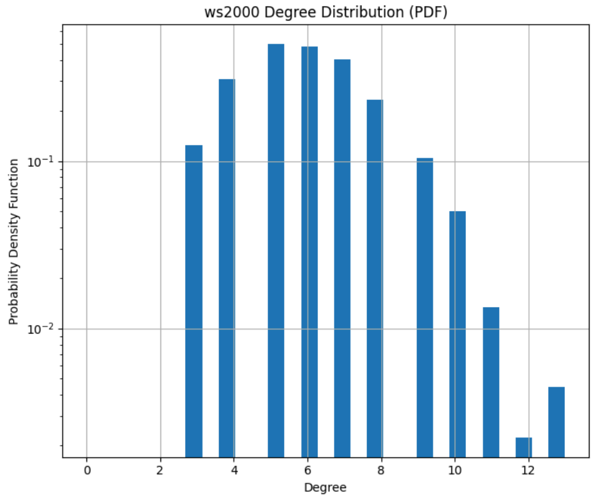
SF_1000_g2 Network

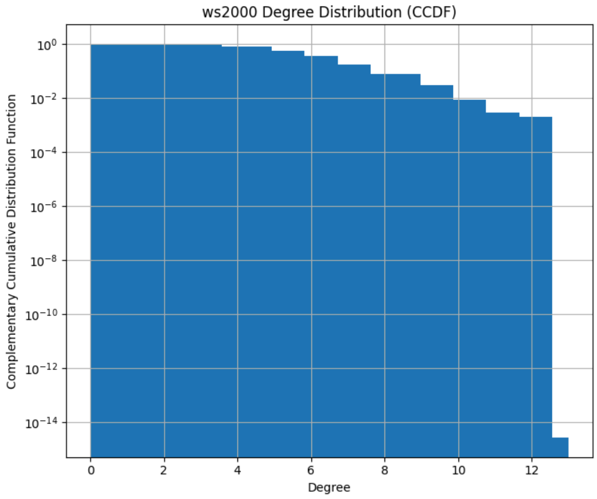




PDF: the scale-free network's PDF showcases the distinct presence of hubs within the network. Nodes with very high degrees are much more prevalent in this model than in random networks, highlighting the central role of hubs in maintaining network connectivity and integrity.
CCDF: the CCDF for the SF_1000_g2 network emphasizes the disparity between the nodes, with a significant fraction of nodes having a high degree of connectivity. This disparity suggests a vulnerability to attacks on the network's hubs but also highlights the efficiency of the network structure, where a small number of nodes ensure extensive network reach.

