# A highly resolved network reveals the role of terrestrial herbivory in structuring aboveground food webs

Kayla R. S. Hale[1,2]*, John David Curlis[1], Giorgia G. Auteri[3], Sasha Bishop[1], Rowan L. K. French[4], Lance E. Jones[5], Kirby L. Mills[1,6], Brian G. Scholtens[7], Meagan Simons[1], Cody Thompson[1,8], Jordon Tourville[9,10], Fernanda S. Valdovinos[11]

1 Department of Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA
2 Department of Integrative Biology, University of Guelph, Guelph, ON, CA
3 Department of Biology, Missouri State University, Springfield, MO, USA
4 Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON, CA
5 Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA
6 School for Environment & Sustainability, University of Michigan, Ann Arbor, MI, USA
7 Department of Biology, College of Charleston, Charleston, SC, USA
8 Museum of Zoology, University of Michigan, Ann Arbor, MI, USA
9 Department of Environmental Biology, SUNY College of Environmental Science and Forestry, Syracuse, NY, USA
10 Research Department, Appalachian Mountain Club, Boston, MA, USA
11 Department of Environmental Science and Policy, University of California, Davis, CA, USA

* Corresponding author; e-mail: kaylasal@umich.edu

**ORCIDs**

Hale, https://orcid.org/0000-0002-2627-9623; Curlis, https://orcid.org/0000-0002-2974-7330; Auteri, https://orcid.org/0000-0002-5579-8078; Bishop, https://orcid.org/0000-0002-0877-3951; French, https://orcid.org/0000-0002-9178-0035; Jones, https://orcid.org/0000-0001-6368-9720; Mills, https://orcid.org/0000-0001-7693-9629; Scholtens, https://orcid.org/0009-0007-8413-4856; Thompson, https://orcid.org/0000-0002-6686-6056; Tourville, https://orcid.org/0000-0003-4005-5914; Valdovinos, https://orcid.org/0000-0002-5270-5286

**Keywords**

Multiplex ecological network, plant-insect interactions, temperate forest ecosystem, niche model, trophic species, scale-dependence

## Abstract

Comparative studies suggest remarkable similarities among food webs across habitats, including systematic changes in their structure with diversity and complexity (scale-dependence). However, historic aboveground terrestrial food webs (ATFWs) have coarsely grouped plants and insects such that these webs are generally small, and herbivory is disproportionately underrepresented compared to vertebrate predator-prey interactions. Furthermore, terrestrial herbivory is thought to be structured by unique processes compared to size-structured feeding in other systems. Here, we present the richest ATFW to date, including ~580,000 feeding links among ~3,800 taxonomic species, sourced from ~27,000 expert-vetted interaction records annotated as feeding upon one of six different resource types: leaves, flowers, seeds, wood, prey, and carrion. By comparison to historical ATFWs and null ecological hypotheses, we show that our temperate forest web displays a potentially unique structure characterized by two properties: a) a large fraction of carnivory interactions dominated by a small number of hyper-generalist, opportunistic bird and bat predators, and b) a smaller fraction of herbivory interactions dominated by a hyper-rich community of insects with variably-sized but highly-specific diets. We attribute our findings to the large-scale, even resolution of vertebrate, insect, and plant guilds in our food web.

## 1) Introduction

Ecosystems contain immense biological complexity. Food webs represent part of this complexity by documenting the feeding interactions (links) between taxa (nodes). Comparative studies of food webs across habitats have revealed robust and nonrandom patterns suggestive of an underlying architecture of life [1–7]. For example, the structure of food webs changes systematically with diversity and complexity (termed “scale-dependence”) but maintains

hierarchy, demonstrated in aquatic and belowground food webs through modular, size-structured pathways of larger consumers feeding on smaller resources [8–13]. Aboveground terrestrial food webs (ATFWs) also exhibit size-structure in predator-prey interactions, but different mechanisms (e.g., chemical composition, trait matching) likely underlie the specialized feeding of insect herbivores on often-larger terrestrial plant resources [14]. However, there are few published ATFWs and even fewer that include high-resolution data for both plant-herbivore and predator-prey interactions across broad taxonomic groups (Table 1). Therefore, whether and how the structure of ATFWs may fundamentally differ from those of other habitats remains unclear [15]. As a step towards answering this question, we construct the most extensive ATFW to date and study the mechanisms by which its increased taxonomic and trophic resolution lead to a unique structure relative to the scale-dependent pattern observed in previous webs.

Constructing food webs is fraught with methodological difficulties [16,17]. Observations of species and interactions depend on the boundaries of the system, the specific spatial (vertical versus horizontal transects, microhabitats) and temporal (seasonal, diurnal, duration) scales of sampling, as well as the taxonomic expertise of the investigators (including ability to detect and identify both consumer and resource species). Many organisms regularly cross ecosystem boundaries as part of their life cycles; for example, some insect species spend larval stages underground or underwater, then move to aboveground habitats after maturation, after which they may migrate to a completely different region for breeding [18]. Species also exhibit adaptive foraging and defensive behaviors, effectively “rewiring” interactions in response to changing biotic and abiotic conditions [14,19,20]. One approach to these problems is to construct an expert-vetted “cumulative” or “meta” food web that pools all species and interactions recorded across time and/or similar habitats [21]. This reduces the likelihood of missing cryptic

88 or rare species and provides a more comprehensive accounting of all potential feeding
89 interactions in the system. Additionally, as human activities alter species' distributions and
90 habitats, "rare" and novel interactions are increasing in frequency [19], making cumulative webs
91 even more important.

92 Even cumulative food webs rely heavily on expertise and long-term and/or regional
93 sampling. Perhaps for this reason, previous high-quality ATFWs have tended to focus either on
94 taxonomic *breadth* or *depth*. Webs with taxonomic *breadth* (Table 1, gray highlighted rows) tend
95 to resolve vertebrates most highly, while aggregating invertebrates and plants into coarse
96 taxonomic or functional groups (e.g., into insect orders or plant tissue categories). This
97 'lumping' strategy *sensu* Briand [22] seeks to describe broad system-level behavior but is largely
98 a result of the technical difficulties associated with documenting and representing the pure
99 volume of plant-insect associations [23]. Nevertheless, classic breadth webs – Coachella Valley
100 [24], St. Martin Island [25], and El Verde Rainforest [23] – have proven highly influential and
101 remain perhaps our best description of ATFWs because they used cumulative approaches with
102 known species lists from long-term fields sites. Indeed, these webs, with the Little Rock Lake
103 web of Martinez [26], contributed to overturning 'empirical generalizations' (such as scale-
104 invariance, low omnivory, etc.) derived from a catalogue of less resolved webs [27–29].

105 In contrast, webs with greater *depth* of resolution tend to have narrower scope (Table 1).
106 These webs generally focus only on a single taxonomic group (e.g., only tetrapods) or on a single
107 energetic pathway (e.g., 'source' or 'sink' webs) [21]. In the same vein, the explosion of
108 ecological networks research in the last two decades has tended to focus on highly specific single
109 interaction types such as frugivory or scavenging, demonstrating the unique structure and
110 importance of these subnetworks for ecosystem dynamics and function [30–33]. However,

111 different subnetworks are rarely recorded in the same system, and as such, it is unknown how
112 they may connect with each other or to their broader food web [30]. An exception is the few
113 'multiplex' networks that report high resolution interactions of different types (i.e., feeding on
114 different resources or with different interaction outcomes) among non-disjoint sets of species
115 [34,35]. These studies bring together interactions that otherwise rarely co-occur in food webs
116 [31–33], especially mutualisms, such as pollination or seed dispersal, which can have a feeding
117 component via consumption of nectar and pollen or seeds and fruits, with other forms of
118 'antagonistic' herbivory such as phloem-feeding by aphids [36–38].

119 Comparative studies attempt to standardize these diverse approaches to constructing food
120 webs in three ways [21,39]. First, they aggregate empirical webs to "trophic species" webs,
121 where taxa with the same set of consumers and resources are grouped into the same node [28].
122 This reduces methodological biases within and between webs by retaining only functionally
123 distinct units with unique trophic niches [2]. Second, they compare the properties of trophic
124 species webs to null expectations provided by the well-known "niche model" of Williams and
125 Martinez [8], which embodies specific ecological hypotheses for the mechanisms structuring
126 food webs [40]. This approach provides scale-dependent expectations for food web properties
127 (i.e., given their richness and complexity), and deviations from null expectations (sometimes
128 called "errors") can be interpreted as rejecting the underlying hypotheses. However, errors are
129 also scale-dependent, meaning that the properties of empirical webs increasingly deviate from
130 niche model expectations with increasing richness [39]. Therefore, third, comparative studies
131 extrapolate from scale-dependent errors to assess whether a focal web exhibits unique properties
132 compared to other webs, given its scale [39,41].

In this study, we leveraged more than a century of research at a biological research station to build the Michigan Temperate Forest food web, the richest ATFW to date. We used a cumulative approach, incorporating public records and occurrence data, supplemented and vetted by experts for local plausibility given species' traits and behaviors. This resulted in ~580,000 feeding links among ~3,800 taxonomic species, represented in a multiplex network according to feeding on different resource types ("prey," "carrion," "leaves," "flowers," "seeds," or "wood"). Using comparative food web methods, we (1) characterized the properties of the Michigan Temperate Forest, (2) studied whether the increased taxonomic and trophic resolution leads to unique structure, given its scale, compared to previous ATFWs and, if so, (3) identified potential mechanisms underlying the structure of more or less resolved webs.

## 2) Methods

### a) Site, species list, and feeding records

The University of Michigan Biological Station (UMBS) was established in 1909 on ~10,000 acres of logged and burned land in northern lower Michigan, USA (45°35.5′ N, 84°43′ W) and represents a strongly seasonal system with historically cold, snowy winters and hot, humid summers. The site has, in recent years, been restored to predominantly dry-mesic, northern hardwood forests with patches of wooded wetlands (hardwood conifer swamp) [42]. A full description of the UMBS site and extended methods are available in the Electronic Supplementary Materials.

Briefly, experts (generally, the authors) vetted and approved species lists of mammals, amphibians, reptiles, vascular plants, birds, insects, and non-insect arthropods from UMBS records. These were accumulated from resident biologists' personal observations, student projects, museum and herbarium specimens, and semi-regular BioBlitz events, in which teams of

biologists roamed the site and identified as many organisms as possible. Hereafter, we refer to all approved taxa as "species," though a small fraction (4.5%) are genera.

The same experts vetted and annotated a list of potential feeding interactions, sourced from region-specific field guides and online databases [43]. Each focal taxon was resolved to species-level, but their interaction partners could be recorded at any taxonomic level (e.g., species *x* eats family *y*). We included all records of direct interactions among species in our system with a bioenergetic flow (i.e., one species consuming another), regardless of lifestage or potential ecological effects (i.e., whether potentially "mutualistic" or "antagonistic", see [14]. Experts approved recorded interactions between species as plausible if the species co-occur (with respect to phenology, activity patterns, microhabitat usage) and have no trait incompatibilities (with respect to acquisition, ingestion, and assimilation). If a partner in a potential interaction was recorded at a coarser taxonomic level than species, the record was approved only if these conditions could also plausibly hold for all local species in that taxonomic unit.

Lastly, experts categorized records by their focal resource type as animal tissues, either (1) live tissues and as prey, or (2) scavenged as carrion, carcasses, or other decaying animal remains, or as plant tissues, grouped as (3) leaves and stems, (4) flowers, nectar, pollen, etc., (5) seeds, fruits, etc., or (6) wood and bark. Hereafter, we refer to these resource types simply as "prey," "carrion," "leaves," "flowers," "seeds," and "wood," respectively.

**b) Network representation**

To translate our list of feeding records into a food web, we began with a "multiplex network" approach (Fig. 1A) in which feeding on different resource types is represented by different types of links between the same set of nodes (taxonomic species). This allowed us to distinguish between the niches of animals feeding on different resources while also accounting

for the fact that such resources are coupled together in the same organism. Specifically, we defined a node for each focal species $i$ in our list. Then, we defined a directed link of type $l$ between nodes $i, j$ if $i$ consumes tissue type $l$ of $j$ or tissue type $l$ of a broader taxonomic group including $j$. Links are binary, indicating the presence or absence of potential feeding, not its frequency, probability, rate, or strength. We retained only unique links, but tracked the most resolved taxonomic level from which each link was sourced. We characterized the complexity of the multiplex network by counting the number of links ($L_l$), consumer species ($A_l$), and resource species ($P_l$) involved in feeding of type $l$, and the fraction of the maximum possible links that were realized as either bipartite connectance $= L_l/(A_l P_l)$ for feeding on plant tissues (leaves, flowers, seeds, or wood) or unipartite connectance $= L_l/(A_l + P_l)^2$ for feeding on animal tissues (prey or carrion).

As a direct comparison to previously published food webs (Table 1), we used the conventional food web approach, where a binary link occurs between nodes $i, j$ if $i$ consumes $j$ in the multiplex network (i.e., consumes any resource type of species $j$). Following convention, we aggregated all webs into "trophic species" versions, wherein taxa with the same sets of consumer and resource species are grouped together (Fig. 1BIII).

To study the contribution of resolving feeding on different resource types to food web structure, we aggregated our multiplex network to a "trophic species multiplex network" (Fig. 1BI-II), wherein all taxa with the same set of consumers and resources, both in terms of taxonomic species and resource types, are grouped into a single node. The difference in resulting richness between this trophic species multiplex network (Fig. 1BII) and that of our trophic species food web (Fig. 1BIII) indicates how many taxonomic species' trophic niches are

differentiated only by feeding on specific resource types (e.g., on the leaves versus the flowers or the live prey versus the carrion of the same resource species).

Hereafter, we discuss the structure of a "food web" as the trophic-species food web with richness denoted as $S$, number of links denoted as $L$, and directed connectance (the fraction of observed links out of the maximum possible, $L/S^2$) denoted as $C$ [26].

**c) Food web structure**

To provide null expectations for the scale-dependent structure of food webs, we used the niche model of Williams & Martinez [8] to simulate $N = 1{,}000$ matching webs using the $S$ and $C$ of our food web and each of the previous webs (Table 1). Traditionally, the model assumes that all nodes are unique trophic species and webs that include nodes with the same consumer and resource set (i.e., duplicate trophic species) are rejected. For webs with low connectance, we slightly relaxed that assumption and allowed niche model webs to be seeded with a slightly higher initial richness as long as (following another trophic species aggregation) $S$ and $C$ matched the empirical trophic-species web.

