Analysis Of An Animal Food Chain Network

Riccardo Lambertini, Laurea Magistrale in Informatica, 0001104008
Leonardo Marzocchi, Laurea Magistrale in Informatica, 0001103878

1 Introduction

In the intricate vastness of delicate and interconnected ecosystems of nature, every relationship among the creatures that inhabit it holds vital importance. Particularly, the balance between prey and predators is one of the foundations upon which the stability of an ecosystem relies. This equilibrium, shaped by millions of years of evolution, is the result of mutual dependencies and intricate interactions, where each species plays a crucial role in maintaining biodiversity and environmental health.
Our project has focused precisely on this intricate network of relationships between prey and predators, aiming to analyze and understand through network analysis the dynamics that regulate their balance. Each node in the analyzed network represents an animal species, while the links between them indicate feeding interactions: who hunts and who is hunted. Through the analysis of this network, we sought to shed light on the consequences that may arise from disturbances or imbalances in these interactions.

2 Problem and Motivation

Our project aims to address several issues in the context of predator-prey interactions within natural ecosystems. Firstly, we focus on the importance of understanding the balance between prey and predators, fundamental for ecosystem stability. This equilibrium, shaped by millions of years of evolution, is essential for maintaining biodiversity and environmental health. We decided to test this balance by simulating extinctions of various animal species considered more significant than others to observe their impact on the rest of the network.
One of the key issues we intend to address is the identification of keystone species within predator-prey networks. These species play crucial roles in maintaining the structure and functionality of the ecosystem, and their conservation is of paramount importance to ensure the resilience of biological communities. Additionally, we aim to explore network dynamics to better understand the structure and complexity of interactions within ecosystems.
Another objective is to understand the impact of human activities on ecosystems and their predator-prey interactions. In particular, the inclusion of humans as predators in the network represents an interesting feature that raises questions about sustainable management of natural resources and biodiversity conservation.
The importance of addressing these issues lies in the need to preserve ecosystem health and ensure species survival. The balance between prey and predators not only influences biological diversity but also has significant impacts on human communities, including food security and economic stability. Furthermore, understanding the dynamics of ecological networks can provide crucial insights for the development of effective and sustainable conservation strategies.

3 Datasets

Our study is based on the following dataset obtained through the Kaggle website (see here). The dataset contains 205 rows and 16 columns. Before utilizing the data within the dataset to create our network, we decided to observe how this data was structured through a data preparation phase. All operations performed on the dataset, graph construction, application of most measurements, and extinction simulation were executed by code using the Python programming language. Below we provide a link to all our code here. Additionally, for a better understanding of the dataset, we provide its structure below.

Column Name	**Meaning**
Animal	Nome dell'animale
Height	Altezza dell'animale in cm
Weigth	Peso dell'animale in kg
Color	Colore dell'animale
Lifespan	Aspettativa di vita dell'animale (giorni, mesi anni)
Diet	Dieta dell'animale
Habitat	Habitat dell'animale
AllPredators	Lista di predatori per quell'animale
Average Speed	Velocità media dell'animale in km/h
Countries Found	Paesi dove si trova quell'animale
Conservation Status	Stato di sconservazione dell'animale (LC, VU, EN, CR, NT, NE, DD)
Family	La famiglia a cui l'animale appartiene
Gestation Period	Quanto dura il periodo di gestazione dell'animale
Top Speed	Velocità massima dell'animale in km/h
Social Structure	Descrizione del tipo di struttura sociale dell'animale
Offspring per Birth	Numero di figli per ogni parto dell'animale

3.1 Data preparation

3.1.1 Predators

Initially, we noticed that each animal in the "AllPredators" column was associated with at most two predators. This limitation seemed too restrictive to us, so we decided to expand the list of predators by consulting reliable sources for each animal in the "Animal" column to identify all relevant predators. After completing this operation, we were able to more comprehensively represent the predator network for each prey.

3.1.2 Outliers Dropping

We performed a data cleaning phase to optimize our analysis. During this process, we eliminated animals that are already extinct, as our goal was to focus exclusively on animals still present in the ecosystem, where conservation actions could still make a difference in species protection. Additionally, we removed animals that do not have listed predators and those that do not have a defined conservation status, as the lack of such information could compromise the completeness of our results. Finally, we eliminated duplicate animals to avoid distortions in our data. At the end of this phase, we proceeded with the creation of the graph based on the following columns:

Animal	AllPredators	Conservation Status

We used only these columns as they were sufficient to answer the questions of our analysis.

3.2 Build the Predator-Prey Network

For constructing the network, we adopted an approach that allowed us to represent predator-prey relationships as directed edges in a directed graph. This means that each node of the graph represents an animal, while each edge indicates a predation relationship from a predator animal to a prey. Before creating the graph, we saved all animals listed in both the "Animal" and "AllPredators" columns, allowing us to identify and avoid node duplication. In this way, during the graph creation, we simply created new edges between existing nodes, rather than generating duplicate nodes for animals already present in the network. This approach ensured the correct representation of predator-prey relationships within our network, enabling an accurate analysis of ecological dynamics.
At the end of the graph construction, we obtained 324 nodes and 505 edges.

