Redundancy in Food Web Graphs via Filtered Entropy Phase Measure

CRL-J

February 21, 2025

Abstract

This paper introduces a approach to quantify redundancy in a certain class of directed graphs by comparing the Shannon entropy of simple-path distributions with the maximal entropy achieved by a tree structure. By defining an energy filtration based on the length of the shortest simple path from a unique root, we derive a filtered entropy measure that highlights how alternative pathways affect network uncertainty. Applications to ecological food webs are discussed.

1 Introduction

Understanding redundancy in complex networks is critical to unraveling their resilience and structural properties. In many real-world systems—such as ecological food webs, communication networks, and supply chains—redundant pathways provide alternative routes that can safeguard the system against disruptions. This paper introduces a novel approach to quantify network redundancy through a filtered Shannon entropy framework, bridging theoretical insights from information theory with tangible biological implications.

We consider a directed graph $G = (V, E)$ that is finite, connected, and features a unique root vertex R (with in-degree zero) such that every other vertex in V is reachable from R via at least one simple path. The *energy* of each vertex is defined as the length of its shortest simple path from R, which naturally induces a filtration of the graph:

$$G(k) = \{v \in V : \mathcal{E}(v) \leq k\}.$$

This energy-based perspective allows us to progressively reveal the network's structure and to investigate how alternative pathways—embodied in redundant edges—emerge as the system unfolds.

At each energy level k, we derive a probability distribution on $G(k)$ based on the counts of distinct simple paths from R. In a tree, where each vertex is reached by a unique path, the distribution is uniform and the entropy is maximized:

$$H_{\text{Tree}}(k) = \log |G(k)|.$$

However, in the presence of redundant links, multiple paths converge on the same vertices, skewing the distribution and lowering the observed entropy $H_G(k)$. We capture this deviation by defining the redundancy measure

$$\phi_G(k) = H_{\text{Tree}}(k) - H_G(k).$$

Beyond its mathematical elegance, our framework has important biological implications. In ecological food webs, redundant pathways can facilitate multiple routes for energy and nutrient transfer, effectively buffering the ecosystem against species loss or environmental disturbances. Empirical observations suggest that ecosystems with higher redundancy often exhibit greater stability and resilience. Moreover, the concepts developed here have potential applications in analyzing robustness in engineered, social, and economic networks.

In the sections that follow, we detail the theoretical underpinnings of our model, provide illustrative examples, and discuss broader applications. By integrating insights from network theory and ecological studies, our work contributes a versatile tool for analyzing the resilience of complex systems.

2 Background

In this section, we review key concepts from network theory and information theory that form the basis for our study of redundancy in food web graphs. We focus in particular on the role of redundant edges—those that do not lie on the shortest simple paths from the root vertex through its predecessors—in shaping the uncertainty and resilience of these networks.

2.1 Network Theory, Directed Graphs, and Redundancy

Graph theory offers a flexible framework for modeling complex systems such as ecological food webs. Here, a food web is represented as a directed graph $G = (V, E)$, where vertices V denote species (or trophic groups) and directed edges E represent feeding relationships. A common special case is the *tree*, a connected acyclic graph in which each vertex (except a unique root R) is reached by a single simple (non-repeating) path. This tree structure serves as a benchmark because it guarantees a uniform distribution of paths from the root, thereby maximizing the network's Shannon entropy.

In contrast, many real-world food webs contain *redundant edges*—additional connections that do not participate in the shortest simple paths from R. These edges create alternative pathways between vertices, meaning that some species can be reached by more than one route. For example, in the food web diagram of Figure 5, certain edges (drawn in red) illustrate these redundancies. Although they do not change the minimal energy required to reach a given species, such redundant links can significantly skew the distribution of simple paths, as they introduce additional routes that converge on the same vertices.

2.2 Shannon Entropy and Path Distributions

Shannon entropy, introduced by Claude Shannon [1], quantifies the uncertainty in a probability distribution. In the context of a food web, we derive a probability distribution from the counts of distinct simple paths from the unique root R to each vertex v. In a tree, where each vertex is accessible via a unique route, the distribution is uniform and the entropy is maximized:

$$H_{\text{Tree}}(k) = \log |G(k)|,$$

where $G(k)$ denotes the subgraph containing all vertices with energy at most k (i.e., vertices reachable within k steps).