We calculated a suite of properties to characterize the Composition, Hierarchy, and Degree Distribution of the empirical food webs [8,39,44–46]. See Supplementary Table 1 for a full list of properties and definitions. For each structural property, we assessed the significance of deviations from null expectations using normalized model errors (NMEs) [44,47]. NMEs are calculated as the difference between the median model and empirical values normalized by either the difference between the median model value and the 97.5 percentile of the model distribution if the empirical value is greater than the model median, or, if the empirical value is less than the model median, by the difference between the 2.5 percentile of the model distribution and the model median. Values >1 or <–1 indicate that the empirical value is significantly higher or

224 lower, respectively, than the null expectation at the 95% confidence level. For the purposes of
225 discussion, we follow previous works to summarize these as a composite *mean* |NME|, though
226 the individual properties are not independent (see Supplementary Methods) [39].

227 To characterize Species Composition, we calculated the *fraction of trophic species* in
228 each of the following categories: *Basal (B)*: with consumers but no resources; *Intermediate (I)*:
229 with both consumers and resources; *Top (T)*: with resources but no consumers; *Herbiv (TL2):* eat
230 only basal species (are strict herbivores, i.e., trophic level [TL] = 2); *Carniv:* eat only other
231 consumers (strict carnivores); *Omniv:* eat both basal and consumer species (omnivores); *Cannib:*
232 eat members of their own species (cannibals).

233 To characterize Link Composition*,* we calculated *HerbLink,* the fraction of total feeding
234 links that are herbivorous (i.e., are on basal resources), and *TL2Link,* the fraction of feeding links
235 from TL2 herbivores.

236 To characterize Hierarchy*,* we calculated *meanTL and maxTL,* the mean and maximum
237 short-weighted trophic level of consumers [46], and *meanTLTop*, the mean trophic level of top
238 consumers.

239 Degree describes the number of resources (in-degree) and consumers (out-degree) of a
240 species. To characterize Degree Distribution, we calculated *meanGen*, the mean in-degree of
241 consumers (i.e., their "generality"), *GenSD*, the normalized variability of generality, *meanVul*,
242 the mean out-degree of resources (i.e., their "vulnerability"), and *VulSD*, the normalized
243 variability of vulnerability. We also calculated these properties for specific subgroups of species
244 to characterize their respective contribution to deviations observed from null expectations.
245 Finally, we used two-sample *Kolmogorov-Smirnov* tests to directly test the null hypothesis that

246 the observed and expected degree distributions were sampled from the same underlying
247 distribution [47]. Full results are reported in Supplementary Table 2.
248 We observed that the niche model tended to generate webs with low $B$ at high $S$, which
249 may drive substantial NMEs simply due to underlying differences in species composition. We
250 directly tested whether this may be the case by simulating $N$ = 1,000 niche model webs with
251 matching $S, C, B$ values to our empirical food web, called the "basal-matched" treatment. We
252 seeded the niche model with $S$ = 2,896, $C$ = 0.057, and set the 927 lowest niche-value species to
253 have feeding ranges of zero (thus forcing them to be basal).
254 All data cleaning and network analyses were performed in MATLAB R2021b [48].

255 **3) Results**

256 **a) Species list and feeding records**

257 Our final species list includes 3,802 local species, representing 2,073 genera in 451
258 families of 85 orders (Fig. 1A). Insects (2,669 spp.) and vascular plants (781 spp.) numerically
259 dominate the community, accounting for ~90% of the taxa, compared to vertebrates (313 spp.)
260 and non-insect arthropods (39 spp.). The richest orders are insects, especially the Lepidoptera
261 (butterflies and moths, 1,168 spp.), Coleoptera (beetles, 512 spp.), Diptera (flies, 390 spp.),
262 Hymenoptera (bees, wasps, and ants, 265 spp.), and Hemiptera (true bugs, 211 spp.). Worldwide,
263 there are more than twice as many named species of Coleoptera as Lepidoptera [49]; therefore,
264 Lepidopterans are likely substantially overrepresented in our list. Vascular plants represent the
265 most taxonomically diverse group on our list, with 38 orders; however, most of these species
266 belong either to Poales (grasses, sedges, and rushes, 149 spp.) or Asterales (composite flowers,
267 96 spp.). Vertebrate species include birds (226 spp., including 127 passeriform birds), mammals
268 (52 spp., including 7 bats), amphibians (18 spp.), and reptiles (17 spp.). Finally, non-insect

arthropods primarily include spiders and mites, but overall, this group is significantly underrepresented in our list, both in terms of richness and taxonomic diversity.

In sum, we recorded 2,541 species of consumers (including 4 carnivorous or parasitic plants) and 3,782 species of resources. Approximately, two-thirds of species (69.7%) included in the records are resolved to taxonomic species level, while 13.7%, 13.1%, and 3.0% of species (primarily insects) have records from genus-, family- or order-level records at best, respectively. We have no feeding records for 19 plants (0.5% of local species), including most of the Lycopodiales (clubmosses, 5 of 6 spp.) and the Polypodiales (ferns, 11 of 19 spp.), which represent two of the major groups of non-seed plants in our system. Additionally, our records include no diet information for 485 species of insects (16.3% of local animals), primarily from the richest orders (183 Lepidopterans, 145 Dipterans, 109 Coleopterans), but also including all Blattodea (cockroaches, 2 spp.), Plecoptera (stoneflies, 2 spp.), Mecoptera (scorpionflies, 4 spp.), and Zygentoma (silverfish, 1 spp.). Some of these do not feed in aboveground terrestrial habits (or at all) during a certain life stage or feed entirely upon resources we excluded (fungi, detritus, lichens, etc.), limiting their potential diet in our food web. However, these gaps in our dataset may also indicate broader gaps in our expertise or the available natural history information for these species.

We constructed our multiplex network from 26,742 approved records of local taxa feeding on prey, carcass, leaf, flower, seed, or wood resources, totaling 588,416 unique feeding interactions (links) between local species. These links primarily consist of feeding on prey (89.2%), especially insect prey, with the remaining links consisting of scavenging carcasses (0.54%) or feeding on plant leaves (6.7%), flowers (2.2%), seeds (1.2%), or wood (0.17%). Though numerous, these interactions are only a small fraction of the possible links. Separating

292 feeding on each type of resource, we calculate that only 5.8% of the interactions are realized
293 among the 3,023 predator and prey species in our carnivory subnetwork, with similarly low
294 connectances for scavenging (0.2% among 1,357 spp.) and the different types of herbivory
295 (leaves: 3.1% among 2,458 spp., flowers: 2.2% among 1,533 spp., seeds: 5.1% among 796 spp.,
296 wood: 9.8% among 200 spp.).

297 Carnivory is the most numerically dominant interaction in our food web, but only
298 approximately one-third (35.9%) of consumers feed on prey. In fact, the 120 most generalist
299 species in our food web (4.7% of consumers) contribute over half (51.5%) of the unique
300 carnivory links in our network, sourced from only 950 (3.6%) records of focal birds and bats
301 thought to feed opportunistically upon entire insect orders. Records of feeding on insect orders
302 by any taxon contribute 87.1% of unique carnivory links overall, meaning that they are not
303 otherwise included by records at lower taxonomic levels. In comparison, feeding between
304 vertebrates accounts for only 1.3% of carnivory links.

305 Herbivory interactions are less numerous than carnivory, but most consumers (85.4%) in
306 our food web feed on plants, with nearly half (46.0%) feeding on a single plant tissue. These are
307 primarily insects, dominated by lepidopterans eating leaves (as caterpillars), but also including
308 hymenopterans and dipterans eating floral resources. Over half of herbivory interactions (52.4%)
309 stem from records of feeding between insects and plants at the genus- and species-level, with
310 only 4.1% of unique herbivory interactions contributed by order-level records across all
311 taxonomic groups. Therefore, in contrast to carnivory, our herbivory records at coarser
312 taxonomic levels do not include or are redundant to interactions from more taxonomically-
313 resolved records.

Over one-third of consumers (39.6%) feed upon more than one type of resource. Around half of these feed on leaves and flowers (18.9% of consumers, primarily lepidopterans and coleopterans). A smaller fraction (11.6%) feed on more than two types of resources, but these represent a more diverse set of insects, mammals, and birds feeding on prey, leaves, and flowers or seeds, or, less frequently, leaves, flowers, and wood. Ants (Formicidae, 48 spp.) uniquely feed on prey, leaves, flowers, and carrion. Among consumers feeding on multiple resource types, we observed significant positive correlations between diet breadths when feeding on prey and plant leaves (Pearson correlation: $r = 0.26$, $p = 2.7 \times 10^{-6}$, $N = 319$), prey and seeds ($r = 0.14$, $p = 0.044$, $N = 202$), leaves and seeds ($r = 0.42$, $p = 7.2 \times 10^{-9}$, $N = 178$), and leaves and wood ($r = 0.24$, $p = 0.017$, $N = 98$; Supplementary Fig. 1). In other words, generalists on one resource type also tend to be generalists on others. In contrast to the tissue specialization by most animals, most plants (91.3% of 781 species) support consumers on more than one of their tissues, with a small set of diverse plants (81 species in 13 orders) sustaining feeding on all four recorded tissue types (Supplementary Fig. 2).

**b) Trophic species composition**

Our final food web consists of $S = 2{,}595$ trophic species and $L = 365{,}951$ links. An additional 29 trophic species and 8,462 links would be distinguished by feeding only on different types of plant tissues (Fig. 1B). Nearly all (92.3%, 2,394) of the trophic species nodes correspond to taxonomic species, including all vertebrates, nearly all non-insect arthropods (94.9%), over two-thirds of plants (70.2%), and over half of insects (55.0%) in our species list. The remaining trophic species are each composed of taxonomic species from a single order. This is likely because entire insect orders, the primary resources in the food web, share the same opportunistic/generalist predators and are therefore only distinguished by their diets (Fig. 1C).

Over half of these taxonomic species (54.1%, 761) are represented in only 27 trophic species groups; these comprise most (83.1%) of the 485 animal species without diet information (that is, the species most likely lacking resolution).

Feeding on leaves contributed to distinguishing the most trophic species in our food web (924 trophic species), followed by feeding on prey, flowers, seeds, wood, and then carrion (417, 322, 13, 5, and 0 respective additional trophic species relative to versions of the food web built with feeding only on the other resource types). Therefore, herbivory interactions, and particularly leaf herbivory, provide the key component for distinguishing species' trophic niches in our system.

**c) Scale-dependence of food web structure**

The structure of our food web deviates significantly from the null expectation provided by the niche model in almost every metric. Beginning with trophic species composition (Fig. 3A-E), approximately one-quarter of trophic species in our web are basal ($B_{obs}$ = 0.245) and the remaining three-quarters are intermediate ($I_{obs}$ = 0.755). Nearly half of trophic species are herbivores that eat only basal trophic species (i.e., are trophic level 2, $TL2_{obs}$ = 0.483), and nearly one-quarter are omnivores that eat both basal and consumer trophic species ($Omniv_{obs}$ = 0.235), while <5% are carnivores that feed strictly on other consumers ($Carniv_{obs}$ = 0.038). This is in stark contrast to the null expectation that <2% of species are basal ($B_{null}$ = 0.018), and the remaining are intermediate ($I_{null}$ = 0.977) with <1% TL2 herbivores ($TL2_{null}$ = 0.006). Moreover, the null model predicts one-third of species as carnivores ($Carniv_{null}$ = 0.315), and the remaining two-thirds as omnivores ($Omni_{null}$ = 0.662). However, our observation of almost no top consumers was not significantly different from the null expectation of the niche model ($T_{obs}$ =

359 3.85 x $10^{-4}$, $T_{null}$ = 0.006), and the fraction of cannibals was also similar to the null expectation
360 ($Cannib_{obs}$ = 0.0312, $Cannib_{null}$ = 0.058).

361 Our web contains a higher fraction of herbivorous links ($HerbLink_{obs}$ = 0.135,
362 $HerbLink_{null}$ = 0.018) – especially a higher fraction by TL2 herbivores – than expected
363 ($TL2Link_{obs}$ = 0.071, $TL2Link_{null}$ = 4.87 x $10^{-5}$, Fig. 3F) and tends to be shorter in terms of mean
364 and max trophic level ($meanTL_{obs}$ = 2.22, $meanTL_{null}$ = 4.37; $maxTL_{obs}$ = 4.62, $maxTL_{null}$ = 6.17).
365 Additionally, the mean trophic level of top species in our web is 2 (i.e., herbivores), significantly
366 lower than expected ($meanTLTop_{obs}$ = 2.0, $meanTLTop_{null}$ = 4.81). This may be a limitation of
367 the dataset rather than a true signal; our temperate forest system does not have megaherbivores,
368 and we expect most insect herbivores to experience parasitoidy or natural enemies in
369 aboveground terrestrial habitats.

370 Our web exhibits a higher average generality among consumers ($meanGen_{obs}$ = 186.7,
371 $meanGen_{null}$ = 143.5, Fig. 3H) and a greater variability in generality overall ($GenSD_{obs}$ = 2.82,
372 $GenSD_{null}$ = 1.19, Fig. 3J). Indeed, the distributions of generality between our web and the
373 simulated niche models are significantly different, both visually (Fig. 2B) and statistically
374 according to Kolmogorov-Smirnov tests (see Supplementary Table 2). These deviations can be
375 attributed to two properties: first, a small core of hyper-generalist omnivores ($meanGenOmniv_{obs}$
376 = 481.3, $meanGenOmniv_{null}$ = 144.4; $GenSDOmniv_{obs}$ = 1.35, $GenSDOmniv_{null}$ = 1.17); second, a
377 long tail of TL2 herbivores with greater generality and variability of generality than expected by
378 the niche model ($meanGenTL2_{obs}$ = 20.8, $meanGenTL2_{null}$ =1.42, Fig. 3I; $GenSDTL2_{obs}$ = 1.30,
379 $GenSDTL2_{null}$ = 0.40, Fig. 3K). However, the TL2 herbivores produced by the niche model are
380 far more specialized than the average consumers in our web (compare to $meanGen_{obs}$).

Our web does not exhibit significant deviations in average vulnerability ($meanVul_{obs}$ = 141.1, $meanVul_{null}$ = 141.1) or variability of vulnerability ($VulSD_{obs}$ = 0.583, $VulSD_{null}$ = 0.583, Fig. 3L) compared to the null expectation, despite significantly different degree distributions (Fig. 2C). However, our web specifically has a lower vulnerability of basal trophic species than expected ($meanVulBasal_{obs}$ = 77.6, $meanVulBasal_{null}$ = 139.7, Fig. 3M) and a higher variability of such vulnerability ($VulSDBasal_{obs}$ = 1.052, $VulSDBasal_{null}$ = 0.580, Fig. 3N).

