

1 THINKING OUTSIDE THE BOX -
2 PREDICTING BIOTIC INTERACTIONS IN
3 DATA-POOR ENVIRONMENTS

4 *DAVID BEAUCHESNE*^{1*}, *PHILIPPE DESJARDINS-PROULX*²,
5 *PHILIPPE ARCHAMBAULT*³, and *DOMINIQUE GRAVEL*²

6 *email: david.beauchesne@ugr.ca

7 ¹ *Université du Québec à Rimouski*

8 ² *Université de Sherbrooke*

9 ³ *Université Laval*

10 September 19, 2016

RUNNING TITLE:
PREDICTING BIOTIC INTERACTIONS IN DATA-POOR ENVIRONMENTS

1 Abstract

Large networks of ecological interactions, such as food webs, are complex to characterize, be it empirically or theoretically. The former requires exhaustive observations, while the latter generally requires ample data to be validated. We therefore wondered whether readily available data, namely empirically described interactions in a variety of ecosystems, could be combined to predict species interactions in data deficient ecosystems. To test this, we built a biotic interactions catalogue from a collection of 94 empirical food webs, detailed predator-prey interaction databases and interactions from the Global Biotic Interactions (GloBI) database. We used an unsupervised machine learning method to predict interactions between any given set of taxa, given pairwise taxonomic proximity and known consumer and resource sets found in the interaction catalogue. Results suggest that pairwise interactions can be predicted with high accuracy. Although conclusions are seemingly dependent on the comprehensiveness of the catalogue knowledge of taxonomy was found to complement well the catalogue and improve predictions, especially as empirical information available diminished. Given its high accuracy, this methodology could democratize the use of food webs and network level descriptors in remote location where empirical data is hard to gather. Network characteristics could then be efficiently evaluated and correlated to levels of environmental stressors in order to improve vulnerability assessments of ecosystems to global changes, opening promising avenues for further research and for management initiatives.

Keywords: Interactions, machine learning, food webs, K-nearest neighbour, taxonomy, St. Lawrence

2 Introduction

Large networks of ecological interactions, such as food webs, are complex to characterize (Polis, 1991; Martinez, 1992; Pascual and Dunne, 2006). Empirical descriptions require exhaustive observations, while theoretical inference generally requires ample data to be validated. For this reason, studies focusing on communities of interacting species remain understudied, even though we acknowledge the importance of considering the reticulated nature of complex networks (Ings et al., 2009; Tylianakis et al., 2008). When time is of the essence, the long term studies required quickly become impractical and the use of network level approaches is relegated to the sideline.

Alternatively, a currently growing approach is to predict interactions using proxies such as functional traits, phylogenies and spatial distributions (e.g. Morales-Castilla et al., 2015; Bartomeus et al., 2016). For example, multiple traits can play a significant role in community dynamics and influence the presence and intensity of biotic interactions, like the influence of body size on predator-prey interactions, a literal take on *big fish eats small fish* (Cohen et al., 2003; Brose et al., 2006; Gravel et al., 2013). However, the time required to gather the necessary data to apply those methods may still be restrictive, or the data be unavailable altogether, so much so that other methods have been developed to fill the gaps in knowledge (e.g. Schrodte et al., 2015).

We therefore wondered whether more readily available data could be used to infer interactions in data deficient ecosystems. There is an increasing amount of data describing worldwide species interactions, some freely available through the Global Biotic Interactions (GloBI) database (Poelen et al., 2014). Another readily available piece of information on species is their taxonomy, through initiatives like the World Register of Marine Species (WoRMS; Bailly et al., 2016). More than simple nomenclature, evolutionary processes are thought

to influence consumer-resource relationships (Mouquet et al., 2012; Rohr and Bascompte, 2014) so that taxonomically related species would be more likely to share similar types of both consumers and resources (Eklöf et al., 2012; Morales-Castilla et al., 2015; Gray et al., 2015). Based on that assumption, taxonomy might be useful in predicting interactions for species lacking detailed information on their biology, but which have a taxonomically related species for which such information is available. The objective of this work is thus to combine empirical biotic interactions originating from a variety of ecosystems with taxonomic relatedness to predict interactions in data deficient ecosystems. As an example, we compare the observed interactions in the southern Gulf of St. Lawrence (SGSL; Savenkoff et al., 2004) with predictions made using our approach.

3 Methods

The objective of our methodology is to predict the interactions between all pairs of taxa within an arbitrary set N_1 , using a set of taxa N_0 with empirically described interactions from which we can extract pairs of consumers and resources and their taxonomy. We couple the use of empirical data with an unsupervised machine learning method to achieve this.

3.1 Biotic interaction catalogue

We built a biotic interaction catalogue to serve as a set of taxa N_0 for with empirically described interactions. The empirical data used to construct the interaction catalogue was gathered in two successive steps. The first consisted of gathering data from a collection of 94 empirical food webs from which we extracted pairwise taxa interactions (see Brose et al., 2005; Kortsch et al., 2015; University of Canberra, 2016 for more information). We also used a detailed

88 predator-prey interaction database describing trophic relationships between ma-
89 rine fishes and their prey (Barnes et al., 2008). From these datasets, only in-
90 teractions between taxa at the taxonomic scale of the family or higher were
91 selected for inclusion in the catalogue. Data used came exclusively from marine
92 and coastal ecosystems and encompassed a wide variety of organisms: fungi, al-
93 gae, parasites, phytoplankton, zooplankton, benthic and pelagic invertebrates,
94 demersal and pelagic fishes, marine birds and marine mammals.

95 As empirical food webs are vastly dominated by non-interactions (96%),
96 these datasets yielded a highly skewed distribution of interactions vs non-interactions.
97 To counterbalance this, the second step of data compilation consisted of extract-
98 ing observed interactions from the Global Biotic Interaction (GloBI) database
99 (Poelen et al., 2014), which describes binary interactions for a wide range of
100 taxa worldwide. We extracted all trphic interactions available on GloBI for
101 species belonging to the families of taxa identified through step 1. Interactions
102 were extracted using the rGloBI package in R (Poelen et al., 2015). As per step
103 1, only interactions between taxa at the taxonomic scale of the family or higher
104 were retained.

105 The nomenclature used between datasets and food webs varied substantially.
106 Taxa names thus had to be verified, modified according to the scientific nomen-
107 clature and validated. This process was performed using the Taxize package in
108 R (Chamberlain and Szöcs, 2013; Chamberlain et al., 2014) and manually veri-
109 fied for errors. The same package was used to extract the taxonomy of all taxa
110 for which interactions were obtained in previous steps. The complete R code
111 and data used to build the catalogue is available at [https://github.com/david-](https://github.com/david-beauchesne/Interaction_catalog)
112 [beauchesne/Interaction_catalog](https://github.com/david-beauchesne/Interaction_catalog).

