Identifying Keystone Species in a Food Web

Yadder Joshua Aceituno González
Maastricht University
`Network Science`

December 12, 2025

Abstract

Ecological networks, specifically food webs, provide a fundamental framework for understanding the stability and biodiversity of ecosystems. This study investigates the structural properties of a predator-prey network constructed from the citizen-science project "Who Eats Whom" on iNaturalist. By modeling interactions as a directed graph ($N = 1,899$), we employ structural metrics, including degree centrality and motif analysis to identify keystone species. Our results highlight the species "Butterflies and Moths" and "Great Blue Heron" as critical nodes for network cohesion.

1 Introduction

Food webs act as roadmap of an ecosystem, describing the complex flow of energy from basal resources to predators. Understanding the architecture of these networks is crucial for identifying keystone species whose removal would precipitate a disproportionate collapse in ecosystem structure. While traditional ecological studies rely on localized field observations, the rise of community science platforms has enabled the analysis of macro ecological patterns on a global scale.

This paper applies network methods to analyze a dataset sourced from iNaturalist, a massive citizen science initiative. By representing species as nodes and predation events as directed edges, we aim to discover the mechanisms that govern ecosystem stability. Specifically, we address three main objectives:

1. **Topological Characterization:** We quantify the structure of the food web using centrality metrics (Degree, Betweenness) to identify influential predator and prey taxa.

2. **Structural Motif Analysis:** We examine local interaction patterns, specifically tri-trophic food chains, to identify species serving as critical bridges for energy flow.

3. **Resilience Testing:** We assess the web robustness against biodiversity loss through simulated extinction cascades, comparing random removal strategies against targeted attacks on different levels.

2 Dataset

The primary dataset utilized in this study was sourced from the iNaturalist project "Who Eats Whom". This project collects predator-prey interactions documented by users around the world. Access to the raw data was obtained via the platform's export functionality.

The dataset gathers approximately 13,000 observations. Each record denotes a binary interaction between two species, categorized respectively as predator or prey. To facilitate analysis, the following specific attributes were extracted for each observation:

- Observation ID
- Species Scientific Name
- Species Common Name
- Taxon Name
- Interaction Type (Predator or Prey)
- Partner Species ID (the species involved in the interaction)

3 Methodology

3.1 Web Construction

Following data gathering step, we constructed a directed graph $G = (V, E)$ to represent the food web. In this structure, the set of vertices V represents distinct species, while the set of edges E represent species interactions. An edge e_{ij} is directed from node i (predator) to node j (prey).

3.2 Centrality Measures and Keystone Species

To identify keystone species, we employed three primary network centrality metrics as defined by Newman [4]:

- **In-Degree:** Quantifies the number of predators consuming a specific species.

- **Out-Degree:** Quantifies the number of distinct prey species a predator consumes.

- **Betweenness Centrality:** Measures the frequency with which a node appears on the shortest paths between other species, these species act as a bridge for energy flow [2].

3.3 Motif Analysis: Community Modules

Beyond node metrics, we analyzed structural motifs to identify local interaction patterns. We adopted the "community module" framework proposed by Holt [3], which defines fundamental sub-graph structures essential for understanding species dynamics within the web:

1. **Food Chain:** A linear interaction ($A \to B \to C$).

2. **Apparent Competition:** A single predator shares two prey species ($B \leftarrow A \to C$).

3. **Exploitative Competition:** Two predators compete for a shared prey ($B \to A \leftarrow C$).

4. **Predation on Competing Prey:** A structure where a top predator consumes two intermediate species, which in turn compete for a shared prey ($B \leftarrow A \to C$ and $B \to D \leftarrow C$).

5. **Intraguild Predation:** A triangular motif where a predator consumes both a prey species and a competitor ($A \to B \to C$ and $A \to C$).

3.4 Extinction Cascades and Robustness

To assess the structural resilience of the food web, we simulated secondary extinction cascades triggered by the removal of identified key species, following the methodology of Dunne et al. [1]. The simulation operates on a bottom-up dependency logic: following the removal of a specific node, any species losing all of its prey is subsequently removed. We evaluated the network's stability by measuring changes in robustness post-simulation.

