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

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

***Supervisor:*** *Athen Ho Ma*

**Tables**

TBD

**Figures**

A graph with different colored squares

AI-generated content may be incorrect.

**Figure 1.** This figure illustrates the performance of the Weisfeiler–Lehman Neural Machine (WLNM) when applied to 30 random food webs under different train/test splits. Each boxplot summarizes four evaluation metrics — Recall, AUC, Precision, and F1-Score — for train ratios ranging from 60% to 80%. The results show that while performance is generally stable across ratios, some metrics such as Recall and Precision exhibit variability depending on the fraction of links used for training. This is important because it highlights the trade-off between training data availability and generalization capacity, guiding the choice of appropriate train/test splits for robust link prediction. The plots were generated by running the WLNM pipeline multiple times on randomly selected food webs, recording evaluation metrics for each run, and summarizing them in boxplots to capture central tendencies and variability.

A screen shot of a chart

AI-generated content may be incorrect.

**Figure 2.** This figure shows how the Weisfeiler–Lehman Neural Machine (WLNM) performs across five ecosystem types when trained with 80% of observed links. Each boxplot represents the distribution of four evaluation metrics — AUC, Precision, Recall, and F1-Score — for a collection of food webs belonging to the same ecosystem category. The comparison reveals that performance is not uniform across ecosystems: for example, lakes and streams tend to show tighter distributions with higher average values, while marine food webs exhibit greater variability. This is important because it demonstrates that predictive accuracy depends not only on model configuration but also on the structural properties of the ecosystems themselves, such as food web size and complexity. The figure was generated by running the WLNM pipeline on subsets of food webs grouped by ecosystem, computing evaluation metrics for each run, and aggregating them into boxplots to visualize central tendency and variability.

A graph with different colored squares

AI-generated content may be incorrect.

**Figure 3.** This figure summarizes the performance of the Weisfeiler–Lehman Neural Machine (WLNM) across all 290 food webs using an 80/20 train/test split. Each boxplot displays the distribution of four evaluation metrics — AUC, Recall, F1-Score, and Precision. The results indicate that WLNM achieves consistently high AUC values, suggesting strong discriminative ability, while Recall and Precision show greater variability across food webs. This is important because it highlights the trade-off between sensitivity (capturing true links) and specificity (avoiding false positives), and how performance can fluctuate depending on network size and structure. The figure was generated by training WLNM on 80% of each food web’s observed links, testing on the remaining 20%, and aggregating evaluation metrics from all runs into boxplots for visualization.

A screenshot of a graph paper

AI-generated content may be incorrect.

**Figure 4.** This figure compares the observed and predicted feeding interactions in the Baxver lake food web. The left panel shows the original adjacency matrix, where each black dot corresponds to an observed predator–prey link, with species ordered by body mass. The right panel overlays the model’s predictions on the same structure: green dots highlight true positives (links correctly predicted), red squares indicate false positives (predicted but not observed), black dots represent training links, and grey crosses mark false negatives (observed but missed). This comparison is important because it provides a visual validation of the model’s performance at the level of individual interactions, showing not only overall accuracy but also where systematic errors occur. The matrices were generated by training the Weisfeiler–Lehman Neural Machine (WLNM) on a subset of links from this food web, then testing the model’s predictions against the withheld interactions, with results displayed in a predator-by-prey grid.

A screenshot of a graph

AI-generated content may be incorrect.

**Figure 5.** This figure presents a comparison between the observed and predicted predator–prey interactions in the TPP food web. On the left, the adjacency matrix displays all observed feeding links as black dots, with prey on the vertical axis and predators on the horizontal axis ordered by body mass. On the right, the predation matrix shows WLNM’s predictions relative to the observed data: true positives (green dots) indicate correctly predicted links, false positives (red squares) are predicted but not observed, training links (black dots) are those used for model fitting, and false negatives (grey crosses) represent observed interactions that the model failed to predict. This visualization is important because it reveals the model’s strengths and weaknesses at the interaction level, making it possible to identify systematic prediction patterns such as clusters of missed links or over-predictions. The figure was generated by training WLNM on a subset of interactions and testing its predictions against withheld links from the same food web.