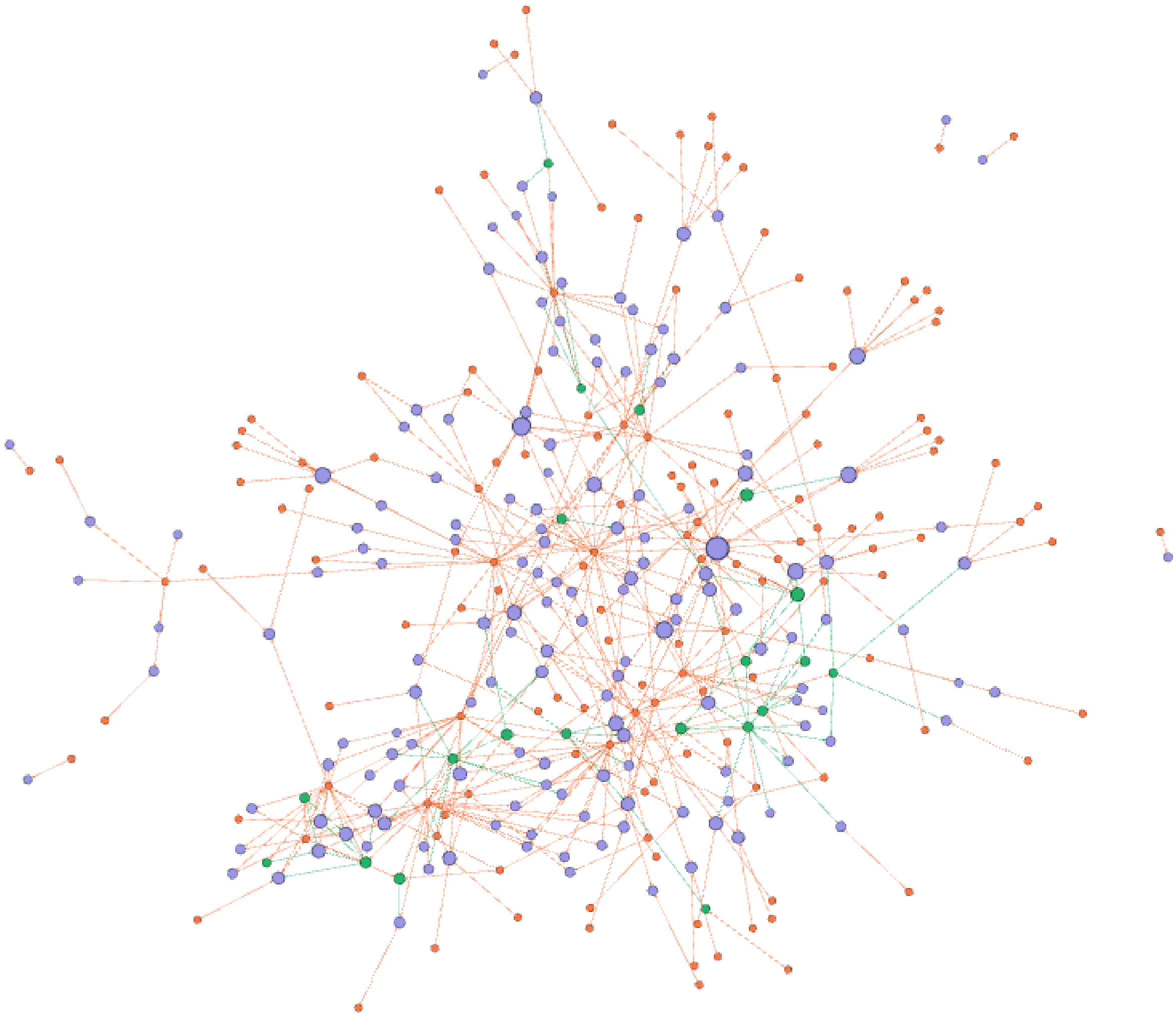

Figura 1: blue = pray, orange = predator, green = both

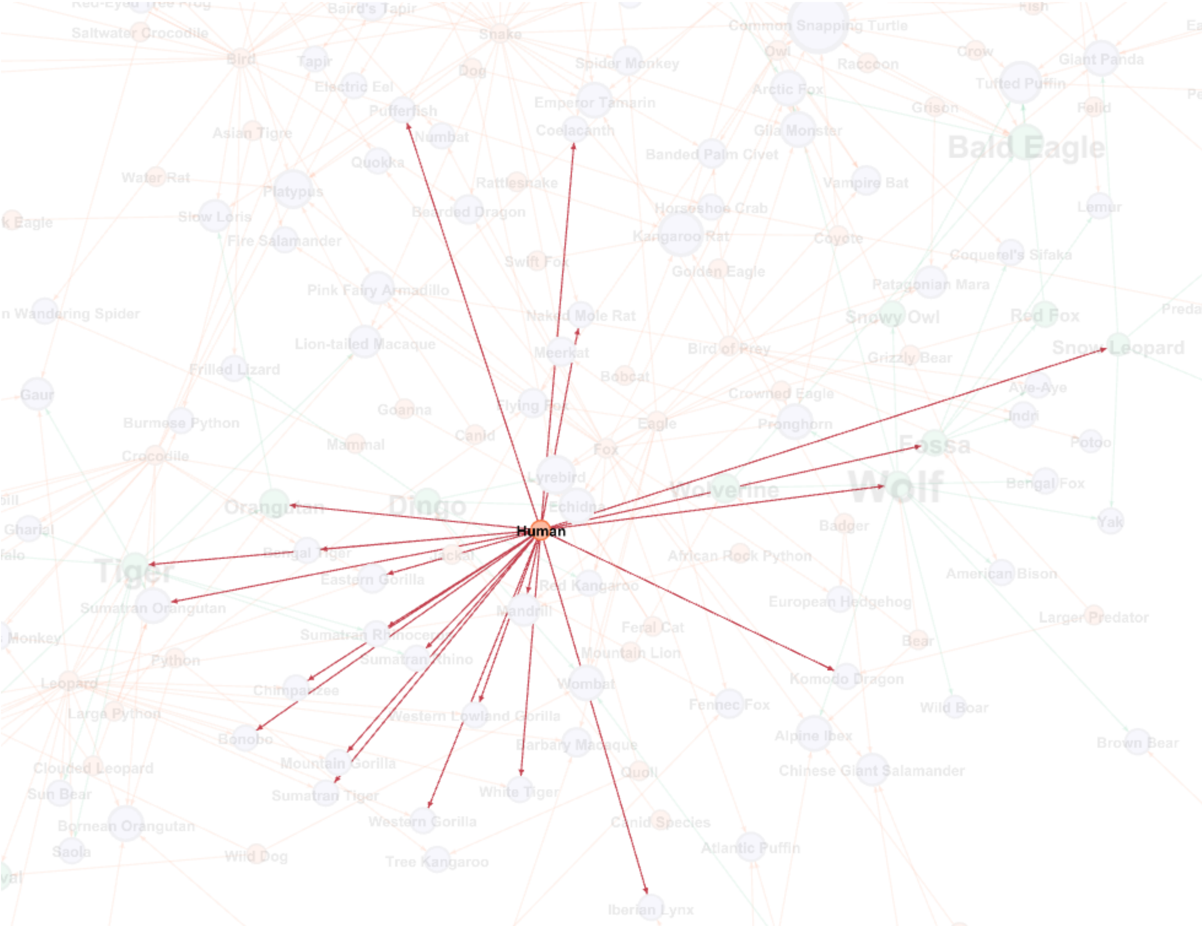

Figura 2: Edges of the Human node

To visualize the complex network of predator-prey interactions we constructed, we used the Gephi software, known for its powerful graph analysis and visualization capabilities. Already at first glance, during the exploration of the network, one can notice an unusual characteristic: the presence of humans among the predators. This unconventional representation may be surprising, but it is important to clarify that the presence of humans as predators is interpreted in terms of hunting relationships compared to other animals they are connected to. This aspect adds a significant element to the understanding of network dynamics, highlighting the role of humans as an integral part of the ecosystem and its interactions.
Furthermore, it is important to note that animals listed only in the "AllPredators" column, without being mentioned in the "Animal" column, do not have incoming edges in the graph. This indicates that these predators have not been identified as prey of other animals within the considered dataset.

4 Validity and Reliability

The validity of our study is quite good and depends on the ability of our dataset to faithfully represent the reality of natural ecosystems. However, it is important to emphasize that our dataset is limited by the number of available data, as there are not many animal species and it was not possible to obtain a very deep food chain. Nonetheless, we have a decent amount of predators for each prey and an up-to-date conservation status as of 20/09/2023.

To ensure good reliability, we only performed a simple data processing phase, adding some predators and removing all animals without predators, those that are extinct, and those without any conservation status. Then, we simply excluded other information that was not relevant to us. Finally, we applied measurements that are widely documented in the fields of mathematics and computer science.

5 Measures and Results

In this chapter, we will list all the measures we have performed in the graph to have a better understanding of the data at our disposal and to interpret them.

5.1 Node Measures

We calculated the degree centrality, which in our case indicates how influential each animal is in the food chain. If an animal has a higher degree centrality, then it is likely to be an important predator.

Tabella 1: Degree Centrality

	Degree Centrality
Snake	0.083591
Leopard	0.077399
Human	0.071207
Bird	0.068111
Orca	0.055727

As we can observe, in our dataset, the nodes with the highest Degree centrality are:

