OIKOS

Research

Why body size matters: how larger fish ontogeny shapes ecological network topology

Korryn Bodner, Chris Brimacombe, Marie-Josée Fortin and Péter K. Molnár

K. Bodner (https://orcid.org/0000-0002-1752-3954) ✉ (k.bodner@mail.utoronto.ca) and P. K. Molnár (https://orcid.org/0000-0001-7260-2674), Dept of Biological Sciences, Univ. of Toronto Scarborough, ON, Canada. – KB, PKM, C. Brimacombe (https://orcid.org/0000-0002-9368-6106) and M.-J. Fortin (https://orcid.org/0000-0002-9935-1366), Dept of Ecology and Evolutionary Biology, Univ. of Toronto, ON, Canada.

Oikos
2022: e08569
doi: 10.1111/oik.08569

Subject Editor: Gregor Kalinkat
Editor-in-Chief: Dries Bonte
Accepted 25 November 2021

Ontogenetic development can strongly shape species interactions. Yet, rarely is stage-structure considered when analyzing species interaction networks, particularly networks that can account for more than feeding relationships. Here, we assess 1) if body size or trophic level regulate the importance of species' ontogeny on their interactions and 2) how including relevant stage-structure affects the topology of species interaction networks. We use a count-based inferential method to create networks from adult and juvenile fish count data and test stage-structure importance by comparing a model that includes stage-structure for all species against models that include stage-structure only for larger fishes and only for piscivorous fishes during network construction. While the inferential method we use cannot differentiate between different types of interactions, it can account for different interaction types within a network as a pairwise interaction is inferred when one species influences the abundance of another. Next, we use graphlet-based techniques to test if including stage-structure alters overall network topology and a linear model to measure if adult-juvenile size differences drive interaction differences at a species-level. We find that the model that includes stage-structure only for larger fishes outperforms other stage-structured models including the model with only piscivore stage-structure, and that larger differences in body size among juveniles and adults lead to greater interaction dissimilarities. Moreover, we find topological differences between inferred networks that only include adults and those that account for the stage-structure of larger species. Overall, our study demonstrates how stage-structured topological changes can be measured using inferred interaction networks and illustrates how larger species' juveniles fundamentally shape the structure of stream fish communities.

Keywords: freshwater fish, inferred network, network topology, Poisson Lognormal, species interactions, stage-structure

Introduction

Species interaction networks are used for conceptualizing complex webs of antagonistic and mutualistic interactions, and can provide core insights into the stability and

NORDIC SOCIETY OIKOS

www.oikosjournal.org

functioning of communities and ecosystems (Thompson et al. 2012, Peralta et al. 2014). Network topology, the architectural description of ecological communities, influences and is influenced by various ecological and evolutionary processes (Guimarães 2020). Identifying the topological differences between networks, therefore, can potentially indicate when ecological systems and communities are fundamentally shaped by different processes.

Most species interaction networks treat individuals as identical (Losapio et al. 2018) with few studies exploring empirically how intraspecific variation influences network topology (Clegg et al. 2018). One of the largest sources of intraspecific variation is ontogenetic variation, which arises due to changes in species' traits or ecology during development (Rudolf and Eveland 2021). As the presence of diet or habitat changes occurring throughout development can modify conditions for species coexistence (Miller and Rudolf 2011) as well as indirectly or directly regulate the dynamics of communities (Osenberg et al. 1992), investigating networks that incorporate ontogenetic variation is useful from both a theoretical and conservation perspective.

Among the few studies that have accounted for ontogenetic differences by including stage-structure in their networks (de Roos and Persson 2013, Nakazawa 2015, Clegg et al. 2018), most contain only a single interaction type, often predation, and ignore that competition, mutualism and parasitism may also be present (but see Ke and Nakazawa 2018). As combining multiple interaction types can non-randomly alter network topology (Kéfi et al. 2015), accounting for different interactions could advance our understanding of how communities may respond to different environmental perturbations. For example, recently theoretical models have shown that stage-structured networks accounting for antagonistic and mutualistic interactions could play a substantial role in stabilizing communities under environmental changes (Ke and Nakazawa 2018). Hence, constructing networks that accommodate both stage-structure and different interaction types could provide a better description of the structure and dynamics of ecological systems.

The inclusion of stage-structure can alter the topology of networks (Clegg et al. 2018) but not all species may contribute equally. For example, larger and/or predator species' stage-structure may disproportionately contribute to ecological network topology due to the size and/or role changes that such species undergo during their lifetime. Size can alter the existence of interactions or influence the type of interactions that an individual is engaged in, such as switching from facilitation (i.e. mutualism or commensalism) to competition as a species grows (Cameron et al. 2019). Species that experience large body size changes commonly undergo diet shifts throughout their lifespan (Werner and Gilliam 1984), which is especially true for predators (Persson 1988). For example, while bluegill and pumpkinseed sunfish compete for a common resource as juveniles, this competitive interaction is non-existent for adults as they occupy different dietary niches (Osenberg et al. 1992). Predator intraspecific variation in size and morphology, in particular, can regulate community structure and ecosystem processes by affecting the structure and strength of complex trophic interactions (Post et al. 2008).

Here, we create species interaction networks to test whether stage-structure influences network topology and whether all species' stage-structure contributes equally to this topology. In freshwater fish communities, various species shift their interactions during development (e.g. *Salmo trutta* (Sánchez-Hernández et al. 2017), *Micropterus dolomieu* (Dauwalter and Fisher 2008)), with piscivore species often undergoing large interaction shifts (Persson 1988). Thus, freshwater stream fish provide a good model system for exploring whether stage-structure can influence the architecture of ecological networks.

Constructing species interaction networks can be challenging as observational data, including direct observations (e.g. a pollinator visiting a plant (Spiesman and Inouye 2013)) and indirect evidence of interactions (e.g. gut-content analysis (McLeod et al. 2020)), may be sparse. For freshwater fish communities, such observations are typically limited to a small number of commercial species (Sánchez-Hernández et al. 2019). As an alternative, inferential methods have been proposed to estimate species interaction networks when limited interaction data are available (Morales-Castilla et al. 2015). EMtree (Momal et al. 2020), an inferential tree-based method, uses species (or age-group) counts and environmental covariates within a joint-species distribution modelling framework to infer networks. Count-based inferential methods have been proposed as an alternative to co-occurrence approaches as they provide more information to make more refined inferences (Blanchet et al. 2020). Here, we adopt EMtree to infer two freshwater stream fish interaction networks: an adult network (nodes of adult fish only) and a stage-structured network (nodes of adult and juvenile fish). As sampling smaller fish often requires different sampling gear (Clavero et al. 2006) and juveniles may inhabit different habitats than adults (St. Mary et al. 2000, Ayllón et al. 2010), we compare adult-only and stage-structured networks to determine if sampling juveniles is worth the additional effort.