113 3.2 Unsupervised machine learning

114 We use the K -nearest neighbor (KNN) algorithm (Murphy, 2012) to predict
 115 pairwise interactions for a set of taxa S . The KNN algorithm predicts missing
 116 entries or proposes additional entries by a majority vote based on the K nearest
 117 (i.e. most similar) entries (see Box 1 for an example). In this case, taxa are
 118 described by a set of resources when considered as a consumer, a set of consumers
 119 when considered as a resource and their taxonomy (i.e. kingdom, phylum, class,
 120 order, family, genus, species). Similarity between taxa was evaluated using
 121 the Tanimoto similarity measure, which compares two vectors x and y with
 122 $n = |\mathbf{x}| = |\mathbf{y}|$ elements, and is defined as the size of the intersection of two sets
 123 divided by their union:

$$\text{tanimoto}(\mathbf{x}, \mathbf{y}) = \frac{|\mathbf{x} \cap \mathbf{y}|}{|\mathbf{x} \cup \mathbf{y}|}, \quad (1)$$

124 where \cap is the intersect and \cup the union of the vectors. Adding a weighting
 125 scheme, we can measure the similarity using two different sets of vectors $\{\mathbf{x}, \mathbf{y}\}$
 126 and $\{\mathbf{u}, \mathbf{v}\}$:

$$\text{tanimoto}_t(x, y, u, v, w_t) = w_t \text{tanimoto}(\mathbf{x}, \mathbf{y}) + (1 - w_t) \text{tanimoto}(\mathbf{u}, \mathbf{v}), \quad (2)$$

127 where w_t the weight (in $[0; 1]$). For our analyses, the first element of the right-
 128 hand side of (2) is a similarity between the sets of resources (or consumers) for
 129 the second is the Tanimoto similarity pertaining to the taxonomy of the same
 130 taxa. When $w_t = 0$ only resource or consumer sets are used to compute similar-
 131 ity, while $w_t = 1$ solely uses taxonomy. This approach to consider the relative
 132 contribution of two sets of vectors to the Tanimoto similarity was developed by
 133 Desjardins-Proulx et al. (2016).

3.3 Predicting interactions

The algorithm was built on a series of logical steps that ultimately predicts a candidate resources list C_R for each taxon in N_1 based on empirical data available and the similarity among consumers and among resources (Figure 1). For all consumer taxa T_C in N_1 , the algorithm first verifies, for all resources in resource set T_R , if they are found in N_0 (Step S1, Figure 1). When it does, all T_R taxa that are also in N_1 are added as predicted resources for T_C (Steps S2-S3). This corresponds to what we refer to as the catalogue contribution to resource predictions. In essence, two taxa in N_1 that are known to interact through empirical data in the catalogue are automatically assumed to interact in N_1 .

Otherwise, the algorithm passes to what we refer to as the predictive contribution to resource predictions (S4 to S16), with candidate resources for T_{C_i} (focal taxa for explanation) identified with the KNN algorithm. For each resource in T_R that were not in N_1 (Step S2), K most similar resources $T_{R'}$ are identified from N_1 (Steps S4). If similar resources $T_{R'}$ have a similarity value above a minimal similarity threshold, they are added to C_R as candidate resources. If not, they are automatically discarded (Steps S5-S7). This minimal threshold is an arbitrary parameters (set to 0.3 in our analysis) used to avoid predicting resources that have very small and insignificant similarity and hence is very unlikely to share consumers and resources with the taxa it is being compared to.

Then for all consumer taxa T_C in N_1 , K most similar consumers $T_{C'}$ are identified from N_0 . This step aims at extracting sets of potential resources T_R from similar types of consumers found in the catalogue (Step S8). Resources T_R are added to candidate resources C_R for T_{C_i} if they are also found in N_1 (Steps S10-S12). Otherwise, Steps S4 to S7 are duplicated to identify

potential similar resources for T_{C_i} in N_1 from the set of resources T_R of similar consumers $T_{C'}$ (Steps S13-S16). A simple working example is presented at Box 1. A comprehensive mathematical description of the algorithm and the parameters used is however available through Figure 1 and the complete R code and data used for the algorithm is available at https://github.com/david-beauchesne/Predict_interactions.

3.4 Algorithm prediction accuracy

We used datasets including more than 50 taxa (Christian and Luczkovich, 1999; Link, 2002; Thompson et al., 2004; Brose et al., 2005; Barnes et al., 2008; Kortsch et al., 2015) to assess the prediction accuracy of the algorithm. Testing accuracy of a particular dataset was done by first removing from the catalogue all pairwise interacting taxa originating from that dataset. Accuracy was evaluated using three different statistics:

1. $Score_y$ is the fraction of interactions correctly predicted:

$$Score_y = \frac{a}{a + c} \quad (3)$$

2. $Score_{\neg y}$ is the fraction of non-interactions correctly predicted:

$$Score_{\neg y} = \frac{d}{b + d} \quad (4)$$

3. TSS, The True Skilled Statistics (TSS) evaluated prediction success by considering both true and false predictions, returning a value ranging from 1 (perfect predictions) to -1 (inverted predictions; Allouche et al., 2006):

$$TSS = \frac{(ad - bc)}{(a + c)(b + d)} \quad (5)$$

where a is the number of links predicted and observed, b is the number predicted but not observed, c is the number of non-interaction predicted but interactions observed and d is the number of non-interaction predicted absent and observed. These three statistics give a different perspective on prediction accuracy, focusing in turn on true interactions and non-interactions, and on both true and false predictions.

For each statistic, we evaluated prediction accuracy 1) for the complete algorithm, 2) for predictions made through the predictive portion of the algorithm (Steps S4-S16; Figure 1) and 3) for the catalogue contribution of the algorithm (Steps S1-S3; Figure 1). We evaluated these steps separately in order to partition the relative contribution of the catalogue and of the predictions made using the KNN algorithm to the overall predictive accuracy of the algorithm. Multiple w_t values were also tested to evaluate whether taxa similarity measured as a function of resource/consumer sets or taxonomy contributed more significantly towards increased predictive accuracy. The same was done with multiple K values.

Finally, we evaluated the influence of the comprehensiveness of the catalogue on prediction accuracy. We selected the arctic marine food web from Kortsch et al. (2015) as a test. This food web was selected as it is highly detailed taxonomically. Furthermore, once removed from the catalogue, almost 100% of its taxa still had information available on sets of consumers and resources, which necessary for testing the impact of catalogue comprehensiveness on prediction accuracy. We iteratively and randomly ($n = 50$ randomizations) removed a percentage of empirical data describing the food web taxa from the catalogue before generating new predictions with the algorithm. We also tested w_t values of 0.5 and 1 to evaluate whether taxonomic similarity could support predictive accuracy in cases when empirical data for species in N_1 in the catalogue is

206 unavailable.

207 4 Results

208 4.1 Biotic interaction catalogue

209 The data compilation process allowed us to build an interaction catalogue com-
210 posed of 276708 pairwise interactions (interactions = 72110; non-interactions =
211 204598). A total of 9712 taxa (Superfamily = 15; Family = 591; Subfamily =
212 29; Tribe = 8; Genus = 1972; Species = 7097) are included in the catalogue,
213 4159 of which have data as consumers and 4375 as resources.

214 4.2 Algorithm predictive accuracy

215 The overall predictive accuracy of the algorithm ranges between 80% to al-
216 most 100% in certain cases (Figure 2). Both interactions and non-interactions
217 are well predicted by the algorithm. TSS scores are lower than $Score_y$ and
218 $Score_{-y}$ due to misclassified interactions and non-interactions. This can also
219 be observed through the effect of varying K values, which increases the number
220 of potential candidate resources for each taxa in the predictive portion of the
221 algorithm. Prediction accuracy increases for interactions, while it decreases for
222 non-interactions, as K values increase.

223 Similarity being predominantly measured with resource/consumer sets (w_t
224 closer to 0) yielded better predictions than when measured with taxonomy (w_t
225 closer to 1; Figure 2). Resource/consumer sets therefore appears to serve as
226 a better measure of similarity between taxa for interactions predictions. It is
227 nonetheless interesting to note that although the predictive contribution of the
228 algorithm decreases as w_t increases, an increased mean and decreased variabil-
229 ity values for the TSS and $Score_y$ statistics is also observed (Figure 2). This

suggests that while using taxonomy for similarity measurements yields lower predictive accuracy, it may also complement the catalogue contribution by predicting interactions not captured through empirical data, effectively increasing the predictive accuracy of the complete algorithm.