- Snake: By far the most important predator, gathering all scaled reptiles belonging to the suborder Serpentes.
- Leopard: Among the most important felines, they are at the top of the food chain regardless of their habitat.
- Human: The third most important predator.
- Bird: Shown as a predator and gathering prey from the bird family.

During the calculation of Eigenvector Centrality, we defined that due to the structure of our dataset, nodes with the highest Eigenvector Centrality are the nodes of the most important prey. For instance, if we take the "bird" node, it has an Eigenvector Centrality of 0, even though birds are one of the most preyed families. Eigenvector Centrality assigns a centrality score to nodes in a network based not only on the number of links they have but also on the centrality of the nodes they are connected to. This is calculated by solving the eigenvector equation for the adjacency matrix of the network.
However, in some networks like the one used in this project, it's possible that some nodes are not directly connected to other nodes. In these cases, calculating eigenvector centrality may produce eigenvectors with null or zero values for those nodes. "Zero-trailing" refers to the process of handling these null or zero values in eigenvectors.

To overcome this issue, we calculated Katz centrality (performed through the formula 1). Although this measure is specifically used to address "zero-trailing," in our case, despite the addition of bias initially given to each node, the imbalance caused by the absence of outgoing edges for many nodes makes the values of Katz centrality very low.

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{od(j)} + \beta \quad (1)$$

The results obtained with both centralities have "Tufted Puffin" as the node with the highest centrality since it is a node without outgoing edges but with only incoming edges (it is a prey).
We can observe in Figure 3 how the trend of the nodes is to converge to zero very rapidly for

Name	Eigenvector Centrality
Tufted Puffin	0.71958
Arctic Fox	0.69416
Bald Eagle	0.01294
Slow Loris	0.01294
Warthog	0.00054
Gerenuk	0.00045
Hummingbird	0.00036
Grevy's Zebra	0.00027
Wildebeest	0.00027
Wombat	0.00027

Tabella 2: Eigenvector Centrality

Name	Katz Centrality
Tufted Puffin	0.30242
Arctic Fox	0.22392
Common Snapping Turtle	0.19389
Hummingbird	0.15833
Warthog	0.15423
Gerenuk	0.14260
Yellow-Eyed Penguin	0.13918
Bald Eagle	0.13911
Kangaroo Rat	0.12550
Slow Loris	0.12544

Tabella 3: Katz Centrality

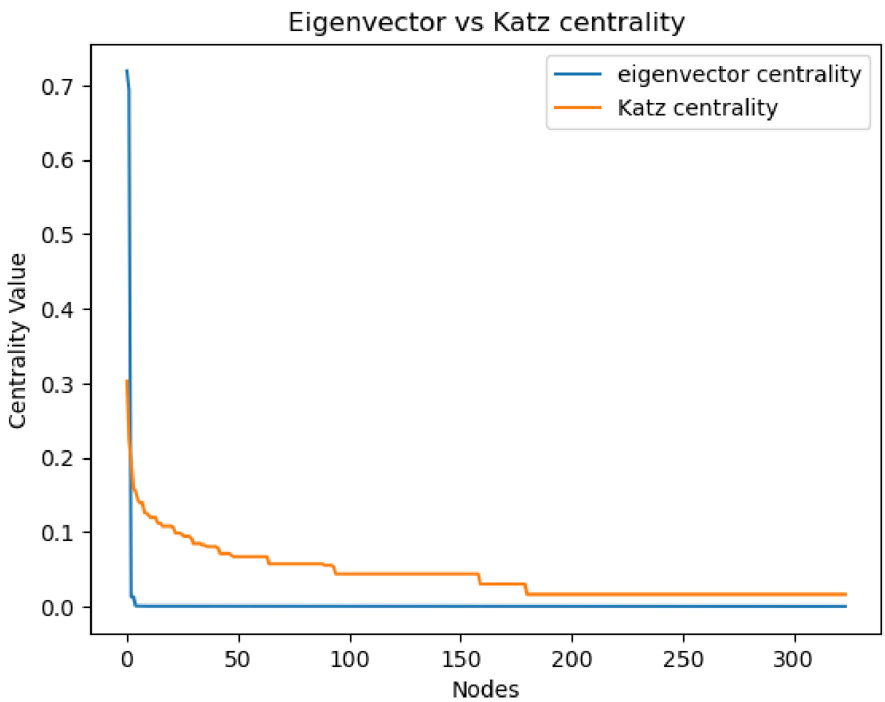

Figura 3: Katz vs eigenvector centrality

both centralities. The difference is that in Katz centrality, no node reaches convergence.