A screenshot of a graph

AI-generated content may be incorrect.

**Figure 6.** This figure compares observed and predicted feeding interactions for the Isolated stream 877 August 2003 food web. The left panel shows the adjacency matrix, with black dots indicating observed predator–prey links arranged by body mass. The right panel overlays WLNM’s predictions: true positives (green dots) are correctly predicted links, false positives (red squares) are predicted but not observed, black dots denote links used in training, and grey crosses indicate observed interactions the model failed to predict. This visualization is important as it highlights how effectively the model captures the feeding structure of this stream ecosystem, while also revealing systematic errors, such as clusters of missed or over-predicted links. The matrices were generated by training WLNM on a portion of the observed interactions and validating predictions against withheld links.

A screenshot of a graph

AI-generated content may be incorrect.

**Figure 7.** This figure presents the observed and predicted feeding interactions in the Big Hopu lake food web. The left panel shows the adjacency matrix of observed interactions, where black dots represent predator–prey links arranged by species body mass. The right panel overlays WLNM’s predictions: true positives (green dots) indicate links that were correctly predicted, false positives (red squares) represent predicted but unobserved links, training links (black dots) were used to fit the model, and false negatives (grey crosses) show observed interactions that the model failed to recover. This visualization is important because it highlights both the successes and errors of the model in capturing the trophic structure of this lake ecosystem. The figure was generated by training WLNM on a subset of observed interactions and testing its predictions against withheld links.

A graph of different colored bars

AI-generated content may be incorrect.

**Figure 8.** This figure shows the predictive performance of the Weisfeiler–Lehman Neural Machine (WLNM) measured by the Area Under the ROC Curve (AUC) across multiple food webs. Each colored bar represents a food web, with the height indicating the mean AUC score over 10 independent runs. Error bars display the minimum and maximum values obtained, providing a sense of variability and robustness. Most food webs achieve high AUC scores close to 1.0, suggesting strong discriminative ability of the model, while a few show lower values, highlighting networks where predictions are more challenging. This analysis is important because it demonstrates not only average model performance but also its consistency across repeated runs, ensuring reliability of the results. The figure was generated by training and testing WLNM 10 times on each food web and summarizing the distribution of AUC values.

A graph of different colored bars

AI-generated content may be incorrect.

**Figure 9.** This figure presents the Precision scores of the Weisfeiler–Lehman Neural Machine (WLNM) across multiple food webs. Each colored bar corresponds to a single food web, with the bar height showing the mean Precision score over 10 independent runs, while error bars represent the minimum and maximum scores obtained. Precision reflects the proportion of predicted links that are actually correct, meaning high values indicate the model is effective at minimizing false positives. While many food webs reach near-perfect Precision, some exhibit more variability, suggesting that prediction difficulty differs depending on food web structure and size. This analysis is important as it complements recall-based evaluations, ensuring that model predictions are not only sensitive but also specific. The figure was generated by running WLNM 10 times per food web, computing Precision in each case, and summarizing results using mean and range values.

A graph of a graph showing the difference between false and true

AI-generated content may be incorrect.

**Figure 10.** This figure displays Precision–Recall (PR) curves for the AEW17 terrestrial belowground food web (N = 145 nodes). The solid blue line corresponds to links included during training (unseen = false), while the dashed orange line represents predictions on unseen links withheld for testing. PR curves are useful because they show the trade-off between Precision (avoiding false positives) and Recall (capturing true positives) across different classification thresholds. The figure reveals that while WLNM performs strongly on seen interactions, its performance declines on unseen links, reflecting the challenge of generalizing predictions to unobserved parts of the network. This analysis is important because it highlights where the model succeeds and where improvements — such as incorporating rare links or species traits — may be needed. The curves were generated by computing Precision and Recall at varying decision thresholds on both training and testing sets.