However, when redundant edges are present, some vertices are reached via multiple distinct paths. If we denote by $Q(v)$ the number of simple paths from R to v and by

$$T(k) = \sum_{v \in G(k)} Q(v)$$

the total number of paths within the subgraph $G(k)$, the probability of reaching vertex v becomes

$$p_v^{(k)} = \frac{Q(v)}{T(k)}.$$

The Shannon entropy at energy level k is then given by

$$H_G(k) = -\sum_{v \in G(k)} p_v^{(k)} \log \left(p_v^{(k)} \right).$$

The presence of redundant edges tends to concentrate the probability distribution on vertices that can be reached via multiple pathways, thereby reducing the overall entropy relative to the tree case.

2.3 Redundancy Measure and Ecological Implications

We quantify the loss in entropy due to redundant edges by defining a redundancy measure:

$$\phi_G(k) := H_{\text{Tree}}(k) - H_G(k).$$

A value of $\phi_G(k) = 0$ indicates a tree-like structure with no redundant links, while $\phi_G(k) > 0$ signifies that redundancy is present. In ecological terms, these redundant pathways can provide alternative routes for energy and nutrient flows. Although such redundancy may lower the entropy by concentrating the path distribution, it can also confer resilience to the ecosystem by buffering against the loss of any single pathway.

The example in Figure 5 demonstrates this concept vividly. The dashed and red edges highlight redundant connections—edges that do not lie on the shortest simple paths but nonetheless contribute to the network's complexity. Their presence, while reducing the maximal entropy expected in a tree-like structure, reflects the intricate and potentially stabilizing interactions found in real food webs.

By integrating these insights from network theory and information theory, our approach provides both a theoretical and practical framework for understanding how redundant edges influence the structural uncertainty and resilience of ecological networks.

3 Theoretical Framework

In this section, we develop the theoretical underpinnings of our approach by introducing the graph model, defining key quantities such as path counts and energy, and formulating the Shannon entropy for directed graphs.

3.1 Graph Model and Definitions

Let $G = (V, E)$ be a finite, connected, directed graph with a unique vertex R of in-degree zero, such that there is at least one simple path (i.e., a path that does not repeat vertices) from R to every vertex $v \in V$. In order to analyze the structure of G progressively, we define the *energy* of a vertex v, denoted $\mathcal{E}(v)$, as the length of the shortest simple path from R to v. This energy concept provides a natural filtration of the graph:

$$G(k) = \{v \in V : \mathcal{E}(v) \leq k\}.$$

The subgraph $G(k)$ contains all vertices that can be reached from R within k steps.

3.2 Path Counting and Probability Distributions

For each vertex $v \in G(k)$, let $Q(v)$ denote the number of distinct simple paths from R to v that lie entirely within $G(k)$. In a tree structure (i.e., a graph without redundant links), we have $Q(v) = 1$ for every v, as there is exactly one simple path from R to each vertex. However, in graphs with redundant links, some vertices may be reachable via multiple paths, so $Q(v) \geq 1$.

Define the total number of simple paths within $G(k)$ as:

$$T(k) = \sum_{v \in G(k)} Q(v).$$

This allows us to define a probability distribution on the vertices of $G(k)$ by setting:

$$p_v^{(k)} = \frac{Q(v)}{T(k)}.$$

This probability represents the likelihood that a randomly chosen simple path from R (within energy level k) ends at vertex v.