To investigate the extent to which deviations in the structure of our food web from both previous webs and null expectations can be attributed simply to differences in species composition, we also compared our web to an ensemble of niche models with a matched fraction of basal trophic species, *B* (Fig. 3A, hollow triangle). This correction automatically fixes *I* (Fig. 3B) and *meanGen* (Fig. 3H) by restricting the available links to the correct number of consumers (Fig. 2B), thereby improving *Omniv, TL2*, and *TL2Links* (Fig. 3D-G, J). However, this causes a compensatory overestimation of total herbivorous links, TL2 generality, and basal vulnerability ($HerbLink_{nullB}$ = 0.344; $meanGenTL2_{nullB}$ = 105.1, Fig. 3I; $meanVulBasal_{nullB}$ = 200.0, Fig. 3M), as well as a further underestimation of the variabilities of TL2 generality and basal vulnerability ($GenSDTL2_{nullB}$ = 0.92, Fig. 3K; $VulSDBasal_{nullB}$ = 0.31, Fig. 3N). We also tested whether the higher *B* and herbivory interactions in our original trophic-species web were attributable to the presence of basal animals. After removing these species and links, we recovered similar properties to the original web ($B_{obs}$ = 0.232, $I_{obs}$ = 0.768, $TL2_{obs}$ = 0.491, $HerbLink_{obs}$ = 0.117).

This finding clarifies why our web deviates from niche model predictions (Fig. 2D). The strict hierarchy of the niche model causes decreasing *B* and *T* with increasing *S* and therefore becomes very unlikely to generate TL2 herbivores. However, even if forced to generate larger *B*, TL2 herbivores remain rare, with low variability in generality. This is because TL2 herbivores

404 must have feeding ranges that only include basal species, which is most likely with narrow
405 ranges (and therefore low niche values). In contrast, the TL2 herbivores in our web are
406 numerous, have diets on average wider and more variable in size than those produced by the
407 niche model, and their trophic species niches are distinguished by highly specific herbivory
408 interactions rather than by their predator interactions.

409 **4) Discussion**

410 The food web for the Michigan Temperate Forest (MTF) presented here is the largest yet
411 published (Table 1, [6,10]) and begins to shine light on the remarkable richness of feeding
412 interactions between plants and animals in aboveground terrestrial systems. Our cumulative
413 approach allowed us to evenly resolve the diets of both vertebrate and insect feeding guilds,
414 revealing a clear pattern among herbivory and carnivory interactions. Herbivory interactions are
415 far rarer (<15% of trophic links) than carnivory interactions, but the former were primarily
416 sourced from the largest guild in the web (~50% of trophic species) – the strict (i.e., trophic-level
417 two) insect herbivores with variably-sized but highly specific diets, both taxonomically and in
418 terms of plant tissue types. In addition, carnivory interactions dominate the web (>85% of
419 trophic links) but were primarily sourced from a small fraction of trophic species (<4.6%) –
420 hyper-generalist birds and bats thought to feed opportunistically on entire insect orders. The
421 combination of these two properties in our food web leads to a structure that qualitatively
422 deviates from previous aboveground terrestrial food webs (ATFWs) of lower taxonomic and
423 trophic resolution, as well as the scale-dependent null expectations of the niche model.

424 Given the richness and complexity of our trophic web, the niche model predicted an even
425 larger fraction of carnivory interactions (>98%), stemming from the minimal fraction of basal
426 trophic species with no consumers ($B$ < 2%) and herbivores ($TL2$ < 1%), ultimately resulting in

427 an overall normalized model error far greater than for any previous webs studied here (MTF
428 mean |*NME*| = 28.9, Fig. 2E). ATFWs using the classic 'lumping' approach (Coachella Valley,
429 St. Martin Island, El Verde Rainforest, Serengeti de Visser [50]) are generally small, with low *B*
430 and *TL2* (due to extremely coarse representation of plants and insects) and high fractions of
431 carnivory links (due to the disproportionate resolution of vertebrate predators). These properties
432 align with the predictions of the niche model, as quantified by normalized model errors (0.42 <
433 mean |NME| < 1.88, Fig. 2E). High-resolution source and sink ATFWs (Scotch Broom [51], UK
434 Grassland [52,53], High Arctic [38,54–57], Norwood Farm [36]) record primarily parasitoids
435 and parasites as the higher trophic-level consumers but generally include herbivory links on
436 multiple plant tissues. The relative lack of generalist or opportunistic predators leads to high
437 fractions of top species and very low omnivory, both in contrast to our web and the expectations
438 of the niche model (0.97 < mean |NME| < 6.21). Most comparable to the MTF in terms of size
439 and resolution is perhaps Messel Forest [41], though it is from the early Eocene and contains
440 many now extinct species. While the MTF has greater *TL2*, omnivores, and carnivory links, as
441 well as a lower fraction of herbivory links, the significantly different distributions of generality
442 are likely the greatest contrast (Supplementary Table 2). Messel Forest has fewer TL2 herbivores
443 with narrower and less variable diets (deviating less from niche model predictions) and does not
444 show an inflection in its (log) degree distribution caused by a group of hyper-generalists. As a
445 result, the Messel Forest web shows an overall better correspondence to null expectations of the
446 niche model (mean |NME| = 5.28), despite a substantially larger richness than classic webs.
447 The niche model is considered to successfully reiterate the properties of natural food
448 webs and is even used to simulate network structures for studies of food web dynamics, though it
449 is known to underestimate *TL2* and overestimate average trophic level [21,44]. This success can

broadly be attributed to two mathematical properties, shared by other generating models of food web structure [1,3,58–60]: first, that species can be strictly ordered along a one-dimensional axis by their niche values ($n_1 < n_2 < \cdots n_S$), and second, that species' feeding ranges ($r_i$) are exponentially-distributed, with a decaying probability of feeding on species of lower niche values. The superior performance of the niche model (given its simplicity) is explained by its third property [8]: it generates "interval" webs, because species feed on all resources within intervals of the niche axis (i.e., they have contiguous diets). These properties are also ecological hypotheses for mechanisms that structure food webs [40]. For the MTF, the largest magnitudes of deviations (|NME| > 95) were in terms of species composition, which we attribute to issues with niche packing (Fig. 2D, but see [61]). This suggests that the MTF does not satisfy the first or third properties, since allowing more herbivores on a packed niche axis would require allowing an additional axis of food selection among herbivores of the same niche value or breaks in their diet contiguity. In fact, the MTF, like many others, is not strictly interval. As such, the more pertinent question is the level of intervality of the MTF relative to other webs of its scale and whether this could cause the large magnitude of deviations we observed. Unfortunately, this is combinatorically intractable for us to assess using current methods [3,62]. Lastly, the MTF does not satisfy the second property of exponentially distributed feeding ranges (Fig. 2B-C). We hypothesize that for the niche model to better reproduce MTF properties, links would need to be preferentially allocated to a small group of generalist species and another axis associated with eating plants would need to be introduced to allow for rich communities of herbivores with variably-sized diets [7].

Whether the structural patterns for the MTF could be general among aboveground terrestrial ecosystems remains to be tested. We observed an overall increase in NMEs with

richness (Fig. 2E, Fig. 3), such that we cannot confidently reject the null hypothesis that our observed structure can be attributed to scale-dependent errors associated with the niche model. However, the MTF appears to exhibit a qualitatively different structure than previous works, including a potentially novel degree distribution. Kolmogorov-Smirnov tests rejected the hypothesis that degree distributions of generality and vulnerability for our web were sampled from the same underlying process as the distributions of each of the other ATFWs studied here ($p < 1.0 \times 10^{-5}$ in all cases, Supplementary Table 2). This could imply that terrestrial food webs do have unique structure – potentially driven by the opportunism of predators and a lack of hierarchy among strict herbivores – which may only be observed in the context of high taxonomic and trophic resolution among a rich community of plants, insects, and vertebrates, such as in our temperate forest system.

As always, there are caveats to these conclusions, primarily associated with our methodological approach and general data limitations. It is possible that the critical hyper-generality of bats and birds in the MTF can be attributed simply to lack of taxonomic resolution regarding their specific foraging preferences (e.g., microlepidopterans, a paraphyletic group with <20 mm wingspans, may be too small to be eaten by vertebrate predators). Yet studies examining species-specific diets of bats and birds have supported hyper-generalism (e.g., [63]), even to the extent that molecular identification of species in bat guano presents a roughly equivalent snapshot of insect biodiversity as traditional blacklight sampling of insects [64]. We also know that many plant and insect species are missing interactions because we were not able to find or verify species-specific data (due to taxonomic or other data limitations) or because records were too vague (e.g., “eating seeds” without further specificity). It is therefore possible that herbivory links were underrepresented relative to carnivory in our observations. Interactions

496 could be more thoroughly refined to account for species' traits and ecological habits, which
497 would likely convert some links currently considered plausible to effectively 'forbidden links'
498 *sensu* Jordano et al. [65,66]. Nevertheless, opportunistic carnivory links would still dominate the
499 web and additional herbivory links would likely be taxonomically specific and trophically
500 distinct, which may even serve to increase trophic species richness by distinguishing plants' or
501 herbivores' trophic niches. In short, a further refinement of feeding interactions would most
502 likely align with the food web structure described here.

503 Compiling taxonomic and feeding records for diverse groups into cumulative food webs
504 has the caveat of potentially introducing diverse and compounding sampling biases, making it
505 difficult to quantify uncertainty in the reported food web structure [16]. In the MTF, for example,
506 the relative representation of some groups can be assessed directly through rarefaction curves
507 over field seasons (e.g., lepidopteran leafminers), while some groups (e.g., non-insect
508 arthropods) include diversity for which we were not able to source even regional lists as
509 reference points (e.g., acariform mites). Likewise, feeding on some types of resources, especially
510 by specific taxonomic groups, is more readily observed or charismatic than others, and therefore
511 better documented. As food web research moves towards synthesizing bigger and more diverse
512 data, an important theoretical question for future work is to develop quantitative methods for
513 characterizing how uncertainty may affect observed network structure (see [16,67] for promising
514 methods).

515 Like previous researchers, we chose to limit our scope to the aboveground portion of our
516 food web. However, we recognize that all ATFWs are intimately and inextricably coupled with
517 belowground soil food webs. A substantial fraction of plant biomass may exist belowground
518 [68], creating habitat structure, and providing food for organisms that consume roots, exudates,

519 and detritus. Decomposers return nutrients generated from aboveground waste to the soil for
520 reuptake by plants, often facilitated by mutualistic symbionts (e.g., mycorrhizae) [69]. Moreover,
521 many consumer species (including some in the MTF) live or feed belowground during certain
522 life stages or times of year. As such, complex food web dynamics belowground have
523 considerable impacts on food web dynamics aboveground, and interactions between these two
524 habitats can significantly influence ecosystem-level processes [70,71]. While this fact has long
525 been recognized as a bias in ATFW research, rarely are above- and belowground food webs
526 recorded in the same system (likely due to the logistical challenges of sampling belowground).
527 Doing so represents a critical research frontier for terrestrial food webs as we seek to understand
528 their structure, dynamics, and emergent ecosystem functions.

529 This study is only our first step towards documenting the immense taxonomic diversity
530 and trophic complexity in the temperate forests of University of Michigan Biological Station
531 (UMBS). Ecological networks have historically been published and analyzed as static structures,
532 encapsulating the biases and practical limitations of their collection. As such, publication in
533 online databases and consistent re-use in meta-analyses by ecologists and network scientists can
534 perpetuate errors [17,72]. Our ability to create a large, highly resolved, expert-vetted ATFW was
535 made possible through decades of research and observations at the University of Michigan
536 Biological Station and highlights the importance of local knowledge, taxon-specific expertise,
537 and collaboration among scientists, students, and members of the public. As insights about the
538 organisms present at UMBS will undoubtedly continue to grow, we consider the Michigan
539 Temperate Forest web to be a living dataset that can be revised and expanded through time. To
540 that end, our database is publicly available [43], and we are soliciting revisions, corrections, and
541 additions that will allow its continual improvement.

**Acknowledgements**

We thank Adam Schubel, Jason Tallant, Aimee Classen, Knute Nadelhoffer, and other current and former University of Michigan BioStation (UMBS) staff for providing the original species lists and hosting the living version of the dataset. We are immensely grateful to Teresa Pegan, Eric Gulson, Simone Oliphant, Nate Sanders, Daniel Swanson, Anton Reznicek, Erika Tucker, and undergraduate research team Kathrine Northman, Taylor Brubaker, John Kelly, Lynnae Gilman, Matthew Palumbo, Lex Newman, and Stephan Verral for contributions to data acquisition and vetting. We acknowledge that the Indian Point Reserve (gifted to UMBS in 1987) includes lands of the Burt Lake Band of Ottawa and Chippewa Native Americans from which they were brutally and illegally evicted in 1900. The University of Michigan Undergraduate Research Opportunities Program (UROP) paid the undergraduate researchers for their time. This work was partially funded by NSF grants DEB-2129757 and DEB-2224915 to F.S.V.

**Data Availability Statement**

Species and interaction data are openly available via the Environmental Data Initiative at https://doi.org/10.6073/pasta/840d70788bde4692a7d6d45f8d04376f. Electronic Supplementary Information is available online.

**Competing Interests**

We have no competing interests.

## 5) References

1. Camacho J, Guimerà R, Amaral LAN. 2002 Analytical solution of a model for complex food webs. *Phys. Rev. E* **65**, 030901. (doi:10.1103/PhysRevE.65.030901)

2. Dunne JA, Williams RJ, Martinez ND. 2002 Food-web structure and network theory: The role of connectance and size. *Proc. Natl. Acad. Sci.* **99**, 12917–12922. (doi:10.1073/pnas.192407699)

3. Stouffer DB, Camacho J, Guimerà R, Ng CA, Amaral LAN. 2005 Quantitative Patterns in the Structure of Model and Empirical Food Webs. *Ecology* **86**, 1301–1311.

4. Stouffer DB, Camacho J, Jiang W, Nunes Amaral LA. 2007 Evidence for the existence of a robust pattern of prey selection in food webs. *Proc. R. Soc. B Biol. Sci.* **274**, 1931–1940. (doi:10.1098/rspb.2007.0571)

5. Riede JO, Rall BC, Banasek-Richter C, Navarrete SA, Wieters EA, Emmerson MC, Jacob U, Brose U. 2010 Scaling of Food-Web Properties with Diversity and Complexity Across Ecosystems. In *Advances in Ecological Research*, pp. 139–170. Elsevier. (doi:10.1016/B978-0-12-381363-3.00003-4)

6. Cirtwill AR, Stouffer DB, Romanuk TN. 2015 Latitudinal gradients in biotic niche breadth vary across ecosystem types. *Proc. R. Soc. B Biol. Sci.* **282**, 20151589. (doi:10.1098/rspb.2015.1589)

7. Williams RJ, Purves DW. 2011 The probabilistic niche model reveals substantial variation in the niche structure of empirical food webs. *Ecology* **92**, 1849–1857.