A graph with a line

AI-generated content may be incorrect.

**Figure 11.** This figure shows Precision–Recall (PR) curves for the AEW03 terrestrial belowground food web (N = 122 nodes). The solid blue line reflects model performance on links included in the training data (unseen = false), while the dashed orange line shows predictions on links withheld during testing (unseen = true). PR curves reveal the balance between Precision — the proportion of predicted links that are correct — and Recall — the proportion of observed links recovered — across classification thresholds. The results indicate that WLNM achieves consistently high Precision and Recall on training data, but its performance drops substantially on unseen links. This highlights the challenge of generalizing link prediction to unobserved interactions in large, complex food webs. The curves were generated by evaluating model predictions over a range of thresholds for both training and test sets.

A diagram of a positive sublocation

AI-generated content may be incorrect.

**Figure 12.** This figure shows the two most frequent positive enclosing subgraphs identified during the WLNM encoding process. Each subgraph centers on a target link (highlighted by the red diamond nodes), with orange circular nodes representing neighboring species included in the enclosing structure. The left subgraph (count = 170) and the right subgraph (count = 174) represent recurring local patterns around predator–prey interactions labeled as positive. Visualizing enclosing subgraphs is important because they reveal the building blocks used by WLNM to learn predictive features: they capture the structural context of interactions, including shared neighbors and connectivity patterns. By analyzing these subgraphs, we can better understand why the model classifies certain pairs as likely interacting. These subgraphs were generated by extracting local neighborhoods around positive links and encoding them with the Weisfeiler–Lehman labeling procedure.

A screenshot of a diagram

AI-generated content may be incorrect.

**Figure 13.** This figure displays two examples of enclosing subgraphs corresponding to negative samples, where the central pair of species (shown as red diamond nodes) does not represent an observed predator–prey interaction. The surrounding orange circular nodes represent the local neighborhood extracted to form the enclosing subgraph. These negative subgraphs are important because they provide contrastive training examples for the Weisfeiler–Lehman Neural Machine (WLNM), allowing the model to learn not only the structural patterns of true interactions but also those of non-interacting pairs. The two examples shown here were randomly selected from the pool of negative samples, highlighting the structural diversity of non-links in food webs. These subgraphs were generated by sampling species pairs without observed interactions, extracting their surrounding neighborhood, and encoding them using the Weisfeiler–Lehman labeling procedure.

A map of the world

AI-generated content may be incorrect.

**Figure 14.** This figure presents a global map showing the geographic distribution of the 290 food webs analyzed in this study. Each point corresponds to a food web location, with color-coded symbols indicating the ecosystem type (terrestrial aboveground, streams, lakes, marine, or terrestrial belowground). The map highlights the strong geographic diversity of the dataset, spanning multiple continents and biomes, which ensures that the results are not biased toward a single region or ecosystem. This diversity is important because it allows the model to be tested under a wide range of ecological contexts, improving the robustness and generality of the findings. The map was generated by geocoding the food web datasets and visualizing them with a geographic information system (GIS) tool, using consistent color and symbol coding to distinguish ecosystem categories.

A black background with white squares and dots

AI-generated content may be incorrect.

**Figure 15.** This figure illustrates the full workflow of the Weisfeiler–Lehman Neural Machine (WLNM) pipeline as applied to food web link prediction. The process begins with splitting the network into training and test sets, followed by negative sampling where ecologically meaningful non-links are selected. Next, enclosing subgraphs are extracted around candidate pairs, encoded using Weisfeiler–Lehman relabeling, and transformed into feature vectors. These representations are then used to train and evaluate a machine learning classifier. The flowchart is important because it provides a clear and systematic overview of how the pipeline operates, showing where ecological constraints (e.g., body mass filtering in negative sampling) are introduced and how subgraph-level encoding leads into predictive modeling. The diagram was generated by formalizing the methodological steps into a process flow to ensure clarity and reproducibility.