We test for relevant stage-structure effects on species interaction networks at various points throughout our analyses. First, during network construction, we compare different stage-structured models including those that contain stage-structure for all species, only for piscivores and only for larger species to determine which stage-structure best explains the sampled adult and juvenile counts. Following the construction of the adult and stage-structured networks, we assess whether these networks exhibit any topological differences. Lastly, if topological differences are detected, we evaluate whether characteristics such as body size and piscivore status (i.e. identification as a piscivore or non-piscivore) can help explain the degree of dissimilarity in interactions between adults and juveniles in the stage-structured network. We hypothesize that including stage-structure will fundamentally alter the topology of the freshwater stream fish networks and that the stage-structure of piscivores and larger non-piscivores will be most important for the stage-structured network given both these types of species are known to experience large shifts in their interactions during ontogeny (Werner and Gilliam 1984, Persson 1988, Cameron et al. 2019).

Methods

Data

Stream fish counts were obtained from the 'Fish electrofishing, gill netting and fyke netting counts' dataset provided by the National Science Foundation's National Ecological Observatory Network (NEON 2020). This dataset includes counts for freshwater stream fish species from 28 aquatic monitoring stations across the United States. Alongside sampling counts, species name, genus, length, weight and life stage were recorded for most sampled fish. Of the 28 aquatic monitoring stations, we considered six stream monitoring locations (Supporting information) where counts were reported per species per age class, for two to four consecutive years between 2017 and 2020 – alongside geographic, environmental and sampling variables, such as latitude, longitude, site name, water temperature, dissolved oxygen, date of sampling and sampling pass identifier (i.e. the identification code for each sampling event). Additional adult feeding behaviours were obtained for each species from NatureServe Explorer (NatureServe 2020). An overview of the species names, their feeding behaviours, their adult and juvenile counts and the sampling month and year can be found in the Supporting information.

Overview

Figure 1 summarizes the progression through the first two major sections of our analyses: 1) network construction and 2) network comparison. In the network construction section, we inferred adult and stage-structured interaction networks by first using Poisson Lognormal (PLN) models (Chiquet et al. 2018, 2019) to generate adult and stage-structured joint species distribution models and then by using the EMtree algorithm to construct adult and stage-structured networks (Momal et al. 2020). For the PLN models, we tested three subsets of stage-structure to determine which best explains the adult and juvenile count data: a) all species (all species model), b) high trophic level species only (piscivore species model) and c) large species only (larger species model). Additionally, we tested a non-stage-structured PLN model fit to adult and juvenile count data to ensure that the use of stage-structured models was necessary. In the network comparison section, we qualitatively and quantitatively assessed the topological similarities and differences between adult and stage-structured networks using the graphlet-based techniques, graphlet correlation distance-11 and graphlet correlation matrix-11 (Yaveroğlu et al. 2014). If the inferred adult and stage-structured networks exhibited topological differences, we performed an additional step (within-network analysis), to test whether size differences or piscivore status best explained interaction differences between juvenile and adult stages within the inferred stage-structured network. This final step was performed using a dissimilarity index and a linear model. The R code used to perform all analyses outlined above is publicly available and can be found on Github (<https://github.com/kbbodner/inferred-stage-structured-fish-networks>).

Network construction

EMtree (Momal et al. 2020) is an inference procedure that uses spanning trees and expectation maximization algorithms to infer conditional dependence networks. The EMtree approach requires a Gaussian covariance matrix, which is obtained from the PLN models created by the PLNmodels R package (Chiquet et al. 2018, 2019). PLN models are multivariate Poisson mixed models that infer species abundances and their joint interactions from species' sampling data. These joint species distribution models combine generalized linear models, which account for environmental and sampling effects, and a Gaussian latent structure, which captures species interactions. In a mixed model framework, the dependency structure between species and/or life stages are represented as correlated random effects.

To construct the inferred networks, the EMtree approach uses the PLN models to generate spanning trees (a subset of the network where nodes are connected by the minimum number of connections (Dale and Fortin 2014)). Here, counts are modelled hierarchically using two hidden layers: the latent Gaussian vectors specified in the PLN models and a random tree where parameters for the latent Gaussian layer are modelled conditionally on spanning trees that were drawn. The final network is inferred by averaging the spanning trees with each edge associated with a probability of being part of the network (Supporting information). A threshold is selected whereby only edges above a specified value are included and a resampling procedure can be adopted to increase network robustness. The expectation maximization algorithms in EMtree provide an efficient exploration of the space of spanning tree structures (Momal et al. 2020). Additional details on the EMtree approach can be found in Momal et al. (2020) and in the Supporting information.

Step 1. PLN models

For all PLN models, we included sampling effort, calculated as the sum of the total counts of fish caught (Paulson et al. 2013), and environmental and geographical covariates to control for the effects of sampling and the main effects of geography and the environment. We limited model explorations to two environmental and geographic covariates in three configurations to avoid convergence issues resulting from model complexity: 1) Site name, 2) Site name + Water temperature and 3) Site name + Dissolved oxygen.

Paired with these environmental factors, we created PLN models to illustrate the effects of excluding and including stage-structure: an adult PLN model constructed with adult