8. Williams RJ, Martinez ND. 2000 Simple rules yield complex food webs. *Nature* **404**, 180–183. (doi:10.1038/35004572)

9. Brose U *et al.* 2006 Consumer-Resource Body-Size Relationships in Natural Food Webs. *Ecology* **87**, 2411–2417.

10. Brose U *et al.* 2019 Predator traits determine food-web architecture across ecosystems. *Nat. Ecol. Evol.* **3**, 919–927. (doi:10.1038/s41559-019-0899-x)

11. Petchey OL, Beckerman AP, Riede JO, Warren PH. 2008 Size, foraging, and food web structure. *Proc. Natl. Acad. Sci.* **105**, 4191–4196. (doi:10.1073/pnas.0710672105)

12. Stouffer DB, Bascompte J. 2011 Compartmentalization increases food-web persistence. *Proc. Natl. Acad. Sci.* **108**, 3648–3652. (doi:10.1073/pnas.1014353108)

13. Potapov AM, Brose U, Scheu S, Tiunov AV. 2019 Trophic Position of Consumers and Size Structure of Food Webs across Aquatic and Terrestrial Ecosystems. *Am. Nat.* **194**, 823–839. (doi:10.1086/705811)

14. Valdovinos FS *et al.* 2023 A bioenergetic framework for aboveground terrestrial food webs. *Trends Ecol. Evol.* **38**, 301–312. (doi:10.1016/j.tree.2022.11.004)

15. Song C, Saavedra S. 2020 Telling ecological networks apart by their structure: An environment-dependent approach. *PLOS Comput. Biol.* **16**, e1007787. (doi:10.1371/journal.pcbi.1007787)

16. De Aguiar MAM *et al.* 2019 Revealing biases in the sampling of ecological interaction networks. *PeerJ* **7**, e7566. (doi:10.7717/peerj.7566)

17. Pringle RM, Hutchinson MC. 2020 Resolving Food-Web Structure. *Annu. Rev. Ecol. Evol. Syst.* **51**, 55–80. (doi:10.1146/annurev-ecolsys-110218-024908)

18. Hu G, Lim KS, Horvitz N, Clark SJ, Reynolds DR, Sapir N, Chapman JW. 2016 Mass seasonal bioflows of high-flying insect migrants. *Science* **354**, 1584–1587.

19. Bartley TJ, McCann KS, Bieg C, Cazelles K, Granados M, Guzzo MM, MacDougall AS, Tunney TD, McMeans BC. 2019 Food web rewiring in a changing world. *Nat. Ecol. Evol.* **3**, 345–354. (doi:10.1038/s41559-018-0772-3)

20. Ceron K, Provete DB, Pires MM, Araujo AC, Blüthgen N, Santana DJ. 2022 Differences in prey availability across space and time lead to interaction rewiring and reshape a predator–prey metaweb. *Ecology* **103**, e3716. (doi:10.1002/ecy.3716)

21. Dunne JA. 2005 The Network Structure of Food Webs. In *Ecological Networks* (eds M Pascual, JA Dunne), pp. 27–92. New York, NY: Oxford University Press. (doi:10.1093/oso/9780195188165.003.0002)

22. Briand F. 1983 Environmental Control of Food Web Structure. *Ecology* **64**, 253–263. (doi:10.2307/1937073)

23. Reagan DP, Waide RB, editors. 1996 *The food web of a tropical rain forest*. Chicago: Univ. of Chicago Press.

24. Polis GA. 1991 Complex Trophic Interactions in Deserts: An Empirical Critique of Food-Web Theory. *Am. Nat.* **138**, 123–155. (doi:10.1086/285208)

25. Goldwasser L, Roughgarden J. 1993 Construction and Analysis of a Large Caribbean Food Web. *Ecology* **74**, 1216–1233. (doi:10.2307/1940492)

26. Martinez ND. 1991 Artifacts or Attributes? Effects of Resolution on the Little Rock Lake Food Web. *Ecol. Monogr.* **61**, 367–392. (doi:10.2307/2937047)

27. Cohen JE. 1989 ECOWeB 1.0: Ecologists' Cooperative Web Bank.

28. Briand F, Cohen JE. 1984 Community food webs have scale-invariant structure. *Nature* **307**, 264–267. (doi:10.1038/307264a0)

29. Schoenly K, Beaver RA, Heumier TA. 1991 On the Trophic Relations of Insects: A Food-Web Approach. *Am. Nat.* **137**, 597–638. (doi:10.1086/285185)

30. Hale KRS, Thébault E, Valdovinos FS. 2023 A general trait-based model for multiplex ecological networks. (doi:10.1101/2023.08.08.552546)

31. Kéfi S *et al.* 2012 More than a meal… integrating non-feeding interactions into food webs. *Ecol. Lett.* **15**, 291–300. (doi:10.1111/j.1461-0248.2011.01732.x)

32. Fontaine C, Guimarães PR, Kéfi S, Loeuille N, Memmott J, Van Der Putten WH, Van Veen FJF, Thébault E. 2011 The ecological and evolutionary implications of merging different types of networks: Merging networks with different interaction types. *Ecol. Lett.* **14**, 1170–1181. (doi:10.1111/j.1461-0248.2011.01688.x)

33. Olff H, Alonso D, Berg MP, Eriksson BK, Loreau M, Piersma T, Rooney N. 2009 Parallel ecological networks in ecosystems. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 1755–1779. (doi:10.1098/rstb.2008.0222)

34. Pilosof S, Porter MA, Pascual M, Kéfi S. 2017 The multilayer nature of ecological networks. *Nat. Ecol. Evol.* **1**, 0101. (doi:10.1038/s41559-017-0101)

35. Kéfi S, Thébault E, Eklöf A, Lurgi M, Davis AJ, Kondoh M, Krumins JA. 2017 Toward Multiplex Ecological Networks: Accounting for Multiple Interaction Types to Understand Community Structure and Dynamics. In *Adaptive Food Webs* (eds JC Moore, PC De Ruiter, KS McCann, V Wolters), pp. 73–87. Cambridge University Press. (doi:10.1017/9781316871867.008)

36. Pocock MJO, Evans DM, Memmott J. 2012 The Robustness and Restoration of a Network of Ecological Networks. *Science* **335**, 973–977. (doi:10.1126/science.1214915)

37. Timóteo S *et al.* 2023 Tripartite networks show that keystone species can multitask. *Funct. Ecol.* **37**, 274–286. (doi:10.1111/1365-2435.14206)

38. Schmidt NM *et al.* 2017 Interaction webs in arctic ecosystems: Determinants of arctic change? *Ambio* **46**, 12–25. (doi:10.1007/s13280-016-0862-x)

39. Dunne JA *et al.* 2013 Parasites Affect Food Web Structure Primarily through Increased Diversity and Complexity. *PLoS Biol.* **11**, e1001579. (doi:10.1371/journal.pbio.1001579)

40. Martinez ND. 2020 Allometric Trophic Networks From Individuals to Socio-Ecosystems: Consumer–Resource Theory of the Ecological Elephant in the Room. *Front. Ecol. Evol.* **8**, 92. (doi:10.3389/fevo.2020.00092)

41. Dunne JA, Labandeira CC, Williams RJ. 2014 Highly resolved early Eocene food webs show development of modern trophic structure after the end-Cretaceous extinction. *Proc. R. Soc. B Biol. Sci.* **281**, 20133280. (doi:10.1098/rspb.2013.3280)

42. Barnett JW, Bergen KM, Currie WS. 2022 Assessing beech bark-diseased forest canopies over landscapes using high resolution open-source imagery in an ecological framework. *For. Ecol. Manag.* **523**, 120486. (doi:10.1016/j.foreco.2022.120486)

43. Hale KRS *et al.* 2023 University of Michigan Biological Station cumulative food web data for terrestrial habitats. (doi:10.6073/PASTA/840D70788BDE4692A7D6D45F8D04376F)

44. Williams RJ, Martinez ND. 2008 Success and its limits among structural models of complex food webs. *J. Anim. Ecol.* **77**, 512–519. (doi:10.1111/j.1365-2656.2008.01362.x)

45. Gauzens B, Rall BC, Mendonça V, Vinagre C, Brose U. 2020 Biodiversity of intertidal food webs in response to warming across latitudes. *Nat. Clim. Change* **10**, 264–269. (doi:10.1038/s41558-020-0698-z)

46. Williams RJ, Martinez ND. 2004 Limits to Trophic Levels and Omnivory in Complex Food Webs: Theory and Data. *Am. Nat.* **163**, 458–468. (doi:10.1086/381964)

47. Pires MM, Prado PI, Guimarães PR. 2011 Do Food Web Models Reproduce the Structure of Mutualistic Networks? *PLoS ONE* **6**, e27280. (doi:10.1371/journal.pone.0027280)

48. Mathworks. MATLAB.

49. Stork NE. 2018 How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? *Annu. Rev. Entomol.* **63**, 31–45. (doi:10.1146/annurev-ento-020117-043348)

50. De Visser SN, Freymann BP, Olff H. 2011 The Serengeti food web: empirical quantification and analysis of topological changes under increasing human impact: Topological changes under human impact. *J. Anim. Ecol.* **80**, 484–494. (doi:10.1111/j.1365-2656.2010.01787.x)

51. Memmott J, Martinez ND, Cohen J. 2000 Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural food web. *J. Anim. Ecol.* **69**, 1–15. (doi:10.1046/j.1365-2656.2000.00367.x)

52. Dawah HA, Hawkins BA, Claridge MF. 1995 Structure of the Parasitoid Communities of Grass-Feeding Chalcid Wasps. *J. Anim. Ecol.* **64**, 708. (doi:10.2307/5850)

53. Martinez ND, Hawkins BA, Dawah HA, Feifarek BP. 1999 Effects of Sampling Effort on

Characterization of Food-Web Structure. *Ecology* **80**.

54. Wirta HK, Vesterinen EJ, Hambäck PA, Weingartner E, Rasmussen C, Reneerkens J, Schmidt NM, Gilg O, Roslin T. 2015 Exposing the structure of an Arctic food web. *Ecol. Evol.* **5**, 3842–3856. (doi:10.1002/ece3.1647)

55. Wirta HK, Hebert PDN, Kaartinen R, Prosser SW, Várkonyi G, Roslin T. 2014 Complementary molecular information changes our perception of food web structure. *Proc. Natl. Acad. Sci.* **111**, 1885–1890. (doi:10.1073/pnas.1316990111)

56. Roslin T, Wirta H, Hopkins T, Hardwick B, Várkonyi G. 2013 Indirect Interactions in the High Arctic. *PLoS ONE* **8**, e67367. (doi:10.1371/journal.pone.0067367)

57. Rasmussen C, Dupont YL, Mosbacher JB, Trøjelsgaard K, Olesen JM. 2013 Strong Impact of Temporal Resolution on the Structure of an Ecological Network. *PLoS ONE* **8**, e81694. (doi:10.1371/journal.pone.0081694)

58. Cattin M-F, Bersier L-F, Banašek-Richter C, Baltensperger R, Gabriel J-P. 2004 Phylogenetic constraints and adaptation explain food-web structure. *Nature* **427**, 835–839. (doi:10.1038/nature02327)

59. Cohen JE, Palka ZJ. 1990 A Stochastic Theory of Community Food Webs. V. Intervality and Triangulation in the Trophic-Niche Overlap Graph. *Am. Nat.* **135**, 435–463.

60. Williams RJ, Anandanadesan A, Purves D. 2010 The Probabilistic Niche Model Reveals the Niche Structure and Role of Body Size in a Complex Food Web. *PLoS ONE* **5**, e12092. (doi:10.1371/journal.pone.0012092)

61. Martinez ND, Lawton JH. 1995 Scale and Food-Web Structure: From Local to Global. *Oikos* **73**, 148. (doi:10.2307/3545903)

62. Eklöf A *et al.* 2013 The dimensionality of ecological networks. *Ecol. Lett.* **16**, 577–583. (doi:10.1111/ele.12081)

63. Clare EL, Fraser EE, Braid HE, Fenton MB, Hebert PDN. 2009 Species on the menu of a generalist predator, the eastern red bat ( *Lasiurus borealis* ): using a molecular approach to detect arthropod prey. *Mol. Ecol.* **18**, 2532–2542. (doi:10.1111/j.1365-294X.2009.04184.x)

64. Maslo B, Mau RL, Kerwin K, McDonough R, McHale E, Foster JT. 2022 Bats provide a critical ecosystem service by consuming a large diversity of agricultural pest insects. *Agric. Ecosyst. Environ.* **324**, 107722. (doi:10.1016/j.agee.2021.107722)

65. Jordano P, Bascompte J, Olesen JM. 2003 Invariant properties in coevolutionary networks of plant–animal interactions. *Ecol. Lett.* **6**, 69–81. (doi:10.1046/j.1461-0248.2003.00403.x)

66. Allesina S, Alonso D, Pascual M. 2008 A General Model for Food Web Structure. *Science* **320**, 658–661. (doi:10.1126/science.1156269)

67. Young J-G, Cantwell GT, Newman MEJ. 2021 Bayesian inference of network structure from unreliable data. *J. Complex Netw.* **8**, cnaa046. (doi:10.1093/comnet/cnaa046)

68. Ma H, Mo L, Crowther TW, Maynard DS, Van Den Hoogen J, Stocker BD, Terrer C, Zohner CM. 2021 The global distribution and environmental drivers of aboveground versus belowground plant biomass. *Nat. Ecol. Evol.* **5**, 1110–1122. (doi:10.1038/s41559-021-01485-1)

69. Bardgett RD, Wardle DA. 2003 Herbivore-mediated linkages between aboveground and belowground communities. *Ecology* **84**, 2258–2268. (doi:10.1890/02-0274)

70. Wardle DA, Bardgett RD, Klironomos JN, Setälä H, Van Der Putten WH, Wall DH. 2004 Ecological Linkages Between Aboveground and Belowground Biota. *Science* **304**, 1629–1633. (doi:10.1126/science.1094875)

71. Jochum M, Eisenhauer N. 2022 Out of the dark: Using energy flux to connect above- and belowground communities and ecosystem functioning. *Eur. J. Soil Sci.* **73**, e13154. (doi:10.1111/ejss.13154)

72. Brimacombe C, Bodner K, Michalska-Smith M, Poisot T, Fortin M-J. 2023 Shortcomings of reusing species interaction networks created by different sets of researchers. *PLOS Biol.* **21**, e3002068. (doi:10.1371/journal.pbio.3002068)

73. Baiser B *et al.* 2019 Ecogeographical rules and the macroecology of food webs. *Glob. Ecol. Biogeogr.* **28**, 1204–1218. (doi:10.1111/geb.12925)

74. Baskerville EB, Dobson AP, Bedford T, Allesina S, Anderson TM, Pascual M. 2011 Spatial Guilds in the Serengeti Food Web Revealed by a Bayesian Group Model. *PLoS Comput. Biol.* **7**, e1002321. (doi:10.1371/journal.pcbi.1002321)