A very different case for our dataset is the Hubs and Authorities Centrality, which allows distinguishing between hub nodes and authority nodes:

- Authorities: These are nodes that are pointed to by many hubs, in our case, a pattern found in prey (a prey is a node with many incoming edges).
- Hubs: These are nodes that indicate the path to the most authoritative authorities, in our case, a pattern found in predators (a predator is a node with many outgoing edges).

If a node is pointed to by many other nodes, it acquires high authority; otherwise, if a node points to many other nodes but is not pointed to by anyone, it acquires authority equal to 0 but has a value different from 0 of hub centrality. The two centralities are calculated with these formulas:

$$x_i = \alpha \sum_j A_{ij} y_i$$

$$y_1 = \beta \sum_j A_{ji} x_i$$

For our dataset, this type of centrality is perfect because through the hubs, we can calculate which are the most important predators and through the authorities, we can calculate which are the prey whose extinction would be more negligible.

	Degree Centrality
Snake	0.11565
Bird	0.06734
Leopard	0.06472
Human	0.04668
Hawk	0.04551

Tabella 4: Hubs

	Degree Centrality
Common Snapping Turtle	0.02637
Galápagos Penguin	0.02364
Gila Monster	0.01975
Lion-tailed Macaque	0.01756
Bearded Dragon	0.01752

Tabella 5: Authorities

From the results, we can see which are the main predators (Table: 4) and the prey hunted by the most important hunters (Table: 5). In the case of Authorities, the extinction of one of these animal species would not imply such a radical change for the most important hubs as they could have other prey to hunt and would not become extinct (this does not take into account other factors such as climate, habitat, and predator habits), as we will actually see in Chapter (5.3), the extinction of nodes with higher authority would cause the extinction of several species with which they share habitat. On the other hand, if we were to remove an animal among the hubs, many species would no longer be hunted (or would have one less predator) and would derive an important advantage from it.

5.2 Network Measures

Given our dataset, it is not possible to appreciate the small-world effect since the graph does not have weights on either the edges or the nodes. To overcome this problem, we could manually weigh the nodes as well as the edges, but the weight on the edges would have no meaning within the scope of our dataset. Each edge connects predators with various prey, weighing the edge would mean indicating how much an animal is preyed upon by a predator, and in the absence of complete data, this would lead to a series of errors in the dataset resulting in inconsistency. Furthermore, calculating a measure like the small-world effect in a social

network graph would help us understand if two people know each other by having common friends, but in our case, the fact that a predator hunts a certain type of prey does not necessarily imply a direct relationship between the predators hunting that prey. We could use this measure by assuming that two hunters who share the same prey may also share the same habitat, but we would also have to consider the geographical distribution and social relationships of that animal (such as territorial sharing), and these information are not present in the dataset.
After calculating the degree centrality (ref: 5.1), we wanted to analyze the degree distribution. This analysis allows us to understand which is the most common degree in our network, the fact that a certain degree occurs more frequently indicates that our dataset is mainly characterized by that degree. In the graph 4, we see how in our dataset there are many instances of small

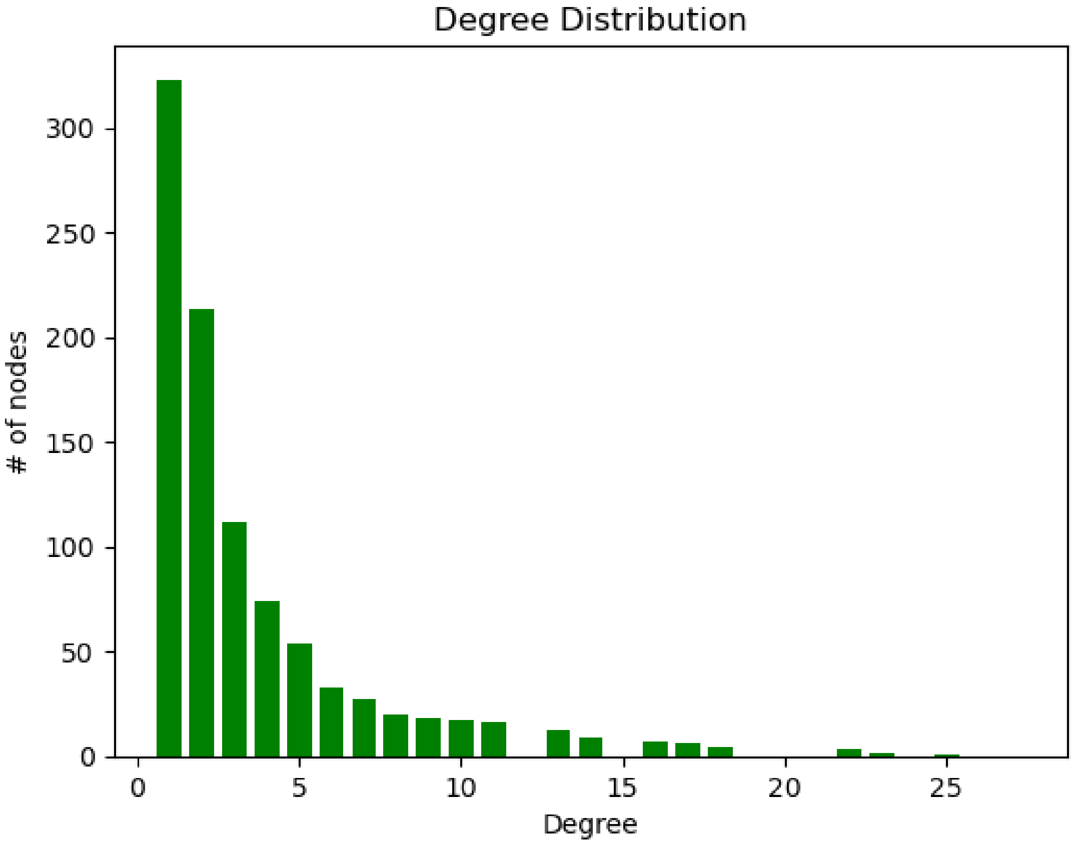

Figura 4: Degree distribution Log-Log graphic

degree and few instances of high degree (> 20). This highlights that there are few instances that are capable of hunting several animals; rather, the various prey are distributed among the various predators.
The shape of the degree distribution can provide valuable information about the structure of the network. Networks can have different forms of degree distributions, but two of the most common are:

