The Role of Network Topology in Error and Attack Tolerance: Implications from a Case Study on Epidemic Spread

Mia Nascimben

November 2024

Abstract

This project has been inspired by the paper *"Error and attack tolerance of complex networks"* by Barabasi and co.[1]. It aims to study how network connectivity and information spreading respond to errors and attacks, with a focus on the role of network topology. The investigation was conducted in several stages: first, network connectivity was assessed by examining the graph diameter. Next, to better support the initial results, an analysis of changes in network structure was performed, evaluating the size of the giant component and the formation of isolated clusters. Finally, the impact on the network's capacity to spread information was evaluated through the simulation of an epidemic, governed by the SIR model. The epidemic's progression was measured using four indicators: duration, total number of infected nodes at the end of the epidemic, peak infection level, and the time-step at which the peak occurs.

Analyses were conducted on three networks: the Erdős-Rényi (ER) network, the Scale-Free (SF) network, and the global air traffic network. Results show that while SF networks are robust to random errors, they are highly vulnerable to targeted attacks due to their reliance on a few central hubs. Conversely, the homogeneous degree distribution of ER networks makes them less affected by specific attacks or errors[1]. Nonetheless, the ER network shows similar, though delayed and less intense, responses to those observed in the SF network under attacks and errors. In line with the hypothesis, the real-world air traffic network aligns with the characteristics of a Scale-Free topology.

Regarding the epidemic simulation outcomes, both types of node removal lead to a decline in epidemic metrics, with the decline being more severe in the case of targeted attacks. The only metric that increases is the time of the infection peak, which is delayed under moderate levels of attack, due to the greater number of steps required for the infection to spread from the infected node to the nearest remaining hubs. In the air transportation network, closing the largest 10% of global airports results in a complete halt of infections. This restriction confines the disease within a limited topological area, safeguarding the broader network from further spread.

The Python code used for running simulations and obtaining the results is freely available on Github at the url:

https://github.com/mianascimben/network-project

1 Introduction

Understanding the resilience of networks and their ability to propagate information is a critical area of research in network science, with broad implications for communication systems, infrastructure, and epidemiology. This study explores how different network topologies respond to errors and targeted attacks, focusing on their effects on connectivity and information dissemination.

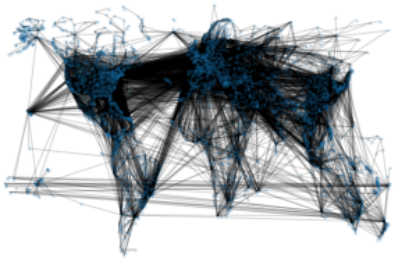


Figure 1: Scheme of the global air traffic network. The blue points represent airports, connected by dark edges.

The investigation encompasses an analysis of three network types: Erdős-Rényi random networks, Scale-Free networks, and the real-world global air transportation network. By assessing metrics such as graph diameter, the size of the giant component, and the fragmentation into isolated clusters, the study evaluates the networks' structural integrity under error and attack scenarios. Furthermore, the study incorporates simulations of epidemic spread using the SIR model to quantify how disruptions affect disease transmission dynamics.

The findings reveal stark contrasts between the robustness of different network topologies. While SF networks are highly resilient to random errors, they are acutely vulnerable to targeted attacks due to their reliance on a small number of hubs. Conversely, ER networks, with their homogeneous degree distribution, exhibit more uniform responses to both errors and attacks but lack the structural efficiency of SF networks. The global air transportation network mirrors SF behavior, demonstrating its reliance on major hubs for connectivity.

Through this analysis, the study highlights the interplay between network topology and resilience, offering insights into mitigating risks and optimizing the design of robust systems.

2 Methods

Here is a brief presentation of the networks analyzed in this study:

Erdős-Rényi ER networks are random graphs created by starting with a set of N nodes and connecting each pair of nodes with a probability p, independent of other connections. The result is a network where the degree distribution follows a binomial distribution, which approximates a Poisson distribution for large N. Most nodes in an ER network have a similar number of connections, and the network has no significant hubs.

Scale-Free SF networks are characterized by a power-law degree distribution,

$P(k) \propto k^{-\gamma}$, meaning that a minority of nodes (hubs) differ in their number of connections by several orders of magnitude compared to most nodes. This structure arises naturally in many real-world systems (e.g., the internet, social networks) through a process called preferential attachment, where new nodes are more likely to connect to already well-connected nodes.