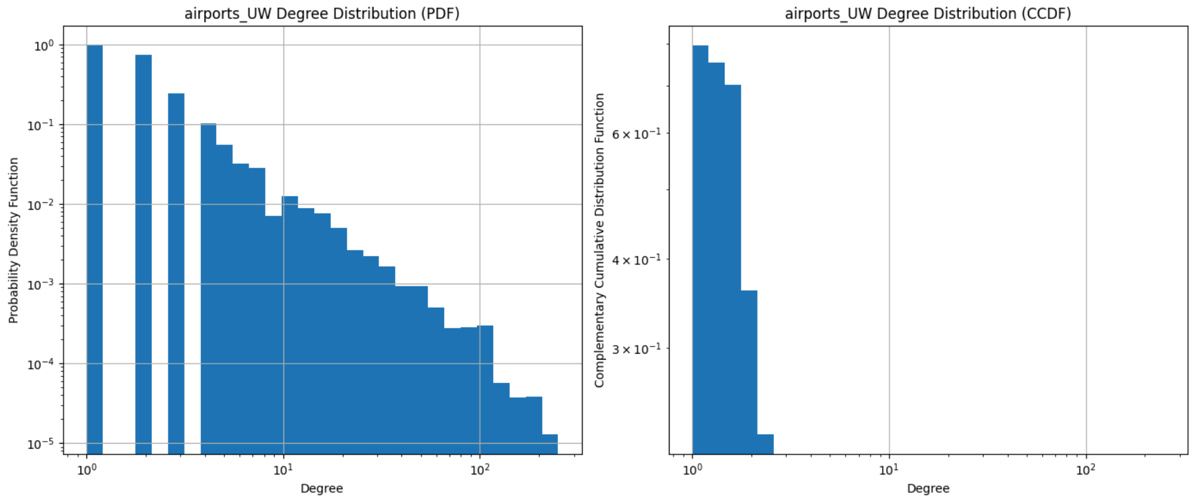
ws2000 Network





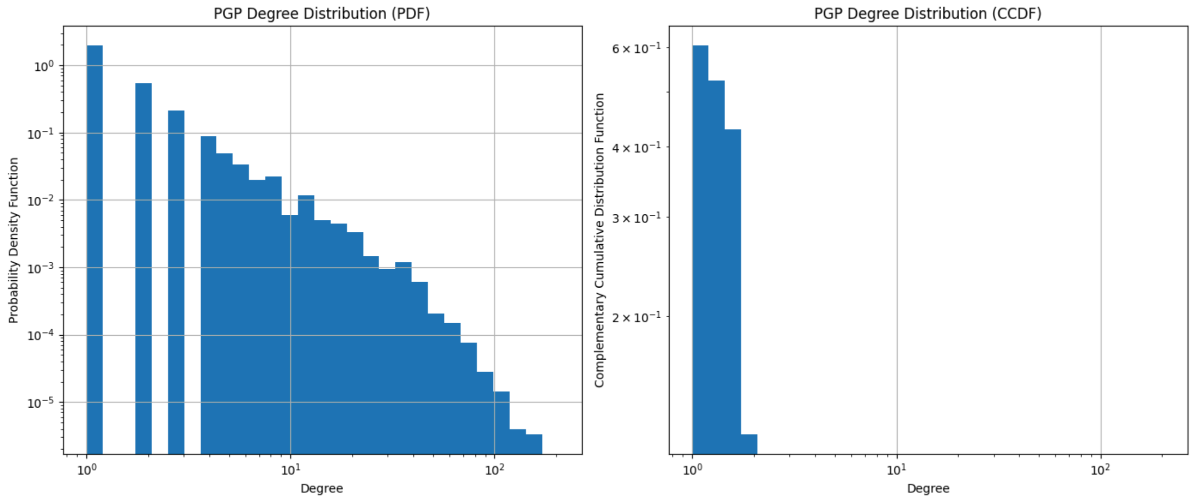
PDF: the ws2000 network exhibits a more bell-shaped distribution, typical of the small-world networks generated by the Watts-Strogatz model. The even distribution of node connectivity signifies a resilient structure against random node failures. The homogeneity in degrees suggests that the network can maintain connectivity even when individual nodes are removed.
CCDF: the near-linear decline on a semi-log scale for the CCDF indicates a uniformity in node connectivity. The small-world nature of the network is revealed through the relatively even distribution of node degrees, which is conducive to swift information transfer and robustness against random failures.

Airports Network



PDF: mirroring the PGP network, the airports network displays a scale-free distribution. However, the slope appears less steep, indicating a more moderate contrast between the highly connected hubs and less connected regional airports. Such a structure suggests efficiency in global connectivity, yet it could lead to potential bottlenecks at the busiest hubs.
CCDF: the CCDF shows a considerable presence of airports with a high degree of connectivity. The criticality of these hubs could imply that disruptions at any major airport could have significant cascading effects on global air traffic, emphasizing the importance of contingency planning for these nodes.

PGP Network



PDF: the PGP network displays a classic scale-free pattern on a log-log scale. The steep downward slope indicates a small number of nodes with high connectivity, contrasted with a vast majority having relatively few connections.
This hierarchy suggests the network's reliance on key nodes for its integrity and function. High-degree nodes could represent well-trusted users within the PGP web, essential for maintaining a cohesive network.
CCDF: the long tail in the CCDF plot for the PGP network confirms the presence of influential nodes that disproportionately ensure network connectivity. The slow decay illustrates how a select few nodes act as critical junctions in the flow of trust and information, with implications for network security and resilience.

The analysis of these networks illuminates several critical points:
Scale-Free Networks (PGP and Airports): the presence of hubs in scale-free networks indicates that while these nodes bolster the network's efficiency and reach, they also present single points of failure.
The resilience strategies for these networks must account for these vulnerabilities, perhaps through the development of redundant connections or increasing the capacity of regional nodes to reduce dependency on hubs.

Small-World Networks: the ws2000 network demonstrates the importance of each node due to the homogeneity in connectivity. Resilience in such networks is inherently high against random failures, but they may be susceptible to systematic failures that exploit the short paths through the network.

Random Networks : the ER5000k8 network lacks the pronounced hubs of scale-free networks, leading to a more evenly distributed network load. These networks may not have the efficiency of scale-free networks in terms of shortest paths, but their uniformity can confer resilience against targeted attacks.

Understanding these properties is crucial for various applications, such as epidemiology, where scale-free networks might allow for rapid disease spread due to their hubs, or communication networks, where small-world properties can facilitate swift information transfer.

In designing resilient infrastructures, recognizing the role of network topology can guide strategies for enhancing connectivity, robustness, and overall system health.