

Network Analysis of Ecological Food Webs: Insights into Community Structure and Link Prediction

Background:

Food webs in ecological networks often show how various complex interactions between species in the ecosystems occur. These networks often showcase interesting hierarchical community structures, where various types of living organisms form nested groups based on their feeding relationships and ecological roles in the system. Traditional methods of analyzing food webs such as species diversity metrics and Shannon diversity index have often focused on metrics such as species diversity, interconnectedness, and trophic levels. However, these approaches often fail to show the special hierarchical organization that emerges from species interactions such as food. This doesn't account for how the food web impacts the roles of species and interaction between species. Recent advancements in network science, particularly in community detection algorithms and hierarchical clustering methods have provided newer tools to uncover and visualize these nested structures in ecological networks. Understanding these hierarchical patterns is very crucial in understanding and comprehending ecosystem dynamics, predicting responses to future disturbances, and creating new conservation strategies. I believe that we can utilize publicly available resources such as food web datasets from the Ecological Archives by ESA and the Web of Life database to get a better understanding of this. This provides data such as species composition lists, interaction matrices, biomass data, and stable isotope data as well as long term data sets and experimental manipulations that have been done in the past to get a better understanding of the food web and its impact on hierarchical structure in different ecosystems. This can help us get a better understanding of energy flow, nutrient cycling and keystone complexes.

Research Question:

How do network analysis techniques such as community detection, link prediction, and motif analysis reveal patterns and insights within food webs? Specifically, can these methods uncover key interactions and predict potential links in these complex biological networks?

Anticipated Findings:

I expect to uncover key community structures within the food webs, identify central species, and predict missing links. The analysis will also provide insights into the overall network's robustness, the role of certain species, and the stability of ecosystems. We aim to demonstrate how computational techniques can be applied to biological networks to offer valuable ecological insights.

Methods:

Datasets

The dataset that were used for analysis consists of 10 food web datasets.

Canadian Freshwater Systems:

- Aishihik Lake
- Cold Lake
- Lake of the Woods
- McGregor River
- Parsnip River
- Sbay Lake Huron
- Smallwood Reservoir

New Zealand Streams:

- AkatoreA
- AkatoreB
- Venlaw

Data preprocessing involved:

1. Converting adjacency matrices to edge lists
2. Removing duplicate edges
3. Handling missing values
4. Creating directed graph representations

Network Analysis Methods

Basic Network Metrics Calculation

- Node count (species richness)
- Edge count (interaction quantity)
- Degree distribution
- Mean degree
- Clustering coefficient
- Mean geodesic distance (MGD)

Community Detection Implementation

- Applied Louvain algorithm
- Calculated modularity scores
- Analyzed community size distributions

Link Prediction Method

- Implemented Common Neighbors algorithm
- Calculated prediction scores
- Identified top potential interactions

Motif Analysis Methodology

- Sampled network motifs (n=50)
- Detected feed-forward loops (FFL)
- Identified feedback loops (FBL)
- Generated random network ensemble (n=50)
- Calculated statistical significance

Visualization Method

- Spring Layout for to visualize the structure, communities, and key species for each ecosystem..

Hypothesis

Network Metrics:

The two largest datasets that was researched was Lake of the woods and Lake Huron. Lake of the woods consists of 178 nodes and 4672 edges specifying 178 species and 4672 interactions among them. Lake Huron consists of 133 nodes specifying species and 3,366 edges specifying species interactions among them. This made sense as the these two are much bigger isolated water bodies compared to other water bodies (ie. Lake Huron being third largest freshwaters and Lake of the Wood being sixth). Whereas, The smallest dataset that was observed was Smallwood Reservoir and Aishihik Lake. Smallwood reservoir consists of 34 nodes (species) and 189 edges (interactions). Aishihik Lake consists of 39 nodes and 248 edges also showcasing a small species and interaction data overall. This finding of the nodes and edges showcased higher number of species and interactions between all species in water bodies such as lake compared to others. Whereas, the Smallwood reservoir and Aishihik lake shared the lowest number of species and

interaction as they are much smaller water bodies compared to Lake of the Woods and Lake Huron.

The mean degree distribution differed diversely based on the ecosystem as well. The AkatoreA and Venlaw has the highest MDG score compared to the rest. The AkatoreA has MDG score of 99.67 and the Venlaw had the MDG score of 80.47. Whereas, the Smallwood Reservoir and Aishihik Lake had the lowest MDG score. The Smallwood Reservoir MDG score consists of 12.72.