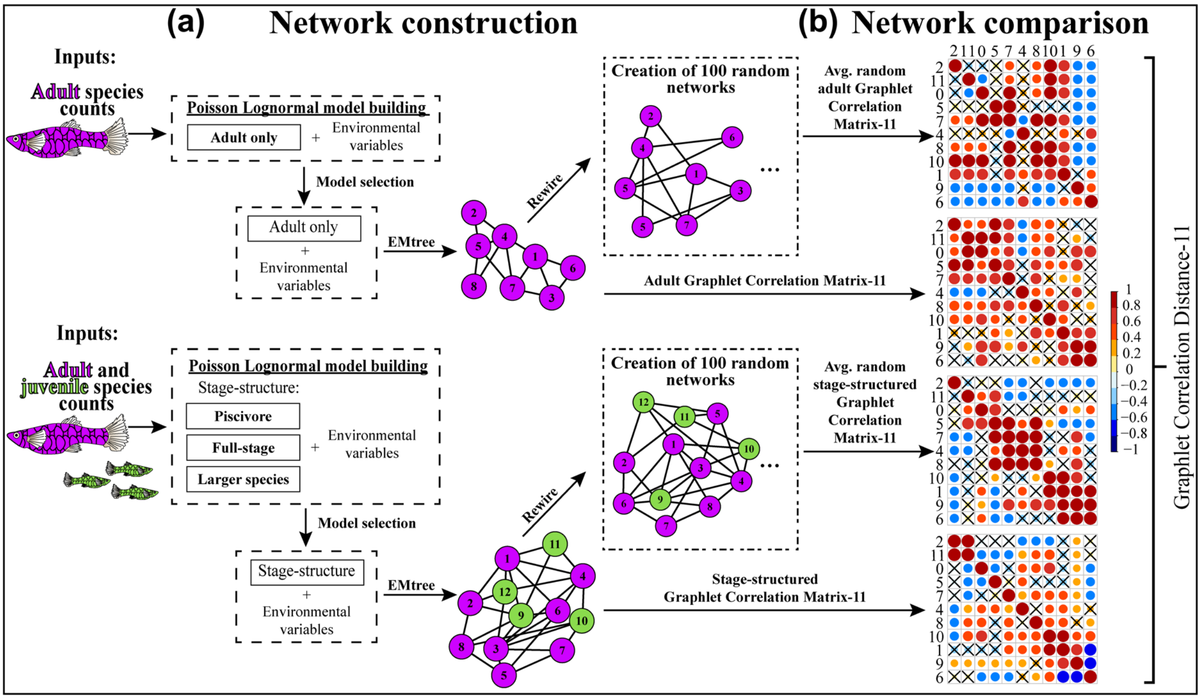


Figure 1. Conceptual diagram illustrating the progression of our analyses through network construction and network comparison. (a) Network construction: Poisson Lognormal (PLN) models were estimated from species count and environmental data, using either adult counts only (adult PLN models) or both adult and juvenile counts (stage-structured PLN models). Separate models were constructed with different environmental covariates and for adults and juveniles, also constructed using the stage-structure of all species, piscivore species or larger species (cf. text for details). Models were tested and PLN models with the best predictive capabilities were selected. While not depicted, non-stage-structured PLN models were also constructed and tested to ensure stage-structured models were the more appropriate models for the adult and juvenile count data. Next, the EMtree algorithm was used to construct an adult and a stage-structured inferred species network from the best-fitting adult and stage-structured PLN models, respectively. Finally, 200 random networks were created by randomly rewiring inferred adult and stage-structured networks. (b) Network comparison: the inferred adult, stage-structured and randomly rewired networks were compared using graphlet correlation matrix-11s, which demonstrate topological patterns of networks based on local network properties in a matrix. Note that random network graphlet correlation matrix-11s were based on average numbers of node appearances across the random networks. Circle size and colour indicate the direction and strength of correlations between orbits, and crosses indicate insignificant relationships. The x- and y-axes of these matrices represent the 11 non-redundant orbits (cf. Fig. 2), with the order of orbits on the x- and y-axes chosen to highlight the correlation patterns within the networks. Additionally, the graphlet correlation distance-11 was calculated, which measures pairwise distances among networks based on their topological differences and provides a single value representing the overall topological similarities and dissimilarities of different networks.

count data only and a set of stage-structured PLN models built using both adult and juvenile count data under different assumptions for the subset of stage-structure: a) all species, b) piscivore species and c) larger species. Stage-structure was included for all species in the all species model, only for species classified as piscivorous (according to adult feeding behaviours in NatureServe 2020) in the piscivore species model, and only for species classified as 'larger' (i.e. if the average adult length was at or above the 50th quantile of all species) in the larger species model. Classifications of species as 'larger' based on membership in the top 25th and 75th quantiles were also explored. We also ensured stage-structured models were most appropriate by comparing their performance to a non-stage-structured model created with adult and juvenile aggregated counts. To create the stage-structured models, species' adult and juvenile counts were separated or amalgamated according to the stage-structure of interest. For example, for the all species model, the PLN model was constructed from separate count measures for all species' juveniles (those classified as 'young-of-year' and 'juvenile'), and adults at each sampling time whereas for the piscivore species model, only piscivore species' juvenile and adult count measurements were separated.

To evaluate PLN model performance, in-sample and out-of-sample assessments were performed. For in-sample assessment, we used a pseudo-R^2 measure, constructed by comparing the log-likelihood of the observed data, the log-likelihood of a saturated model (a model that contains one parameter per observation) and the null model (a Poisson regression GLM with no latent structure; see Chiquet et al. 2018 for more details). For out-of-sample assessment, we calculated the prediction error when one season of species count data were withheld (19 sampling points per species within the season of one year). Prediction error was measured using the cumulative root mean squared error (RMSE) and the RMSE calculated only including species for which counts were above zero (RMSE obs > 0). Models with lower RMSE scores were considered better performing and if two measures disagreed, we used RMSE obs > 0 since the deciding metric as it is not influenced by species' absences. We relied mainly on these RMSE metrics for model selection given that models validated on external data are typically more accurate and reliable than those only assessed via in-sample performance (Bodner et al. 2020). PLN models were constructed and evaluated using the PLNmodels package (Chiquet et al. 2018, 2019).

Step 2. EMtree network construction

After selecting the best-fitting PLN models for the adult-only and the juvenile and adult data, we used the EMtree package (Momal et al. 2020) in R ver. 4.0.3 (<www.r-project.org>) to infer two networks: an adult and a stage-structured network. If the non-stage-structured PLN model performed better than our selected stage-structured model, a non-stage structured network would be created using only aggregated counts. The networks produced by EMtree were composed of nodes representing either species or their life stage as well as undirected links, which represented the interactions between them. To reduce spurious connections, a higher minimum probability threshold for including a species' interaction can be selected as a cut-off. We selected the highest threshold before which any node lost all its connections, a general conservative guideline (Bassett et al. 2006). To increase network robustness, each network was resampled 100 times.

Network comparison

Graphlet analysis

Graphlet correlation matrix-11 and graphlet correlation distance-11 (Yaveroğlu et al. 2014) were used to measure the type and degree of topological differences between different networks. We selected these approaches over alignment-, spectral- and other graphlet-based techniques because they tend to perform best at discriminating undirected network topologies (Tantardini et al. 2019). The graphlet correlation matrix is a symmetric correlation matrix that captures each network's local topology based on graphlet node positions (Fig. 2; Yaveroğlu et al. 2014), whereas the graphlet correlation distance measures the pairwise distances between the local topology of all networks by comparing their graphlet correlation matrices. The graphlet correlation matrix-11 and graphlet correlation distance-11 measures use graphlets of 2–4 nodes and 11 non-redundant orbits (Fig. 2), where orbits are defined as groups of nodes that are topologically symmetrical in a graphlet (Pržulj 2006) and are non-redundant if they cannot be constructed using equations of other orbits (see the Supporting information for an example of a redundancy equation).

The graphlet correlation matrix-11 is an 11 × 11 matrix representing 11 non-redundant orbits where each cell is a Spearman correlation measuring the covariance of two orbits using the number of times nodes occupy these orbit positions (see the Supporting information for a visual example). For example, if cell (*i*, *j*) contained a strong positive correlation, this would indicate that nodes that rarely appeared in orbit *i* would also rarely appear in orbit *j* or those that appeared often in *i* would also appear often in *j*. After calculating graphlet correlation matrix-11s for each of the networks, a graphlet correlation distance-11, denoted as GCD-11 in the equation below, can be calculated for each pair of networks. The calculation uses the Euclidean distance of the upper triangular values of the graphlet correlation matrix-11s as follows:

$$\text{GCD-11}\left(K_1,K_2\right) = \sqrt{\sum_{i=1}^{11}\sum_{j=i+1}^{11}\left(\text{GCM-11}_{K_1}\left(i,j\right)-\text{GCM-11}_{K_2}\left(i,j\right)\right)^2}$$

where GCM-11_{K_1} and GCM-11_{K_2} are the graphlet correlation matrix-11s for networks K_1 and K_2 (Yaveroğlu et al. 2014). Higher graphlet correlation distance-11 scores indicate greater topological differences between networks.

