

1      **Supplementary Methods for Diverse interactions and ecosystem engineering can  
2      stabilize community assembly**

3      Justin D. Yeakel,<sup>1,2</sup> Mathias M. Pires,<sup>3</sup> Marcus A. M. de Aguiar,<sup>3</sup> James L.  
4      O'Donnell,<sup>4</sup> Paulo R. Guimarães Jr.,<sup>5</sup> Dominique Gravel,<sup>6</sup> and Thilo Gross<sup>7,8,9,10</sup>

5      <sup>1</sup>*University of California, Merced, Merced, CA 95340, USA*

6      <sup>2</sup>*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

7      <sup>3</sup>*Universidade Estadual de Campinas, Campinas - SP, Brazil*

8      <sup>4</sup>*University of Washington, Seattle, WA 98195, USA*

9      <sup>5</sup>*Universidade de São Paulo, São Paulo, Brazil*

10     <sup>6</sup>*Université de Sherbrooke, Sherbrooke, QCJ1K0A5, Canada*

11     <sup>7</sup>*University of California, Davis, Davis, CA 95616, USA*

12     <sup>8</sup>*Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung*

13     <sup>9</sup>*Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB),*

14     *Ammerländer Heerstrasse 231, 26129 Oldenburg, Germany*

15     <sup>10</sup>*University of Oldenburg, ICBM, 26129 Oldenburg, Germany*

16      **SUPPLEMENTARY METHODS**

17      **Appendix 1: Building the source pool**

18      Here and henceforth, we refer to the assembly model  
19      presented in the main text as the ENIgMa model (E:eat,  
20      N:need, Ig:ignore, Ma:make). To initiate the ENIgMa  
21      assembly model, we must first construct the source pool,  
22      where each ecological entity (species + modifiers) is de-  
23      fined by their potential interactions. The model is initial-  
24      ized by creating  $S$  species and  $M = \eta S$  modifiers, such  
25      that  $N = S + M$  is the expected total number of enti-  
26      ties (prior to considering engineering redundancies) and  
27       $\eta$  is the expected number of modifiers made per species  
28      in the community, where the expectation is taken across  
29      replicates. For each pair of species (x,y) there is a prob-  
30      ability  $p_e$  that x eats y and probability  $p_n$  that x needs  
31      y. For each pair of species x and modifier m, there is a  
32      probability  $q_e$  that species x eats modifier m and a prob-  
33      ability  $q_n$  that species x needs modifier m. For simplicity  
34      we assume throughout that  $p_e = q_e$  and that  $p_n = q_n$ ,  
35      such that the probability of drawing trophic and service  
36      interactions for both species-species and species-modifier  
37      interactions is the same.

38      Without engineering redundancies (i.e. each modifier  
39      that a species makes is unique), the expected number of  
40      modifiers is  $M = \eta S$  where  $\eta$  is the mean number of  
41      modifiers made per species  $S$ . If we allow for engineering  
42      redundancies, the realized number of modifiers  $M' < M$ .  
43      To determine the number of modifiers in the pool, for  
44      each species a set number of modifiers is drawn, where  
45       $M_i \sim \text{Poiss}(\eta)$ . The expected proportion of species that  
46      are engineers (species that make modifiers) is thus  $1 -$   
47       $e^{-\eta}$ , where  $e$  is Euler's number. If a particular modifier is  
48      randomly and independently drawn for a given engineer  
49      from a complete list of all possible modifiers, such that  
50      multiple species – with some probability – can make the

51      same modifier, the expected number of modifiers becomes

$$M' = \eta S \left(1 - \frac{1}{e}\right). \quad (\text{S1})$$

52      The frequencies of eat and need interactions,  $p_e$  and  
53       $p_n$  respectively, are assigned a priori (see Supplementary  
54      Appendix 2 for different model parameterizations). The  
55      frequency of engineering (make) interactions can be cal-  
56      culated as

$$p_m = \frac{\eta S}{(S + M')^2} = \frac{\eta}{S(1 + \eta - \frac{1}{e})^2}. \quad (\text{S2})$$