**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. 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.

**Precision Score Distribution**

In addition to AUC, we evaluated **precision scores** across 10 randomized runs. 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.

**Precision–Recall Trade-offs**

Precision–recall curves 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.

**Species Most Prone to False Positives**

To interpret model behavior, we analyzed **false positive links by species**. 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**.

**Methodology**

We develop a WLNM-based link prediction framework tailored to **directed trophic networks**. The pipeline (Fig. 15) integrates ecological priors during data splitting and negative sampling, encodes local directed structure via WL canonicalization, and evaluates models with metrics suited to class imbalance.

**Problem setting and notation**

Let a food web be a directed graph with adjacency . Nodes represent taxa; edges denote a trophic interaction . Each node optionally carries metadata: role , body mass , and name . Given a training graph and a candidate pair set , the task is to estimate .

**Overview of the pipeline (Fig. 15)**

1. **Inputs.** A list of food webs and, for each, a directed adjacency matrix with optional node metadata (role, mass, taxonomy).
2. **Train–test partitioning.** For each food web, we generate one or more train/test splits under a chosen strategy:
   * **Random**: uniformly sample test edges from .
   * **Degree-stratified**: “high→low” or “low→high”, partitioning nodes by total degree to stress generalization across degree regimes.
   * **Rare-link**: prioritize edges whose endpoints have low total degree; a fraction of such “rare” links is forced into training to mitigate extreme sparsity.

Optionally, a **connectivity safeguard** rejects test removals that break reachability from prey to predator; we disable this on very small graphs for efficiency.

1. **Negative sampling with ecological constraints.** We form non-edge pairs as negatives. To reduce biologically implausible pairs, we apply a **role-compatibility filter** (e.g., or , consistent with directed trophic flow) before random sampling. We balance the positive/negative ratio by a factor aa and report when scarcity forces relaxation.
2. **Local subgraph extraction and canonicalization.** For each candidate pair , we extract a **directed enclosing subgraph** up to a small depth (two hops from the endpoints; capped at nodes). We then (i) remove the direct edge between endpoints (to avoid label leakage), (ii) apply **Weisfeiler–Lehman (WL) color refinement** to derive stable node colors, and (iii) compute a **canonical labeling** consistent with the WL colors. The canonically ordered subgraph is vectorized by taking the upper-triangular entries of a symmetrized, distance-weighted adjacency, yielding a fixed-length representation of size .
3. **Classifier.** We use a small feed-forward network on the subgraph vectors to score candidate links. The architecture is lightweight (three ReLU blocks and a softmax head) to keep the focus on structural encoding rather than high-capacity modeling.
4. **Model selection and evaluation.** We compute ROC–AUC and **Average Precision (AP)** (recommended under class imbalance). A decision threshold is selected on the test scores via a predefined grid to report Precision/Recall/F1 (the thresholding step is post-hoc and does not influence AUC/AP). We aggregate metrics across food webs and, when relevant, by ecosystem type.

**Design choices (why these pieces?)**

* **Directed locality**: Enclosing subgraphs centered at preserve asymmetry essential to trophic flow and reduce confounding triangles from undirected symmetrization.
* **WL canonicalization**: Canonical ordering eliminates permutations, allowing a compact fixed-length representation without resorting to heavy graph neural nets.
* **Ecology-aware negatives**: Role filtering reduces obviously impossible pairs, improving learning signal and ecological plausibility.
* **Split strategies**: Degree-stratified and rare-link splits probe robustness beyond uniformly sampled links—closer to realistic “hard” discoveries in sparse food webs.

**Train–test protocols**

We report results under three protocols:

* **Random split (baseline):** sample the held-out proportion of edges uniformly.
* **Degree-stratified split:** sort nodes by total degree and hold out interactions predominantly crossing high/low partitions (two directions: and ).
* **Rare-link split:** compute s(u,v)=deg(u)+deg(v); up to a fraction ρ\rho of the lowest-score edges are forced into training before filling the remainder uniformly.