75. Harris LD, Paur, Leonard Francis. 1972 Quantitative food web analysis of a shortgrass community.

76. O'Connor LMJ, Pollock LJ, Braga J, Ficetola GF, Maiorano L, Martinez-Almoyna C, Montemaggiori A, Ohlmann M, Thuiller W. 2020 Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *J. Biogeogr.* **47**, 181–192. (doi:10.1111/jbi.13773)

77. Wang X, Ibáñez I. 2022 The Contrasting Effects of Local Environmental Conditions on Tree Growth between Populations at Different Latitudes. *Forests* **13**, 429. (doi:10.3390/f13030429)

78. Nave LE, Gough CM, Perry CH, Hofmeister KL, Le Moine JM, Domke GM, Swanston CW, Nadelhoffer KJ. 2017 Physiographic factors underlie rates of biomass production during succession in Great Lakes forest landscapes. *For. Ecol. Manag.* **397**, 157–173. (doi:10.1016/j.foreco.2017.04.040)

79. Ricart RD, Pearsall DR, Curtis PS. 2020 Multidecadal shifts in forest plant diversity and community composition across glacial landforms in northern lower Michigan, USA. *Can. J. For. Res.* **50**, 126–135. (doi:10.1139/cjfr-2019-0138)

80. Scott A. Chamberlain. 2013 Taxize Source Code And Program Files.

## 6) Tables & Figures

**Table 1. Properties of aboveground terrestrial food webs.** A collection of classic and more recent food webs used for comparison to our new Michigan Temperate Forest web for the University of Michigan Biological Station. We include the classic webs (published prior to 2000) traditionally considered 'highly resolved;' this excludes the many historical webs in the ECoWeB catalogue of low richness and variable resolution [27]. We exclude modern webs in "container habitats" (e.g., under a log, in a tree hole) as well as newer parasitoid-host webs [6,73]. The High Arctic web was assembled from data in Appendix S2 of Wirta et al. [54], originally from Roslin et al. and Rasmussen et al. [56,57]. Column definitions: *Web:* Traditional name and reference. *Richness*: Number of 'species' in original publication (but see *Extent*). *S*: Number of trophic species. *C:* Directed connectance. *L:* Number of links. *B:* Fraction of basal trophic species (with no consumers). *Taxonomic Groups*: Taxa and resolution in the food web. *Basal Nodes:* Types of trophic species or functional groups at the base of the food web. *Types of Consumers:* Types of feeding interactions included in the food web. *Extent:* Notes on the space and time of food web construction. Unless otherwise noted, webs are *cumulative metawebs,* built from records pooled across time (including published literature) and similar habitats (usually contiguous field sites). Other definitions: "lumped by taxa": grouped to order or family except potentially for key species. *Source-web:* web recording the food chain(s) up from a set of resources. *Sink-web*: web recording the food chain(s) down from a set of consumers. "*Breadth"* webs (highlighted gray) are those that include multiple taxonomic groups and energy pathways but with lower resolution. "*Depth"* webs are those with higher resolution but missing key structural components, including some non-traditional food webs like the Pocock Farm multiplex network. Abbreviations: "incl." including, "inverts" invertebrates. See text for other definitions.

775

| Web | Rich-ness | S | C | L | B | Taxonomic Groups | Basal Nodes | Types of Consumers | Extent |
|---|---|---|---|---|---|---|---|---|---|
| Coachella Valley [24] | 30* | 29 | 0.312 | 262 | 0.103 | vertebrates, arthropods, other inverts, microbes lumped by taxa, size, & trophic role | plants & plant products; detritus; carrion | predators, parasitoids, detritivores, scavengers | *lumps 138 known vertebrate, 55 arachnid, 174 vascular plant, & ~2-3000 insect spp. |
| St. Martin Island [25] | 44 | 42 | 0.116 | 205 | 0.143 | vertebrates resolved to spp.; fungi, arthropods, other inverts, lumped by taxa & lifestage | fruits & seeds; nectar & floral; leaves; roots; wood; detritus | predators, parasites, detritivores, herbivores (incl. nectar & fruit feeders) | "Anolis-centered" |
| UK Grassland [53] | 87 | 67 | 0.027 | 119 | 0.149 | grasses & wasps resolved to spp. | grass spp. | herbivores, parasitoids | source-web; direct observations |
| Scotch Broom [51] | 154 | 83 | 0.032 | 220 | 0.012 | broom, insects, fungi, bacteria, & arachnids resolved to spp. | broom (*Cytisus scoparius*) | predators, parasites, parasitoids, herbivores (incl. seed, sap, & root feeders) | source-web; single field site |
| Serengeti – de Visser [50] | 350* | 85 | 0.075 | 545 | 0.071 | vertebrates resolved to spp.; inverts incl. insects, gastropods, etc. lumped by taxa & size | decaying matter; plant juices; fruits & nectar; grains & seeds; grass & herbs; trees & shrubs | predators, herbivores (incl. seed, nectar & fruit feeders), scavengers | *lumps >322 vertebrate spp., 28 invertebrate orders, 7 resource groups |
| Serengeti – Baskerville [74] | 161 | 105 | 0.042 | 463 | 0.695 | plants & mammals resolved to spp. | plant spp. | predators (carnivores), herbivores | |
| Shortgrass Prairie [75] | 133 | 105 | 0.034 | 379 | 0.657 | plants & vertebrates resolved to spp.; arthropods lumped by taxa | plant spp.; unidentified seeds | predators, herbivores (incl. seed feeders) | sampling over 2 years at 2 sites |
| El Verde Rainforest [23] | 156* | 155 | 0.063 | 1509 | 0.181 | vertebrates & some inverts incl. insects, gastropods, etc. resolved to spp.; other inverts, lumped by taxa & lifestage | resource categories, e.g., leaves; dead wood; roots; bacteria; algae; slime molds; basal prey taxa | predators, parasites, scavengers, detritivores, herbivores (incl. leaf, seed, fruit, nectar, pollen, flower, wood, lichen, root, & sap feeders), fungivores | *lumps >2,601 known heterotroph & >214 autotroph plant spp. |
| High Arctic [54,56,57] | 214 | 161 | 0.025 | 646 | 0.348 | plants, spiders, insects, other inverts, & birds resolved to spp. | plant & basal prey spp. | predators, parasitoids, herbivores, pollinators (nectar & pollen feeders) | source-webs coupled to sink-web following [38]; Zackenberg Valley |
| Norwood Farm [36] | 560 | 327 | 0.011 | 1199 | 0.297 | plants, insects, fleas, ticks, birds, rodents resolved to spp. | plant spp. | parasitoids, herbivores, flower visitors (nectar & pollen feeders), seed feeders, ectoparasites | source-web; recorded as multiplex network; sampled over 2 years at a 125 ha farm |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Messel Forest [41] | 700 | 646 | 0.014 | 5758 | 0.307 | plants, microbes, fungi, arthropods, vertebrates, other inverts; >50% taxa resolved to genus or spp. | Detritus, carrion, plant spp. | predators, herbivores (incl. seed, wood, root, pollen, & leaf feeders) | taxa deposited in a lake basin during the Eocene (cumulative) |
| European Tetrapods [76] | 1152 | 1054 | 0.043 | 47510 | 0.763 | vertebrates resolved to spp. | basal prey spp. | predators (carnivores) | European continent; not all spp. co-occur locally; 17 diet categories are provided as traits |
| Michigan Temperate Forest | 3802 | 2597 | 0.054 | 366385 | 0.245 | vertebrates, plants, insects, & other inverts resolved to spp. | plant & basal prey spp. | predators, parasitoids, ectoparasites, scavengers, herbivores (incl. leaf, seed & fruit, nectar & pollen, & wood feeders) | UMBS field station; recorded as multiplex network |

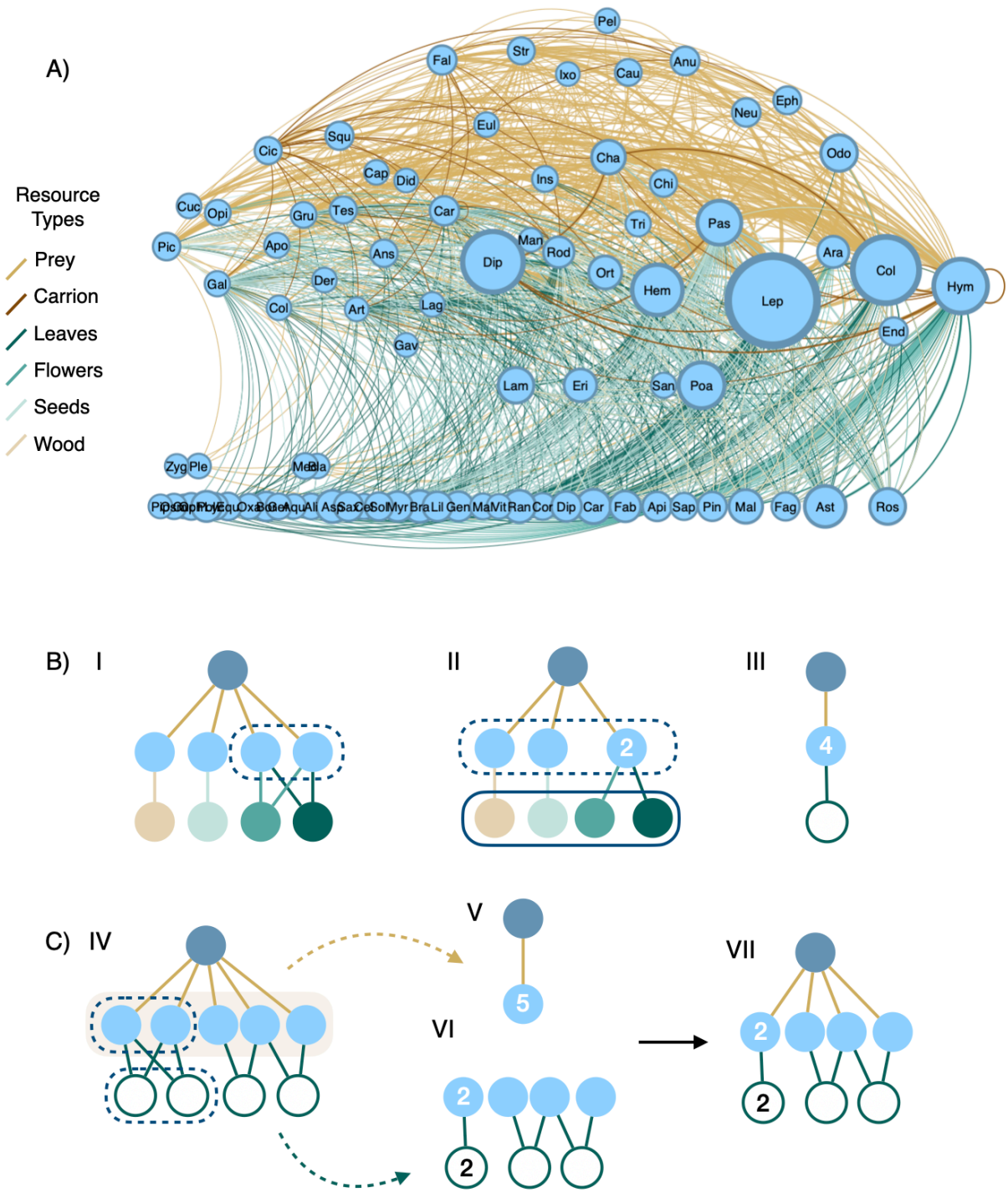



**Figure 1 Representations of the Michigan Temperate Forest food web.**
**(A)** Simplified visualization of the multiplex food web. Each node represents a taxonomic order (labeled by its first three letters), with width scaled to the number of local species, ranging from one (e.g., order Gaviiformes, represented only by the Common Loon) to >1,000 (order Lepidoptera). Each link is a feeding interaction between orders, with width scaled to the total

number of feeding interactions between species in each order. Links colored by the resource type consumed as follows: tan – live prey and animal tissues; brown – scavenged carrion; dark green – leaves and stems; green – floral resources; light green – seeds and fruits; and beige – wood and bark. Self-links indicate feeding among species within the order, including cannibalism. Nodes are ordered horizontally by their number of consumers (in-degree), increasing from left to right, and vertically by increasing trophic level (TL) from basal resources on the bottom (TL = 1) to carnivores at the top. Three carnivorous/parasitic plant orders were assigned TL = 1.75 and four basal animal orders were assigned trophic levels 1.25 for visualization. The 85 orders shown here represent 3,082 taxonomic species. **(B)** Illustrated effect of aggregating the multiplex network into 'trophic species' – species with same set of consumers and resources. Numbers indicate how many nodes were aggregated into a trophic species. Here, one generalist predator feeds on four herbivores of the same plant species, with its different tissue types (wood, seeds, flowers, and leaves) illustrated as different nodes for clarity (colors follow panel A). There are six taxonomic multiplex species total in this example. The 'multiplex network' representation **(I)** can differentiate among the trophic niches of herbivores feeding on different plant tissues. Aggregating into a 'trophic-species multiplex network' **(II)** groups the two herbivores that feed on the same plant tissues into one trophic species, resulting in five multiplex trophic species total. Further aggregating all four herbivores into one trophic species feeding on the plant species (white node), results in three trophic species total. This is the traditional representation of 'trophic-species food webs' **(III)**, where only taxonomic identity differentiates species' trophic niches. **(C)** Illustration of the disproportionate effect of herbivory on trophic species resolution in the Michigan Temperate Forest (MTF) web. The generality of predators (especially birds and bats) in the MTF web **(IV)** means that prey species (especially insect herbivores) tend to be aggregated into fewer trophic species when considering carnivory interactions alone **(V)**. The specificity of herbivores in the MTF web results in more trophic species resolved when considering herbivory interactions alone **(VI)**. This leads to a pattern of numerous generalized carnivory interactions and fewer but more specific herbivory interactions among the more numerous plant and herbivore species **(VII)**.