To find the clustering coefficient, methods such as Louvain algorithm was used to calculate modularity scores and analyzed community size distributions. The clustering coefficient revealed shocking findings such that the species that interact A to B, B to C may not interact A to C on a host/ prey level depending on the type of water body (ie. fast moving body of water such as streams compared to lake, rivers or reservoir). This was highlighted in the finding for clustering coefficient for datasets with the highest and lowest clustering coefficient. The highest clustering coefficient realized was AkatoreA at 0.78 and AkatoreB. Whereas, the rest of Canadian freshwater streams were all at 0. The clustering coefficient for AkatoreA was 0.78 suggesting that for most species, their prey or predators are also interacting with each other. The clustering coefficient for AkatoreB was 0.75. The clustering coefficient for Venlaw was 0.73 which was also very high. Additionally, "Clustering coefficients in empirical food webs are generally higher than would be expected in random networks of similar size and connectance" (Dunne et al., 2002). This helps explain why the New Zealand streams showed such high clustering compared to Canadian systems. Whereas, the rest of the clustering coefficient for Canadian Freshwater Systems (river, lakes, and reservoir) was at 0 suggesting less likelihood for species interconnectedness. This clustering coefficient in the data predicted that New Zealand water streams has a more tight knit groups and are more interconnected compared to the Canadian fresh water systems.

Community Structure

The clustering coefficient between the water bodies directly tied to community detection results. "Food webs often display modular structure, with groups of species more strongly connected to each other than to other species" (Gini et al., 2022). This modular structure is particularly evident in the New Zealand streams, where we found multiple distinct communities (44 in AkatoreA, 28 in AkatoreB, and 32 in Venlaw) compared to the single community structure in Canadian systems. In AkatoreA, there were 44 communities found. AkatoreB had 28 communities. Venlaw had 32 communities. These results make sense as New Zealand stream is a fast moving water bodies consisting various interactions between species. Whereas, the rest of Canadian freshwater systems all had 1 community in total.

Link Prediction

A link prediction score quantifies the probability of a missing or future connection between two species in a network based on various measures (e.g., network topology, species traits, or observed interactions). Methods such as common neighbor algorithm was used to calculate link prediction for all of the data sets from both Canadian freshwater systems and Newzealand streams. The highest predicted links scores was found in Lake of the woods at (146) between Pearl Dace (*Semotilus margarita*) with both White Sucker (*Catostomus commersoni*) and Fathead Minnow (*Pimephales promelas*), suggesting strong potential for shared parasites despite different fish species. Whereas, Aishihik lake showed a much lower prediction score of (31) for Arctic Grayling (*Thymallus arcticus*) interactions with both Longnose Sucker (*Catostomus catostomus*) and Northern Pike (*Esox lucius*) suggesting limited potential for shared parasites despite habitat overlap. Whereas, the AkatoreA showed a very moderate prediction score of (85) between freshwater clam (*Sphaerid*) and both midge (*Paucispinigera approximata*) and amphipod (*Paracalliope fluviatilis*), reflecting balanced interaction potential across different taxonomic groups. These prediction showed that Canadian lake systems show either very high or low prediction link scores, whereas, Newzealand streams had a more moderate but consistent prediction links overall.

Motif Analysis

To analyze recurring interaction patterns, methods such as feed-forward loops (FFLs) and feedback loops (FBLs) detection were used to identify significant motifs across both Canadian freshwater systems and Newzealand streams. The motif analysis revealed interesting findings such that the presence of significant interaction patterns varies tremendously depending on the type of water body (such as fast moving streams compared to lakes, rivers, or reservoirs). This was highlighted in the motif counts and statistical significance across datasets, with New Zealand streams showing highly significant patterns while Canadian systems showing little to none. The highest motif counts were found in AkatoreA with 296,856 FFLs (z-score: 20.61, $p < 0.001$) and 85,932 FBLs with z-score: 8.69, $p < 0.001$) suggesting 296,856 patterns and standard deviation of 20.61 with less than 0.1% chance that this pattern occurred randomly. Similarly, AkatoreB showed 89,544 FFLs (z-score: 15.88) and 22,386 FBLs (z-score: 9.23), while Venlaw demonstrated 164,368 FFLs (z-score: 20.52) and 27,781 FBLs (z-score: 6.58), all with $p < 0.001$. In stark contrast, all Canadian freshwater systems showed zero counts for both FFLs and FBLs with neutral z-scores (0.0) and p-values (1.0), suggesting no significant recurring interaction patterns. These motif patterns in the data suggest that New Zealand streams maintain complex, non-random interaction structures while Canadian lakes, rivers, and reservoirs display simpler, more linear relationship patterns, reflecting fundamental differences in how species organize

their feeding relationships in different water body types.