To test whether noise could impede our ability to detect differences in our networks, we created random networks (100 rewired adult networks and 100 rewired stage-structured networks) that each retained the same number of nodes and degree for each node as their inferred counterpart. While the preservation of these characteristics creates random networks that still retain characteristics of the inferred networks, it ensures that all differences found between the inferred and random networks are due to differences in their local network topology rather than to differences in size or node

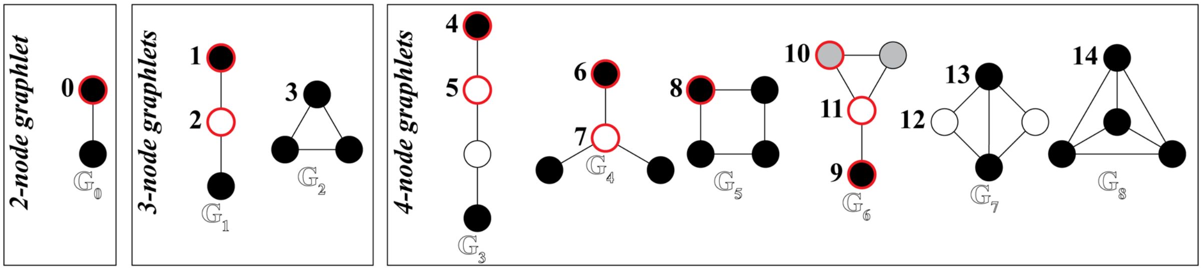


Figure 2. The nine graphlets (G_n), containing 4-nodes or fewer and their 15 labelled automorphism orbits. Nodes of the same shade within a graphlet belong to the same automorphism orbit (i.e. these nodes are topologically symmetrical in the graphlet). The 11 orbits used by the graphlet correlation distance-11 are highlighted in red.

degree. Graphlet analysis was performed on all random networks separately and metric multi-dimensional scaling (Cox and Cox 2000) was performed to reduce the dimensionality to 3-dimensions to visually inspect networks. To calculate graphlet correlation distance-11 scores for groups, we averaged the pairwise difference scores. For graphlet correlation matrix-11 calculations for rewired random networks, we averaged the number of node appearances across each adult and stage-structured random network and used the averages to calculate the correlations for the average random adult and the average random stage structured graphlet correlation matrix-11. Rewiring and metric multi-dimensional scaling were performed in R ver. 4.0.3 (<www.r-project.org>) using the packages igraph (Csardi and Nepusz 2006) and car (Fox and Weisberg 2019), respectively. All graphlet counting was performed in Python ver. 2.7.13 (Python Software Foundation 2016) with code from Yaveroğlu et al. (2014).

Within-network analysis: dissimilarity of adult-juvenile connections

Following the creation of our stage-structured network, a linear regression was used to quantify the relationship between both size and piscivore status and the dissimilarity of the adult and juvenile connections. For size-based explanatory variables, we used the average size and the proportional size differences of the adults and juveniles (i.e. the average length of adults – average length of juveniles; average length of juveniles/average length of adults) and for piscivore status we used whether adults of the species were classified as piscivores. We only calculated the size and piscivore-based measures for adults and juveniles defined as separate nodes. The dissimilarity measure adopted was the Jaccard dissimilarity index (Jaccard 1900), *J*, which is defined as follows:

$$J = \frac{b + c}{a + b + c}$$

where *a* is the number of common nodes connected to both the adults and juveniles of a species, *b* is the number of nodes connected to only the adults and *c* is the number of nodes connected to only the juveniles. Nodes represent either an adult or juvenile of a species or simply another species depending on how species' adult and juvenile counts were aggregated in the stage-structured network.

Results

Network construction

When comparing models with different environmental covariates, the adult PLN models Site name, Site name + Water temperature and Site name + Dissolved oxygen, had similar pseudo-R^2 values (0.82, 0.83 and 0.84 respectively) but the adult PLN models Site name and Site name + Dissolved oxygen had lower RMSE and RMSE obs > 0 values (Table 1). For both the stage-structured models and non-stage-structured models (built using aggregated juvenile and adult counts), almost all PLN models fit with different environmental covariates had a pseudo-R^2 value within 0.02 from each other (the exception being the larger species model with Site name + Water temperature with a difference of 0.04 from the highest pseudo-R^2; Table 1, Supporting information). According to most RMSE and RMSE obs > 0, the most appropriate environmental model across non-stage-structured and stage-structured models was Site name + Dissolved oxygen (see Table 1 and the Supporting information for RMSE values).

When testing which type of stage-structure was the most appropriate, our stage-structured models performed similarly during in-sample assessment (Table 1) but exhibited substantial differences in performances during out-of-sample assessment. The best-performing stage-structured model according to out-of-sample assessment was the larger species model with RMSE = 4.01 and RMSE obs > 0 = 10.00, followed closely

Table 1. The in-sample assessment (pseudo-R^2) and out-of-sample assessment (RMSE (root mean square error), RMSE obs > 0) for each adult and stage-structured Poisson Lognormal model. Out-of-sample assessment was calculated by removing fish counts from the most recent season and then predicting those counts for each species or species' life stage. In total, 19 sampled counts from across six stream locations from the last sampled season per site were used to validate each species or species' life stage counts. The predictive ability of each model was assessed using two measures: (1) RMSE using all species' counts and (2) RMSE obs > 0 using only species that were present (i.e. observed counts > 0).

Count data	Stage-structure	Variables	Pseudo-R^2	RMSE	RMSE obs > 0
Adult	None (31 species)	Site name	0.82	3.45	8.59
		Site name + Water temperature	0.83	3.96	9.07
		Site name + Dissolved oxygen	0.84	3.48	8.61
Adult and juvenile	All species (57 species/stage-classes)	Site name	0.89	6.21	13.34
		Site name + Water temperature	0.87	5.33	13.58
		Site name + Dissolved oxygen	0.88	5.07	13.34
	Piscivore species (39 species/stage-classes)	Site name	0.88	5.99	10.02
		Site name + Water temperature	0.87	5.16	10.53
		Site name + Dissolved oxygen	0.87	4.21	10.13
	Larger species (44 species/stage-classes)	Site name	0.87	5.84	9.50
		Site name + Water temperature	0.85	5.19	10.19
		Site name + Dissolved oxygen	0.87	4.01	10.00

by piscivore species (RMSE = 4.21; RMSE obs > 0 = 10.03), and then all species (RMSE = 5.07; RMSE obs > 0 = 13.34). The non-stage-structured model and the 25% and 75% larger species models fit to the same adult and juvenile count data performed better than the all species model but worse than both the larger species and piscivore species models (see Supporting information for RMSE metrics).

