Link Prediction in Ecological Networks using WLNM with Directed Graph Support and Ecological Sampling

**Student:** *Jorge Eduardo Castro Cruces*

**Supervisor:** *Athen Ho Ma*

**Overview:** This document describes the methodology and results of applying the Weisfeiler-Lehman Neural Machine (WLNM) for link prediction across various ecosystem types, incorporating directed graph support and ecologically-informed negative sampling strategies.

# Introduction

Ecological networks, particularly food webs, represent the complex interactions among species through predation relationships. Understanding the structure and dynamics of these networks is essential for addressing fundamental ecological questions, such as energy flow, ecosystem stability, and resilience to perturbations. However, many empirical food webs remain incomplete due to observational limitations, leading to missing links that obscure our understanding of ecosystem functioning.

Link prediction methods offer a computational approach to infer these missing interactions by leveraging network topology and domain-specific knowledge. While recent advances in network science have introduced powerful machine learning techniques for link prediction, their application to ecological systems has been limited. In particular, existing approaches often neglect important characteristics of food webs, such as directionality of trophic interactions and ecological constraints like feeding roles or body mass hierarchies.

In this work, it is applied and extended the Weisfeiler-Lehman Neural Machine (WLNM), a subgraph-based learning model originally proposed for social and information networks, to the domain of ecological networks. Contributions are threefold:

1. **Directed Graph Support**: Modification of the WLNM framework to preserve and exploit directionality in trophic interactions, which is critical for modeling ecological flows.
2. **Ecological Negative Sampling**: Introduction to a biologically informed negative sampling strategy that filters implausible interactions based on species roles, improving the interpretability of predictions.
3. **Cross-Ecosystem Evaluation**: Evaluation of the model across a diverse set of food webs spanning multiple ecosystem types, providing a comprehensive assessment of its predictive performance.

This study demonstrates that integrating domain-specific constraints into modern link prediction frameworks enhances their ecological relevance and offers a scalable tool for exploring incomplete or uncertain food web data.

# Methodology

## Dataset Preparation

We evaluated the proposed method on a diverse set of ecological food webs representing distinct ecosystem types, including freshwater, marine, estuarine, and terrestrial environments. Each food web was represented as a directed binary trophic matrix where denotes a predation link from species (prey) to species (predator).

Accompanying each food web, we utilized metadata on species taxonomy and body mass (where available), and an ecological role classification (consumer, resource, or mixed), used to guide ecologically plausible negative sampling.

## Link Prediction Framework

We employed the Weisfeiler-Lehman Neural Machine (WLNM) framework, modified to handle **directed networks**, **ecologically-informed negative sampling**, and **species metadata**. The prediction pipeline consists of subgraph extraction, vector encoding, classification, and evaluation, described as follows.

### Data Splitting and Strategies

Each network was randomly partitioned into training and testing subsets, preserving 80% of links for training. An optional degree-based strategy was incorporated, assigning nodes to training or testing partitions based on total degree, enabling the assessment of high-to-low or low-to-high generalization regimes.

The splitting function (DivideNet.m) ensures no overlap between training and testing links, removes self-loops, and generates sparse binary adjacency matrices for each subset.

### Ecological Negative Sampling

To ensure ecological realism, we constrained the sampling of negative links using species roles. Only biologically implausible links — such as consumer-to-consumer or resource-to-resource — were considered valid negatives. This filter reduces false positives arising from structurally possible but ecologically implausible interactions.

The sampling routine (sample\_neg.m) generates both training and testing negatives in a balanced ratio with positive links (by default, 1:1). For full unseen evaluation, the method optionally predicts over all links absent in the training matrix.

### Subgraph Extraction and Encoding

For each candidate link, an **enclosing subgraph** was extracted via a breadth-first expansion over at most two hops. The process collects up to nodes surrounding the source-target pair, generating a directed induced subgraph.

This subgraph is then transformed into a **fixed-length feature vector** via a sequence of operations:

* Canonical node ordering using a Weisfeiler-Lehman color refinement and canonical labeling (g\_label.m).
* Weighted adjacency extraction, where link weights decay with expansion depth.
* Upper-triangular flattening into a vector of dimension .

These steps are implemented in graph2vector.m and subgraph2vector.m.