For small graphs we disable reachability checks; otherwise, we reject a removal if it destroys a path from prey to predator in the training graph.

**Subgraph encoding details**

We cap the node budget at (typically ) using a bounded, directed expansion (successors of and predecessors of ), then apply WL refinement and canonical relabeling. To stabilize magnitudes, we optionally weight edges by inverse hop distance before symmetrization; the upper-triangular vector is used as input. This yields a permutation-invariant yet simple representation well suited to small MLPs.

**Metrics and reporting**

We report **ROC–AUC** and **AP**; for interpretability we add **Precision/Recall/F1** at a fixed grid-selected threshold. To summarize across heterogeneous webs, we (i) average per-web metrics and (ii) provide **ecosystem-level aggregates** with mean ± s.d. (or 95% CIs via bootstrap) to reflect variability across systems.

**Reproducibility and ablations (for Supplementary Material)**

* Exact hyperparameters, threshold grid, and training settings.
* Variants: undirected “original WLNM” extraction vs. our directed version; with/without role-filtered negatives; with/without connectivity safeguard; sensitivity to , , and .
* Runtime scaling with and .

**Algorithm 1 — High-level training/evaluation loop**

**Input:** Food-web set , split strategy , subgraph size , train ratio , rare fraction (if rare).  
**Output:** Per-web and aggregated metrics.

1. **for each** food web **do**
2. Generate ,using strategy (apply reachability safeguard when enabled).
3. Construct positive pairs from and .
4. Sample negative pairs with role-compatibility filtering and ratio .
5. **for each** candidate pair **do**
6. Extract a directed enclosing subgraph (depth ≤ 2, ≤ nodes); remove if present.
7. Apply WL refinement and canonical labeling; vectorize the ordered subgraph.
8. **end for**
9. Train a small MLP on training vectors; score test vectors.
10. Compute ROC–AUC, AP; select a reporting threshold to compute Precision/Recall/F1.
11. **end for**
12. Aggregate metrics across food webs and, if applicable, by ecosystem.

**Threshold Optimization and Evaluation**

We evaluate model predictions on the test set using:

* **AUC (Area Under the ROC Curve)**,
* **Precision**, **Recall**, and **F1-score** across varying thresholds,
* Optimal classification threshold is selected to maximize F1-score.

Predicted vs. actual links are compared to generate:

* Confusion matrices with **True Positives (TP), False Positives (FP)**, and **False Negatives (FN)**,
* CSV reports including species names and body masses for interpretability.

**Parallelization and Logging**

* The entire pipeline can be parallelized across experiments using MATLAB's *parfor*.
* Results are logged in two formats:
  + Terminal logs (for debugging),
  + CSV result logs (for analysis and figure generation).

**Summary of Key Innovations**

|  |  |
| --- | --- |
| **Feature** | **Description** |
| **Directed Subgraphs** | Subgraph extraction and encoding maintain link directionality. |
| **Role-Based Negative Sampling** | Avoids biologically implausible negatives using ecological role labels. |
| **WL Labeling + Canonical Encoding** | Enables isomorphism-invariant link representation. |
| **Species Metadata Integration** | Supports downstream ecological interpretation. |
| **TP/FP/FN Export** | Enables detailed inspection of model errors at the species level. |

**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.

**List of all WLNM modifications**

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.

**References**

Abbasi Moud, A. (2022). Recent advances in utility of artificial intelligence towards multiscale colloidal based materials design. *Colloid and Interface Science Communications*, *47*, 100595. https://doi.org/10.1016/j.colcom.2022.100595

Aminzadeh, B., & Khansefid, M. (2010). A case study of urban ecological networks and a sustainable city: Tehran’s metropolitan area. *Urban Ecosystems*, *13*(1), 23–36. https://doi.org/10.1007/s11252-009-0101-3

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, *438*(7066), 303–309. https://doi.org/10.1038/nature04141