Global Air Transportation A real-world network where nodes represent airports, and edges correspond to airlines connecting them. This study does not consider weighted links[1]. The dataset is available on Kaggle ([2]). Since the corpus of the dataset extends beyond the aims of this study, the data cleaning process involved selecting only the information about airline source-to-destination airports and removing all routes and airports that are not successfully registered. The cleaned dataset is visualized in Figure 1.

Tolerance analysis

Errors and *attacks* refer to the strategies used to determine which nodes have to be removed; the key difference is that errors involve the random removal of nodes, while attacks deliberately target and remove the most connected nodes in a network.

Changes in network connectivity have been explored by examining the behavior of network diameter, d, after introducing a small fraction, f, of errors or attacks. d measures how topologically near two nodes are: the smaller d is, the shorter the shortest path between them. In this report, d follows the definition of Réka Albert and co.[1], which is the average length of the shortest paths between any two nodes in the network:

$$d = \sum_{i,j \in V i \neq j} \frac{a(i,j)}{N(N-1)}$$

where $a(i,j)$ is the shortest path between the node i and j, V is the set of nodes in the network, and N is the network size.

An attack or an error implies, together with the node, the removal of all its links, which may cause the separation of a cluster of nodes from the main cluster. To investigate further the modifications in network structure caused by the application of errors or attacks, it is useful to look at the variation of both the size of the giant component, S, and the average size of all the connected components except the largest one, $< s >$. These features provide information on the fragmentation process; S explores the number of nodes that become isolated, while $< s >$ reveals their organization into smaller and bigger clusters following fragmentation.

To compare the results from the ER and SF networks, they have been created with the same number of nodes, $N = 1000$, and approximately the same number

[1]The weight could be calculated, for example, from the number of passengers in transit.

of links. For the ER network, the expectation value of the binomial distribution, i.e., the average degree for each node, has been set to $< k >= 4$.

Epidemic simulation

The epidemics have been simulated using the SIR model. Through this model, each individual within the network can be in one of three stages: Susceptible (S), Infected (I), or Recovered (R). Susceptible nodes can be infected only by the infected nodes they are attached to; once infected, they may recover and acquire immunization. The epidemic ends when all the infected nodes move to the recovered stage. The epidemic dynamics are represented by the infective and recovery curves, which correspond respectively to the count of infected and recovered cases over time (Figure 2).

The SIR model has been implemented in the code by creating an array for the nodes' states, $\vec{s}$, where the term s_i refers to the state of the i^{th} node. Each node can be in one of the three stages:

$$\begin{cases} s_i = 1 & \text{if infected} \\ s_i = 0 & \text{if susceptible} \\ s_i = -1 & \text{if recovered} \end{cases} \tag{1}$$

The spreading of the infection, as well as the recovery process, is governed by probability: at each time step, t, an infected node can transmit the disease with a probability μ per unit of time, and it can recover with a rate ν[2].
The transmission and recovery processes in a given network G are described as follows: during the transmission process, an infected node i considers only its *susceptible neighbors*. These neighbors are identified as the nodes $j \in N_G(i)$ (where $N_G(i)$ represents the neighborhood of i in G) that satisfy the condition:

$$s_i + s_j = 1, \quad \text{with } s_i = 1.$$

For each susceptible neighbor j, the probability of infection transmission from i to j is modeled using a *binomial distribution*: $\text{Bin}(1, \mu) = \mu$. Once the infection process is completed, the recovery phase begins. Each infected node has a probability of recovering, which is also modeled using a binomial distribution: $\text{Bin}(1, \nu) = \nu$. To prevent a node from being infected and recovering within the same time step, nodes that have just been infected during the current time step are excluded from the recovery process.

For all the epidemics simulated over different networks, the values for probabilities μ and ν have been kept constant, as well as the number of infected cases at the first time step. In particular, $\mu = 0.2$, $\nu = 0.05$, and the number of initially infected equals 1. As the simulation of errors, attacks, and epidemics

[2]The value $R_0 = \frac{\mu}{\nu}$ is called the *basic reproduction number* and represents the average number of infections each node generates before recovering. In the simplest epidemic model, an outbreak is strong enough to start if $R_0 > 1$. However, it is important to note that the above condition can be violated due to the probabilistic nature of the infection spreading.