4 Results

4.1 Analysis of the Largest Component

Following the network construction, the web showed a highly fragmented structure, comprising a total of 2,688 separated components. The majority of these were isolated nodes or small interactions. This fragmentation is attributable to data ambiguities, where observations lacked precise predator or prey attribution.

To ensure robustness, we extracted the largest connected component for subsequent analysis. This component contains $N = 1,899$ nodes, representing approximately 70% of the original network volume. Table 1 summarizes the structural metrics of this primary component.

Table 1: Structural metrics of the largest connected component.

Metric	Value
Nodes (N)	1,899
Edges (E)	2,313
Connectance (Density)	0.06%
Average Degree (k)	1.21
Average Path Length (L)	2.49

The calculated connectance (density) of 0.06% indicates a highly sparse network structure, a common characteristic of large-scale ecological networks [1]. The low average degree ($k \approx 1.21$) suggests that, on average, species in this dataset interact with very few species. This sparsity is partly driven by the high prevalence of basal species (plants) which, by definition in this directed graph, have an out-degree of zero.

To isolate the structural dynamics of higher species, we recalculated the average degree after excluding basal plants. This adjustment resulted in increase average degree of $\langle k \rangle = 2.56$, indicating that species maintain interactions with approximately two to three other species.

In the figure 1, we visualize the food web structure of the largest component. The colors of the nodes represent the category that the species belongs to. We easily can notice how basal-plants (green nodes) appear everywhere in the food web, this indicates that they are food sources for different species. Also, some clusters of predators and prey are clearly visible, highlighting the modular structure of the ecosystem.
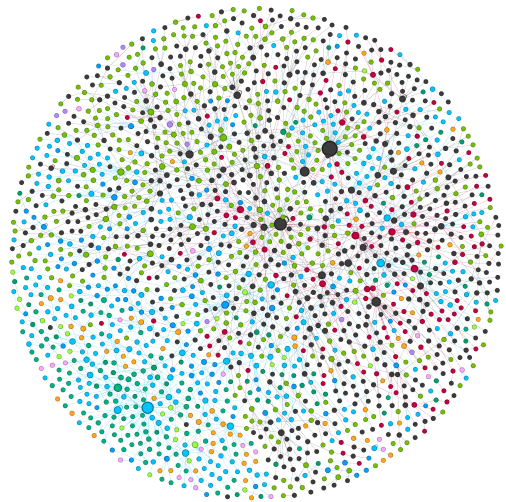

Figure 1: Food web visualization.

4.2 Identification of Keystone Species

Having established the web structure, we calculated node level metrics to identify potential keystone species. Figure 2 illustrates the top five species ranked by In-Degree and Out-Degree.
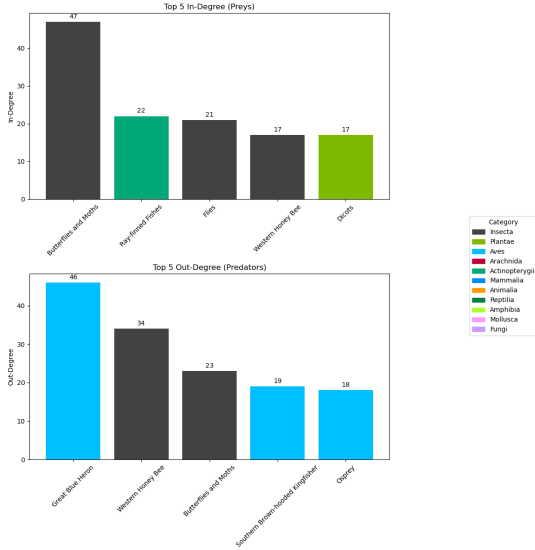

Figure 2: Top 5 Species ranked by In-Degree (up) and Out-Degree (down).

4.2.1 Prey Consumption (In-Degree)

Analysis of the In-Degree distribution identifies "Butterflies and Moths" as the most consumed species, with an in-degree of $k_{in} = 47$. Three of the top five identified prey nodes belong to the category *Insecta*. This metric aligns with ecological expectations, as insects serve as resources for a diverse of predators, including arachnids, aves, and mammals.