### Neural Classifier

We used a feedforward neural network to classify each subgraph as a link (1) or non-link (0). The architecture includes:

* Input layer: dimension
* Three hidden layers: [32 → 32 → 16 units], with ReLU activations
* Output layer: Softmax over two classes (link/non-link)

Training was performed using stochastic gradient descent with momentum (sgdm) over 200 epochs, with a mini-batch size of 128. The model was trained exclusively on the training subgraphs and evaluated on a disjoint test set.

## Evaluation Metrics and Logging

Model performance was assessed using the **Area Under the ROC Curve (AUC)**, as well as classification metrics computed at an optimal decision threshold:

* Precision
* Recall
* F1 Score

The optimal threshold was selected to maximize F1 score over a range . For each experiment, true positive (TP), false positive (FP), and false negative (FN) links were saved along with taxonomic and body mass metadata.

All experiments were repeated across randomized splits, and results were logged to CSV files for reproducibility. Intermediate outputs such as scores, subgraph encodings, and confusion matrices were stored in structured folders for post-hoc analysis.

# Results

## Predictive Performance Across Ecosystems

We evaluated the WLNM model on 30+ food webs representing diverse ecological systems. The model consistently achieved high predictive performance across datasets, with **mean AUC values exceeding 0.90** for the majority of networks (Figure 1). Notably, several food webs such as Brook Trout Lake, Grand Caricaie, and AEM17 achieved **perfect or near-perfect AUC (1.00 ± 0.00)**, indicating exceptional discriminative ability.

AUC variability was more pronounced in mid-sized and sparse food webs such as CGP1 and Broad Stream, likely due to reduced link density and higher noise-to-signal ratios. These results suggest that WLNM maintains robust performance even in relatively challenging ecological scenarios.

A graph of different colored bars

AI-generated content may be incorrect.**Figure 1:** AUC score distribution per food web (mean ± min/max across 10 runs)

## Precision Score Distribution

In addition to AUC, we evaluated **precision scores** across 10 randomized runs (Figure 2). While AUC reflects global ranking quality, precision emphasizes how often high-confidence predictions are correct — a critical metric for ecological applications where false positives can mislead inference.

* **High precision** was observed in food webs such as Brook Trout Lake, AEM17, and Grand Caricaie, which achieved **mean precision scores of 1.00**, indicating near-perfect predictions at the chosen threshold.
* **Lower and more variable precision** was observed in webs like CGP1 and Broad Stream, where standard deviation exceeded 0.3. This suggests the presence of highly ambiguous or noisy interactions, challenging the classifier’s confidence calibration.

These results affirm that **ecological metadata and role-aware sampling** reduce spurious link predictions in well-structured food webs.

A graph of different colored bars

AI-generated content may be incorrect.**Figure 2:** Mean precision score across networks (± standard deviation)

## Precision–Recall Trade-offs

Precision–recall curves (Figure 3) highlight the model's sensitivity across the full recall spectrum. Food webs with strong trophic hierarchies, such as SF12, GUBP3, and Indian Lake, exhibited **steep precision drop-offs at high recall**, indicating difficulty in recovering less obvious interactions without increasing false positives.

Conversely, ecosystems like Brook Trout Lake and Ythan Estuary maintained high precision even at high recall, underscoring the model’s ability to generalize well in structurally cohesive networks.

We also compared performance under two test regimes:

* **Unseen=false** (solid lines): test links sampled randomly
* **Unseen=true** (dashed lines): test links sampled among all non-observed links

The model performed better under unseen=false as expected, but still retained reasonable generalization under the more challenging full-unseen setup.

A graph of a graph

AI-generated content may be incorrect.

A graph of a graph

AI-generated content may be incorrect.**Figure 3:** Precision–Recall curves for all food webs; solid = random test links, dashed = full unseen

## Species Most Prone to False Positives

To interpret model behavior, we analyzed **false positive links by species** (Figure 4). Across many networks, species with high generality (e.g., omnivores or basal resources) were overrepresented in FP predictions. For instance:

* In Twin Lake East, species like Disphania ambigua appeared frequently in FP links.
* In CGP1, generalist predators like Lota lota and Sander vitreus accounted for many FP interactions.
* Networks like Brook Trout Lake and Indian Lake showed **no significant FP-prone nodes**, aligning with their high overall performance.

This analysis suggests that **model overconfidence** in well-connected or taxonomically ambiguous species may inflate FP rates, despite strong global metrics.