Given its performance during model testing, the larger species model fitted with Site name + Dissolved oxygen was the PLN model selected for network construction of the inferred stage-structured network. Given that Site name and Site name + Dissolved oxygen models fit to adult-only data had similar performance, to maintain consistency with the stage-structured model selection, the Site name + Dissolved oxygen PLN model was selected for the adult network construction.

In the next step, the adult Site name + Dissolved oxygen and the larger species stage-structured Site name + Dissolved oxygen models were paired with EMtree to build the inferred species interaction networks. Various potential threshold cut-offs for classifying interactions as present were tested and the most appropriate threshold was 0.7, as higher values resulted in disconnected networks. The resulting inferred adult network contained 31 nodes and 96 edges whereas the inferred stage-structured network contained 44 nodes and 123 edges (Supporting information). The average number of connections were similar across networks with an average of 2.8 and 3.1 connections for the stage-structured and adult networks, respectively.

Network comparison

The graphlet correlation distance-11 scores between the inferred stage-structured, inferred adult and randomly rewired networks ranged from 0.80 to 3.62 and included within-group distance measures for the random stage-structured and random adult networks (Fig. 3a). The within-group random stage-structured networks and the within-group random adult networks had the lowest graphlet correlation distance-11 scores (0.80 and 0.99, respectively) indicating high topological similarities within each network type. In contrast, the inferred stage-structured and adult networks were approximately twice that distance from their respective random counterparts (1.95 and 1.99, respectively), indicating greater topological dissimilarities. The greatest dissimilarity was between the random adult networks and the random stage-structured networks (3.62) with the second greatest dissimilarity between the inferred stage-structured network and the inferred adult networks (2.61).

The graphlet correlation matrix-11s provided more in-depth insight into the topological differences between networks (Fig. 3b). We expect that very different networks should generally have very different orbit dependencies, and thus very different matrices (Yaveroğlu et al. 2014). As our average random networks maintained the same number of nodes and degree sequences as their inferred counterparts, it is unsurprising that their graphlet correlation matrix-11s contained many positive correlations given that the matrices of the inferred networks also had substantially more positive than negative correlations. However, in comparison with the average random networks, our inferred networks produced fewer strong correlations and fewer significant interactions. In our inferred networks, strong positive correlations were retained between all orbits representing internal nodes (2, 11, 7) and between most orbits representing external nodes (1, 9, 6), except for orbit 4. Interestingly, for both inferred stage-structured and adult networks, there were almost no relationships between orbits representing external nodes and those representing internal nodes.

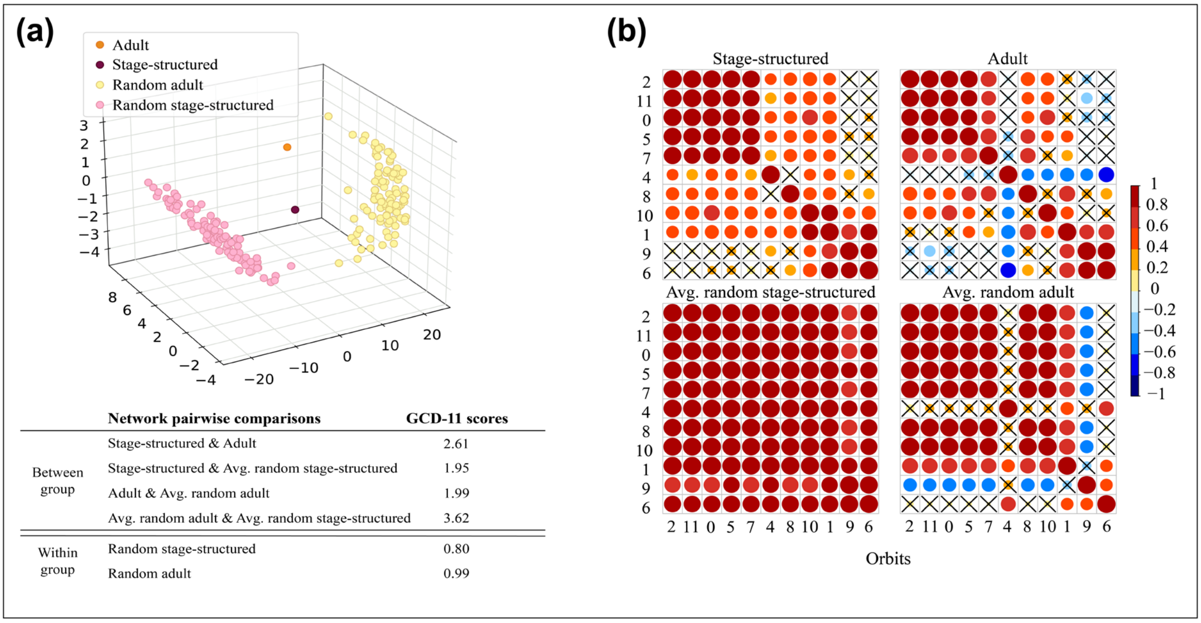
The graphlet correlation matrix-11s revealed topological differences between the inferred adult and inferred stage-structured networks. First, there were more significant correlations in the inferred stage-structured network than the inferred adult network. Second, there were negative relationships in the adult network that did not appear in the inferred stage-structured network. In the inferred stage-structured network, orbit 4, which represents the outside node of a 4-node chain, retained a moderately positive relationship with all orbits except for orbits 8 and 6. In contrast, in the inferred adult network, orbit 4 was insignificant for all internal nodes, but had a moderate to large negative correlation with the other orbits.

Within-network analysis: dissimilarity of adult-juvenile connections

Within-network analysis was performed on the larger species stage-structured network. We found all ‘larger’ species had a Jaccard dissimilarity index of 0.50 or above, indicating a moderate to high dissimilarity of interactions between their juvenile and adult stages (Fig. 4). The two largest species in our analysis, *Salvelinus fontinalis* and *Salmo trutta*, scored 1 and 0.92 respectively, signifying no overlap and very minimal overlap in their connections across their life stages. Using a linear model, we found a positive relationship between the Jaccard dissimilarity index of a species and the average length difference between their juvenile and adult stages ($\beta_0 = 0.5$; $\beta_1 = 0.006$, $p < 0.05$). The model had an adjusted R^2 of 0.31 when all species classified as ‘larger’ species were included, and an adjusted R^2 of 0.63 when the two smallest piscivores, *Semotilus atromaculatus* and *Lepomis megalotis* were removed. Adding piscivore status as an explanatory variable resulted in a worse-fitting model (Supporting information). The linear model containing the proportional size differences between juveniles and adults explained less than 2% of the variability of the dissimilarity index and was not significant (Supporting information).