- **Poisson distribution**: In a network with a Poisson distribution, most nodes have a similar degree, and there are few nodes with extremely high degrees. This suggests a more uniform and random structure of the network.
- (**Power-law distribution**): In a network with a power-law distribution, there is a long "tail" in the degree distribution, meaning that there are some nodes with a very high

number of connections (the so-called "hubs") and many nodes with few connections. This distribution indicates a scale-free structure of the network, where some nodes play a critical role in maintaining network connectivity.

To verify if there are scale-free behaviors, we decided to project the same graph on a log-log scale 5.
To more precisely calculate how well our distribution approximates the power-law distribution, we calculated α using the formula:
if $p_d = Cd^{-\alpha}$

$$P_d = C\sum_{d'=d}^{\infty} d'^{-\alpha} \simeq C\int_{d}^{\infty} d'^{-\alpha}\partial d' = \frac{C}{\alpha-1}d^{-(\alpha-1)} \tag{2}$$

The results obtained are as follows:
$\alpha = 3.0061$
$p-value = 0.0952$

Looking at the graph 5, we can visually recognize a straight line, which is indicative of a scale-free network. The degree distribution appears to follow a power-law, as indicated by the calculated α (3.006), suggesting that the network might be scale-free as it closely approaches the condition $2 < \alpha < 3$, which defines whether a distribution follows a power-law. This implies that there are some nodes with a very high number of connections, while the majority of nodes have a relatively low number of connections. The p-value (0.095) associated with the power-law is higher than the usual significance level (0.05), suggesting that the distribution might not be extremely well-fitted to a power-law.
In summary, we can say that the network seems to have some highly influential nodes (with high degree centrality), and that the degree distribution approximately follows a power-law, although with a certain degree of statistically significant uncertainty.

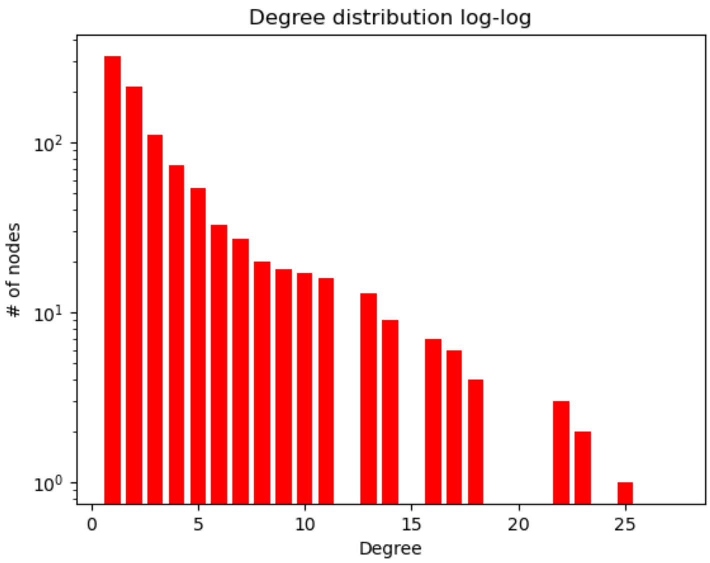

Figura 5: Degree distribution Log-Log graphic

An interesting aspect of scale-free networks is that they are resilient to random failures, as the random removal of nodes tends to not significantly affect the global connectivity of the network. However, they are vulnerable to targeted attacks on the so-called "hubs", nodes with the highest number of connections, as removing these nodes can compromise the structure of the network. Translated to our dataset, this means that the extinction of low-degree animals will not result in significant changes in the network, however, the extinction of more prominent predators such as humans, snakes, or leopards will bring about significant changes.

5.3 Simulations

One of the objectives of our study was to simulate the extinction of certain animal species in order to observe the impact of such events. To do this, it was necessary to identify some key animals that would then be removed within the network. Regarding the meaning of "key animal", we decided to adopt three different strategies and observe the subsequent consequences resulting from the elimination of nodes identified as particularly relevant.

5.3.1 In-Degree Strategy

The first methodology adopted is based on the in-degree value of each node. In this case, we considered key animals to be those types of animals that had a high in-degree, and therefore were preyed upon by many others, as well as those present in sub-networks, as even though they have a low value, they are essential given the small size of their belonging network. To carry out this type of operation, we used the measurement previously applied to our network and conducted an analysis to obtain all the present sub-networks. Through this analysis, we identified the following sub-networks:

Animal	Predator
Amazon Rainforest Frog	Leimadophis Epinephelu
Galapagos Tortois	Galapagos Hawk
Galliformes	Vary
Hagfish	Variuos Predators
Shortfin Mako Shark	Larger Sharks

As we can observe from the table, most of the sub-networks are quite vague as they do not clearly define the species of animals falling into the predator category. For this reason, the results related to "Amazon Rainforest Frog" and "Galapagos Tortoise" will be more relevant compared to the other three; consequently, we have only removed the nodes related to the former two animals. Once all the key animals were identified, we removed the corresponding nodes and visualized the new structure of the network.