3.3 Shannon Entropy in Graphs

Using the probability distribution $\{p_v^{(k)}\}_{v \in G(k)}$, we define the Shannon entropy at energy level k as:

$$H_G(k) = -\sum_{v \in G(k)} p_v^{(k)} \log\left(p_v^{(k)}\right).$$

For a tree-like graph, where every vertex is reached by a unique path, the distribution is uniform over $G(k)$. In this ideal case, the maximum possible entropy is:

$$H_{\text{Tree}}(k) = \log\left(|G(k)|\right).$$

We then define a redundancy measure by comparing the observed entropy $H_G(k)$ with the maximal entropy $H_{\text{Tree}}(k)$:

$$\phi_G(k) = H_{\text{Tree}}(k) - H_G(k).$$

The quantity $\phi_G(k)$ serves as a quantitative indicator of redundancy:

- $\phi_G(k) = 0$ indicates that the network is tree-like up to energy level k, with no redundant pathways.

- $\phi_G(k) > 0$ indicates the presence of redundant links, where multiple simple paths converge onto the same vertices, thereby reducing the entropy.

By analyzing $\phi_G(k)$ across increasing energy levels, we obtain a dynamic picture of how redundancy—and thus structural complexity—evolves throughout the network. This framework is not only theoretically appealing but also provides practical insights into identifying critical hubs, bottlenecks, and overall network resilience.

4 Examples

In this section, we illustrate our theoretical framework with several examples.

4.1 Example 1: Directed Tree

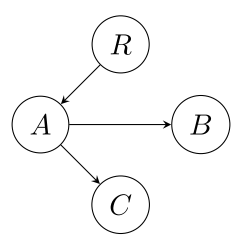
Consider a simple directed tree with vertices

$$\{R, A, B, C\},$$

and edges

$$R \to A, \quad A \to B, \quad A \to C.$$

In this tree, every vertex (except R) is reached by exactly one simple path, yielding a uniform distribution of paths.



Define the energy $\mathcal{E}(v)$ as the length of the shortest simple path from R to v. Then:

$$\mathcal{E}(R) = 0, \quad \mathcal{E}(A) = 1, \quad \mathcal{E}(B) = 2, \quad \mathcal{E}(C) = 2.$$

Let $G(2) = \{R, A, B, C\}$. Since the graph is a tree, we have

$$Q(v) = 1 \quad \text{for all } v \in G(2).$$

Thus, the total number of paths is

$$T(2) = 1 + 1 + 1 + 1 = 4.$$

The probability distribution on $G(2)$ is

$$p_v = \frac{1}{4}, \quad \text{for } v \in \{R, A, B, C\}.$$

The filtered Shannon entropy is then:

$$H_G(2) = -\sum_{v \in G(2)} \frac{1}{4} \log\left(\frac{1}{4}\right) = \log 4.$$

Since this is the maximal entropy for 4 vertices, we have:

$$H_{\text{Tree}}(2) = \log 4.$$

Thus, the redundancy measure is

$$\phi_G(2) = H_{\text{Tree}}(2) - H_G(2) = \log 4 - \log 4 = 0.$$

This confirms that the tree has no redundancy.

4.2 Example 2: Graph with a Redundant Link

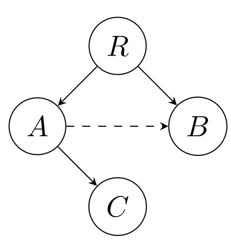
Now, consider a graph with vertices

$$\{R, A, B, C\},$$

and edges

$$R \to A, \quad A \to B, \quad A \to C, \quad R \to B.$$

In addition to the tree structure, a new edge$R \to B$ is introduced, making the edge $A \to B$ redundant.



The energies are:

$$\mathcal{E}(R) = 0, \quad \mathcal{E}(A) = 1, \quad \mathcal{E}(B) = 1, \quad \mathcal{E}(C) = 2.$$

For the filtration at $k = 1$, we have:

$$G(1) = \{R, A, B\}.$$

Now, count the simple paths (restricted to $G(1)$):

$$Q(R) = 1, \quad Q(A) = 1 \ (R \to A), \quad Q(B) = 2 \ (\text{via } R \to B \text{ and } R \to A \to B).$$

Thus, the total number of paths is:

$$T(1) = 1 + 1 + 2 = 4.$$

The probability distribution is:

$$p_R = \frac{1}{4}, \quad p_A = \frac{1}{4}, \quad p_B = \frac{2}{4} = \frac{1}{2}.$$