Conclusion

This research on the 10 food web datasets from various aquatic ecosystem such as Canadian Freshwater System and Newzealand streams gave me various insight into the species interactions and network structures based on the size and type of water body. It seems that larger water body such as Lake of the Woods and Lake Huron, with their greater number of species and interactions, showed a more complex and interconnected networks, highlighting their more diverse ecological dynamics. Whereas, smaller water bodies like Smallwood Reservoir and Aishihik Lake had simpler webs with fewer species and interactions, showing their more limited ecological complexity. The network analysis further showed that New Zealand streams had significantly higher clustering coefficients, indicating that species interactions were more tightly interconnected, with predators and prey often interacting with each other across different species. Meanwhile, Canadian freshwater systems exhibited lower clustering, suggesting more linear and less interconnected interactions. These differences showed how the impact of ecosystem size, type, and location affect the structural complexity of different types of ecological networks.

Future Research Suggestion

How do multilayer ecological networks, which incorporate different types of interactions (such that of predator-prey), affect the overall stability and resilience of ecosystems?

This question could lead to a deeper understanding of the complexity of ecological systems by exploring the interdependencies between different ecological relationships.

Bibliography

1. Arai, H. P., & Mudry, D. R. (1983). Protozoan and metazoan parasites of fishes from the headwaters of the Parsnip and McGregor Rivers, British Columbia: A study of possible parasite transfaunations. *Canadian Journal of Fisheries and Aquatic Sciences*, 40, 1676-1684.
2. Arthur, J. R., Margolis, L., & Arai, H. P. (1976). Parasites of fishes of Aishihik and Stevens Lakes, Yukon Territory, and potential consequences of their interlake transfer through a proposed water diversion for hydroelectrical purposes. *Journal of the Fisheries Research Board of Canada*, 33, 2489-2499.
3. Bangham, R. V. (1955). Studies on fish parasites of Lake Huron and Manitoulin Island. *American Midland Naturalist*, 53, 184-194.
4. Chinniah, V. C., & Threlfall, W. (1978). Metazoan parasites of fish from the Smallwood Reservoir, Labrador, Canada. *Journal of Fish Biology*, 13, 203-213.
5. Dechtiar, A. O. (1972). Parasites of fish from Lake of the Woods, Ontario. *Journal of Fisheries Research Board of Canada*, 29, 275-283.
6. 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 of the United States of America*, 99(20), 12917-12922.
7. Jaarsma, N. G., de Boer, S. M., Townsend, C. R., Thompson, R. M., & Edwards, E. D. (1998). Characterising food webs in two New Zealand streams. *New Zealand Journal of Marine and Freshwater Research*, 32, 271-286.
8. Johnson, J. C., Borgatti, S. P., Luczkovich, J. J., & Everett, M. G. (n.d.). Network role analysis in the study of food webs: An application of regular role coloration.
9. Leong, T. S., & Holmes, J. C. (1981). Communities of metazoan parasites in open water fishes of Cold Lake, Alberta. *Journal of Fish Biology*, 18, 693-713.
10. Thompson, R. M., Edwards, E. D., McIntosh, A. R., & Townsend, C. R. (2001). Allocation of effort in stream food web studies – The best compromise? *Marine and Freshwater Research*, 52(3), 339-345.
11. Thompson, R. M., & Townsend, C. R. (1999). The effect of seasonal variation on the community structure and food-web attributes of two streams: Implications for food-web science. *Oikos*, 87, 75-88.
12. Thompson, R. M., & Townsend, C. R. (2000). Is resolution the solution? The effect of taxonomic resolution on the calculated properties of three stream food webs. *Freshwater Biology*, 43, 1-10.
13. Thompson, R. M., & Townsend, C. R. (2003). Impacts on stream food webs of native and exotic forest: An intercontinental comparison. *Ecology*, 84, 145-161.
14. Townsend, C. R., Thompson, R. M., McIntosh, A. R., Kilroy, C., Edwards, E. D., & Scarsbrook, M. R. (1998). Disturbance, resource supply and food-web architecture in streams. *Ecology Letters*, 1, 200-209.
15. Wikipedia. (n.d.). Lake Huron. https://en.wikipedia.org/wiki/Lake_Huron

16. Wikipedia. (n.d.). Lake of the Woods. https://en.wikipedia.org/wiki/Lake_of_the_Woods