## Structural Discrepancies: Adjacency vs Predation Matrices

To assess the alignment between inferred links and ecological realism, we compared each food web’s **adjacency matrix** (raw predicted links) with its **predation matrix** (TP links recovered from the test set).

Figure 5 displays these pairwise comparisons, highlighting how many of the model’s predictions correspond to biologically validated interactions:

* In structured webs like Brook Trout Lake or Indian Lake, the adjacency matrix aligns tightly with the TP matrix — most predicted links matched actual predation relationships.
* In contrast, food webs like CGP1 or Dempsters Stream exhibit a **larger set of false positives**, where adjacency matrices contain additional links not supported by test data. This may be due to:
  + Weak trophic stratification
  + Sparse training data
  + Role ambiguity (e.g., species with dual consumer/resource roles)

This analysis underscores the need to incorporate **biological constraints during training** and supports our approach of using **role-filtered negative sampling**.

A chart of different colored bars

AI-generated content may be incorrect.**Figure 4:** Top species by false positive frequency across networks

A screenshot of a graph

AI-generated content may be incorrect.

A paper with black dots

AI-generated content may be incorrect.

A graph of a number of graphs

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

**Figure 5:** Adjacency vs Predation matrix visual comparison per food web

# Discussion

This study demonstrates the effectiveness of a graph-based neural link prediction approach, WLNM, when applied to a broad range of ecological networks. By extending the original framework with domain-informed enhancements — including directed graph support and ecologically-constrained negative sampling — we achieved high predictive performance across a wide spectrum of food webs.

## Accuracy and Robustness Across Ecosystems

Our results show that WLNM performs consistently well, with AUC and precision scores exceeding 0.90 in most cases. Ecosystems with well-defined trophic hierarchies and higher connectance (e.g., Brook Trout Lake, AEM17) yielded perfect or near-perfect classification metrics. These results suggest that the combination of structural information and local subgraph encoding captures essential patterns of trophic interaction, particularly in densely observed systems.

More variable performance was observed in certain networks, such as CGP1 and Broad Stream, where AUC and precision scores exhibited higher standard deviation. These discrepancies are likely attributable to a combination of factors, including network sparsity, the presence of poorly resolved omnivory, and a limited number of well-defined consumer-resource pairs. The model’s reliance on structural features alone may be insufficient in such cases, emphasizing the importance of integrating additional ecological priors (e.g., habitat type, functional group).

## Ecological Relevance of Predictions

A critical challenge in ecological prediction lies not only in identifying potential links but ensuring that predicted interactions are biologically meaningful. To address this, we introduced a role-aware negative sampling strategy that constrained link prediction to ecologically plausible candidate pairs.

This strategy significantly reduced the rate of false positives, especially in species with ambiguous feeding roles. Furthermore, our analysis of adjacency versus predation matrices confirmed that in high-performing networks, predicted links largely overlap with test-set ground truth. In contrast, networks with noisy metadata or weak trophic structure exhibited a higher proportion of spurious links — a reminder that link prediction must be interpreted through an ecological lens.

## Model Interpretability and FP Analysis

While global metrics provide a high-level view of model quality, our false positive (FP) analysis revealed that a disproportionate number of errors are concentrated in a few taxa, often species with high degree, intermediate trophic level, or inconsistent role annotation. These patterns suggest that even sophisticated graph-based models are prone to **topological overfitting** in the absence of strong ecological constraints.

Such findings highlight the value of **metadata-aware interpretability** tools, which not only assess model output but help ecologists identify sources of uncertainty and refine datasets accordingly.

## Limitations and Future Work

Several limitations must be acknowledged. First, while the model incorporates directionality and basic metadata (roles, mass), it does not yet leverage full species traits or spatial co-occurrence data. Incorporating such information — perhaps via node embeddings or hybrid models — could further improve prediction accuracy and ecological plausibility.

Second, the WLNM architecture remains a fixed subgraph classifier. Recent advances in graph neural networks (GNNs), such as GraphSAGE or SEAL, may offer more expressive and scalable alternatives, especially for large or dynamic food webs.

Finally, a critical next step is to evaluate the **transferability** of trained models across ecosystems — that is, can a model trained on one system reliably predict links in another? This would move the approach beyond retrospective validation toward true ecological forecasting.