Bi, Z., Sun, J., Xie, Y., Gu, Y., Zhang, H., Zheng, B., Ou, R., Liu, G., Li, L., Peng, X., Gao, X., & Wei, N. (2024). Machine learning-driven source identification and ecological risk prediction of heavy metal pollution in cultivated soils. *Journal of Hazardous Materials*, *476*, 135109. https://doi.org/10.1016/j.jhazmat.2024.135109

Brose, U., Archambault, P., Barnes, A. D., Bersier, L. F., Boy, T., Canning-Clode, J., Conti, E., Dias, M., Digel, C., Dissanayake, A., Flores, A. A. V., Fussmann, K., Gauzens, B., Gray, C., Häussler, J., Hirt, M. R., Jacob, U., Jochum, M., Kéfi, S., … Iles, A. C. (2019). Predator traits determine food-web architecture across ecosystems. *Nature Ecology and Evolution*, *3*(6), 919–927. https://doi.org/10.1038/s41559-019-0899-x

Burkle, L. A., Marlin, J. C., & Knight, T. M. (2013). Plant-Pollinator Interactions over 120 Years: Loss of Species, Co-Occurrence, and Function. *Science*, *339*(6127), 1611–1615. https://doi.org/10.1126/science.1232728

Camerano, L. (1880). *Dell’equilibrio dei viventi merce la reciproca distruzione*. stamperia reale della ditta GB Paravia e Comp.

CHEN Jianyang, YIN Haiwei, KONG Fanhua, & YAO Guipeng. (2015). Study on the complex eco-network development around Taihu Lake, China. *Acta Ecologica Sinica*, *35*(9). https://doi.org/10.5846/stxb201306101647

Darwin, C. (1959). The Origin of Species, 1859-1959. *Bios*, *30*(2), 67–72.

Du, C., Cui, J., Wang, D., Li, G., Lu, H., Tian, Z., Zhao, C., Li, M., & Zhang, L. (2024). Prediction of aquatic vegetation growth under ecological recharge based on machine learning and remote sensing. *Journal of Cleaner Production*, *452*, 142054. https://doi.org/10.1016/j.jclepro.2024.142054

Egerton, F. N. (2016). Part 57: Aspects of Limnology in America, 1930s to about 1990, Led by Hutchinson and Hasler. *The Bulletin of the Ecological Society of America*, *97*(3), 228–282. https://doi.org/10.1002/bes2.1242

Forup, M. L., Henson, K. S. E., Craze, P. G., & Memmott, J. (2008). The restoration of ecological interactions: plant–pollinator networks on ancient and restored heathlands. *Journal of Applied Ecology*, *45*(3), 742–752. https://doi.org/10.1111/j.1365-2664.2007.01390.x

Friberg, N., Bonada, N., Bradley, D. C., Dunbar, M. J., Edwards, F. K., Grey, J., Hayes, R. B., Hildrew, A. G., Lamouroux, N., Trimmer, M., & Woodward, G. (2011). *Biomonitoring of Human Impacts in Freshwater Ecosystems* (pp. 1–68). https://doi.org/10.1016/B978-0-12-374794-5.00001-8

Gao, X., Liang, X., Guo, Z., Yuan, Z., Li, G., & Likulunga, L. E. (2024). The analysis of ecological network characteristics of Baiyangdian Basin in China using the complex network theory. *Ecological Indicators*, *167*, 112650. https://doi.org/10.1016/j.ecolind.2024.112650

Gray, C., Baird, D. J., Baumgartner, S., Jacob, U., Jenkins, G. B., O’Gorman, E. J., Lu, X., Ma, A., Pocock, M. J. O., Schuwirth, N., Thompson, M., & Woodward, G. (2014). FORUM: Ecological networks: The missing links in biomonitoring science. *Journal of Applied Ecology*, *51*(5), 1444–1449. https://doi.org/10.1111/1365-2664.12300

Green, D. G., & Sadedin, S. (2005). Interactions matter—complexity in landscapes and ecosystems. *Ecological Complexity*, *2*(2), 117–130. https://doi.org/10.1016/j.ecocom.2004.11.006