Discussion

Most species interaction networks treat individuals as identical (Losapio et al. 2018) with few studies exploring how intraspecific variation affects ecological network topology (Clegg et al. 2018). Here, we explored how



	Network pairwise comparisons	GCD-11 scores
Between group	Stage-structured & Adult	2.61
	Stage-structured & Avg. random stage-structured	1.95
	Adult & Avg. random adult	1.99
	Avg. random adult & Avg. random stage-structured	3.62
Within group	Random stage-structured	0.80
	Random adult	0.99

Figure 3. Graphlet correlation distance-11 and graphlet correlation matrix-11 for the inferred adult, stage-structured, randomly rewired adult and randomly rewired stage-structured networks. (a) Graphlet correlation distance-11 plot and table where the multi-dimensional scaling plot depicts graphlet distances between the inferred stage-structured, inferred adult and random networks in three dimensions and the table summarizes the graphlet correlation distance-11 pairwise or average pairwise distances (GCD-11 scores) either between networks of different types (e.g. distance between inferred adult and inferred stage-structured is 'between group') or between networks of the same type (e.g. distance between all the random adult networks is 'within group'). The distance comparisons between random networks in both the 'between group' and 'within group' represent the average distance of all pairwise comparisons. (b) Graphlet correlation matrix-11s are presented for the inferred stage-structured, inferred adult, averaged random stage-structured and averaged random adult networks where x- and y-axes represent the 11 non-redundant orbits (Fig. 2). The order of orbits on the x- and y-axes highlights the correlation patterns of orbits within the networks. The circle size and colour indicate the direction and strength of correlations between orbits (calculated using the number of node appearances in each orbit) and crosses indicate insignificant relationships. For random networks, correlations were calculated using the average number of occurrences per node.

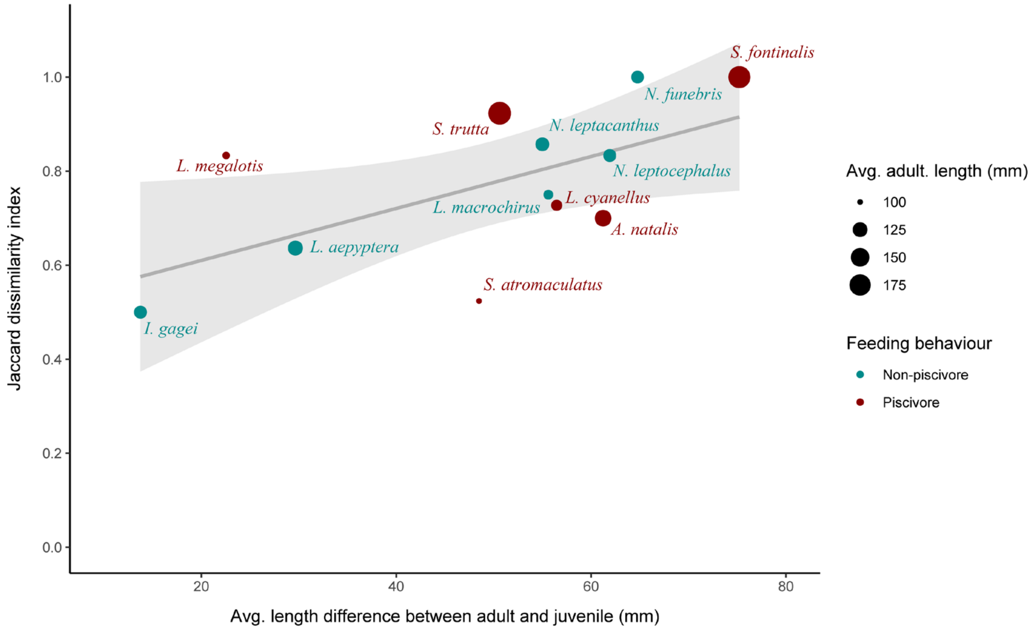


Figure 4. Relationship between the average length difference between 'larger' adults and juveniles and their Jaccard dissimilarity index. Turquoise indicates non-piscivore feeding behaviour in adults, dark red indicates piscivore feeding behaviour in adults and the size of the point indicates the average adult length (mm) of each species.

species characteristics such as body size and trophic position interact with species' stage-structure to influence interactions, and how this stage-structure can produce topological differences at a network level. Specifically, we find that species body size and the size differences of juveniles and adults drive stage-structured species interactions, and that explicitly considering the ontogenetic variation of larger fishes fundamentally alters the estimated topology of freshwater stream fish networks.

In our freshwater stream system, larger fish stage-structure influenced interactions, suggesting body size and traits correlated with body size play a strong role in shaping interactions. This is not unexpected as species size influences both feeding and competitive interactions (Grabowska et al. 2016). Piscivore stage-structure was also important for explaining interactions as six out of the eight piscivores were classified as 'larger'. However, while body size commonly signalled stage-structured importance, the consequences of the resulting interaction shifts depend on the interaction type and species role: predators may experience increased extinction risk if they specialize in different resources at different life stages (Rudolf and Lafferty 2011); prey may experience decreased predation pressure if they outgrow predators (Pessarrodona et al. 2019); and competitors may experience different levels of competition at different life stages as they undergo niche shifts. For all larger species, understanding relevant interactions and their potential consequences requires a detailed perspective that considers different development periods.

At the network level, adult and stage-structured networks had key topological differences (Fig. 3). These differences are expected as ontogenetic diversity has been shown to increase network complexity and the potential for indirect interactions (Rudolf and Eveland 2021). One prominent difference between the adult and stage-structured graphlet correlation matrix-11s is the relationship of a four-node chain external position (orbit 4) with other external node positions. The negative correlation in the adult network graphlet correlation matrix suggests species switch from an external to non-external positions. This unintuitive result is likely due to missing juveniles in the dataset, which may occupy this external chain position acting as prey for larger piscivorous fish, or other external positions as competitors for smaller fish. While previous studies have shown that stage-structure can influence food web topology (Clegg et al. 2018), here we demonstrated that this influence is driven by larger species and that topological changes can be found in networks that account for more than only feeding interactions. Given these topological differences, researchers should therefore increase efforts to include the juveniles of larger fish species when sampling species to build interaction networks.