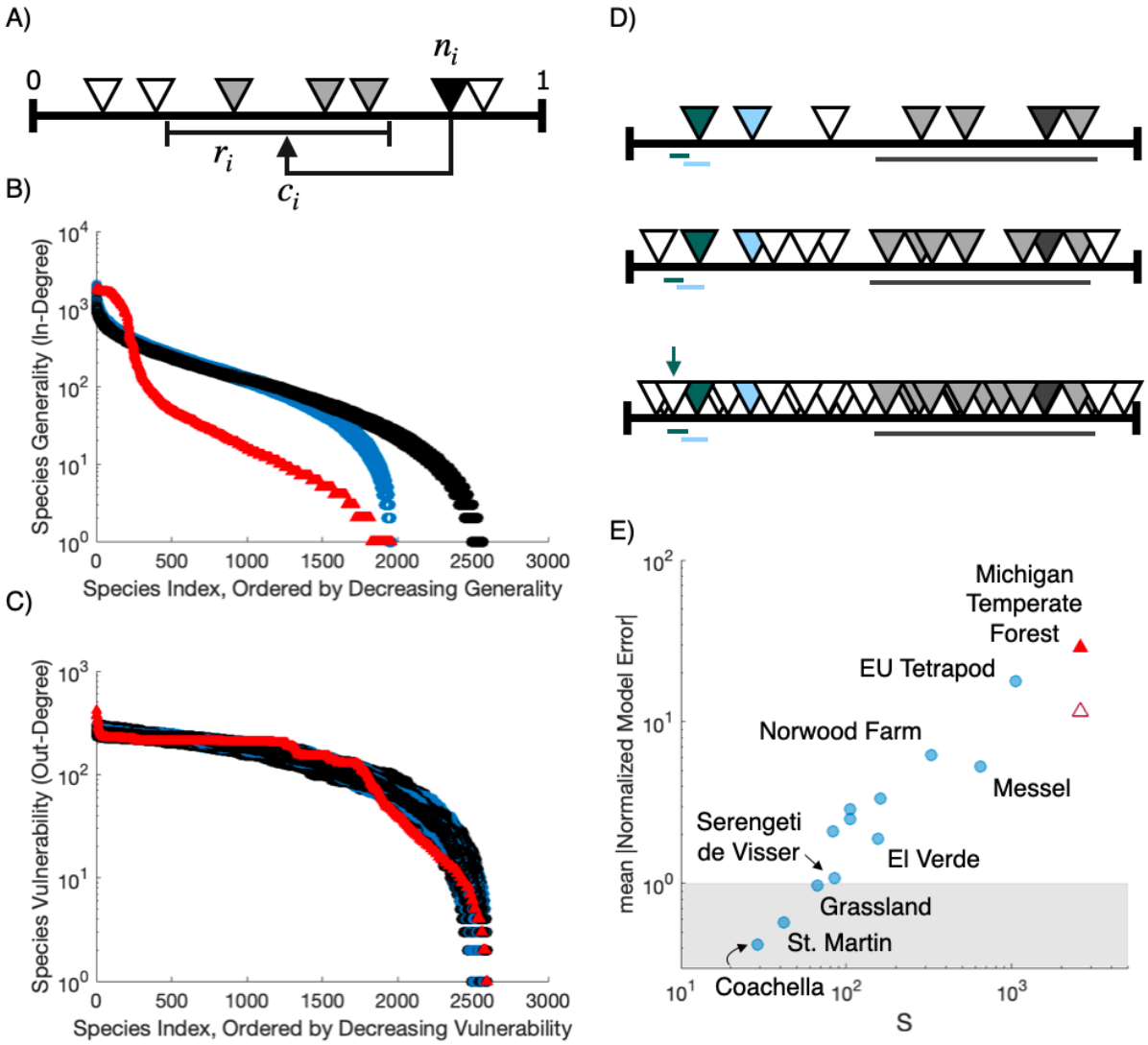



**Figure 2. The Michigan Temperate Forest food web structure deviates from null expectations of the niche model due to problems of niche packing at high richness.** **(A)** We use Williams & Martinez' (2000) niche model to simulate an ensemble of null-model food webs with trophic species richness ($S$) and directed connectance ($C$) matched to the empirical food web. For each species $i$, the model randomly draws a niche value from a uniform distribution ($n_i \sim U(0,1)$) and assigns $i$ a feeding range along the niche axis ($r_i = n_i x$, where $x$ is $\beta$-distributed with $E[x] = 2C$) centered at a value ($c_i$) below its niche value ($c_i \in [r_i/2, min(n_i, 1 - r_i/2)]$). In this way, the higher the niche value of a species, the wider its feeding range. Then, $i$ feeds on all species $j$ with niche values in that range or is considered a basal species if the range is empty. Only trophic species with a unique set of consumers and resources are permitted. To guarantee a basal species, the feeding range of the species with the lowest niche value is set to zero ($r_i := 0$).
**(B)** The observed in-degree distributions for the Michigan Temperate Forest web (red triangles) compared to in-degree distributions from $N = 1{,}000$ simulated niche model webs with matched

$S$,$C$ (black circles) or matched $S,C$, and fraction of basal species, $B$, achieved by setting more species' feeding ranges to zero, blue circles). Degree is on a log scale ($y$-axis), with species ($x$-axis) ordered by decreasing degree. In-degree is generality, the richness of diet. The empirical web has a core of high-degree generalists (birds and bats) and a long tail of lower-degree trophic-level 2 herbivores (insects), leading to a significantly different distribution of generality than predicted by the niche model.
**(C)** Observed and simulated out-degree distributions. Out-degree is vulnerability, the richness of predators.
**(D)** The strict one-dimensional hierarchy of the niche model does not accommodate the rich community of trophic-level 2 herbivores ($TL2$) observed in the empirical food web. As $S$ increases, the niche axis becomes increasingly and uniformly packed with species, represented here as triangles. Horizontal bars represent the feeding ranges of each colored species. At high $S$, the chance that species will have empty feeding ranges decreases, decreasing $B$ (e.g., the green species). The fraction of TL2 herbivores also decreases with $S$ because herbivores' feeding ranges must be perfectly placed to include a rare basal species but exclude any consumers. For example, the blue species feeds only on the green basal species but becomes a carnivore when the green species is switched to a consumer at high $S$. The fraction of top species (i.e., without predators, $T$) also decreases with $S$ because the broad feeding ranges and tendency to cannibalism of high niche-value species increasingly cover the niche axis with predators and excludes those species from being top (e.g., the gray species). **(E)** Mean absolute normalized model errors (NMEs) across all traditional niche model properties. Formatting follows Fig. 3 but absolute rather than signed errors are shown, and points are on a log-log scale.

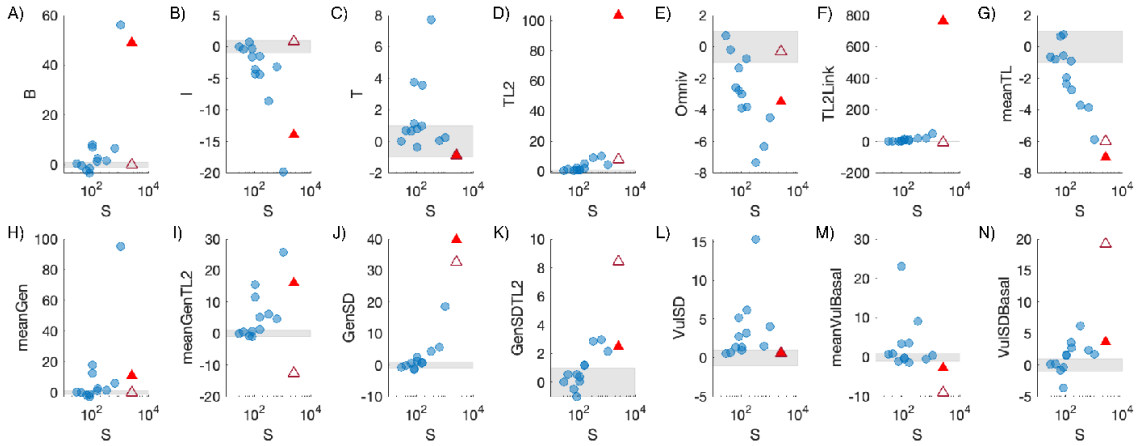



**Figure 3. Structural properties of the Michigan Temperate Forest web deviate from previous aboveground terrestrial food webs and null model expectations.** Points are normalized model errors (NMEs) calculated for empirical food webs compared to null expectations from an ensemble of $N = 1{,}000$ niche model food webs simulated with matched trophic species richness ($S$) and directed connectance ($C$). NMEs > 0 or < 0 indicate that observed properties are greater or less than expected, respectively. |NMEs| > 1 (outside of the gray box) are significantly different from null expectations at the 95% confidence level. Blue dots represent previously published food webs (Table 1). Red triangles are the Michigan temperate forest food web compared to the niche model with matched $S$, $C$ (filled) or matched $S$, $C$, and $B$, the fraction of species with no resources (i.e., basal species, achieved by setting species' feeding ranges to zero, hollow). Plots show that NMEs broadly scale with $S$, but the magnitudes of deviation associated with our web are greater than previously observed. See Methods for full definitions of food web properties. Fraction of **(A)** basal, **(B)** intermediate, **(C)** top, **(D)** strict herbivore (i.e., trophic level = 2), and **(E)** omnivorous species. **(F)** Fraction of feeding links by TL2 herbivores. **(G)** Average trophic level. Average generality of **(H)** consumers and **(I)** TL2 herbivores and normalized standard deviation of generality of **(J)** consumer species and **(K)** TL2 herbivores. Normalized standard deviation of vulnerability of **(L)** resource species and **(N)** basal species. **(M)** Average vulnerability of basal species.

# A highly resolved network reveals the role of terrestrial herbivory in structuring aboveground food webs

Kayla R. S. Hale, John David Curlis, Giorgia G. Auteri, Sasha Bishop, Rowan L. K. French, Lance E. Jones, Kirby L. Mills, Brian G. Scholtens, Meagan Simons, Cody Thompson, Jordon Tourville, Fernanda S. Valdovinos

**Electronic Supplementary Information**

**Supplementary Methods**

**i) Site description**

The University of Michigan Biological Station (UMBS) includes ~10,000 acres of land that is used for teaching and research in northern lower Michigan, USA (45°35.5′ N, 84°43′ W). The property is composed predominantly of dry-mesic, northern hardwood forests with patches of wooded wetlands (hardwood conifer swamp) flanked by two lakes [42]. This is a strongly seasonal, temperate system with historically cold, snowy winters (average minimum temperature of –12.1ºC in January) and hot, humid summers (average maximum temperature of 26.2ºC in July; [77]). Soil composition varies across microhabitat patches from sandy outwash plain to moraine. Prior to acquisition by the University of Michigan in 1909, the landscape was almost completely logged and burned, with only a few old-growth forest tracts remaining. Therefore, the vegetative landscape is relatively uniform in age, with variation in stand structure and composition attributable to glacial landforms and differences in soil composition [78,79].

UMBS is contiguous to other forested habitats, allowing free movement in and out for mobile species, including those that seasonally migrate. Such transient organisms, coupled with those that are only active or present aboveground during certain times of year, lead to strong effects of seasonality on species composition. Though their impacts may only be temporary, we

included these species in our analyses because they may represent critical consumers or resources for local species at certain times of year. In particular, most species of ectotherms at UMBS are inactive during winter months, and many migratory bird species are present only during the spring and summer. By pooling observations into a cumulative web, we avoid sensitivity of food web structure to this spatial and temporal transience.

**ii) Species list**

To construct the food web, we began from taxonomic lists of mammals, amphibians, reptiles, vascular plants, birds, insects, and non-insect arthropods. These lists represent an accumulation of decades of records at UMBS, from resident biologists' personal observations, student papers and research projects, regional lists, museum specimens, online databases (such as eBird and iNaturalist), and semi-regular BioBlitz events, in which teams of biologists roamed the site and identified as many organisms as possible.

Where possible, we updated records to the most recently accepted species names according to the Integrated Taxonomic Information System (ITIS) using the "taxize" package in R [80]. We excluded taxa that could not be resolved to at least genus-level (e.g., parasitoid wasp family Diapriidae). For taxa that could only be resolved to genus, we excluded those that are highly speciose in the Nearctic (> 20 species, e.g., the bloodworm genus *Chironomus*) and those that had congenerics already included in the food web (e.g., six species of masked bee *Hylaeus* were included, but another unidentified species was excluded). Finally, we treated variants and subspecies (e.g., deer mouse *Peromyscus maniculatus gracilis*) as their binomial species name. Hereafter, we refer to all taxa occurring at UMBS as "species," though a small fraction (4.5%) are genera.

921 The species lists were vetted and approved by experts (generally, the authors) that have
922 extensive knowledge of the communities and natural histories of the organisms that occur in the
923 region. In determining which species to include or exclude, we excluded any species that do not
924 occur at UMBS or do not have a significant lifestage or feeding behavior in aboveground
925 terrestrial habitats [76]. We defined "aboveground terrestrial habitats" as land above the soil
926 layer (including leaf litter and above), on terrestrial plants (growing potentially over but not
927 exclusively in water), or in air (above other terrestrial habitats). Thus, we excluded species that
928 live and feed primarily on the surface of aquatic habitats, such as waterlilies (Nymphaceae) and
929 water bugs (Belostomatidae). Additionally, given our primary focus on even resolution among
930 aboveground plants, insects, and vertebrates, we chose to exclude some major groups, such as
931 bryophytes, fungi, lichens, and molluscs.

932 **iii) Feeding interactions**

933 To assemble a list of potential feeding interactions between species, we referenced
934 region-specific field guides, online databases (e.g., Animal Diversity Web, Birds of the World),
935 and central repositories that automatically scrape data from museum records and the web (e.g.,
936 Global Biotic Interactions). Trained undergraduate researchers searched sources by species name
937 and recorded predators, diet, and other potential feeding interactions with as much resolution as
938 possible. Each focal taxon was resolved to species-level, but their interaction partners could be
939 recorded at any taxonomic level (e.g., species *x* eats Family *y*). If we could not find information
940 on a given focal species, we did not extrapolate from closely related species. We included all
941 records of direct interactions among species in our system with a bioenergetic flow (i.e., one
942 species consuming another), regardless of lifestage or potential ecological effects (i.e., whether

943 potentially “mutualistic” or “antagonistic”, see the recent framework for studying aboveground
944 terrestrial food webs, [14]).

945 Our records were supplemented, revised, and annotated by the same experts who vetted
946 the species lists. Where possible, records were assigned to experts for both the focal species and
947 its partner’s group (that is, two separate experts). We provide the final list of annotated feeding
948 records and their experts online [43].

949 Experts broadly categorized records by their focal resource type as animal tissues, both
950 (1) live tissues and as prey, and (2) scavenged as carrion, carcasses, or other decaying animal
951 remains, and plant tissues, grouped as (3) leaves and stems, (4) flowers, nectar, pollen, etc., (5)
952 seeds, fruits, etc., and (6) wood and bark. Hereafter, we refer to these resource types simply as
953 “prey,” “carrion,” “leaves,” “flowers,” “seeds,” and “wood.” We excluded records of detritus
954 resources that could not be resolved taxonomically.