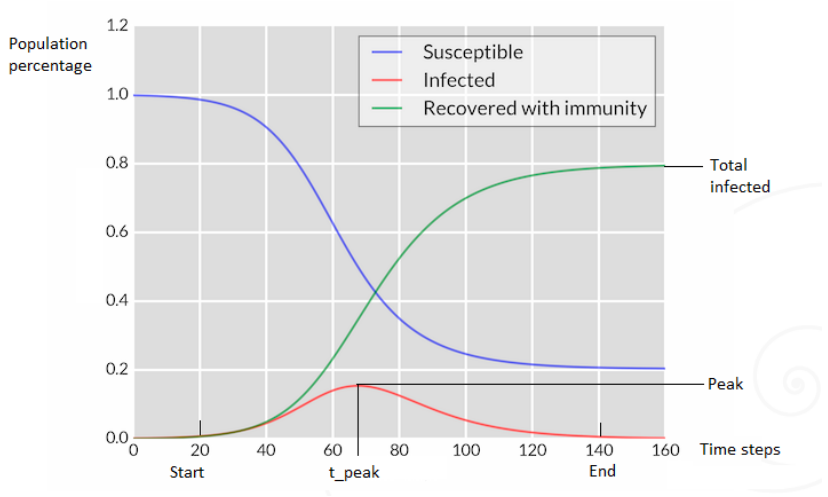


Figure 2: Scheme showing the dynamics of epidemic features given by the SIR model.

relies on random processes, to obtain the same results shown in the following report, ensure that the seed for reproducibility is set to 102.

The reasons that lead network structure to influence epidemic dynamics are researched and discussed through the analysis of epidemic features, which are extracted from the epidemic dynamics, as shown in Figure (2).
To minimize the effects of randomness, epidemic properties were analyzed by averaging results across multiple simulations. Specifically, 100 epidemic simulations were run on each network after each structural alteration caused by errors or attacks. Key information about the epidemic properties was extracted from each simulation and averaged over all runs.
The analyzed features are:

Duration It has been examined whether and how the increasing frequency of errors or attacks, along with the network's topology, influenced the duration of the epidemics. Although a threshold could be set for identifying the starting and ending time for calculating the epidemic duration, in this study, the trivial case where $threshold = 0$ has been considered.

Final infection cases The total number of infection cases at the end of the epidemic simulation measures the efficiency of disease propagation. This value is obtained by summing the infected and recovered portions of the population at the final time step. Given the transmission and recovery probabilities, a high final infection count indicates that the disease (or, in general, the information) has propagated widely. However, this metric alone does not provide insight into the structural dynamics of how the disease spreads within the network.

Infection peak and its timing To better understand the role of network structure in the infection process, the behavior of the epidemic peak and its corresponding time step are analyzed. Structural changes in the network can cause a delay or advance in the epidemic peak. For instance, if the first infected node is

a hub, the likelihood of a large number of infections in a short time is high; this means the infective curve is expected to show a relevant peak after a few time steps. However, if the first infected node is poorly connected, the epidemic will take some time to trigger and reach its peak. The results will report a delayed peak in comparison to the previous case. Additionally, the peak value is likely to be dampened as the delay in the outbreak has led to the recovery process starting, reducing the number of infective cases.

3 Results

Connectivity

Figure 3 shows the results for the network diameter (d) of the three graphs under study as a function of error and attack frequencies. The plots illustrate how different network topologies respond to node removals and highlight the contrasting effects of random failures (errors) and targeted attacks.

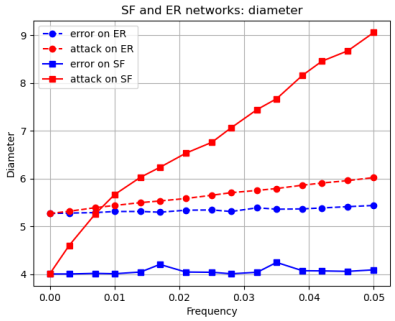
In Figure 3a), the diameter of the ER network demonstrates strong tolerance to small fractions of both attacks and errors. The ER network's homogeneous degree distribution ensures that all nodes contribute similarly to the network's structure. As a result, the removal of even the most connected nodes has little effect, as alternative short paths are readily available. However, as the frequency of nodes removed increases, ER graph continues to have a strong tolerance to errors, while, under attacks, recreates the SF behavior.