Heleno, R., Devoto, M., & Pocock, M. (2012). Connectance of species interaction networks and conservation value: Is it any good to be well connected? *Ecological Indicators*, *14*(1), 7–10. https://doi.org/10.1016/j.ecolind.2011.06.032

Hill, S. L., Keeble, K., Atkinson, A., & Murphy, E. J. (2012). A foodweb model to explore uncertainties in the South Georgia shelf pelagic ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanography*, *59–60*, 237–252. https://doi.org/10.1016/j.dsr2.2011.09.001

Huettmann, F., Craig, E. H., Herrick, K. A., Baltensperger, A. P., Humphries, G. R. W., Lieske, D. J., Miller, K., Mullet, T. C., Oppel, S., Resendiz, C., Rutzen, I., Schmid, M. S., Suwal, M. K., & Young, B. D. (2018). Use of Machine Learning (ML) for Predicting and Analyzing Ecological and ‘Presence Only’ Data: An Overview of Applications and a Good Outlook. In *Machine Learning for Ecology and Sustainable Natural Resource Management* (pp. 27–61). Springer International Publishing. https://doi.org/10.1007/978-3-319-96978-7\_2

HUTCHINSON, H. W. (1958). Village and Plantation Life in Northeastern Brazil. *Soil Science*, *85*(1), 56. https://doi.org/10.1097/00010694-195801000-00019

Ignatieva, M., Stewart, G. H., & Meurk, C. (2011). Planning and design of ecological networks in urban areas. *Landscape and Ecological Engineering*, *7*(1), 17–25. https://doi.org/10.1007/s11355-010-0143-y

Ings, T. C., Montoya, J. M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C. F., Edwards, F., Figueroa, D., Jacob, U., Jones, J. I., Lauridsen, R. B., Ledger, M. E., Lewis, H. M., Olesen, J. M., Van Veen, F. J. F., Warren, P. H., & Woodward, G. (2009). Review: Ecological networks – beyond food webs. *Journal of Animal Ecology*, *78*(1), 253–269. https://doi.org/10.1111/j.1365-2656.2008.01460.x

Kazemi, S. M., Goel, R., Jain, K., Kobyzev, I., Sethi, A., Forsyth, P., & Poupart, P. (2019). *Representation Learning for Dynamic Graphs: A Survey*.

Kremen, C. (2005). Managing ecosystem services: what do we need to know about their ecology? *Ecology Letters*, *8*(5), 468–479. https://doi.org/10.1111/j.1461-0248.2005.00751.x

Layer, K., Hildrew, A. G., Jenkins, G. B., Riede, J. O., Rossiter, S. J., Townsend, C. R., & Woodward, G. (2011). *Long-Term Dynamics of a Well-Characterised Food Web* (pp. 69–117). https://doi.org/10.1016/B978-0-12-374794-5.00002-X

Layer, K., Hildrew, A. G., & Woodward, G. (2013). Grazing and detritivory in 20 stream food webs across a broad pH gradient. *Oecologia*, *171*(2), 459–471. https://doi.org/10.1007/s00442-012-2421-x

Layer, K., Riede, J. O., Hildrew, A. G., & Woodward, G. (2010). *Food Web Structure and Stability in 20 Streams Across a Wide pH Gradient* (pp. 265–299). https://doi.org/10.1016/B978-0-12-381363-3.00005-8

Ledger, M. E., Brown, L. E., Edwards, F. K., Hudson, L. N., Milner, A. M., & Woodward, G. (2013). *Extreme Climatic Events Alter Aquatic Food Webs* (pp. 343–395). https://doi.org/10.1016/B978-0-12-417199-2.00006-9

Leman, A. (2018). THE REDUCTION OF A GRAPH TO CANONICAL FORM AND THE ALGEBRA WHICH APPEARS THEREIN. *Null*. https://doi.org/null