57      The frequency of the null interaction is then calculated  
58      by  $p_\emptyset = 1 - (p_e + p_n)$  for species-species interactions  
59      and  $p_\emptyset = 1 - (p_e + p_n + p_m)$  species-modifier interac-  
60      tions, respectively. Pairwise interactions are established  
61      randomly, such that the source pool matrix has no im-  
62      bued structure apart from the number of species, the  
63      number of modifiers, and the frequency of each direc-  
64      tional interaction type. Each source pool is provided a  
65      'basal resource' (the first row/column). A species with  
66      a trophic interaction to this resource is identified as an  
67      autotroph (or mixotroph depending on its other trophic  
68      interactions). If they do not have service dependencies  
69      with other species/modifiers, it is these species that are  
70      uniquely able to initiate assembly.

71      When engineering redundancies are allowed, the ex-  
72      pected number of unique versus redundant modifiers in  
73      the source pool can be determined analytically. The to-  
74      tal number of modifiers is  $M' = \eta S(1 - e^{-1})$ , and can  
75      be subdivided into modifiers that have a unique engineer  
76      and those that have multiple engineers. The number of  
77      modifiers with a single engineer is  $M'_{\text{unique}} = \eta S e^{-1}$ . The  
78      number of modifiers made by multiple engineers is calcu-  
79      lated as  $M' - M'_{\text{unique}}$ , such that

$$M'_{\text{redundant}} = M' - M'_{\text{unique}} = \eta S \frac{e - 2}{e}, \quad (\text{S3})$$

such that the proportion of redundant modifiers  $\phi$  is

$$\phi = \frac{M' - M'_{\text{unique}}}{M'} = \frac{e - 2}{e - 1} \approx 0.418. \quad (\text{S4})$$

Accordingly, we find that the number of redundant modifiers increases linearly with  $\eta$ , while the proportion of modifiers that are redundant is fixed. Figure S8a,b shows both analytical expectations and numerically-derived measures for  $M'_{\text{redundant}}$  and  $\phi$ , respectively.

As described in Methods, the assembly process can be simulated efficiently with an event-driven simulation utilizing a Gillespie algorithm. Generally, one computes the rates  $r_j$  of all possible events  $j$  in a given step. One then selects the time at which the next event happens by drawing a random number from an exponential distribution with mean  $1/\sum_j r_j$ . At this time, an event occurs that is randomly selected from the set of possible events such that the probability of event  $a$  is  $r_a/\sum_j r_j$ . The effect of the event is then realized and the list of possible events is updated for the next step.

In our framework, at the beginning of each simulation step we compute: 1) all species in the pool and absent from the community that have trophic and service dependencies met by those species in the community: these species are subject to colonization; 2) all species in the community that do not have a competition strength that is highest for at least one of their resources: these species are subject to primary extinction; 3) all species in the community that do not meet their eat and/or need dependencies: these species are subject to secondary extinction; 4) all modifiers in the community that no longer have an engineer: these modifiers are subject to elimination. We then select one of the four events with a probability proportional to the number of entities that satisfy the criteria for each event. The rates at which each event occurs change at each step, equal to the number of entities that meet the criteria for each event at that point in time. The species/modifier that colonizes or is eliminated from the community is randomly chosen once the event-type is determined.

For example, if the community is empty, and 50 species are able to colonize, the probability of drawing ‘colonization’ is 1, and the colonizer would be randomly drawn from the 50 capable of colonizing. Another example: if 20 species are able to colonize, 10 species are not superior competitors for any one of their resources, and 30 species do not meet their dependencies,  $\sum_j r_j = 60$ , and  $r_{\text{colonize}} = 1/3$ ,  $r_{\text{primary extinction}} = 1/6$ , and  $r_{\text{secondary extinction}} = 1/2$ . In this case, the most probable event is a secondary extinction. After this single event takes place, the community is updated depending on which event occurred, and the simulation proceeds to the next step. This algorithm is known to offer a much better approximation to the true stochastic continuous time process than a simulation in discrete time steps, while providing a much higher numerical efficiency<sup>1</sup>.