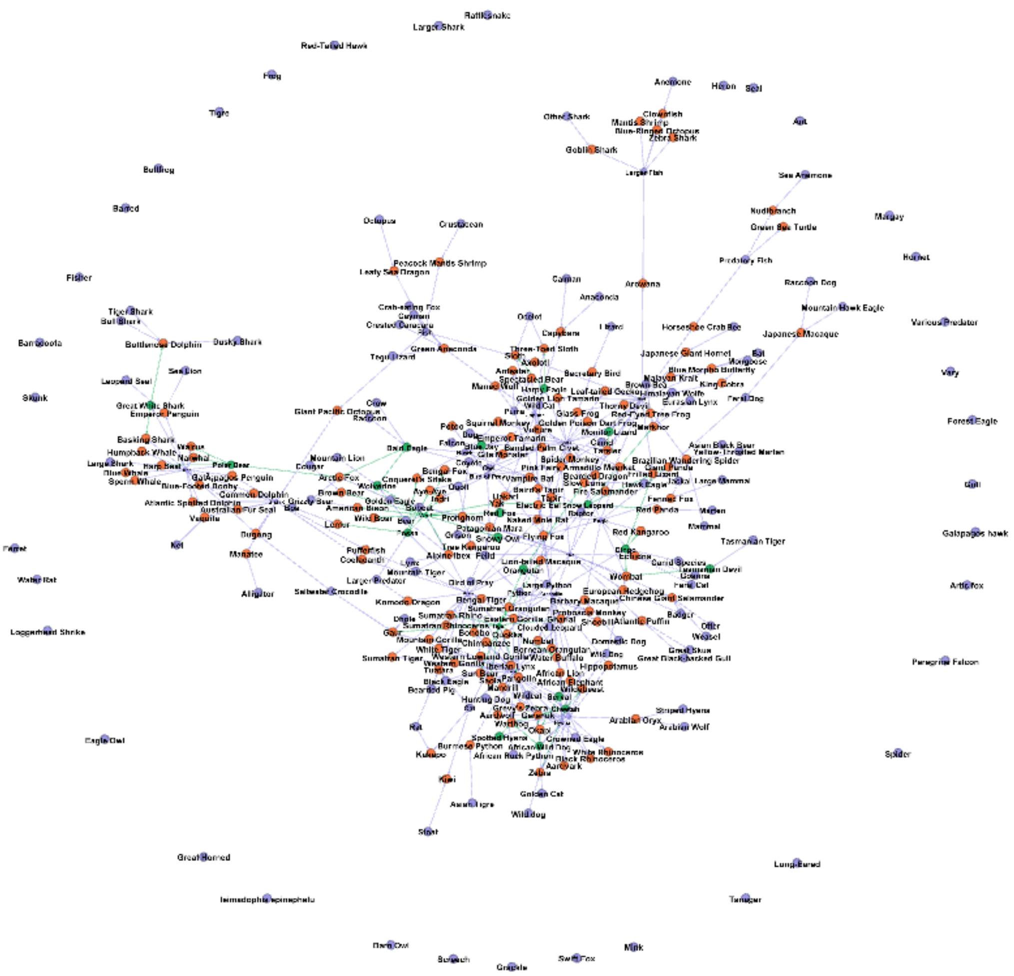

Figura 6: In-Degree Strategy

As can be seen, this extinction has had significant consequences for various animals as they found themselves without prey to hunt. In particular, we can observe that after this extinction, 88 edges were removed, and 32 nodes no longer have any relationships.

5.3.2 Betweenness Centrality Strategy

Regarding this other strategy, we relied on the value of betweenness centrality, as animals with a high value represent key points within a food chain, acting as bridges between multiple species of animals. This may mean that these types of animals are extremely relevant as their potential extinction would cause damage to both their predators and the habitat in which they live, as it could lead to an excessive proliferation of their prey.
The eliminated animals were the top 10 with the highest betweenness centrality (Table 6).

Animal	Beetwennes Centrality Value
Wolf	0.000214
Tiger	0.000155
Bald Eagle	0.000119
Dingo	0.000097
Fossa	0.000077
Wolverine	0.000061
Orangutan	0.000039
Snow Leopard	0.000034
Blue Jay	0.000029
Cheetah	0.000024

Tabella 6: Deleted nodes based on beetwenness centrality

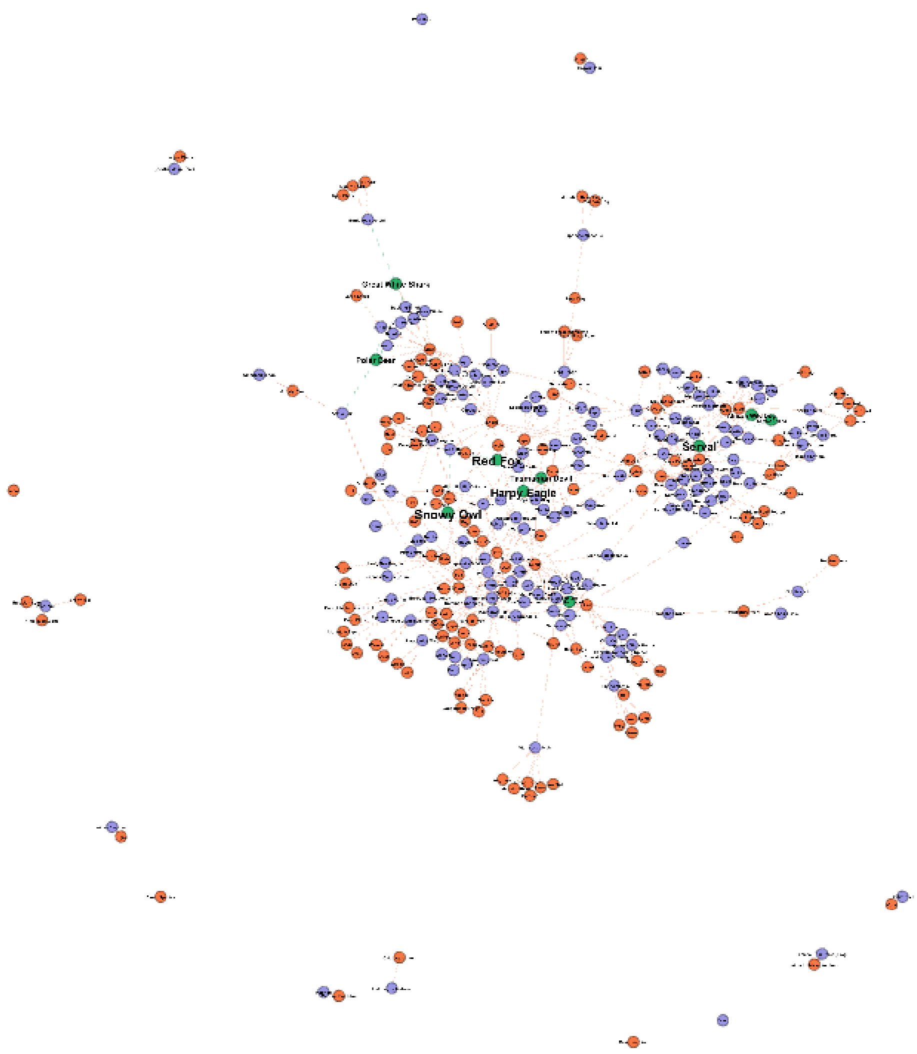

Figura 7: Betweenneess Centrality Strategy

As can be observed, this extinction also had repercussions on certain animals. In this case, the nodes left without any connection were 6, while the removed edges were 70. This is due to the fact that many of these were large predators.