The partitioning of the catalogue and predictive portions of the algorithm shows that it is dependent on the comprehensiveness of the catalogue for high prediction accuracy (Figures 2, 3). As the amount of empirical data available in the catalogue decreases so does the overall accuracy of the algorithm (Figures 3). The predictive contribution of the algorithm however slows down the decrease in the prediction efficiency of the algorithm. Prediction accuracy still remains around 75% with only 40% of N_1 taxa found in the catalogue (Figures 3). Furthermore, the use of taxonomy for similarity measurements is more efficient as empirical data becomes scarcer and no different than resource/consumer sets for the complete algorithm when ample data is available (Figures 3).

4.3 Southern Gulf of St. Lawrence

As an example, we predict interactions in the southern Gulf of St. Lawrence (SGSL) in eastern Canada. The empirical data and taxa list come from Savenkoff et al. (2004). They present a list of 29 functional groups for a total of 80 taxa presented at least at taxonomical scale of the family. Other coarser functional groups were not used for this example (see Table S1 in Supplementary information (SI) and Savenkoff et al. (2004) for a complete description of documented groups). As their analysis was performed on the functional groups rather than the taxa themselves, we used the algorithm to predict interactions between all 80 taxa selected. We then aggregated them back to their original functional groups to compare with interactions presented in Savenkoff et al. (2004). In total, there were empirical data available in the catalogue for 78% of SGSL taxa (62/80).

256 The algorithm correctly predicted close to 80% of interactions ($a = 135/170$)
 257 and non-interactions ($d = 354/455$) extracted from Savenkoff et al. (2004). It
 258 also predicted an additional 101 interactions that were not noted in Savenkoff
 259 et al. (2004) and failed to predict 36 observed interactions that were, resulting in
 260 a TSS score of 0.57. A visual comparison of results obtained from the algorithm
 261 with interactions noted in Savenkoff et al. (2004) is available at Figure 4. The
 262 network presented is centered on the observed and predicted interactions of the
 263 capelin (*Mallotus villosus*) and piscivorous small pelagic feeders (e.g. *Scomber*
 264 *scombrus* and *Illex illecebrosus*).

265 **5 Discussion**

266 **5.1 Algorithm accuracy**

267 We show that out of the box interaction inference for a set of taxa with incom-
 268 plete or unavailable preexisting information can be achieved with high accuracy
 269 using a combination of empirical data describing biotic interactions and tax-
 270 onomic relatedness. Although the efficiency of the algorithm is dependent on
 271 the comprehensiveness of the interactions catalogue, taxonomic proximity acts
 272 as a complement to increase the number of observed interactions correctly pre-
 273 dicted. Taxonomic proximity also supports the efficiency of the algorithm when
 274 catalogue comprehensiveness decreases.

275 **5.2 Usefulness of taxonomic relatedness**

276 While we found that taxonomy could be useful as a complement to predictions
 277 made using empirical data, the accuracy of predictions made using the KNN
 278 algorithm could be improved. While evolutionary history plays a significant role
 279 in influencing consumer-resource trait matching and food web structure (Mou-

280 quet et al., 2012; Rohr and Bascompte, 2014), phylogenetic constraints do not
 281 account efficiently for certain traits such as body size (Eklöf and Stouffer, 2016).
 282 Including traits like body size and metabolism as an additional component of
 283 this algorithm could thus help increasing overall prediction accuracy, especially
 284 in cases where the catalogue lacks data on taxa for which interactions have to be
 285 predicted. Although promising, such an approach would undermine the premise
 286 under which this method was built and which constitutes its main strength, *i.e.*
 287 predicting interactions in data deficient environments using readily available
 288 data.

289 **5.3 Interactions classification**

290 That $Score_y$ and $Score_{\neg y}$ are inversely proportional means that non-interactions
 291 are misclassified as interactions in the process of increasing $Score_y$, consequently
 292 decreasing $Score_{\neg y}$. This could either stem from the algorithm poorly predicting
 293 non-interactions or from the empirical data itself. Accuracy evaluation assumes
 294 that non-interactions from empirical food web are observed data, yet it is usually
 295 not the case. Most empirical webs have a strong focus attributed to higher order
 296 consumer species and often uneven effort made to thoroughly detail species
 297 interactions (Dunne, 2006). Furthermore, the methodologies used to obtain
 298 consumer-resource data, often relying on gut content analyses, which is efficient
 299 at observing interactions, may be inefficient to detect absence of interactions in
 300 natural systems (Dunne, 2006). This is especially true with our methodology,
 301 where we predict interactions between species whose co-occurrence may have
 302 been observed in the other ecosystems we are using to predict interactions.
 303 Misclassified interactions could thus be real, albeit unobserved through empirical
 304 data available.

305 5.4 Southern Gulf of St. Lawrence

306 The St Lawrence example (Figure 4 and SI) provides great material to discuss
307 predictions in greater detail. The algorithm fails to predict 20% of interactions
308 presented in Savenkoff et al. (2004). Interactions that failed to be predicted
309 were mainly centered on invertebrate species (e.g. polychaetes and mollusks)
310 and large functional groups described by coarse taxonomic categories (e.g. di-
311 atoms) alongside few species in Savenkoff et al. (2004) (e.g. piscivorous small
312 pelagic feeders; Table S3). As we focused on the taxa at least at the scale of
313 family, it is likely that their functional groups had a broader range of possible
314 interactions included than what the algorithm could predict using only a few
315 taxa. Furthermore, the efficiency of the algorithm greatly depends on the un-
316 derlying empirical data that defines the catalogue. If the empirical data used
317 to build the catalogue focuses on higher order consumers, it should come as no
318 surprise that the algorithm would be afflicted by the same limitations.

319 The algorithm also predicts substantially more interactions than those pre-
320 sented in Savenkoff et al. (2004) (Figure 4; Table S2). The catalogue is not
321 currently built to take into account life stages of species. Considering life stages
322 and the fact that they are not explicitly considered in the catalogue could
323 explain additional interactions that seem suspicious at first, like the surprise
324 amount of additional interactions predicted for small piscivorous pelagic feeders
325 as consumers (Figure 4). Due to the aggregated nature of the SGSL web, we
326 believe the TSS score to be an underestimate of the efficiency of the algorithm.

327 5.5 Perspectives

328 Overall, we believe our method performs well and offers promising avenues for
329 further applied research and management initiatives. Interaction strength and
330 species co-occurrence are major attributes affecting the probability of observing

331 interactions. Interaction strength is instrumental to understanding community
332 dynamics, stability and robustness (Laska and Wootton, 1998; Morales-Castilla
333 et al., 2015), while the co-occurrence of species encloses valuable information
334 on interactions and is a pre-requisite for them to exist (Cazelles et al., 2016).
335 Considering them in our methodology would be highly valuable to correctly
336 assess interactions in a given ecosystem and predict the spatial distribution of
337 interaction networks.

338 Given its high efficiency and simplicity, our methodology could broaden the
339 use and the accessibility of food webs and network level descriptors for integra-
340 tive management initiatives such as cumulative impacts assessments and sys-
341 tematic planning (Giakoumi et al., 2015; Beauchesne et al., 2016), especially for
342 remote locations where empirical data is hard to gather. Network characteristics
343 could be efficiently evaluated and correlated to levels of multiple environmental
344 stressors to assess the vulnerability of ecosystems to global changes (Albouy et
345 al., 2014). We believe that the development of such predictive approaches could
346 represent the first much needed steps towards the use of ecological networks in
347 systematic impacts assessments.

348 **6 Acknowledgements**

349 We thank the Fond de Recherche Québécois Nature et Technologie (FRQNT)
350 and the Natural Science and Engineering Council of Canada (CRSNG) for fi-
351 nancial support. This project is also supported by Québec Océan, the Quebec
352 Centre for Biodiversity Science (QCBS), and the Notre Golfe and CHONeII net-
353 works. We also wish to thank K. Cazelles for the help, constructive comments
354 and suggestions.