955 Experts were asked to assess the plausibility of the recorded interactions occurring at
956 UMBS. Interactions between species *x* and *y* were considered plausible if the two species co-
957 occur (with respect to phenology, activity patterns, microhabitat usage) and have no trait
958 incompatibilities (it would not be possible for *x* to acquire, ingest, or assimilate *y*). Interactions
959 were considered plausible even if they could be considered inefficient (*x* can consume *y*, but this
960 is rare or unlikely, especially if more rewarding and easily acquired foods than *y* are available).
961 Interactions that are non-consumptive (e.g., nesting or hunting locations) or occur outside of
962 aboveground terrestrial habitats were excluded even if they occur between local species (e.g.,
963 least sandpiper *Calidris minutilla* preys upon toad *Anaxyrus americanus* tadpoles, but this occurs
964 exclusively in aquatic habitats).

965 Experts were also asked to assess the plausibility of the taxonomic level at which
966 interactions were recorded. For example, when Wilson's warbler (*Cardellina pusilla*) is recorded
967 to eat beetles (order Coleoptera), it is unclear if: (case 1) *C. pusilla* eats every species of
968 Coleoptera, potentially or opportunistically, including all local Coleopterans, or (case 2) *C.*
969 *pusilla* eats (at least) one species of Coleoptera that was not identified in the original record and
970 which may or may not be local. Either case may be possible, depending on the biology of the
971 species. At one extreme (case 1), the lack of taxonomic resolution in the record could reflect an
972 ecologically relevant lack of discrimination by the focal species (e.g., opportunism due to
973 sensory capabilities in distinguishing between predators or resources), with the probability of
974 interactions occurring between any given partner species determined broadly by their
975 abundances. At the other extreme (case 2), the lack of taxonomic resolution could simply reflect
976 a lack of knowledge, representing a summary of potentially highly specific interactions. In the
977 first case, experts accepted the interaction, while in the second, experts rejected the interaction
978 unless it could be plausibly revised for local taxa.
979 In the original sources, interactions were often recorded following traditional disciplinary
980 categorizations. In the absence of other information, we assumed that animals recorded as
981 "eating," "feeding on," "consuming," "parasitizing," or "hosted by" plants were feeding on
982 leaves. We assumed that feeding on wood or bark was specifically recorded as such. We
983 assumed that animals recorded as "pollinator of," "visiting flowers of," "visiting," or "mutualist
984 of" plants were feeding on flowers. Other mutualisms involving potential feeding such as seed
985 dispersal (through frugivory, scatterhoarding, etc.) and ant protection (of hemipterans or plants)
986 were noticeably underrepresented in our records. We assumed animals recorded as "disperser of"
987 plants were feeding on seeds. We assumed that interactions between animals recorded as "kills,"

988 “predates upon,” “host of,” “parasite of,” or “parasitoid of” involved consumption of animal
989 tissue, and that “scavenges” reflected feeding on carrion. When necessary, these default
990 assumptions were revised by experts to reflect true feeding on another resource type or a lack of
991 feeding altogether.

992 **iv) Sampling biases**

993 Vertebrates and vascular plants are likely well-sampled in our system. Butterflies and
994 moths are reasonably well-represented as they have been the focus of repeated target sampling.
995 For some groups of insects (especially beetles, wasps, and true bugs), regional datasets suggest
996 that as many as a few thousand more species occur but currently remain undetected at UMBS.
997 Broadly, the species most likely missing from our list are those that are specialists, cryptic,
998 and/or rare. Regional datasets also suggest that we are missing a smaller number of predatory
999 non-insect arthropods (including up to 200 spp. of spiders, harvestmen, centipedes, millipedes,
1000 and terrestrial isopods). Among the major insect groups, true flies are likely the most
1001 underrepresented. This is not necessarily surprising as flies are extremely ecologically diverse
1002 and require specialized collecting – groups such as the fungus-associated flies (a notable
1003 exclusion in our list) would not necessarily be detected in malaise traps. We are also aware we
1004 are missing potentially many species of groups such as acariform mites, which were both
1005 underrepresented in our system and for which we did not have a regional reference list. Finally,
1006 for the purposes of limiting the scope of our work, we excluded microbes, fungi, detritus, non-
1007 vascular plants (including up to 900 spp. of mosses, liverworts, and lichens), and many non-
1008 insect invertebrate groups (including up to 100 spp. of gastropods) from our original species list
1009 and focused only on the species interactions that occur primarily in aboveground terrestrial
1010 habitats.

### v) Food web structure

To simulate niche model food webs with matching $S$ and $C$ to the empirical food webs with low connectance, we seeded the algorithm with $S_{seed} = S + 4$ for the High Arctic, UK Grassland, Scotch Broom, Shortgrass Prairie, and Messel Forest webs, and $S_{seed} = S + 10$, $C_{seed} = C \times 1.01$ for the Norwood Farm web. At each replication, we aggregated the simulated web to trophic species and discarded it if it did not match the target $S$, $C$. For the purposes of discussion, we followed previous works to summarize deviations from null expectations as a composite *mean absolute Normalized Model Error* (|NME|) of traditional food web properties. We calculated *mean* |NME| as the average of the absolute value for the NME of each of the following properties: *Top (T)*, *Intermediate (I), Basal (B), Cannib, Herbiv (TL2), TrophicOmniv, LinkSD, GenSD, VulSD, PathLen, meanTL, MaxSim.* See full definitions in Supplementary Table 1, following [8,26,39,44–47]. Notice the nontraditional notation here and in the main text to accommodate new and additional focal properties.

## Supplementary Data & Code

`taxonomic_web_analysis_EDI.m` is a MATLAB script that builds and analyzes the taxonomic version of the Michigan Temperate Forest (MTF) network, using multiplex network and food web representations for different focal levels of taxonomic resolution.

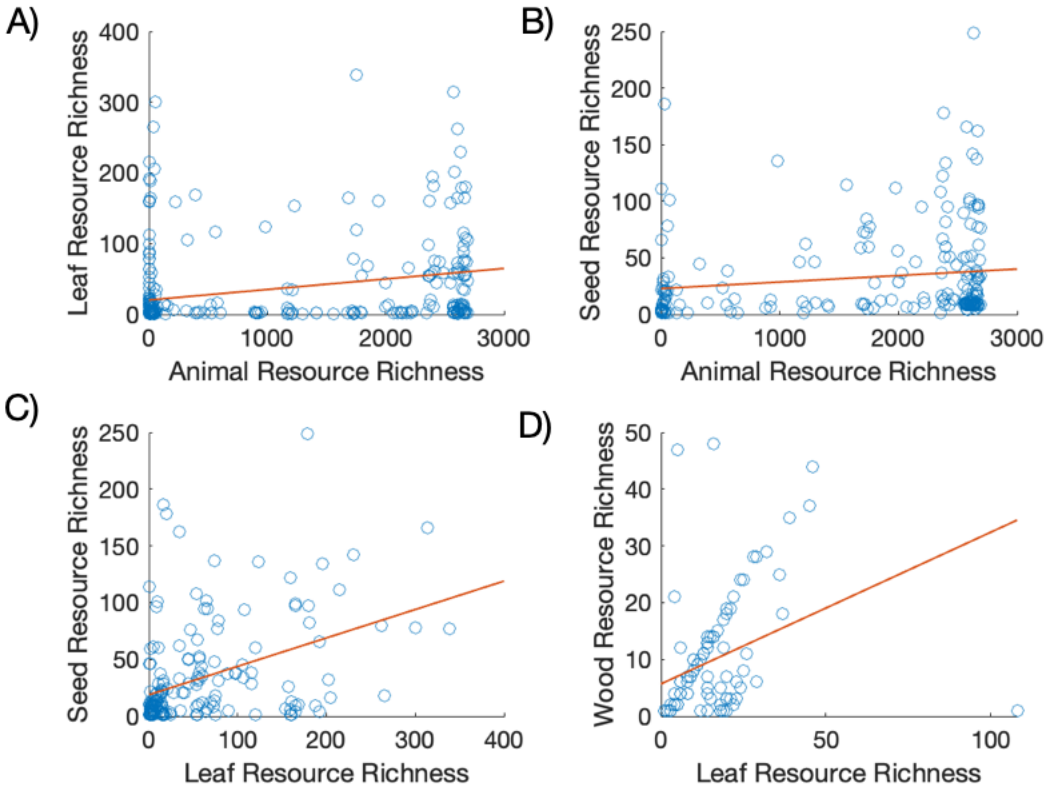
`simulated_matching_niche_EDI.m` is a MATLAB script that runs the leave-one-out analysis on the MTF network, creates trophic-species food web representations of historical empirical aboveground terrestrial food webs (including the MTF), simulates matching niche models, analyzes their properties, and reproduces the main text figures.

1034 **Supplementary Table 1 Full definitions of food web properties**

| **Food Web Property** | **Definition** |
|---|---|
| *S* | 'Trophic species' richness, i.e., number of nodes in the food web after aggregating taxa with the same set of consumers and resources |
| *L* | Number of links in the trophic-species food web |
| *C* | Directed connectance, $C = L/S^2$ |
| *LD* | Link density, $L/S$ |
| **Species Composition** | |
| *Basal (B)* | Fraction of trophic species with no resources |
| *Intermediate (I)* | Fraction of trophic species with consumers and resources |
| *Top (T)* | Fraction of trophic species with no consumers |
| *Herbiv (TL2)* | Fraction of trophic species that eat only basal species, i.e., are strict herbivores with trophic level 2 |
| *Omniv* | Fraction of trophic species that eat both basal species and consumers, i.e., are omnivores |
| *Carniv* | Fraction of trophic species that eat other consumers, i.e., are carnivores |
| *Cannib* | Fraction of trophic species that eat themselves, i.e. are cannibals |
| *TrophicOmniv* | Fraction of trophic species that have a non-integer trophic level, i.e., that eat across trophic levels |
| **Link Composition** | |
| *HerbLink* | Fraction of *L* that are feeding on a basal resource, i.e., are herbivorous |
| *TL2Link* | Fraction of *L* by TL2 herbivores onto basal resources |
| *MaxSim* | Mean of the maximum trophic similarity between each pair of trophic species, as quantified by the Jaccard overlap |
| **Hierarchy** | |
| *TL* | Trophic level, calculated as short-weighted trophic level (*swTL*), the mean of trophic level calculated using species' shortest path to a basal resource (*short TL*) and trophic level calculated using the average of species' prey's trophic level (*prey-averaged TL*) [46] |
| *meanTL* | Mean TL of all trophic species |
| *maxTL* | Max TL of all trophic species |
| *meanTLTop* | Mean TL of top trophic species |
| *PPRatio* | Predator-prey ratio, calculated as $(T + I)/(B + I)$ |
| *logNumPaths* | Log10 of the total number of shortest directed path lengths between each pair of trophic species in the food web for which a path exists |
| *PathLen* | Mean length of the path lengths defined above |
| *PathSD* | Standard deviation of the path lengths defined above |
| *avgOmni* | Mean of the variance in *prey-averaged TL* for each consumer's set of resources |

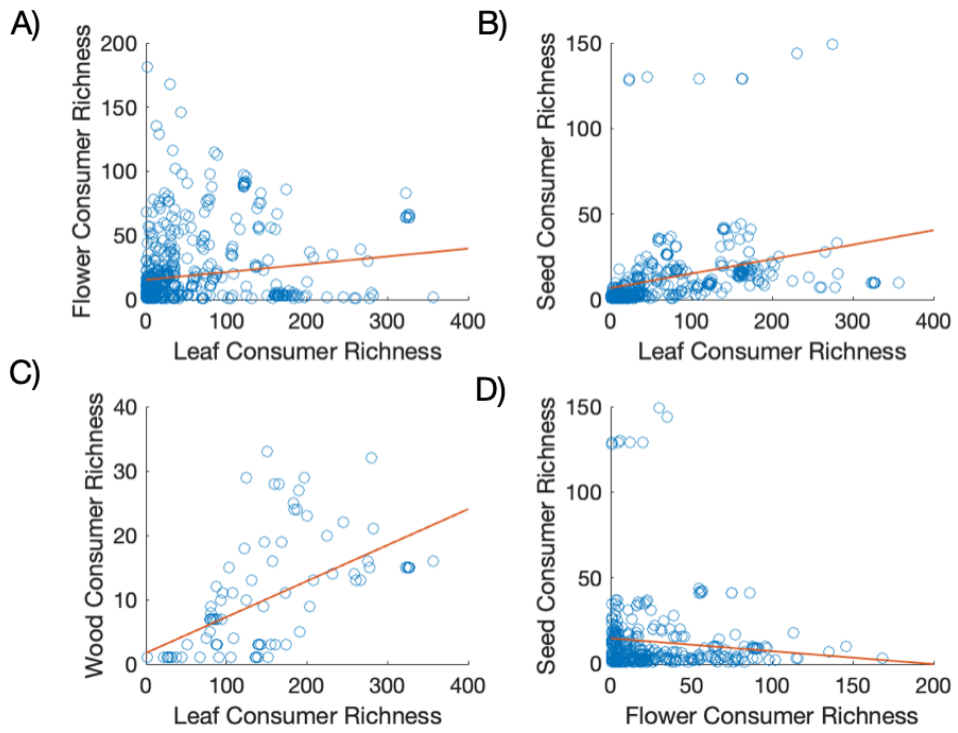
| **Degree Distribution** | |
|---|---|
| *In-degree* | Number of resources, i.e., generality |
| *Out-degree* | Number of consumers, i.e., vulnerability |
| *meanGen* | Mean in-degree for consumers, (i.e., trophic species with in-degree > 0) |
| *meanGenTL2* | Mean in-degree for TL2 herbivores |
| *meanGenOmniv* | Mean in-degree for omnivores |
| *GenSD* | Standard deviation of in-degree for all trophic species, normalized by *LD* |
| *GenSDTL2* | Standard deviation of in-degree for TL2 herbivores, normalized by *meanGenTL2* |
| *GenSDOmniv* | Standard deviation of in-degree for omnivores, normalized by *meanGenOmniv* |
| *meanVul* | Mean out-degree for resources, (i.e., trophic species with out-degree > 0) |
| *meanVulBasal* | Mean out-degree for basal trophic species |
| *VulSD* | Standard deviation of out-degree for all trophic species, normalized by *LD* |
| *VulSDBasal* | Standard deviation of out-degree for basal trophic species, normalized by *meanVulBasal* |
| *LinkSD* | Standard deviation of total degree (in-degree + out-degree) for all trophic species, normalized by *LD* |
| *Kolmogorov-Smirnov test* | Tests the hypothesis that data in two samples (here, two degree distributions) come from populations with the same underlying distribution. We test in-degree and out-degree distributions separately. If this hypothesis is rejected at the 5% confidence level for the two-sided comparison of an empirical web and >950 of its $N$ = 1,000 simulated niche models, then we say their degree distributions "significantly differ" [47] |
| ***mean* \|*NMEs*\|** | Mean of the absolute values of 12 traditional food web properties: *Top (T)*, *Intermediate (I), Basal (B), Cannib, Herbiv (TL2), TrophicOmniv, LinkSD, GenSD, VulSD, PathLen, meanTL, MaxSim*. |

**Supplementary Table 2 Normalized model errors and Kolmogorov-Smirnov test results for empirical food webs**




**Supplementary Figure 1 Correlations between consumer diet breadths feeding on different resources**

Scatter plots showing the relationship between consumers' diet breadths in terms of number of resource species when feeding on **(A)** animals and plant leaves ($r = 0.26$, $p = 2.7 \times 10^{-6}$, $N = 319$), **(B)** animals and plant seeds ($r = 0.14$, $p = 0.044$, $N = 202$), **(C)** plant leaves and seeds ($r = 0.42$, $p = 7.2 \times 10^{-9}$, $N = 178$), and **(D)** plant leaves and wood ($r = 0.24$, $p = 0.017$, $N = 98$). Each point is a species that feeds on both focal resources. Lines are least squares fits. Only relationships with significant Pearson's correlations (r) are shown.