5.3.3 Conservation Status Strategy

Regarding the last strategy for identifying key nodes, we relied on the conservation status of various animals. We decided to implement this strategy to replicate a possible realistic scenario of a not too distant future, as we chose to eliminate only animals with a conservation status of "Endangered" and "Critically Endangered". Before observing the result of this operation, we would like to highlight the evidence of human impact in our network. Indeed, it is easily noticeable that most of the animals connected to humans are in a high state of danger. This observation is extremely significant and raises important concerns about the sustainability of our interactions with the natural environment. This phenomenon highlights the crucial role of humans as agents of significant change in global ecosystems. Human activities such as hunting and habitat destruction have significantly contributed to the decline of animal populations and the loss of biodiversity.
The presence of numerous endangered animals among those connected to humans in our network underscores the urgency of adopting more effective and sustainable conservation measures to protect these vulnerable species. It is essential to promote awareness of the impact of our actions on the environment and to adopt management and conservation practices that promote the recovery of animal populations and the restoration of damaged ecosystems.
Furthermore, this data emphasizes the importance of promoting peaceful and sustainable coexistence between humans and other species with whom we share the planet. It is necessary to adopt eco-friendly management approaches that take into account the needs of animal species and promote the conservation of the natural environment for the benefit of all life forms on Earth.
As can be noted, in this case as well, this strategy had a significant impact on other animals, as the number of eliminated connections is 136 and the number of nodes left without any connection is 18.

5.3.4 Measures On The New Networks

Regarding the measures applied to the new networks generated after extinction simulations, we analyzed the connectance and the edge-to-node ratio to assess any changes in the network structure following the removal of key nodes. However, it's important to note that no significant changes were observed compared to the original network.
Connectance, representing the fraction of all possible connections present in the network that are actually realized, remained substantially stable after extinction simulations. This suggests that, despite the removal of some key nodes, the network still retains a high degree of interconnection among animal species.
Similarly, the edge-to-node ratio, which provides a measure of connection density in the network relative to the total number of nodes, showed no significant variations compared to the original network configuration. This indicates that, despite the loss of some nodes, the network still maintains a balanced distribution of connections among the remaining nodes. This is likely due to the presence of various animals from different habitats and geographical areas in our network.
Another consideration regarding the results obtained is the number of nodes resulting without any connection after extinction execution. Specifically, we had 32 nodes for the first extinction, 6 for the second, and 18 for the third. This could be due to the observation we made through the hubs and authorities measure, where most authorities are prey and most hubs are predators. When we conducted the first extinction, we mainly removed large authorities, which consequently were targeted by many hubs, both small and large. Therefore, when these authorities

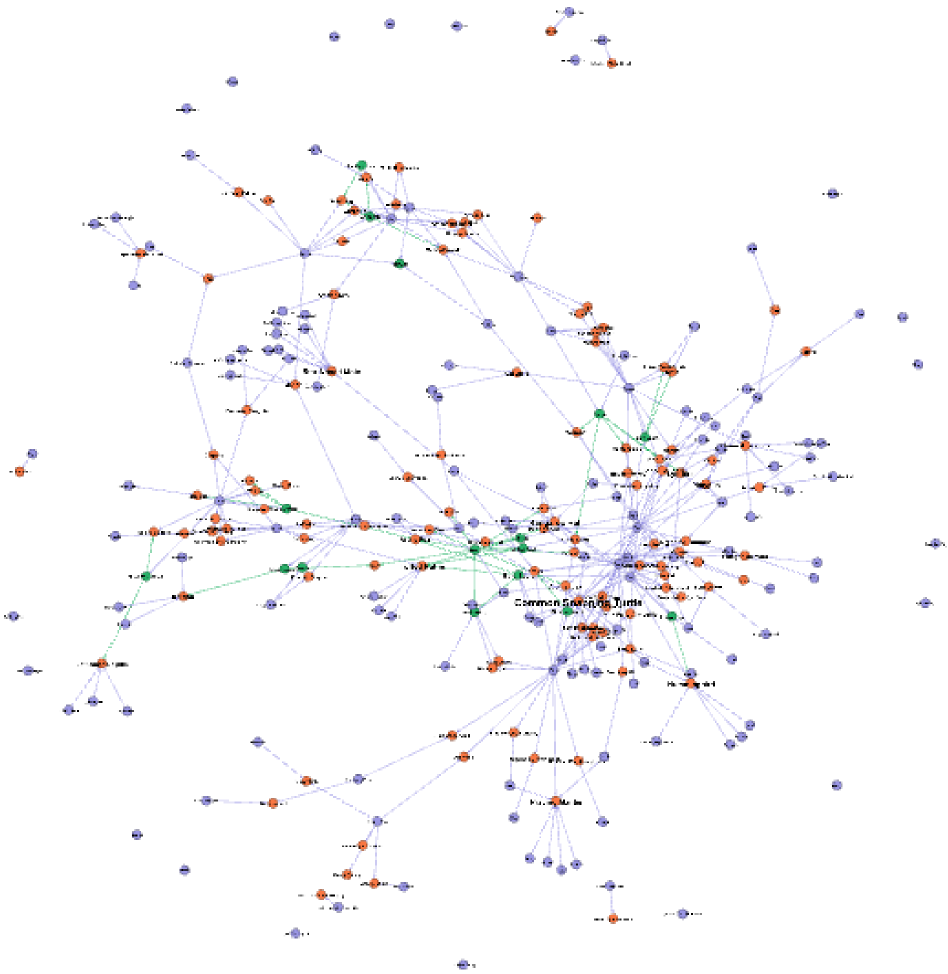

Figura 8: Conservation Status Strategy

were removed, many relationships involving small hubs were also eliminated. This occurred with the other two extinctions as well, but to a lesser extent, as with the second extinction, small authorities were eliminated, and with the third, both small and large authorities were eliminated.