References

- Albouy, Camille, Laure Velez, Marta Coll, Francesco Colloca, François Le Loc’h, David Mouillot, and Dominique Gravel (2014). “From projected species distribution to food-web structure under climate change”. In: *Global Change Biology* 20.3, pp. 730–741. ISSN: 13541013. DOI: [10.1111/gcb.12467](https://doi.org/10.1111/gcb.12467). URL: <http://doi.wiley.com/10.1111/gcb.12467>.
- Allouche, Omri, Asaf Tsoar, and Ronen Kadmon (2006). “Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)”. In: *Journal of Applied Ecology* 43.6, pp. 1223–1232. ISSN: 00218901. DOI: [10.1111/j.1365-2664.2006.01214.x](https://doi.org/10.1111/j.1365-2664.2006.01214.x). URL: <http://doi.wiley.com/10.1111/j.1365-2664.2006.01214.x>.
- Bailly, N et al. (2016). *World Register of Marine Species (WoRMS)*. \url=http://www.marinespecies.org. URL: <http://www.marinespecies.org>.
- Barnes, C., D. M. Bethea, R. D. Brodeur, J. Spitz, V. Ridoux, C. Pusineri, B. C. Chase, M. E. Hunsicker, F. Juanes, A. Kellermann, J. Lancaster, F. Ménard, F.-X. Bard, P. Munk, J. K. Pinnegar, F. S. Scharf, R. A. Rountree, K. I. Stergiou, C. Sassa, A. Sabates, and S. Jennings (2008). “Predator and prey body sizes in marine food webs”. In: *Ecology* 89.3, pp. 881–881. DOI: [10.1890/07-1551.1](https://doi.org/10.1890/07-1551.1). URL: <http://doi.wiley.com/10.1890/07-1551.1>.
- Bartomeus, Ignasi, Dominique Gravel, Jason M. Tylianakis, Marcelo A. Aizen, Ian A. Dickie, and Maud Bernard-Verdier (2016). “A common framework for identifying linkage rules across different types of interactions”. In: *Functional Ecology*, n/a–n/a. ISSN: 02698463. DOI: [10.1111/1365-2435.12666](https://doi.org/10.1111/1365-2435.12666). URL: <http://doi.wiley.com/10.1111/1365-2435.12666>.
- Beauchesne, David, Cindy Grant, Dominique Gravel, and Philippe Archambault (2016). “L’évaluation des impacts cumulés dans l’estuaire et le golfe du Saint-Laurent : vers une planification systémique de l’exploitation des ressources”.

382 In: *Le Naturaliste canadien* 140.2, p. 45. ISSN: 0028-0798. DOI: [10.7202/](https://doi.org/10.7202/1036503ar)
383 [1036503ar](https://doi.org/1036503ar). URL: <http://id.erudit.org/iderudit/1036503ar>.

384 Brose, Ulrich, Lara Cushing, Eric L. Berlow, Tomas Jonsson, Carolin Banasek-
385 Richter, Louis-Félix Bersier, Julia L. Blanchard, Thomas Brey, Stephen
386 R. Carpenter, Marie-France Cattin Blandenier, Joel E. Cohen, Hassan Ali
387 Dawah, Tony Dell, Francois Edwards, Sarah Harper-Smith, Ute Jacob, Roland
388 A. Knapp, Mark E. Ledger, Jane Memmott, Katja Mintenbeck, John K.
389 Pinnegar, Björn C. Rall, Tom Rayner, Liliane Ruess, Werner Ulrich, Philip
390 Warren, Rich J. Williams, Guy Woodward, Peter Yodzis, and Neo D. Mar-
391 tinez (2005). “Body sizes of consumers and their resources”. In: *Ecology*
392 86.9, pp. 2545–2545. ISSN: 0012-9658. DOI: [10.1890/05-0379](https://doi.org/10.1890/05-0379). URL: [http:](http://doi.wiley.com/10.1890/05-0379)
393 [//doi.wiley.com/10.1890/05-0379](http://doi.wiley.com/10.1890/05-0379).

394 Brose, Ulrich, Tomas Jonsson, Eric L. Berlow, Philip Warren, Carolin Banasek-
395 Richter, Louis-Félix Bersier, Julia L. Blanchard, Thomas Brey, Stephen
396 R. Carpenter, Marie-France Cattin Blandenier, Lara Cushing, Hassan Ali
397 Dawah, Tony Dell, Francois Edwards, Sarah Harper-Smith, Ute Jacob, Mark
398 E. Ledger, Neo D. Martinez, Jane Memmott, Katja Mintenbeck, John K.
399 Pinnegar, Björn C. Rall, Thomas S. Rayner, Daniel C. Reuman, Liliane
400 Ruess, Werner Ulrich, Richard J. Williams, Guy Woodward, and Joel E.
401 Cohen (2006). “Consumer-resource body-size relationships in natural food
402 webs”. In: *Ecology* 87.10, pp. 2411–2417. DOI: [10.1890/0012-9658\(2006\)](https://doi.org/10.1890/0012-9658(2006)87[2411:CBRINF]2.0.CO;2)
403 [87\[2411:CBRINF\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[2411:CBRINF]2.0.CO;2). URL: [http://doi.wiley.com/10.1890/0012-](http://doi.wiley.com/10.1890/0012-9658(2006)87[2411:CBRINF]2.0.CO;2)
404 [9658\(2006\)87\[2411:CBRINF\]2.0.CO;2](http://doi.wiley.com/10.1890/0012-9658(2006)87[2411:CBRINF]2.0.CO;2).

405 Cazelles, Kévin, Miguel B. Araújo, Nicolas Mouquet, and Dominique Gravel
406 (2016). “A theory for species co-occurrence in interaction networks”. In:
407 *Theoretical Ecology* 9.1, pp. 39–48. ISSN: 1874-1738. DOI: [10.1007/s12080-](https://doi.org/10.1007/s12080-)

015-0281-9. URL: <http://link.springer.com/10.1007/s12080-015-0281-9>.

Chamberlain, Scott A. and Eduard Szöcs (2013). “taxize: taxonomic search and retrieval in R”. In: *F1000Research* 2. ISSN: 2046-1402. DOI: [10.12688/f1000research.2-191.v1](https://doi.org/10.12688/f1000research.2-191.v1). URL: <http://f1000research.com/articles/2-191/v1>.

Chamberlain, Scott A., Eduard Szocs, Carl Boettiger, Karthik Ram, Ignasi Bartomeus, and John Baumgartner (2014). *taxize: Taxonomic information from around the web*. URL: <https://github.com/ropensci/taxize>.

Christian, Robert R. and Joseph J. Luczkovich (1999). “Organizing and understanding a winter’s seagrass foodweb network through effective trophic levels”. In: *Ecological Modelling* 117.1, pp. 99–124. ISSN: 03043800. DOI: [10.1016/S0304-3800\(99\)00022-8](https://doi.org/10.1016/S0304-3800(99)00022-8).

Cohen, Joel E, Tomas Jonsson, and Stephen R Carpenter (2003). “Ecological community description using the food web, species abundance, and body size.” In: *Proceedings of the National Academy of Sciences of the United States of America* 100.4, pp. 1781–6. ISSN: 0027-8424. DOI: [10.1073/pnas.232715699](https://doi.org/10.1073/pnas.232715699). URL: <http://www.ncbi.nlm.nih.gov/pubmed/12547915> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC149910>.