Liu, H., Niu, T., Yu, Q., Yang, L., Ma, J., & Qiu, S. (2022). Evaluation of the Spatiotemporal Evolution of China’s Ecological Spatial Network Function–Structure and Its Pattern Optimization. *Remote Sensing*, *14*(18), 4593. https://doi.org/10.3390/rs14184593

Malmqvist, T. (2008). Environmental rating methods: selecting indoor environmental quality (IEQ) aspects and indicators. *Building Research & Information*, *36*(5), 466–485. https://doi.org/10.1080/09613210802075841

McCann, K. S. (2000). The diversity–stability debate. *Nature*, *405*(6783), 228–233. https://doi.org/10.1038/35012234

Montoya, J. M., Pimm, S. L., & Solé, R. V. (2006). Ecological networks and their fragility. *Nature*, *442*(7100), 259–264. https://doi.org/10.1038/nature04927

Reddy, S. (2024). *Examine the effects of climate change in an ecological network using machine learning*. Queen Mary University of London.

Skarding, J., Gabrys, B., & Musial, K. (2021). Foundations and Modeling of Dynamic Networks Using Dynamic Graph Neural Networks: A Survey. *IEEE Access*, *9*, 79143–79168. https://doi.org/10.1109/ACCESS.2021.3082932

Solomon, S. (2007). IPCC (2007): Climate Change The Physical Science Basis. *AGU Fall Meeting Abstracts*, *2007*, U43D-01.

Swenson, S. C., & Milly, P. C. D. (2006). Climate model biases in seasonality of continental water storage revealed by satellite gravimetry. *Water Resources Research*, *42*(3). https://doi.org/10.1029/2005WR004628

Thomas, J. M., Beddar-Wiesing, S., Moallemy-Oureh, A., & Nather, R. (2021). *A Note on the Modeling Power of Different Graph Types*.

Thompson, R. M., Brose, U., Dunne, J. A., Hall, R. O., Hladyz, S., Kitching, R. L., Martinez, N. D., Rantala, H., Romanuk, T. N., Stouffer, D. B., & Tylianakis, J. M. (2012). Food webs: reconciling the structure and function of biodiversity. *Trends in Ecology & Evolution*, *27*(12), 689–697. https://doi.org/10.1016/j.tree.2012.08.005

Tylianakis, J. M., Laliberté, E., Nielsen, A., & Bascompte, J. (2010). Conservation of species interaction networks. *Biological Conservation*, *143*(10), 2270–2279. https://doi.org/10.1016/j.biocon.2009.12.004

Tylianakis, J. M., Tscharntke, T., & Lewis, O. T. (2007). Habitat modification alters the structure of tropical host–parasitoid food webs. *Nature*, *445*(7124), 202–205. https://doi.org/10.1038/nature05429

Webb, L., Whetton, P., & Barlow, E. (2008). Climate change and winegrape quality in Australia. *Climate Research*, *36*, 99–111. https://doi.org/10.3354/cr00740

Zhang, L., Peng, J., Liu, Y., & Wu, J. (2017). Coupling ecosystem services supply and human ecological demand to identify landscape ecological security pattern: A case study in Beijing–Tianjin–Hebei region, China. *Urban Ecosystems*, *20*(3), 701–714. https://doi.org/10.1007/s11252-016-0629-y

Zhang, M., & Chen, Y. (2017). Weisfeiler-lehman neural machine for link prediction. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F129685*, 575–583. https://doi.org/10.1145/3097983.3097996

Zhang, S., Zhang, W., Bu, Z., & Zhang, X. (2024). ClusterLP: A novel Cluster-aware Link Prediction model in undirected and directed graphs. *International Journal of Approximate Reasoning*, *172*, 109216. https://doi.org/10.1016/j.ijar.2024.109216

Zhang, Z., Hu, B., Jiang, W., & Qiu, H. (2023). Spatial and temporal variation and prediction of ecological carrying capacity based on machine learning and PLUS model. *Ecological Indicators*, *154*, 110611. https://doi.org/10.1016/j.ecolind.2023.110611