The filtered entropy is:

$$H_G(1) = -\left(\frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{2}\log\frac{1}{2}\right) \approx 1.0397.$$

For comparison, if the distribution were uniform on $G(1)$ (3 vertices), the maximum entropy would be:

$$H_{\text{Tree}}(1) = \log 3.$$

Thus, the redundancy measure at energy level 1 is:

$$\phi_G(1) \approx 1.098 - 1.0397.$$

4.3 Example 3: Larger Network with a Critical Hub

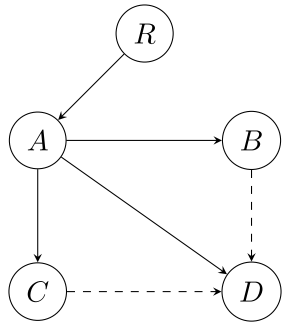
Consider a network with vertices

$$\{R, A, B, C, D\},$$

and edges

$$R \to A, \quad A \to B, \quad A \to C, \quad B \to D, \quad C \to D, \quad A \to D.$$

The edge $C \to D$ is redundant as is $B \to D$.



The energies are:

$$\mathcal{E}(R) = 0, \quad \mathcal{E}(A) = 1, \quad \mathcal{E}(B) = 2, \quad \mathcal{E}(C) = 2, \quad \mathcal{E}(D) = 2 \quad .$$

Thus, $G(2) = \{R, A, B, C, D\}$. The path counts are:

$$Q(R) = 1, \quad Q(A) = 1, \quad Q(B) = 1, \quad Q(C) = 1.$$

For D, the simple paths are:

1. $R \to A \to D$,
2. $R \to A \to B \to D$,
3. $R \to A \to C \to D$.

Hence, $Q(D) = 3$ and

$$T(2) = 1 + 1 + 1 + 1 + 3 = 7.$$

The probability distribution is:

$$p_R = \frac{1}{7}, \quad p_A = \frac{1}{7}, \quad p_B = \frac{1}{7}, \quad p_C = \frac{1}{7}, \quad p_D = \frac{3}{7}.$$

Thus, the filtered entropy is:

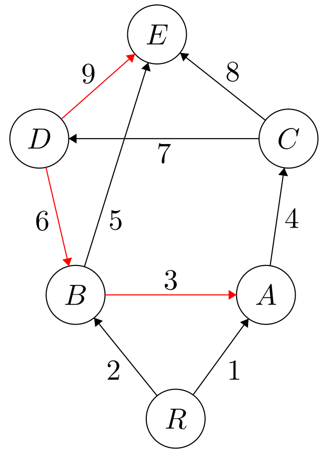
$$H_G(2) = -\left(4 \cdot \frac{1}{7} \log \frac{1}{7} + \frac{3}{7} \log \frac{3}{7}\right) \approx 1.47$$

The maximal (tree) entropy for 5 vertices is:

$$H_{\text{Tree}}(2) = \log 5.$$

So the redundancy measure is:

$$\phi_G(2) \approx 1.609 - 1.47$$



5 Entropy Analysis of a Food Web Graph

In this section, we analyze the food web graph shown in Figure 5 from [5] using our filtered entropy framework. The graph G has six vertices:

$$V = \{R,\ A,\ B,\ C,\ D,\ E\},$$

and the following directed edges:

- Edge 1: $R \to A$
- Edge 2: $R \to B$
- Edge 3: $B \to A$ (redundant)
- Edge 4: $A \to C$
- Edge 5: $B \to E$
- Edge 6: $D \to B$ (redundant)
- Edge 7: $C \to D$
- Edge 8: $C \to E$
- Edge 9: $D \to E$ (redundant)

For this graph:

$$
\begin{aligned}
\mathcal{E}(R) &= 0, \\
\mathcal{E}(A) &= 1, \quad \text{via } R \to A, \\
\mathcal{E}(B) &= 1, \quad \text{via } R \to B, \\
\mathcal{E}(C) &= 2, \quad \text{via } R \to A \to C, \\
\mathcal{E}(E) &= 2, \quad \text{via } R \to B \to E, \\
\mathcal{E}(D) &= 3, \quad \text{via } R \to A \to C \to D.
\end{aligned}
$$

Next, we count the number $Q(v)$ of distinct simple paths from R to each vertex (avoiding repeated vertices), taking into account redundant links:

- R**:** $Q(R) = 1$ (trivially).
- A**:**
  - $R \to A$ (Edge 1),
  - $R \to B \to A$ (Edges 2 and 3).