Desjardins-Proulx, Philippe, Timothée Poisot, and Dominique Gravel (2016). “Ecological interactions and the Netflix problem”. In:

Dunne, JA (2006). “The network structure of food webs”. In: *networks: linking structure to dynamics in food webs*. URL: <https://books.google.ca/books?hl=en&%7B%5C%26%7Dlr=%7B%5C%26%7Ddid=bF3JoZgoo24C%7B%5C%26%7Ddoi=fnd%7B%5C%26%7Dpg=PA27%7B%5C%26%7Ddq=The+Network+Structure+>

434 [of+Food+Webs%7B%5C&%7Dots=00mi%7B%5C_%7DKSWEi%7B%5C&%7Dsig=](#)
 435 [kVzdEtE5toSzP0Kq7bXnBwHghaY.](#)

436 Eklöf, Anna, Matthew R. Helmus, M. Moore, and Stefano Allesina (2012). “Rel-
 437 evance of evolutionary history for food web structure”. In: *Proceedings of the*
 438 *Royal Society of London B: Biological Sciences* 279.1733.

439 Eklöf, Anna and Daniel B. Stouffer (2016). “The phylogenetic component of
 440 food web structure and intervality”. In: *Theoretical Ecology* 9.1, pp. 107–
 441 115. ISSN: 1874-1738. DOI: [10.1007/s12080-015-0273-9](#). URL: [http://link.springer.com/10.1007/s12080-015-0273-9](#).

443 Giakoumi, Sylvaine, Benjamin S. Halpern, Loïc N. Michel, Sylvie Gobert, Maria
 444 Sini, Charles-François Boudouresque, Maria-Cristina Gambi, Stelios Kat-
 445 sanevakis, Pierre Lejeune, Monica Montefalcone, Gerard Pergent, Christine
 446 Pergent-Martini, Pablo Sanchez-Jerez, Branko Velimirov, Salvatrice Vizzini,
 447 Arnaud Abadie, Marta Coll, Paolo Guidetti, Fiorenza Micheli, and Hugh P.
 448 Possingham (2015). “Towards a framework for assessment and management
 449 of cumulative human impacts on marine food webs”. In: *Conservation Biol-*
 450 *ogy* 29.4, pp. 1228–1234. ISSN: 08888892. DOI: [10.1111/cobi.12468](#). URL:
 451 [http://doi.wiley.com/10.1111/cobi.12468](#).

452 Gravel, Dominique, Timothée Poisot, Camille Albouy, Laure Velez, and David
 453 Mouillot (2013). “Inferring food web structure from predator-prey body size
 454 relationships”. In: *Methods in Ecology and Evolution* 4.11. Ed. by Robert
 455 Freckleton, pp. 1083–1090. ISSN: 2041210X. DOI: [10.1111/2041-210X.12103](#).
 456 URL: [http://doi.wiley.com/10.1111/2041-210X.12103](#).

457 Gray, Clare, David H. Figueroa, Lawrence N. Hudson, Athen Ma, Dan Perkins,
 458 and Guy Woodward (2015). “Joining the dots: An automated method for
 459 constructing food webs from compendia of published interactions”. In: *Food*
 460 *Webs* 5, pp. 11–20. ISSN: 23522496. DOI: [10.1016/j.fooweb.2015.09.001](#).

461 Ings, Thomas C., José M. Montoya, Jordi Bascompte, Nico Blüthgen, Lee Brown,
 462 Carsten F. Dormann, François Edwards, David Figueroa, Ute Jacob, J. Iwan
 463 Jones, Rasmus B. Lauridsen, Mark E. Ledger, Hannah M. Lewis, Jens M.
 464 Olesen, F.J. Frank van Veen, Phil H. Warren, and Guy Woodward (2009).
 465 “Review: Ecological networks - beyond food webs”. In: *Journal of Ani-*
 466 *mal Ecology* 78.1, pp. 253–269. ISSN: 00218790. DOI: [10.1111/j.1365-](https://doi.org/10.1111/j.1365-2656.2008.01460.x)
 467 [2656.2008.01460.x](https://doi.org/10.1111/j.1365-2656.2008.01460.x). URL: [http://doi.wiley.com/10.1111/j.1365-](http://doi.wiley.com/10.1111/j.1365-2656.2008.01460.x)
 468 [2656.2008.01460.x](http://doi.wiley.com/10.1111/j.1365-2656.2008.01460.x).
 469 Kortsch, Susanne, Raul Primicerio, Maria Fossheim, Andrey V. Dolgov, and
 470 Michaela Aschan (2015). “Climate change alters the structure of arctic ma-
 471 rine food webs due to poleward shifts of boreal generalists”. In: *Proceedings*
 472 *of the Royal Society of London B: Biological Sciences* 282.1814.
 473 Laska, Mark S. and J. Timothy Wootton (1998). “Theoretical concepts and
 474 empirical approaches to measuring interaction strength”. In: *Ecology* 79.2,
 475 pp. 461–476. DOI: [10.1890/0012-9658\(1998\)079\[0461:TCAEAT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1998)079[0461:TCAEAT]2.0.CO;2).
 476 URL: [http://doi.wiley.com/10.1890/0012-9658\(1998\)079\[0461:](http://doi.wiley.com/10.1890/0012-9658(1998)079[0461:TCAEAT]2.0.CO;2)
 477 [TCAEAT\]2.0.CO;2](http://doi.wiley.com/10.1890/0012-9658(1998)079[0461:TCAEAT]2.0.CO;2).
 478 Link, J (2002). “Does food web theory work for marine ecosystems?” In: *Ma-*
 479 *rine Ecology Progress Series* 230, pp. 1–9. ISSN: 0171-8630. DOI: [10.3354/](https://doi.org/10.3354/meps230001)
 480 [meps230001](https://doi.org/10.3354/meps230001). URL: [http://www.int-res.com/abstracts/meps/v230/p1-](http://www.int-res.com/abstracts/meps/v230/p1-9/)
 481 [9/](http://www.int-res.com/abstracts/meps/v230/p1-9/).
 482 Martinez, Neo D. (1992). “Constant connectance in community food webs”. In:
 483 *American Naturalist* 139.6, pp. 1208–1218. URL: [http://www.jstor.org/](http://www.jstor.org/stable/2462337)
 484 [stable/2462337](http://www.jstor.org/stable/2462337).
 485 Morales-Castilla, Ignacio, Miguel G. Matias, Dominique Gravel, and Miguel B.
 486 Araújo (2015). “Inferring biotic interactions from proxies”. In: *Trends in*

487 *Ecology & Evolution* 30.6, pp. 347–356. ISSN: 01695347. DOI: [10.1016/j.](https://doi.org/10.1016/j.tree.2015.03.014)
488 [tree.2015.03.014](https://doi.org/10.1016/j.tree.2015.03.014).

489 Mouquet, Nicolas, Vincent Devictor, Christine N. Meynard, Francois Munoz,
490 Louis-Félix Bersier, Jérôme Chave, Pierre Couteron, Ambroise Dalecky, Colin
491 Fontaine, Dominique Gravel, Olivier J. Hardy, Franck Jabot, Sébastien Lavergne,
492 Mathew Leibold, David Mouillot, Tamara Münkemüller, Sandrine Pavoine,
493 Andreas Prinzing, Ana S.L. Rodrigues, Rudolf P. Rohr, Elisa Thébault, and
494 Wilfried Thuiller (2012). “Ecophylogenetics: advances and perspectives”. In:
495 *Biological Reviews* 87.4, pp. 769–785. ISSN: 14647931. DOI: [10.1111/j.1469-](https://doi.org/10.1111/j.1469-185X.2012.00224.x)
496 [185X.2012.00224.x](https://doi.org/10.1111/j.1469-185X.2012.00224.x). URL: [http://doi.wiley.com/10.1111/j.1469-](http://doi.wiley.com/10.1111/j.1469-185X.2012.00224.x)
497 [185X.2012.00224.x](http://doi.wiley.com/10.1111/j.1469-185X.2012.00224.x).

498 Murphy, Kevin P. (2012). *Machine learning : a probabilistic perspective*. MIT
499 Press, p. 1067. ISBN: 9780262018029.