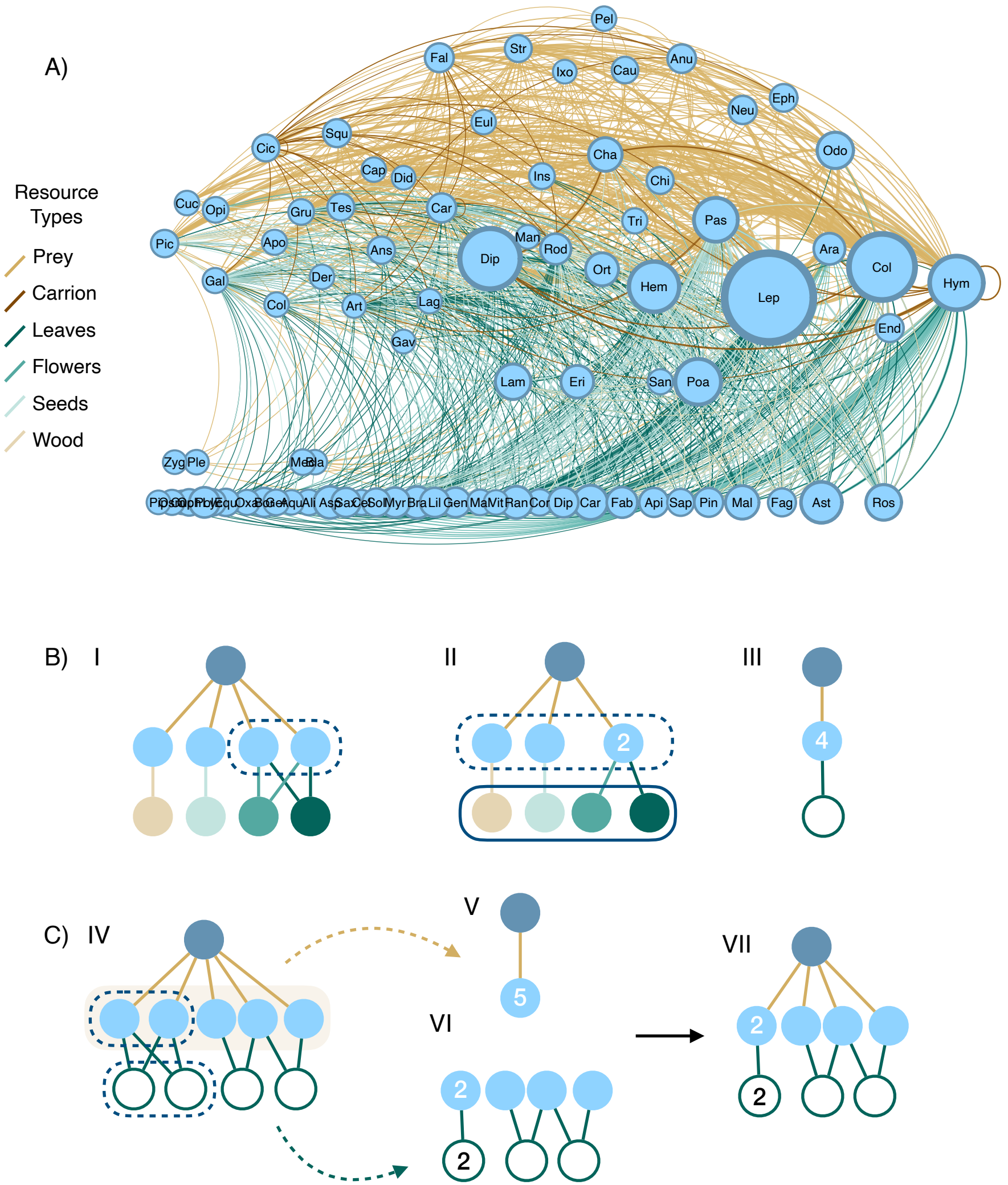


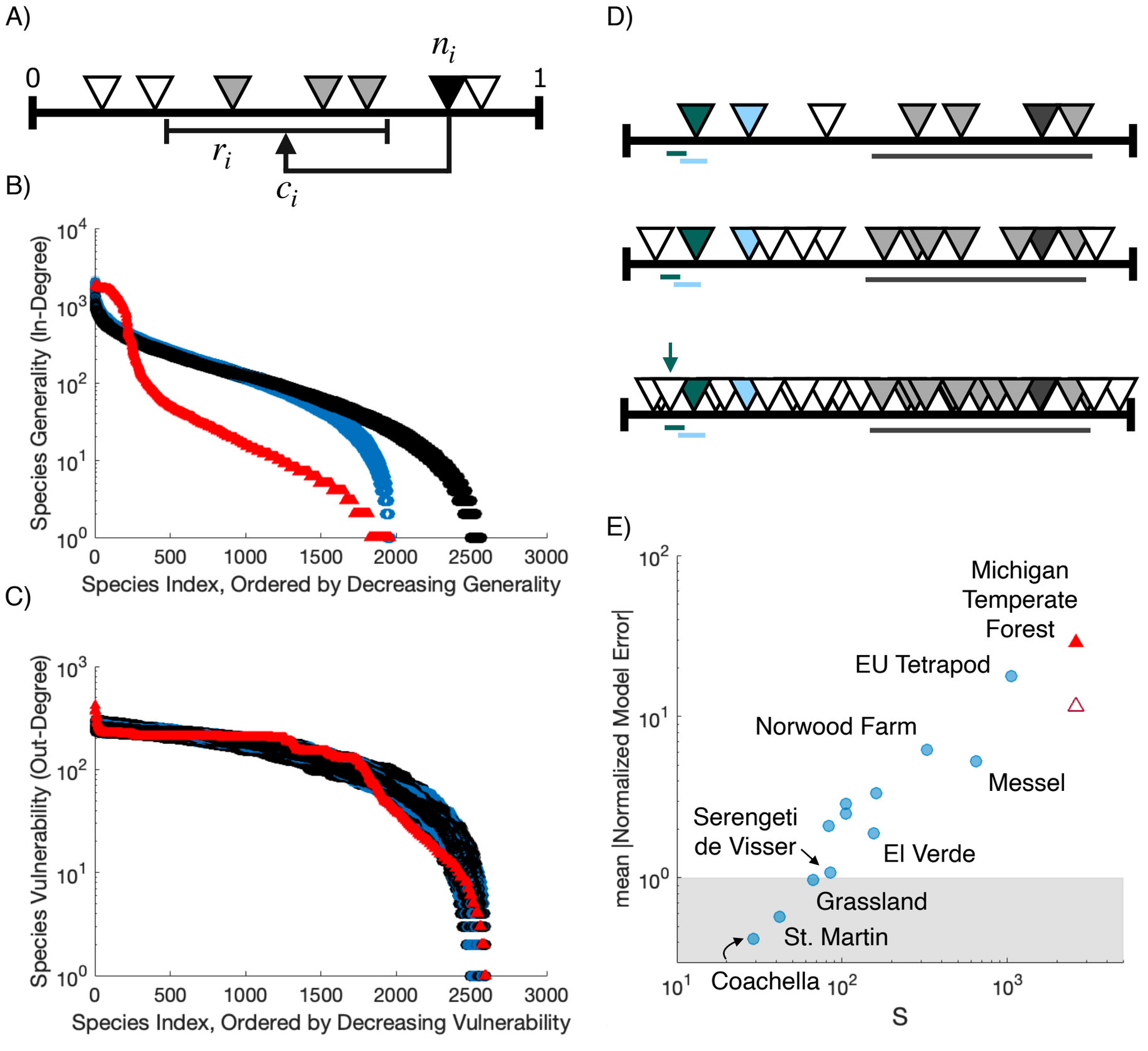

**Supplementary Figure 2 Correlations between animal richnesses hosted by plants on different tissues**

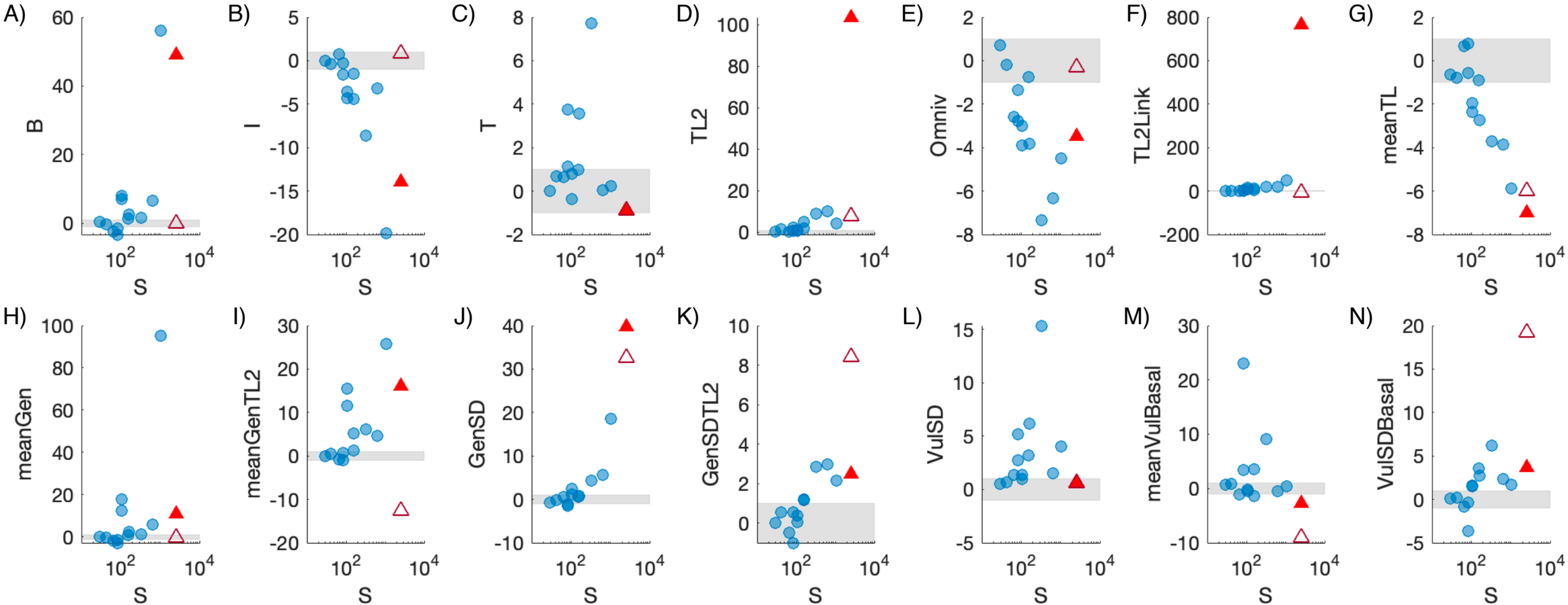
Scatter plots showing the relationship between the richness in terms of number of consumers species when feeding on plant species' **(A)** leaves and flowers ($r = 0.17$, $p = 1.8 \times 10^{-5}$, $N = 649$), **(B)** leaves and seeds ($r = 0.35$, $p = 8.1 \times 10^{-17}$, $N = 548$), **(C)** leaves and wood ($r = 0.61$, $p = 1.7 \times 10^{-11}$, $N = 98$), and **(D)** flowers and seeds ($r = -0.11$, $p = 0.013$, $N = 510$). Each point is a plant species with consumers that feed on both focal resources. Lines are least squares fits. Only relationships with significant Pearson's correlations (r) are shown. This shows that the plants most consumed by leaf-herbivores are also most likely to host rich communities of consumers eating their other tissues. But there may be a trade-off in plants' ability to support flower and seed eaters (potential pollinator and seed dispersal mutualists, respectively) or a deterrence effect between seed- and flower-eaters in our system (primarily birds and insects, respectively).

## Lay Summary

Terrestrial ecosystems are immensely complex, including diverse species feeding in diverse ways. Plants and insects engage in highly-specific interactions that have often been excluded from terrestrial food webs due to limited sampling and expertise. Here, we use records from a biological research station accumulated over ~100 years to construct a food web for a Michigan temperate forest. We report ~580,000 interactions among ~3,800 species in a living dataset that is openly available to be supplemented and revised. This is the largest and most detailed food web yet reported and provides a valuable tool for ecosystem management and research.


A)
Resource Types
Prey
Carrion
Leaves
Flowers
Seeds
Wood
Pel
Fal
Str
Anu
Ixo
Cau
Eph
Neu
Eul
Squ
Cic
Cha
Odo
Cap
Did
Ins
Chi
Cuc
Opi
Gru
Tes
Car
Tri
Pas
Pic
Apo
Man
Rod
Ans
Dip
Ara
Col
Hym
Gal
Der
Ort
Hem
Lep
Col
Art
Lag
End
Gav
Lam
Eri
San
Poa
Zyg
Ple
Car
Fab
Api
Sap
Pin
Mal
Fag
Ast
Ros
B)
I
II
2
III
4
C)
IV
V
5
VI
2
2
VII
2
2


A)
0
1
$n_i$
$r_i$
$c_i$
B)
Species Generality (In-Degree)
Species Index, Ordered by Decreasing Generality
C)
Species Vulnerability (Out-Degree)
Species Index, Ordered by Decreasing Vulnerability
D)
E)
mean |Normalized Model Error|
S
Michigan Temperate Forest
EU Tetrapod
Norwood Farm
Messel
Serengeti
de Visser
El Verde
Grassland
St. Martin
Coachella


A)
B
S
B)
I
S
C)
T
S
D)
TL2
S
E)
Omniv
S
F)
TL2Link
S
G)
meanTL
S
H)
meanGen
S
I)
meanGenTL2
S
J)
GenSD
S
K)
GenSDTL2
S
L)
VulSD
S
M)
meanVulBasal
S
N)
VulSDBasal
S