The ability of inferred network approaches to infer interactions without directly or indirectly observing them creates opportunities to build networks for a greater range of species and interaction types (Faisal et al. 2010). However, inferred networks constructed using these methods have a higher possibility of including spurious interactions and of ignoring true pairwise interactions compared to those empirical networks built on observed species interactions (Blanchet et al. 2020). To reduce potential issues, we used a joint species distribution model framework to control for abiotic factors, reducing the possibility of producing spurious relationships (D'Amen et al. 2018) and species counts, improving inference by providing richer information than presence–absence data (Blanchet et al. 2020). Moreover, while our inferred networks account for more than just trophic interactions (as connections are inferred when one species or life stage influences the abundance of another; Momal et al. 2020), they are undirected and therefore cannot help identify the different types of interactions included in the network. Consequently, if two species/life stages engage in more than one type of interaction, we would only be aware that an interaction exists. Overall, while we can assume non-trophic interactions are likely contributing to the topology of our interaction networks, their specific contributions cannot be quantified using our adopted methodology. However, if longer stationary time series data are available with fixed, equal sampling intervals, empirical dynamic models could be adopted to resolve some of these issues as they provide interaction direction and strength (Ushio et al. 2018).

Due to limited spatio-temporal sampling coverage in our data, we had to construct inferred networks using temporally and spatially aggregated species counts, but we limited temporal aggregation to four consecutive years where minimal yearly variability was detected (Brimacombe et al. 2021), and implicitly controlled for spatial aspects by including a site covariate in our PLN models. As adult and juvenile species counts are continuously being collected by NEON, future studies should have sufficient data to construct stage-structured spatial and/or temporal networks. Such studies could illuminate whether stage-structure differentially influences network topology across these gradients, thereby exploring the context-dependency of freshwater stream fish stage-structured networks.

A key challenge is to identify the topological properties useful for distinguishing ecological networks (Michalska-Smith and Allesina 2019). Despite the ability of graphlet correlation distance- and graphlet correlation matrix-11 to summarize and identify different network topologies (Fortin et al. 2021), they have been rarely used in ecological research to date. Previous ecological studies have often instead characterized network topology using motifs, a technique that also uses subgraphs to measure network topology (Stouffer et al. 2007, Baiser et al. 2016, Clegg et al. 2018). However, motifs count subgraphs as any subset of connections between nodes, which reduces the ability to capture topological similarities between networks (Yaveroğlu 2013) whereas graphlet correlation distance- and graphlet correlation matrix-11 overcome this limitation by requiring that all connections be accounted for when counting subgraphs (Yaveroğlu et al. 2014). Increasingly adopting these powerful techniques could increase our ability to detect general topological patterns across systems and decipher how topological differences between networks translate to differences in dynamics and functioning.

The presence of ontogenetic shifts can affect the stability of consumer resource dynamics, modify conditions for coexistence and alter the direction and strength of trophic cascades (Miller and Rudolf 2011). Accounting for such intraspecific variation can improve our understanding of ecological communities, including how anthropogenic changes affect ecosystem resilience (Nakazawa 2011). Ontogenetic shifts are not limited to predator–prey interactions, so accounting for these shifts across interaction types can help networks portray more realistic images of ecosystems (Ke and Nakazawa 2018). In our system, we found that larger species' stage-structure most affected ecological network topology suggesting that not all species' stage-structure contributes equally to influence interactions. With this result we emphasize not only the potential generality of species' body size as an indicator of stage-structure importance in freshwater stream fish communities, but also the potential utility of traits in general to act as indicators for when stage-structure may be shaping species interaction networks. Lastly, the use of inferred networks and graphlet-based techniques allowed us to construct, characterize and compare adult and stage-structured networks. By using approaches such as EMtree and graphlet correlation distance- and graphlet correlation matrix-11 to study how species traits and ontogeny shape freshwater stream fish networks, we not only illuminate how ontogenetic variation can shape species' interactions but also illustrate a potential template for future network creations and comparisons.

Acknowledgements – We thank the staff at the National Ecological Observatory Network (NEON) for their help with data wrangling, and we are grateful for their publicly available data. We also thank Cedric B. Hunter and Gracie F. Z. Wild for providing engaging discussions on the subjects of both fish and networks. Finally, thank you to Dr. Daniel Newman and the other members of the Dissertation Article Writing Group – your feedback and support were essential for this project.

Funding – Funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant to PKM (no. 06301) and MJF (no. 5134), Canada Research Chair to MJF, Ontario Graduate Scholarship (OGS) to CB, Canadian Foundation for Innovation John R. Evans Leader Funds (no. 35341) and Ministry of Research, Innovation and Sciences Ontario Research Funds to PKM.

Author contributions

Korryn Bodner: Conceptualization (lead); Methodology (equal); Visualization (equal); Writing – original draft (lead); Writing – review and editing (lead). **Chris Brimacombe**: Conceptualization (supporting); Methodology (equal); Visualization (equal); Writing – original draft (supporting); Writing – review and editing (supporting). **Marie-Josée Fortin**: Conceptualization (supporting); Methodology (supporting); Resources (equal); Supervision (equal); Writing – original draft (supporting); Writing – review and editing (supporting). **Péter K. Molnár**: Conceptualization (supporting); Methodology (supporting); Resources (equal); Supervision (equal); Writing – original draft (supporting); Writing – review and editing (supporting).

Data availability statement

Data are available from the Github Digital Repository: <https://github.com/kbbodner/inferred-stage-structured-fish-networks> (Bodner et al. 2021).

Supporting information

The supporting information associated with this article is available from the online version.

References

Ayllón, D. et al. 2010. Ontogenetic and spatial variations in brown trout habitat selection. – Ecol. Freshwater Fish 19: 420–432.

Baiser, B. et al. 2016. Motifs in the assembly of food web networks. – Oikos 125: 480–491.

Bassett, D. S. et al. 2006. Adaptive reconfiguration of fractal small-world human brain functional networks. – Proc. Natl Acad. Sci. USA 103: 19518–19523.

Blanchet, F. G. et al. 2020. Co-occurrence is not evidence of ecological interactions. – Ecol. Lett. 23: 1050–1063.

Bodner, K. et al. 2020. Making predictive modelling ART: accurate, reliable and transparent. – Ecosphere 11: e03160.

Bodner, K. et al. 2021. Data from: Why body size matters: how larger fish ontogeny shapes ecological network topology. – Github Digital Repository, <https://github.com/kbbodner/inferred-stage-structured-fish-networks>.

Brimacombe, C. et al. 2021. Inferred seasonal interaction rewiring of a freshwater stream fish network. – Ecography 44: 219–230.

Cameron, H. et al. 2019. Size and density mediate transitions between competition and facilitation. – Ecol. Lett. 22: 1879–1888.

Chiquet, J. et al. 2018. Variational inference for probabilistic poisson PCA. – Ann. Appl. Stat. 12: 2674–2698.

Chiquet, J. et al. 2019. Variational inference of sparse network from count data. – In: 36th Int. Conf. on Machine Learning, ICML 2019, pp. 1988–1997.

Clavero, M. et al. 2006. Monitoring small fish populations in streams: a comparison of four passive methods. – Fish. Res. 78: 243–251.

Clegg, T. et al. 2018. The impact of intraspecific variation on food web structure. – Ecology 99: 2712–2720.