500 Pascual, M and JA Dunne (2006). *Ecological networks: linking structure to dy-*
501 *namics in food webs*. URL: [https://books.google.ca/books?hl=en%](https://books.google.ca/books?hl=en%7B%5C%7Dlr=%7B%5C%7Ddid=YpQRDAAAQBAJ%7B%5C%7Ddoi=fnd%7B%5C%7Dpg=PP1%7B%5C%7Ddq=Pascual+and+Dunne+2006+interactions%7B%5C%7Dots=K4a5d62r9X%7B%5C%7Dsig=01fs%7B%5C%7DfXV1pgP6IeP1jBIb3B61rU)
502 [7B%5C%7Dlr=%7B%5C%7Ddid=YpQRDAAAQBAJ%7B%5C%7Ddoi=fnd%7B%5C%](https://books.google.ca/books?hl=en%7B%5C%7Dlr=%7B%5C%7Ddid=YpQRDAAAQBAJ%7B%5C%7Ddoi=fnd%7B%5C%7Dpg=PP1%7B%5C%7Ddq=Pascual+and+Dunne+2006+interactions%7B%5C%7Dots=K4a5d62r9X%7B%5C%7Dsig=01fs%7B%5C%7DfXV1pgP6IeP1jBIb3B61rU)
503 [%7Dpg=PP1%7B%5C%7Ddq=Pascual+and+Dunne+2006+interactions%7B%](https://books.google.ca/books?hl=en%7B%5C%7Dlr=%7B%5C%7Ddid=YpQRDAAAQBAJ%7B%5C%7Ddoi=fnd%7B%5C%7Dpg=PP1%7B%5C%7Ddq=Pascual+and+Dunne+2006+interactions%7B%5C%7Dots=K4a5d62r9X%7B%5C%7Dsig=01fs%7B%5C%7DfXV1pgP6IeP1jBIb3B61rU)
504 [5C%7Dots=K4a5d62r9X%7B%5C%7Dsig=01fs%7B%5C_%7DfXV1pgP6IeP1jBIb3B61rU](https://books.google.ca/books?hl=en%7B%5C%7Dlr=%7B%5C%7Ddid=YpQRDAAAQBAJ%7B%5C%7Ddoi=fnd%7B%5C%7Dpg=PP1%7B%5C%7Ddq=Pascual+and+Dunne+2006+interactions%7B%5C%7Dots=K4a5d62r9X%7B%5C%7Dsig=01fs%7B%5C%7DfXV1pgP6IeP1jBIb3B61rU).

505 Poelen, Jorrit H., Stephen Gosnell, and Sergey Slyusarev (2015). *rglobi: R In-*
506 *terface to Global Biotic Interactions*. URL: [https://cran.r-project.org/](https://cran.r-project.org/package=rglobi)
507 [package=rglobi](https://cran.r-project.org/package=rglobi).

508 Poelen, Jorrit H., James D. Simons, and Chris J. Mungall (2014). “Global biotic
509 interactions: An open infrastructure to share and analyze species-interaction
510 datasets”. In: *Ecological Informatics* 24, pp. 148–159. ISSN: 15749541. DOI:
511 [10.1016/j.ecoinf.2014.08.005](https://doi.org/10.1016/j.ecoinf.2014.08.005).

- Polis, GA (1991). “Complex trophic interactions in deserts: an empirical critique of food-web theory”. In: *American naturalist* 138.1, pp. 123–155. URL: <http://www.jstor.org/stable/2462536>.
- Rohr, Rudolf P. and Jordi Bascompte (2014). “Components of Phylogenetic Signal in Antagonistic and Mutualistic Networks”. In: *The American Naturalist* 184.5, pp. 556–564. DOI: [10.1086/678234](https://doi.org/10.1086/678234). URL: <http://www.journals.uchicago.edu/doi/10.1086/678234>.
- Savenkoff, Claude, Hugo Bourdages, Douglas P. Swain, Simon-Pierre Despatie, J. Mark Hanson, Red Méthot, Lyne Morissette, and Mike O. Hammil (2004). *Input data and parameter estimates for ecosystem models of the southern Gulf of St. Lawrence (mid-1980s and mid-1990s)*. Tech. rep. Mont-Joli, Québec, Canada: Canadian Technical Report of Fisheries, Aquatic Sciences 2529, Department of Fisheries, and Oceans, p. 105.
- Schrodtt, Franziska, Jens Kattge, Hanhuai Shan, Farideh Fazayeli, Julia Joswig, Arindam Banerjee, Markus Reichstein, Gerhard Bönisch, Sandra Díaz, John Dickie, Andy Gillison, Anuj Karpatne, Sandra Lavorel, Paul Leadley, Christian B. Wirth, Ian J. Wright, S. Joseph Wright, and Peter B. Reich (2015). “BHPMF - a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography”. In: *Global Ecology and Biogeography* 24.12, pp. 1510–1521. ISSN: 1466822X. DOI: [10.1111/geb.12335](https://doi.org/10.1111/geb.12335). URL: <http://doi.wiley.com/10.1111/geb.12335>.
- Thompson, Ross M., Kim N. Mouritsen, and Robert Poulin (2004). “Importance of parasites and their life cycle characteristics in determining the structure of a large marine food web”. In: *Journal of Animal Ecology* 74.1, pp. 77–85. DOI: [10.1111/j.1365-2656.2004.00899.x](https://doi.org/10.1111/j.1365-2656.2004.00899.x). URL: <http://doi.wiley.com/10.1111/j.1365-2656.2004.00899.x>.

538 Tylianakis, Jason M., Raphael K. Didham, Jordi Bascompte, and David A.
539 Wardle (2008). “Global change and species interactions in terrestrial ecosys-
540 tems”. In: *Ecology Letters* 11.12, pp. 1351–1363. ISSN: 1461023X. DOI: 10 .
541 1111/j . 1461-0248 . 2008 . 01250 . x. URL: [http://doi.wiley.com/10 .](http://doi.wiley.com/10 . 1111/j . 1461-0248 . 2008 . 01250 . x)
542 [1111/j . 1461-0248 . 2008 . 01250 . x](http://doi.wiley.com/10 . 1111/j . 1461-0248 . 2008 . 01250 . x).
543 University of Canberra (2016). *Food Web Database - University of CANBERRA*.
544 URL: <http://globalwebdb.com/>.

DRAFT

6.1 Box 1

The algorithm follows a series of logical steps to predict resources for all taxa in an arbitrary set of taxa N_1 using a set of taxa N_0 with empirically described interactions from which we can extract sets of consumers and resources and their taxonomy. In this example, we are predicting interactions for a fictitious $N_1 = \{T_1, T_9, T_{10}, T_{11}, T_{12}\}$ using N_0 with information on 12 taxa. This catalogue holds information on consumer or resource for 10 taxa and the taxonomy for all 12 taxa in the list.