  Hence, $Q(A) = 2$.
- B**:** $Q(B) = 1$ via $R \to B$ (Edge 2).
- C**:**
  - $R \to A \to C$ (Edges 1 and 4),
  - $R \to B \to A \to C$ (Edges 2, 3, and 4).

  Thus, $Q(C) = 2$.
- D**:**
  - $R \to A \to C \to D$ (Edges 1, 4, and 7),
  - $R \to B \to A \to C \to D$ (Edges 2, 3, 4, and 7).

  So, $Q(D) = 2$.
- E**:** E can be reached via several routes:
  1. $R \to B \to E$ (Edges 2 and 5),
  2. $R \to A \to C \to E$ (Edges 1, 4, and 8),
  3. $R \to B \to A \to C \to E$ (Edges 2, 3, 4, and 8),

4. $R \to A \to C \to D \to E$ (Edges 1, 4, 7, and 9),
5. $R \to B \to A \to C \to D \to E$ (Edges 2, 3, 4, 7, and 9),
6. $R \to A \to C \to D \to B \to E$ (Edges 1, 4, 7, 6, and 5).

Therefore, $Q(E) = 6$.

The total number of simple paths is:

$$T = \sum_{v \in V} Q(v) = 1 + 2 + 1 + 2 + 2 + 6 = 14.$$

Probability Distribution and Shannon Entropy

We assign each vertex a probability:

$$p(v) = \frac{Q(v)}{T}.$$

Thus, we have:

$$p(R) = \frac{1}{14}, \quad p(A) = \frac{2}{14}, \quad p(B) = \frac{1}{14}, \quad p(C) = \frac{2}{14}, \quad p(D) = \frac{2}{14}, \quad p(E) = \frac{6}{14}.$$

The Shannon entropy of the graph is then

$$H_G = -\sum_{v \in V} p(v) \log p(v).$$

Substituting the values:

$$H_G = -\left(\frac{1}{14}\log\frac{1}{14} + \frac{1}{7}\log\frac{1}{7} + \frac{1}{14}\log\frac{1}{14} + \frac{1}{7}\log\frac{1}{7} + \frac{1}{7}\log\frac{1}{7} + \frac{3}{7}\log\frac{3}{7}\right) \approx 1.5746$$

If G were a tree (with a unique path to each vertex), the maximum entropy for 6 vertices would be:

$$H_{\text{Tree}} = \log 6 \approx 1.7918$$

Thus, the redundancy measure is

$$\phi_G = H_{\text{Tree}} - H_G \approx 1.7918 - 1.5746 \approx 0.2172$$

The calculations above show that redundant links (e.g., the alternative pathway $R \to B \to A$ for reaching A) result in a probability distribution that is more skewed compared to a uniform (tree) distribution. Consequently, the Shannon entropy H_G is lower than the maximum entropy H_{Tree} achievable in a non-redundant network. The redundancy measure $\phi_G \approx 0.2172$ quantitatively reflects the impact of these redundant pathways on the overall uncertainty of energy flow in the food web.This example underscores how an entropy-based approach can capture the structural complexity of ecological networks, complementing the framework developed in Allesina et al. (2009) [5].

References

[1] Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423.

[2] Paine, R. T. (1966). Food web complexity and species diversity. *American Naturalist*, 100(910), 65–75.

[3] Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20), 12917–12922.

[4] Allesina, S., & Tang, S. (2012). Stability criteria for complex ecosystems. *Nature*, 483(7388), 205–208.

[5] Allesina, S., Bodini, A. & Pascual, M. (2009). Functional links and robustness in food webs. *Philosophical Transactions of the Royal Society B*, 364, 1701–1709. doi:10.1098/rstb.2008.0214.