4.2.2 Predator Dominance (Out-Degree)

In the other hand, the Out-Degree analysis highlights the "Great Blue Heron" as the dominant predator in the network, exhibiting an out-degree of $k_{out} = 46$. Similar to the prey distribution, the predator hierarchy is dominated by a specific category: three of the top five predators are members of the cateogry *Aves* (Birds). This result reflects the feeding strategies typical of many avian species, which often consume across multiple species, consuming fish, small mammals, and invertebrates.

4.2.3 Community Modules: The Food Chain Motif

The evaluation of motifs provides a complementary method for identifying keystone species. In this study, we specifically focused on the Food Chain motif ($A \rightarrow B \rightarrow C$). This tri-node structure is critical as the intermediate species (B) functions as a obligated bridge between preys and predators.

From a bio-energetic perspective, these bridge species are essential for vertical energy transfer. The removal of such species in the food web may damage the bottom-up propagation of energy.

Figure 3 highlights the top five species that function as critical bridges in these chains. Specifically, these species support the highest number of obligated predators that rely exclusively on that single species for sustenance. Consequently, the loss of these key species would guarantee the secondary extinction of the dependent predators.
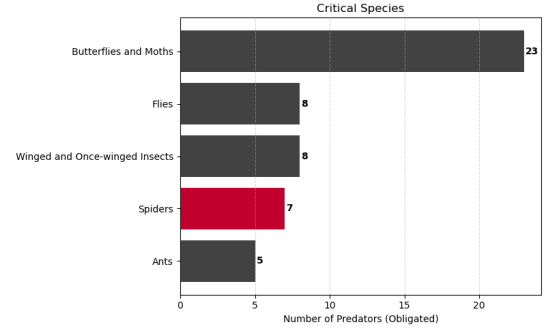

Figure 3: Top 5 Critical Bridge Species by Obligated Predator Count

4.3 Network Robustness and Extinction Simulation

To evaluate the structural resilience of the food web, we simulated three extinction scenarios: (1) Random removal, (2) Targeted removal of top predators, and (3) Targeted removal of top prey species.

We quantified robustness using the metric R, defined as the area under the extinction curve [1]. In this framework, the theoretical maximum robustness is $R = 0.50$, which reflects the effect of secondary extinctions. Values closer to 0.50 indicate high stability, while lower values imply rapid system collapse.

4.3.1 Scenario 1: Random Extinction

Figure 4 presents the results of the stochastic removal strategy. The network showed a robustness value of $R = 0.42$. This suggests that the food web maintains relative stability under random perturbations.

4.3.2 Scenario 2: Removal of Top Predators

In the second scenario, we simulated the targeted removal of top predators. As shown in Figure 5, this resulted in a expected high robustness value of $R = 0.46$, indicating minimal secondary extinctions.

This result highlights a limitation of the second cascade model. Since top predators generally do not serve as food sources for other species, their removal does not
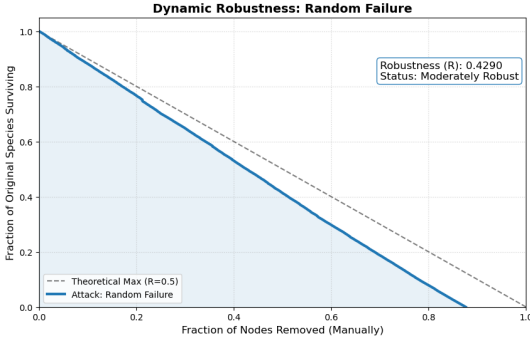
Figure 4: Network robustness under random species removal ($R = 0.42$).

Figure 5: Network robustness under targeted removal of top predators ($R = 0.46$).

4.3.3 Scenario 3: Removal of Top Preys

Finally, Figure 6 illustrates the consequences of removing key prey species. This scenario proved the most damaging, yielding a significantly reduced robustness value of $R = 0.30$.

This drop confirms that the network is highly sensitive to the loss of important preys ("bottom-up" control). While the system exhibits higher stability than some comparable studies—which have reported values as low as $R \approx 0.15$ for similar attacks, the 30% reduction compared to the random baseline emphasizes the critical role of these consumed species.