## Appendix 2: Model parameterizations

Simulations described in the main text have default parameterizations of  $S = 200$ ,  $p_e = 0.01$ ,  $c_n = \pi$ ,  $c_e = \sqrt{2}$ ,  $c_v = 1$ , and 4000 iterations (time-steps). Replicates are defined as the independent assembly of independently drawn source pools with a given parameterization.

**Assembly without ecosystem engineering** Here we set the average number of modifiers made per species  $\eta = 0$  and the probability of need interactions in the species pool  $p_n = 0.002$ .

**Structure and dynamics of mutualisms** Again we used the default parameterizations but set  $\eta = 0$ , while varying  $p_n \in [0, 0.002]$ . We note that while the density of eat interactions is not changed as service interaction frequency increases, what would previously have been a trophic interaction is more likely to be a mutualism (trophic in one direction, service in the other) if an incoming eat interaction is paired with an outgoing need interaction. In other words, as service interaction frequency increases, eat interactions are not substituted with need interactions (maintaining the fixed trophic interaction frequency), however null interactions are substituted with need interactions.

**Assembly with ecosystem engineering** Here we used the default parameterizations but varied  $\eta \in [0, 2]$  and  $p_n \in [0, 0.002]$ .

## Appendix 3: Comparison to Niche Model

We compared certain structural features of ENIgMa at steady state to those of the Niche Model<sup>2</sup>. Comparisons were restricted to networks constructed in the absence of engineering because engineers introduce indirect effects that are not considered in static food web models, and may make such comparisons irrelevant. While there are many similarities, there are also some important differences, some of which are highlighted in the main text. While we consider a comparison of our framework with other food web models such as the Niche Model relevant, we emphasize that the motivations underlying both are distinct. Our approach is intended to provide a deeper understanding into how multitype dependencies between species and the environment impact the dynamics of community assembly. While capturing general qualitative features of empirical systems demonstrates that the dynamics we consider are ecologically relevant, the goal of our approach is distinct from that of static food web models, which aim to maximize structural similarities between model and empirical systems<sup>2,3</sup>.

We compared steady state ecological networks that emerge from ENIgMa (described in Methods, main text) with food webs constructed from the Niche Model<sup>2</sup> with similar species richness and connectance. Because species richness and connectance of the Niche Model are often altered by eliminating disconnected species, we compared i) species richness, ii) connectance, iii) mean species de-

187 gree, iv) standard deviation of out-degree distributions, 220 state species richness (Fig. S6a; species richness is shown  
 188 and v) standard deviation of in-degree distributions av- 220 in blue, modifier richness is shown in red).  
 189 eraged across 1000 replicates for each model.