In the SF network, d is highly sensitive to attacks while remaining almost unaffected by errors. For these networks, hubs play a critical role in maintaining short paths between nodes due to their high degree. When a hub is removed, all its connections (a significant portion of the total links, given the power-law degree distribution) are eliminated. This results in longer paths between remaining nodes, leading to a rapid increase in the diameter even at low attack frequencies. On the other hand, random errors are unlikely to target hubs, as these nodes constitute a minority of the network. As a result, the diameter remains largely unaffected under errors.

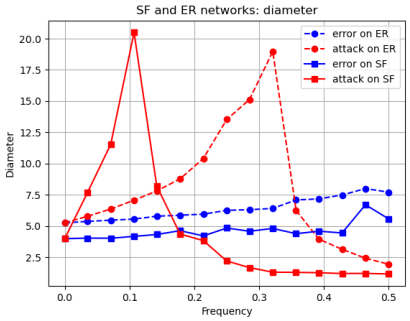
Figures 3c) and 3d) present the results for the real-world air traffic network, which exhibits trends consistent with SF networks. From a practical perspective, these results underline the need to develop robust protocols to protect critical hubs in communication networks against targeted attacks, while also highlighting the potential to exploit network vulnerabilities to limit the spread of diseases, misinformation, or digital viruses.

As the frequency of node removal increases, both the ER and SF networks eventually undergo fragmentation, leading to a decline in d. This occurs as highly connected nodes are removed, causing the network to split into isolated clusters. The diameter d of the entire network is then calculated as the average diameter of these isolated clusters:

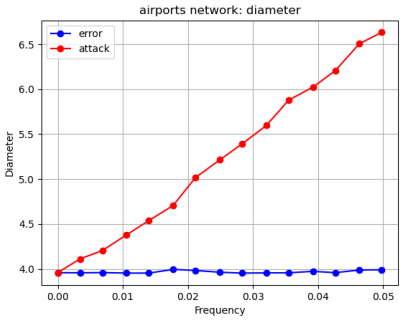
$$d = \frac{\sum_{m=1}^{M} d_{c,m}}{M},$$



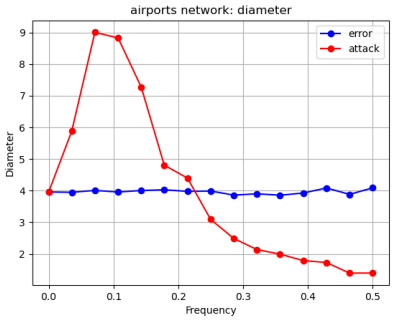
(a) ER & SF



(b) ER & SF



(c) Air-Traffic Graph



(d) Air-Traffic Graph

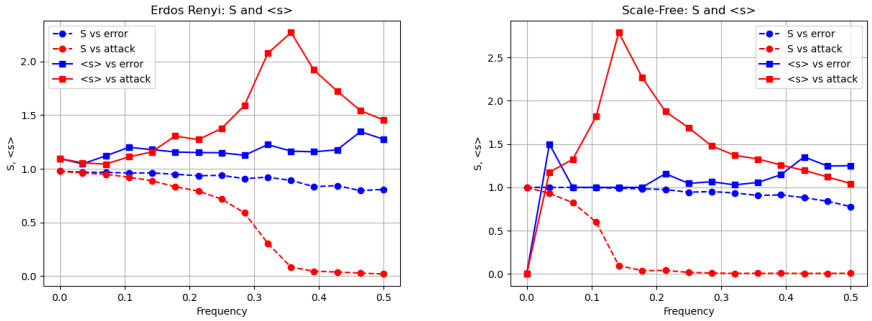
Figure 3: The plots show the tolerance of the diameter (d) under increasing fractions of attacks (red lines) and errors (blue lines) for the ER (circles) and SF (squares) networks. The parameters for building the graphs have been set to $\langle k \rangle = 4$, $N = 1000$. On the left, a zoomed view of the frequency axis is shown.

where M is the total number of isolated clusters and $d_{c,m}$ is the diameter of the m-th cluster. Since smaller clusters have shorter diameters, d decreases as the fragmentation progresses. This fragmentation process and its implications are discussed further in Section 3.