5 Discussion

A primary challenge encountered in this study was the inherent sparsity of the observational data. While individual records from the iNaturalist platform are well-

Figure 6: Network robustness under targeted removal of top prey species ($R = 0.30$).

documented, they do not constitute a complete representation of the species' dynamics. This incompleteness is attributable to the nature of predation events; many predators exhibit elusive behavior, making direct observation difficult. Consequently, the dataset is skewed towards species that are easy to observe, such as insects, arachnids, and plants, rather than elusive vertebrate predators..

To mitigate the impact of missing links, a future work that could help is the implementation of a Graph Neural Network to infer latent interactions among species outside the largest component. The efficacy of this link prediction task relies heavily on the quality of node embeddings. By enriching the food web with biological attributes such as body mass, size, habitat preference, and dietary classification the GNN can learn representations of species to accurately predict unobserved species relationships.

Despite these data limitations, the application of node metrics (In-Degree, Out-Degree, and Betweenness Centrality) successfully identified critical species within the existing food web. Our analysis highlighted "Butterflies and Moths" as the most significant prey resource, while the "Great Blue Heron" was identified as the dominant predator.

Our simulations indicate that the targeted removal of keystone prey precipitates the most significant decline in network robustness. However, despite this abrupt drop, the network maintained a higher degree of stability ($R = 0.30$) compared to similar empirical studies, which often report values near 0.15 under identical attack strategies. This resilience is likely attributable to the fact the network is heavily populated by everywhere-groups, specifically insects, arachnids, and aves, which possess high connectivity, allowing the system to absorb the loss of specific nodes without total collapse.

Conversely, the predator removal simulations highlighted the limitations of purely structural analysis. While the structural integrity of the web appeared unaffected ($R = 0.46$), this static metric fails to account for "top-down" regulation. In biological reality, predators are essential for controlling prey populations. Their removal

trigger "bottom-up" starvation cascades in this simulation. However, ecologically, the loss of these keystone regulators would likely lead to "top-down" instability (e.g., prey overpopulation), a dynamic that requires dynamic population modeling rather than static structural analysis.
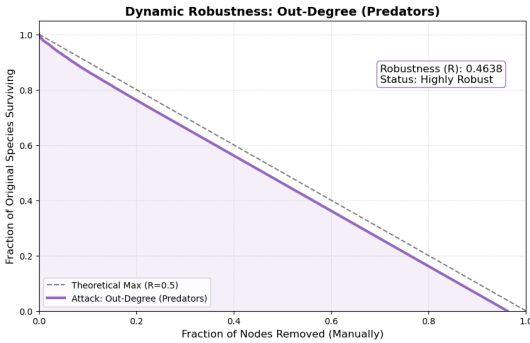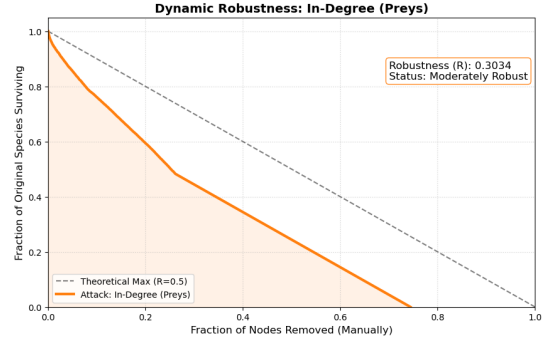
typically leads to population growth among lower species, which can subsequently destabilize the ecosystem through resource depletion. Consequently, accurately modeling predator loss requires a more complex simulation framework that incorporates node weights (population size) rather than simple connections.

6 Conclusions

This research demonstrated the efficacy of network science in elucidating food web dynamics, allowing us to successfully identify keystone species using established topological metrics. Our analysis specifically highlighted the critical role of insects and birds in maintaining ecosystem stability, confirming that the removal of insect prey precipitates a significant decline in network robustness.

However, these findings must be interpreted within the context of data limitations; the sparsity of the dataset depicted in the exclusion of numerous poorly connected species, a constraint that highlights the need for improved data completeness to enhance the generalizability of future studies.

References

[1] Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters*, 2002.

[2] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[3] Robert D Holt. Community modules. In *Multitrophic Interactions in Terrestrial Systems*, pages 333–349. Blackwell Science London, 1997.

[4] Mark Newman. *Networks: An Introduction*. Oxford University Press, 2010.