190 We found that all measures resulted in fairly simi- 221 The role of specialists does and does not change with  
 191 lar values between ENIgMa and the Niche Model food 222 the introduction of engineering, depending on how spe-  
 192 webs with a some important differences (Figs. S2,S3). 223 cialization is defined. As in the main text, a specialist is  
 193 While similar, ENIgMa produces consistently lower val- 224 defined when its generality index  $G_i < 1$  relative to the  
 194 ues of connectance, mean species degree, as well as stan- 225 steady state link density. When engineered modifiers are  
 195 dard deviations of the in- and out-degree distributions. 226 included, we account for a trophic interaction between  
 196 This means that the food webs produced by ENIgMa 227 a species and another's modifier as an interaction that  
 197 are more sparsely connected with less variance between 228 occurs between those two species indirectly through the  
 198 species. These results were expected, as the Niche Model 229 modifier intermediary. So if a species  $B$  makes a modifier  
 199 assumes systematically increasing dietary ranges with 230  $M$ , and  $A$  eats  $M$ , then we set  $A$  to (indirectly) eat  $B$ .  
 200 higher niche values, whereas the trophic interactions as- 231 This accounting of both direct and indirect trophic inter-  
 201 ssigned to species in the source pool of ENIgMa are drawn 232 actions between species can then be compared to i) the  
 202 independently. An important difference between the 233 direct trophic link density of the community, or ii) the  
 203 Niche Model and ENIgMa is that we do not distinguish 234 direct + indirect trophic link density of the community,  
 204 between traditional consumers and parasites. A different 235 and some insights can be gained from both approaches.  
 205 framework known as the Inverse Niche Model<sup>4</sup> has been  
 206 proposed to address parasitic interactions. The Inverse  
 207 Niche Model assumes increasing specialization with feed-  
 208 ing hierarchies, which would serve to lower the average  
 209 generality of species (lower degree). In addition, the In-  
 210 verse Niche model outputs lower standard deviations of  
 211 in- and out-degree distributions. Together these trends  
 212 suggest that the qualitative structural differences that we  
 213 observe for the assembly and Niche model may reflect an  
 214 important structural distinction between food webs that  
 215 do and do not include parasitic species.

#### 216 Appendix 4: The structure of engineered food webs

217 We examined whether and to what extent the structure 220 state (mean proportion specialists at steady state is 0.04;  
 218 of food webs was altered when engineers are introduced 221 Fig. S6b). This means that the indirect links that define  
 219 into the community. Because trophic links can now exist 222 trophic interactions between species and modifiers in-  
 220 between species-modifiers as well as species-species, there 223 crease the link-density of the network relative to that de-  
 221 are different ways of accounting for structure, making di- 224 fined only by direct trophic interactions. In words, modi-  
 222 rect comparisons with non-engineered food webs some- 225 fiers serve to connect otherwise disconnected species, for-  
 223 what difficult. We note that we exclude service interac- 226 malizing the otherwise indirect relationships that struc-  
 224 tions in this case to best match the structural analysis 227 ture the role of engineers in the community. In the second  
 225 described in the main text and shown in Fig. ???. While 228 case, where  $G_i$  is determined relative to  $L_{\text{indirect}}^*/S^*$ , we  
 226 the inclusion of engineers ( $\eta = 2$ ) does have an impact on 229 find that the changes in both functional and potential  
 227 stability in terms of primary versus secondary extinction 230 specialists over the course of assembly (Fig. S6c) follow  
 228 rates, there is not a strong effect of engineering on steady 231 those observed for non-engineered food webs (Fig. ??b).

232 Finally, we observe that while the number of trophic  
 233 levels increase in the presence of species-modifier inter-  
 234 actions, the overall trophic structure of the community  
 235 advances over the course of assembly in much the same  
 236 way as it does without engineers (Fig. S6d). Trophic lev-  
 237 els are calculated with respect to indirect species inter-  
 238 actions through modifier intermediaries. Because species  
 239 at any trophic level can engineer modifiers used as re-  
 240 sources by other species, the mean trophic level of the  
 241 community is systematically elevated.

- 
- 273 [1] Gillespie, D. T. Exact stochastic simulation of coupled 279 of empirical food webs. *Ecology* **92**, 1849–1857 (2011).  
 274 chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977). 280 [4] Warren, C. P., Pascual, M., Lafferty, K. D. & Kuris, A. M.  
 275 [2] Williams, R. J. & Martinez, N. D. Simple rules yield 281 The inverse niche model for food webs with parasites.  
 276 complex food webs. *Nature* **404**, 180–183 (2000). 282 *Theor. Ecol.* **3**, 285–294 (2010).  
 277 [3] Williams, R. J. & Purves, D. W. The probabilistic niche 278 model reveals substantial variation in the niche structure