6 Conclusion

In conclusion, through the analysis of a dataset containing information about certain animals and their predators, we delineated the intricate predator-prey relationships, aiming to construct a network that could reflect, as much as possible, the complexity of ecological interactions. Once this network was constructed, we utilized a series of measurements to examine its structure and identify key nodes within it.

The conducted measurements allowed the identification of several key species, crucial for maintaining ecosystem balance. Through three successive extinction simulations, we applied different criteria to identify these key species: initially, relying on the high in-degree of prey and prey present in sub-networks, subsequently considering animals with the highest betweenness centrality, and finally focusing on species at risk of extinction.

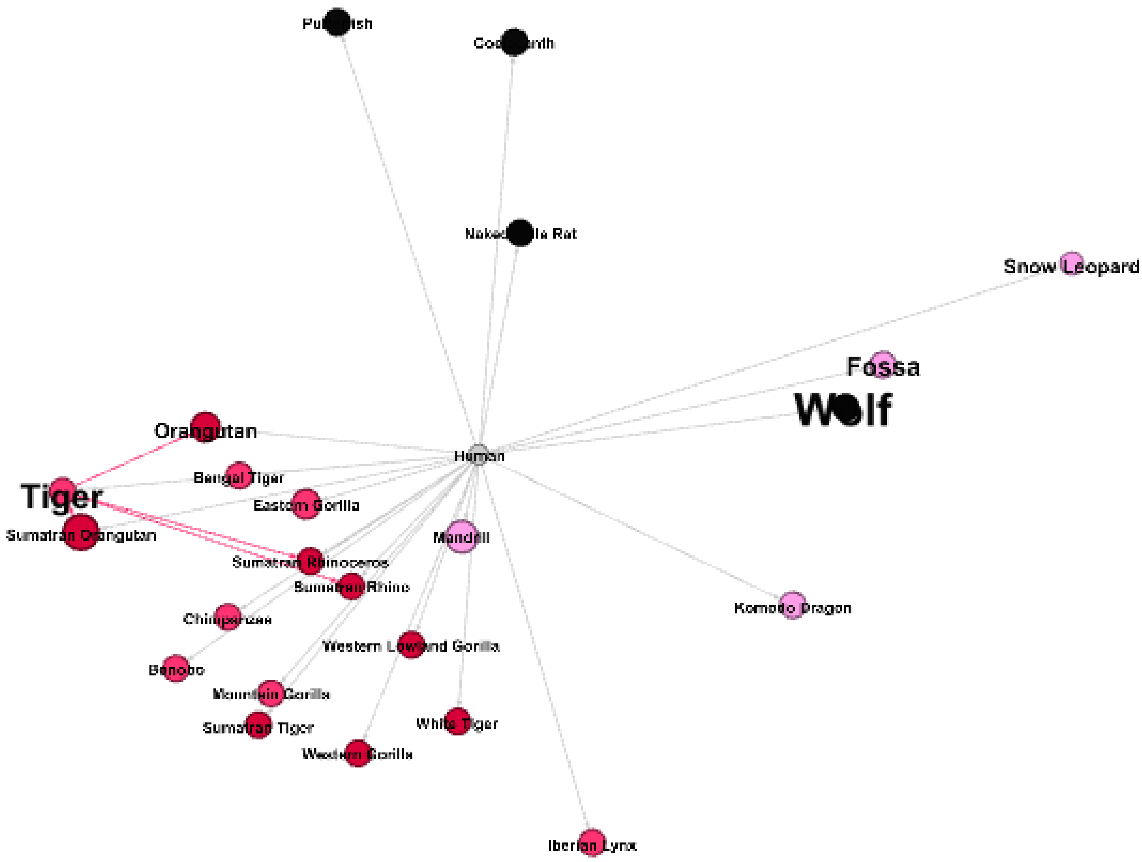

Figura 9: Relationship between humans and animals. Dark Red = Critically Endangered, Red = Endangered, Pink = Vulnerable

This study seeks to provide insight into interactions within various ecosystems, enabling a deeper understanding of ecological dynamics. Furthermore, the identification of key species and the assessment of extinction risk could aid in guiding conservation strategies, allowing for the concentration of resources and efforts towards the most vulnerable or relevant species.
Regarding future developments, this study offers numerous opportunities to deepen the understanding of predator-prey interactions and to undertake further research in the realm of ecological conservation. For instance, one could opt to delve deeper into the analysis of interactions within the network, focusing on specific species subgroups or particularly vulnerable ecosystems, to identify emerging patterns and more complex relationships. Additionally, integrating spatial and temporal information in further studies could enhance understanding of how ecological dynamics may vary by location and over time, considering factors such as the effects of climate change or habitat modifications.

7 Critique

Unfortunately, our study has limitations stemming from the nature of the dataset used. The lack of detailed data and the limited number of animals included in the dataset have constrained the depth of the analysis. The scarcity of information on the majority of predators, such as their conservation status, the absence of data regarding the predators of predators, and thus the depth of the species' food chain, has limited our ability to fully understand ecological dynamics and predator-prey relationships.

Moreover, another criticism concerns the unweighted nature of the graph we constructed. This considerably simplified the analysis of the consequences of extinction on other species since we did not account for the strength of relationships between species. In a complex ecological context where interactions can be influenced by multiple factors, the lack of a weighted assessment of relationships can limit the accuracy of our predictions and conclusions.

Additionally, predator-prey relationships are not limited solely to the direct prey of a predator. Some animals may contribute to each other's survival even if they are not directly linked by the food chain. This complexity of ecological interactions underscores the importance of considering a broader range of factors and relationships within the ecosystem to gain a comprehensive understanding of the dynamics at play.

In conclusion, while acknowledging the value of our study in contributing to the understanding of predator-prey interactions, it is important to be aware of the intrinsic limitations due to the nature of the dataset and the model used. For a more in-depth and accurate analysis of ecological dynamics, it will be necessary to gather more comprehensive data and employ more sophisticated models that account for the complexity of relationships within natural ecosystems.