N_0 taxa ID	taxonomy	resource	consumer
T_1	$\{a, b, c\}$	$\{T_2, T_3, T_{12}\}$	$\{T_4\}$
T_2	$\{e, f, g\}$		$\{T_1, T_5\}$
T_3	$\{i, j, k\}$		$\{T_5\}$
T_4	$\{m, n, o\}$	$\{T_1, T_5\}$	
T_5	$\{a, b, d\}$	$\{T_8, T_9\}$	$\{T_4\}$
T_6	$\{i, q, r\}$	$\{T_2, T_8\}$	$\{T_4\}$
T_7	$\{e, f, h\}$		$\{T_1, T_6\}$
T_8	$\{s, t, u\}$		$\{T_5, T_6\}$
T_9	$\{s, t, v\}$		$\{T_5\}$
T_{10}	$\{i, j, l\}$		
T_{11}	$\{m, n, p\}$		
T_{12}	$\{q, r, s\}$		$\{T_1\}$

Similarity between all pairs of taxa in N_0 is measured for consumer, resource and taxonomic proximity using equation 1. The upper triangular matrix represents similarity measured with taxa sets of resources/consumers, while the lower triangular represents taxonomic similarities. For consumer/resource set similarities, values of 0 mean that similarity equals 0 for both similarity measurements.

$$\frac{\text{tanimoto}(T_Cx, T_Cy)}{\text{tanimoto}(T_Rx, T_Ry)}$$

$$\text{tanimoto}(T_Tx, T_Ty)$$

From these, the algorithm goes through logical steps (Figure 1) to identify

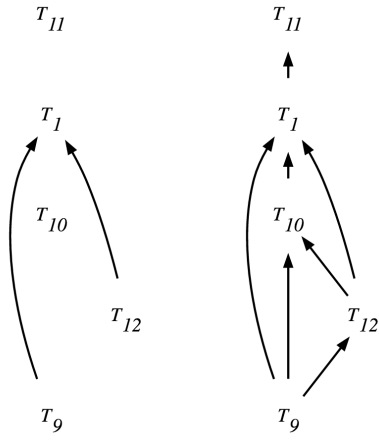
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
T_1	-	0	0	0	0/1	0.3/1	0	0	0	0	0	0
T_2	0	-	0/0.5	0	0	0	0/0.3	0/0.3	0/0.5	0	0	0/0.5
T_3	0	0	-	0	0	0	0	0/0.5	0/1	0	0	0
T_4	0	0	0	-	0	0	0	0	0	0	0	0
T_5	0.5	0	0	0	-	0.3/1	0	0	0	0	0	0
T_6	0	0	0.2	0	0	-	0	0	0	0	0	0
T_7	0	0.5	0	0	0	0	-	0/0.3	0	0	0	0/0.5
T_8	0	0	0	0	0	0	0	-	0	0	0	0
T_9	0	0	0	0	0	0	0	0.5	-	0	0	0
T_{10}	0	0	0.5	0	0	0.2	0	0	0	-	0	0
T_{11}	0	0	0	0.5	0	0	0	0	0	0	-	0
T_{12}	0	0	0	0	0	0.5	0	0.2	0.2	0	0	-

560 a candidate resource list C_R for each taxon in N_1 using either empirical data
561 directly or K most similar taxa with equation 2. Going through the process for
562 T_1 , using $K = 1$ and $w_t = 1$:

Steps	Catalogue	Prediction
1 $I(T_1, T_R)$ in N_0 ?		
2 T_R in N_1 ?		
4-7 $T_2 = \text{no} \rightarrow t(T_2, T_{R'}, w_t) = \text{NA}$	$\{\}$	$\{\}$
4-7 $T_3 = \text{no} \rightarrow t(T_2, T_{R'}, w_t) = T_{10} = 0.5$	$\{\}$	$\{T_{10}\}$
3 $T_{12} = \text{yes}$	$\{T_{12}\}$	$\{T_{10}\}$
8 $t(T_1, T_{C'}, w_t) = T_5 = 0.5$		
9 $I(T_5, T_R)$ in N_1 ?		
13-16 $T_8 = \text{no} \rightarrow t(T_8, T_{R'}, w_t) = T_9 = 0.5$	$\{T_{12}\}$	$\{T_9, T_{10}\}$
10-12 $T_9 = \text{yes}$	$\{T_9, T_{12}\}$	$\{T_9, T_{10}\}$

563 The logical steps allow us to predict a set of resources for $T_1 = \{T_9, T_{10},$
564 $T_{12}\}$. Doing it for all taxa in N_1 with $w_t = 0$ and 1 predicts the following
565 networks:

$$w_t = \mathbf{0} \qquad w_t = \mathbf{1}$$



566 6.2 Figures

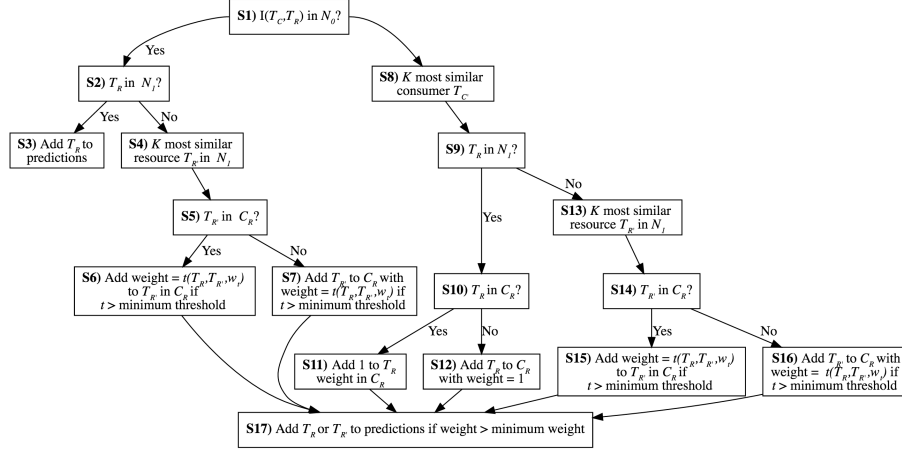


Figure 1: Description of logical steps used by the algorithm to suggest a list of candidate resources (C_R) for each consumer taxa (T_C) in an arbitrary set of N_1 for which interactions are predicted, using a set of taxa N_0 with empirically described interactions. Interactions between consumer and resource taxa are denoted as $I(T_C, T_R)$. K is the number of most similar neighbours selected for the KNN algorithm; t stands for tanimoto in equation 1; w_t is the weight given to sets of resources and consumers in equation 2; the minimum threshold is an arbitrary value setting the minimal similarity value accepted for taxa to be considered as close neighbours in the KNN algorithm; the weight is the value added to a candidate resource each time it is added to C_R ; the minimum weight is the minimal weight value accepted for candidate resources to be selected as predicted sources in the algorithm.