Fragmentation

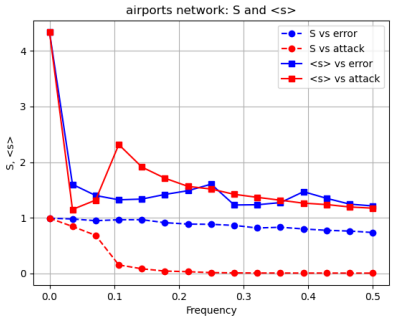
The analysis of the size of the giant component (S) and the average size of the smaller components ($\langle s \rangle$) under node removal provides valuable insights into the fragmentation process of different network types. The results, shown in Figure 4, highlight how network topology and removal strategies influence resilience to disruption.

In the ER network (Figure 4a)), a phase transition characteristic of random graphs is observed. The fragmentation of the giant component is closely tied



(a)

(b)



(c)

Figure 4: The plots show the trend of the giant component size (circles), S, and the averaged sizes of all the other components (squares), $< s >$, under increasing the fraction of removed nodes for errors (blue lines) and attacks (red lines). a) shows the results for the ER network, b) for the Scale Free graph and c) for the air traffic network. The parameters of the networks are the same as Figure 3.

to the average degree $\overline{k}$, which represents the average number of connections per node[3]. As predicted by network theory, the giant component disintegrates when $\overline{k}$ drops below the critical threshold of 1:

$$\begin{cases} S > 0 & \text{for } \overline{k} > 1, \\ S \approx 0 & \text{for } \overline{k} < 1. \end{cases}$$

At this critical point, where $\overline{k} = 1$, the network fragments into smaller clusters, leading to a peak in $\langle s \rangle$. Random errors, which remove nodes indiscriminately, reduce $\overline{k}$ more gradually due to the homogeneous degree distribution of the ER network. This delays the onset of fragmentation compared to targeted attacks, which prioritize the removal of the most connected nodes, causing a sharper decline in $\overline{k}$ and a faster transition to network collapse.

In the SF network (Figure 4b)), fragmentation occurs at a lower critical frequency (f_c) than in the ER network. This is due to the inhomogeneous degree distribution of SF networks, where connectivity depends heavily on a small number of hubs. Attacks on these hubs rapidly decrease $\overline{k}$, dismantling the giant component with only a small fraction of nodes being removed. Conversely, random errors rarely affect hubs and therefore result in a slower decline in $\overline{k}$ and delayed fragmentation.

The real-world air traffic network (Figure 4c)) mirrors the behavior of SF networks due to its scale-free topology. The initial high value of $\langle s \rangle$ is attributed to the presence of a densely populated isolated cluster, which skews the average size. As nodes are removed, this cluster suddenly fragments into numerous smaller components, leading to a decrease in $\langle s \rangle$. These results underscore the role of hubs in maintaining the structural integrity of the air traffic network and its susceptibility to targeted disruptions.

Epidemic

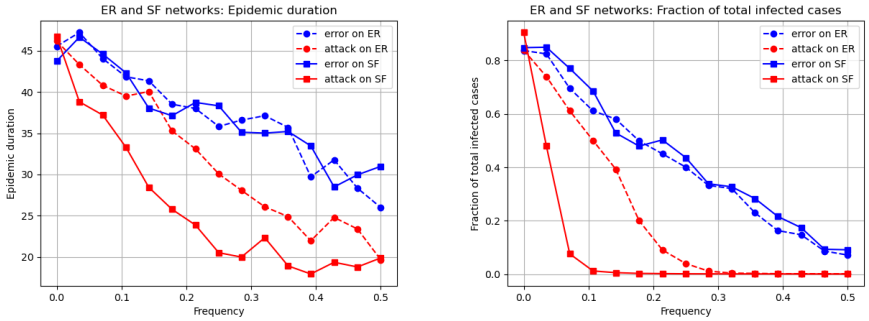
The results of the epidemic analysis for the ER and SF networks are shown in Figure 5[3]. The same analysis conducted on the real-world air traffic network produced the results in Figure 6, which align with those observed for networks with a SF topology.

For all networks, the epidemic's progression is strongly influenced by the removal strategy (errors vs. attacks) and the network's structural properties. The four metrics analyzed — epidemic duration, total number of infected nodes, infection peak, and the time of the infection peak — provide a comprehensive view of how disruptions affect epidemic dynamics.

Epidemic duration (Figure 5a)): The epidemic duration decreases steadily with increasing node removals. Errors produce a similar reduction in both ER and SF networks. Attacks, however, lead to a more drastic decrease in epidemic duration, especially in SF networks, due to the targeted removal of hubs, which are critical for disease propagation. Interestingly, the ER network, despite its homogeneous structure, exhibits a delayed but similar response under attacks, replicating the trends observed in SF networks.