Cox, T. and Cox, M. 2000. Multidimensional scaling, 2nd edn. – Chapman and Hall/CRC.

Csardi, G. and Nepusz, T. 2006. The igraph software package for complex network research. – InterJ. Complex Syst. 1695: 1–9.

D'Amen, M. et al. 2018. Disentangling biotic interactions, environmental filters and dispersal limitation as drivers of species co-occurrence. – Ecography 41: 1233–1244.

Dale, M. R. T. and Fortin, M.-J. 2014. Spatial analysis. Page spatial analysis: a guide for ecologists, 2nd edn. – Cambridge Univ. Press.

Dauwalter, D. C. and Fisher, W. L. 2008. Ontogenetic and seasonal diet shifts of smallmouth bass in an Ozark stream. – J. Freshwater Ecol. 23: 113–121.

de Roos, A. M. and Persson, L. 2013. Population and community ecology of ontogenetic development. – Princeton Univ. Press.

Faisal, A. et al. 2010. Inferring species interaction networks from species abundance data: a comparative evaluation of various statistical and machine learning methods. – Ecol. Inform. 5: 451–464.

Fortin, M. J. et al. 2021. Network ecology in dynamic landscapes. – Proc. R. Soc. B 288: 20201889.

Fox, J. and Weisberg, S. 2019. An R companion to applied regression. – SAGE Publications Inc., Thousand Oaks, CA, USA.

Grabowska, J. et al. 2016. Interspecific competition for a shelter between non-native racer goby and native European bullhead under experimental conditions – effects of season, fish size and light conditions. – Limnologica 56: 30–38.

Guimarães, P. R. 2020. The structure of ecological networks across levels of organization. – Annu. Rev. Ecol. Evol. Syst. 51: 433–460.

Jaccard, P. 1900. Contribution au problème de l'immigration postglaciaire de la flore alpine. – Bull. Soc. Vaud. Sci. Nat. 36: 87–130.

Ke, P. J. and Nakazawa, T. 2018. Ontogenetic antagonism–mutualism coupling: perspectives on resilience of stage-structured communities. – Oikos 127: 353–363.

Kéfi, S. et al. 2015. Network structure beyond food webs: mapping non-trophic and trophic interactions on Chilean rocky shores. – Ecology 96: 291–303.

Losapio, G. et al. 2018. Plant life history stage and nurse age change the development of ecological networks in an arid ecosystem. – Oikos 127: 1390–1397.

McLeod, A. M. et al. 2020. Effects of species traits, motif profiles and environment on spatial variation in multi-trophic antagonistic networks. – Ecosphere 11: e03018.

Michalska-Smith, M. J. and Allesina, S. 2019. Telling ecological networks apart by their structure: a computational challenge. – PLoS Comput. Biol. 15: e1007076.

Miller, T. E. X. and Rudolf, V. H. W. 2011. Thinking inside the box: community-level consequences of stage-structured populations. – Trends Ecol. Evol. 26: 457–466.

Momal, R. et al. 2020. Tree-based inference of species interaction networks from abundance data. – Methods Ecol. Evol. 11: 621–632.

Morales-Castilla, I. et al. 2015. Inferring biotic interactions from proxies. – Trends Ecol. Evol. 30: 347–356.

Nakazawa, T. 2011. Alternative stable states generated by ontogenetic niche shift in the presence of multiple resource use. – PLoS One 6: e14667.

Nakazawa, T. 2015. Ontogenetic niche shifts matter in community ecology: a review and future perspectives. – Popul. Ecol. 57: 347–354.

National Ecological Observatory Network 2020. Data product DP1.20107.001. Fish electrofishing, gill netting and fyke netting counts. – <http://data.neonscience.org>.

NatureServe 2020. NatureServe Explorer (web application). – NatureServe, Arlington, VI. <https://explorer.natureserve.org/>, accessed 1 Jan 2020.

Osenberg, C. W. et al. 1992. Two-stage life histories in fish: the interaction between juvenile competition and adult performance. – Ecology 73: 255–267.

Paulson, J. N. et al. 2013. Differential abundance analysis for microbial marker-gene surveys. – Nat. Methods 10: 1200–1202.

Peralta, G. et al. 2014. Complementarity and redundancy of interactions enhance attack rates and spatial stability in host–parasitoid food webs. – Ecology 95: 1888–1896.

Persson, L. 1988. Asymmetries in competitive and predatory interactions in fish populations. – In: Size-structured populations. Springer, pp. 203–218.

Pessarrodona, A. et al. 2019. Consumptive and non-consumptive effects of predators vary with the ontogeny of their prey. – Ecology 100: e02649.

Post, D. M. et al. 2008. Intraspecific variation in a predator affects community structure and cascading trophic interactions. – Ecology 89: 2019–2032.

Pržulj, N. 2006. Biological network comparison using graphlet degree distribution. – Bioinformatics 23: e177–e183.

Python Software Foundation 2016. Python language reference <www.python.org>.

Rudolf, V. H. W. and Eveland, L. 2021. Ontogenetic diversity buffers communities against consequences of species loss. – J. Anim. Ecol. 90: 1492–1504.

Rudolf, V. H. W. and Lafferty, K. D. 2011. Stage structure alters how complexity affects stability of ecological networks. – Ecol. Lett. 14: 75–79.

Sánchez-Hernández, J. et al. 2017. Community structure affects trophic ontogeny in a predatory fish. – Ecol. Evol. 7: 358–367.

Sánchez-Hernández, J. et al. 2019. Causes and consequences of ontogenetic dietary shifts: a global synthesis using fish models. – Biol. Rev. 94: 539–554.

Spiesman, B. J. and Inouye, B. D. 2013. Habitat loss alters the architecture of plant–pollinator interaction networks. – Ecology 94: 2688–2696.

St. Mary, C. M. et al. 2000. Stage structure, density dependence and the efficacy of marine reserves. – Bull. Mar. Sci. 66: 675–690.

Stouffer, D. B. et al. 2007. Evidence for the existence of a robust pattern of prey selection in food webs. – Proc. R. Soc. B 274: 1931–1940.

Tantardini, M. et al. 2019. Comparing methods for comparing networks. – Sci. Rep. 9: 17557.

Thompson, R. M. et al. 2012. Food webs: reconciling the structure and function of biodiversity. – Trends Ecol. Evol. 27: 689–697.

Ushio, M. et al. 2018. Fluctuating interaction network and time-varying stability of a natural fish community. – Nature 554: 360–363.

Werner, E. E. and Gilliam, J. F. 1984. The ontogenetic niche and species interactions in size-structured populations. – Annu. Rev. Ecol. Syst. 15: 393–425.

Yaveroğlu, Ö. N. 2013. Graphlet correlations for network comparison and modelling: world trade network example. – Imperial College, London.

Yaveroğlu, Ö. N. et al. 2014. Revealing the hidden language of complex networks. – Sci. Rep. 4: 4547.