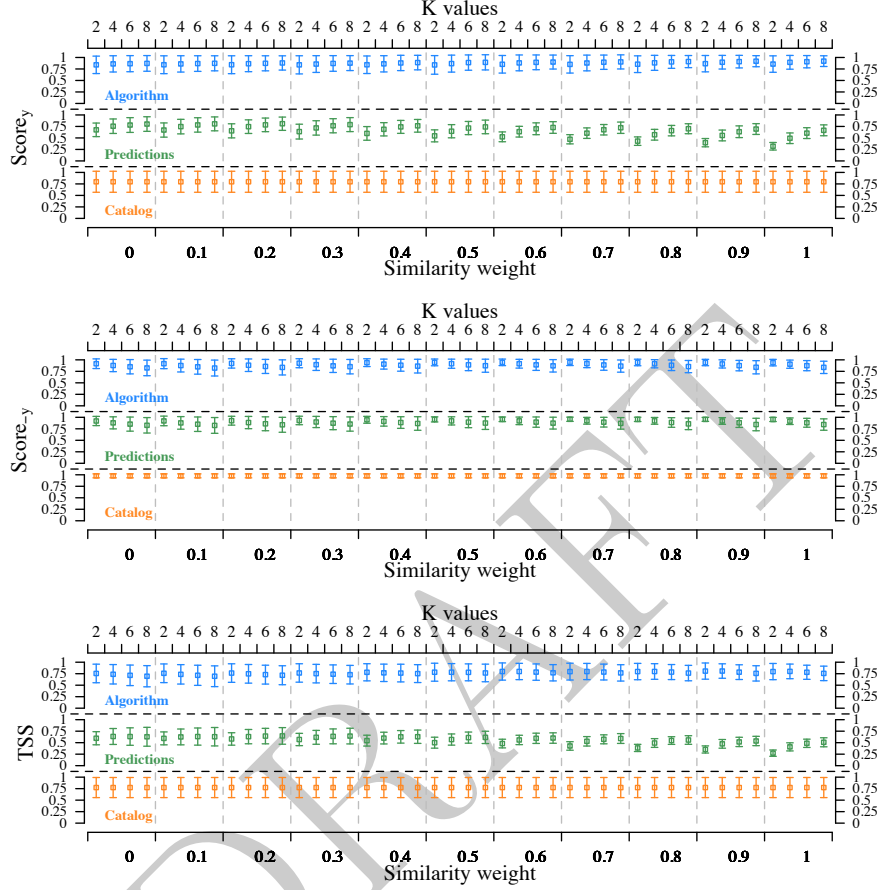


Figure 2: Representation of the three statistics (*i.e.* $Score_y$, $Score_{-y}$ and TSS) used to evaluate the accuracy of the algorithm as a function of K values tested (*i.e.* 2, 4, 6 and 8 most similar neighbours, top x -axis) and weight for taxonomy (bottom x -axis), which varies between 0 and 1. A weight of 0 means that similarity is measured only using set of resources/consumers for each taxa, while a weight of 1 means that similarity is based solely on taxonomy. For each statistic, the topmost panel presents prediction accuracy for the complete algorithm, the middle panel corresponds to predictions made through the predictive portion of the algorithm (Steps S4-S16; Figure 1) and the bottom panel presents the catalogue contribution for the algorithm (Steps S1-S3; Figure 1). Note that the sum of the predictive and catalogue contributions can be over 100% as there is overlap between predictions made through both. The 7 datasets used for this analysis contained over 50 taxa (Thompson2005; Christian and Luczkovich, 1999; Link, 2002; Brose et al., 2005; Barnes et al., 2008; Kortsch et al., 2015)

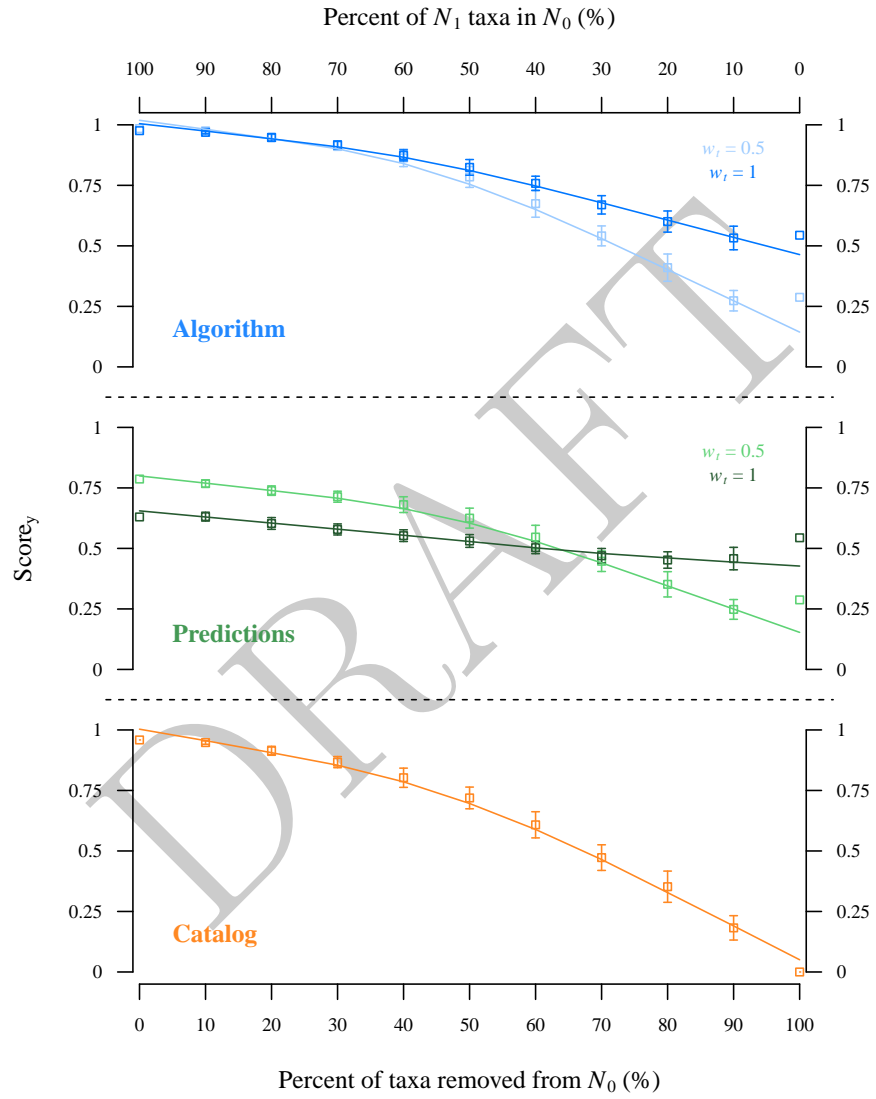


Figure 3: Caption on next page.

Figure 3: Representation of $Score_y$ as a function of catalogue comprehensiveness, *i.e.* the amount of information on sets of consumer and resources available in the catalogue. The sensitivity of the algorithm to data accuracy was evaluated with the arctic food web from Kortsch et al. (2015). This food web was highly detailed taxonomically. Once removed from the catalogue, almost 100% of its taxa still had information available on sets of consumers and resources, which necessary for testing the impact of catalogue comprehensiveness on prediction accuracy. A random percentage of data available in the catalogue for taxa in the food web (*i.e.* 0 to 100%) was iteratively removed ($n = 50$ randomizations) before generating new predictions with the algorithm. w_t values of 0.5 and 1 were evaluated to verify the usefulness of taxonomy in supporting predictive accuracy. The topmost panel presents prediction accuracy for the complete algorithm, the middle panel corresponds to predictions made through the predictive portion of the algorithm (Steps S4-S16; Figure 1) and the bottom panel presents the catalogue contribution for the algorithm (Steps S1-S3; Figure 1). Note that the sum of the predictive and catalogue contributions can be over 100% as there is overlap between predictions made through both.

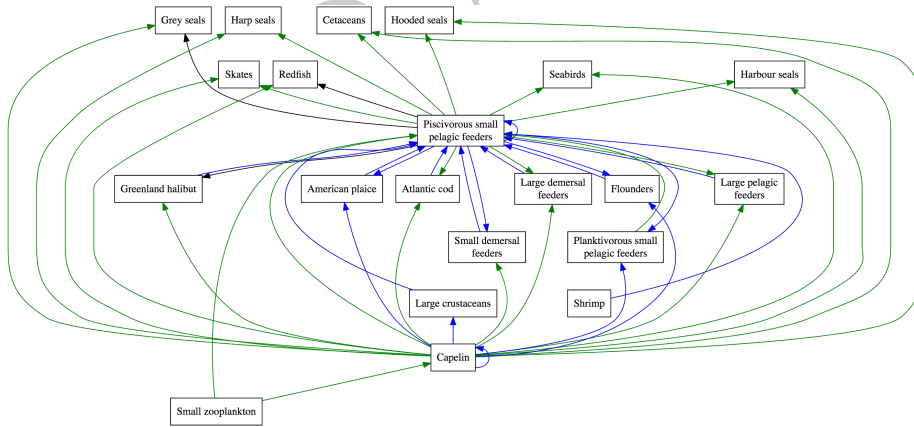


Figure 4: Example of predicted interactions with the network of the southern Gulf of St. Lawrence (Savenkoff et al., 2004), centered around the interactions of the capelin (*Mallotus villosus*) and piscivorous small pelagic feeders (*e.g.* *Scomber scombrus* and *Illex illecebrosus*). Edge with colors green were both predicted and observed (26), black were observed only (3) and blue were predicted only (19). Arrows are pointed towards consumers.