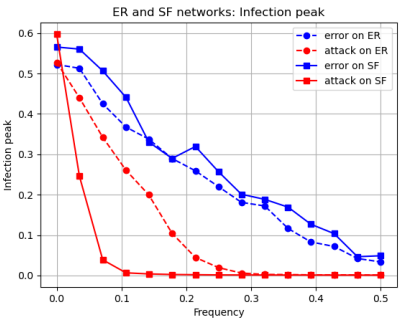
Total number of infected nodes (Figure 5b)): The total number of infected nodes during the epidemic is highly dependent on the fraction of removed nodes. Errors produce a gradual, almost linear decrease in the total infections for both ER and SF networks. This behavior reflects the uniform reduction of connectivity in the network. In contrast, attacks cause a sharp decline in total infections, particularly in SF networks, where the removal of a small fraction of nodes ($f \approx 0.1$) is sufficient to reduce infections near zero. This rapid drop highlights the vulnerability of SF networks due to their reliance on a small number of highly connected nodes. The ER network shows a similar decline under attacks but at a higher removal frequency ($f \approx 0.28$) compared to SF.

[3]The parameters for the creation of the artificial networks are the same as those used in the previous simulations.

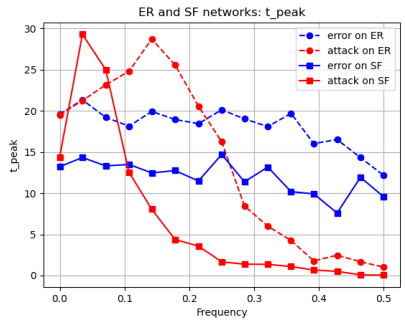


(a) Epidemic duration

(b) Total infected



(c) Peak

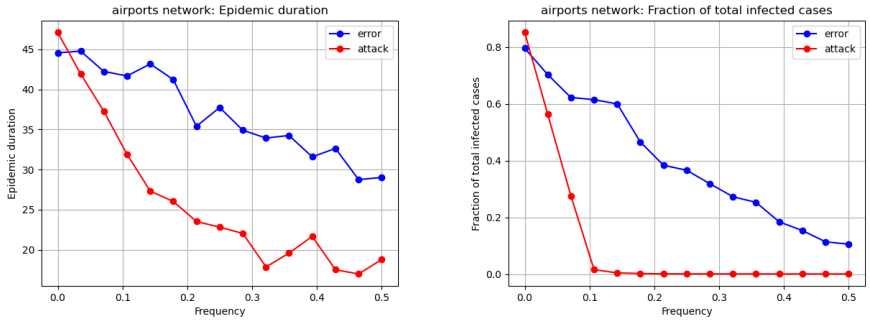


(d) t_{peak}

Figure 5: The plots show how epidemic features are affected by increasing the fraction of removed nodes (errors and attacks are represented by blue and red lines respectively) on ER and SF networks. From a) to d) plots respectively display the average behavior of the duration of the epidemic, the number of total infected nodes, the value of the infection peak and the time at which this latter happens. The epidemic parameters have been set as following: $\nu = 0.05$, $\mu = 0.2$, the total number of time-steps equal to 50, and the starting number of infected equal to 1.

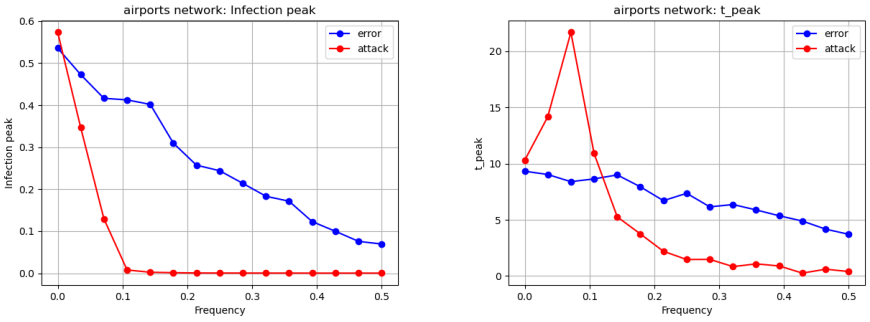
Infection peak (Figure 5c)): The infection peak behaves as expected, given its relationship to the total number of infections. Since the peak represents the maximum fraction of infected nodes at any time during the epidemic, it decreases proportionally with the total number of infected nodes.

Timing of the peak (Figure 5d)): One of the key observations is that the high connectivity of SF networks, in the absence of node removals, allows for a faster spread of the epidemic compared to ER networks. Specifically, the infection peak in SF networks is reached approximately five steps earlier than in ER networks, as seen from the first data points (for ER graph $t_{peak}(f = 0) \approx 19$ steps, while for SF network $t_{peak}(f = 0) \approx 14$ steps). This behavior is directly linked to the presence of hubs in SF networks, which facilitate rapid disease



(a) Epidemic duration

(b) Total infected



(c) Peak

(d) t_{peak}

Figure 6: The plots show how epidemic features are affected by increasing the fraction of removed nodes (errors and attacks are represented by blue and red lines respectively) on the global air traffic network. From a) to d) plots respectively display the average behavior of the duration of the epidemic, the number of total infected nodes, the value of the infection peak and the time at which this latter happens. The epidemic parameters have been set as following: $\mu = 0.2$, $\nu = 0.05$, total number of time-steps equal to 50, and the starting number of infected equal to 1.

transmission once infected.

In SF networks, the rapid propagation of the epidemic is driven by the high degree of the hubs, which ensures that the first infected node is likely to be topologically close to a hub (indicated by the network's small diameter). Once the hubs are removed, the network diameter increases, meaning that it takes more steps for the infection to reach a sufficiently connected node to trigger the peak. If attacks continue, the number of connections in the network decreases to the point where new infections drop below the number of recoveries, resulting in an earlier and smaller infection peak.

For ER networks, the delayed t_{peak} can be explained using a similar framework as for SF networks. However, the difference lies in the homogeneity of

the ER network, where connectivity is distributed more uniformly. As a result, t_{peak} starts to decrease only at higher attack frequencies in ER networks compared to SF networks. This is because the removal of the most connected nodes does not significantly impact the overall connectivity of the ER network, which maintains a relatively small diameter. This allows the infection to continue spreading efficiently even after the removal of key nodes.

Random errors have little effect on t_{peak} in either network, highlighting the robustness of both ER and SF networks to random disruptions.

4 Conclusion

This study highlights how network topology fundamentally influences the resilience of networks to errors and attacks, as well as the dynamics of information or disease propagation. The analysis of both artificial and real-world networks reveals critical insights into the interplay between network structure and robustness.

Scale-Free networks, characterized by their reliance on a small number of highly connected hubs, exhibit a dual nature: they are highly resilient to random errors but acutely vulnerable to targeted attacks. The removal of a small fraction of hubs rapidly dismantles the network, leading to a sharp fragmentation of the giant component and a collapse in connectivity. This vulnerability extends to the propagation of epidemics, where attacks cause a drastic reduction in epidemic metrics, such as duration, total infections, and infection peaks. However, errors have a minimal impact, reflecting the robustness of SF networks to random disruptions.

Conversely, Erdős-Rényi networks, with their homogeneous degree distribution, demonstrate more uniform responses to both errors and attacks. They are less reliant on individual nodes, which delays fragmentation under targeted attacks. Despite this, at higher attack frequencies, ER networks replicate the behavior of SF networks, showing a sharp decline in connectivity and epidemic metrics.

The real-world air traffic network mirrors the behavior of SF networks, underscoring the importance of hubs in maintaining connectivity. The removal of the largest 10% of airports effectively halts epidemic spread, demonstrating the critical role of key nodes in disease containment. These findings highlight the need for robust strategies to protect hubs in real-world networks while leveraging their vulnerabilities for targeted interventions, such as controlling the spread of diseases or misinformation.

The epidemic simulations further illustrate the cascading effects of network fragmentation on disease dynamics. While random errors lead to gradual declines in infection metrics, targeted attacks cause abrupt disruptions. One key observation is the delayed infection peak under moderate attacks, caused by the increased steps required for infections to propagate through the fragmented network. However, as attacks persist, the infection peak becomes both smaller and earlier, reflecting the inability of the network to sustain widespread propagation.

References

[1] R. e. a. Albert, "Error and attack tolerance of complex networks," *Nature vol. 406,6794*, 2000.

[2] T. Devastator, "Global air transportation network: Mapping the world's air traffic," 2020. Accessed: August 18, 2024.

[3] P. Erd˝os and A. R´enyi, "On random graphs," *Publicationes Mathematicae 6, 290–297*, 1959.