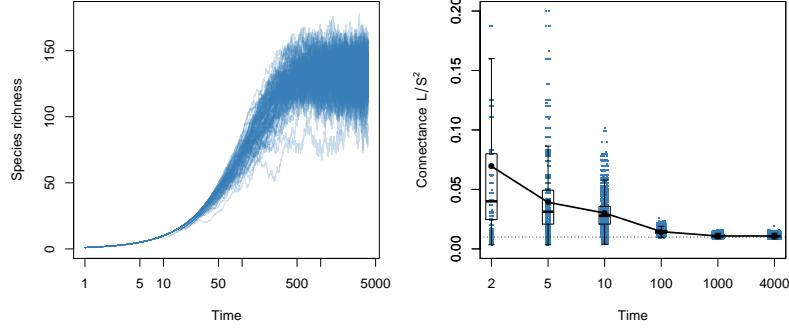


Figure S1. Left: Assembly of communities over time results in steady state species richness by ca. time-step 250. Right: Trophic connectance early in assembly is high because a small number of species interact with each other such that the proportion of realized interactions (out of all possible interactions) is closer to unity. Over time, connectance decays as species richness increases, and the density of trophic interactions declines. Box edges define the 25th and 75th percentiles of simulation results, and whiskers represent 1.5× the interquartile range.

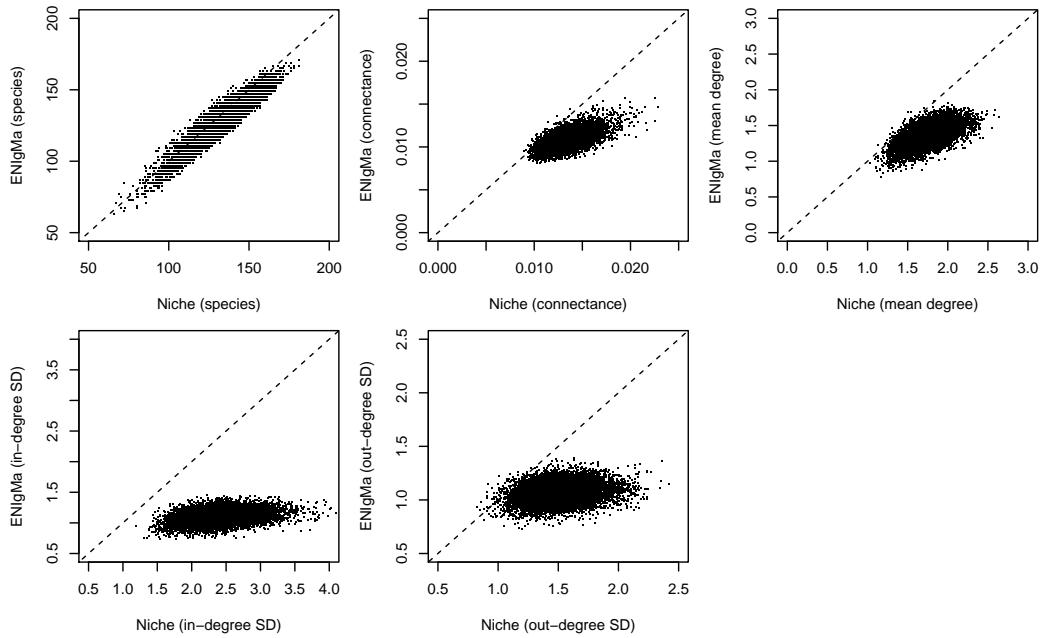


Figure S2. Comparisons of raw structural measures for the assembly (y-axis) and Niche model (x-axis). If the models produce similar structures, metrics will tend to fall on the 1:1 line (drawn). While the values for both models are similar, connectance, mean degree, and the standard deviation of in- and out-degree are all lower for the assembly model relative to those measures for the Niche model.

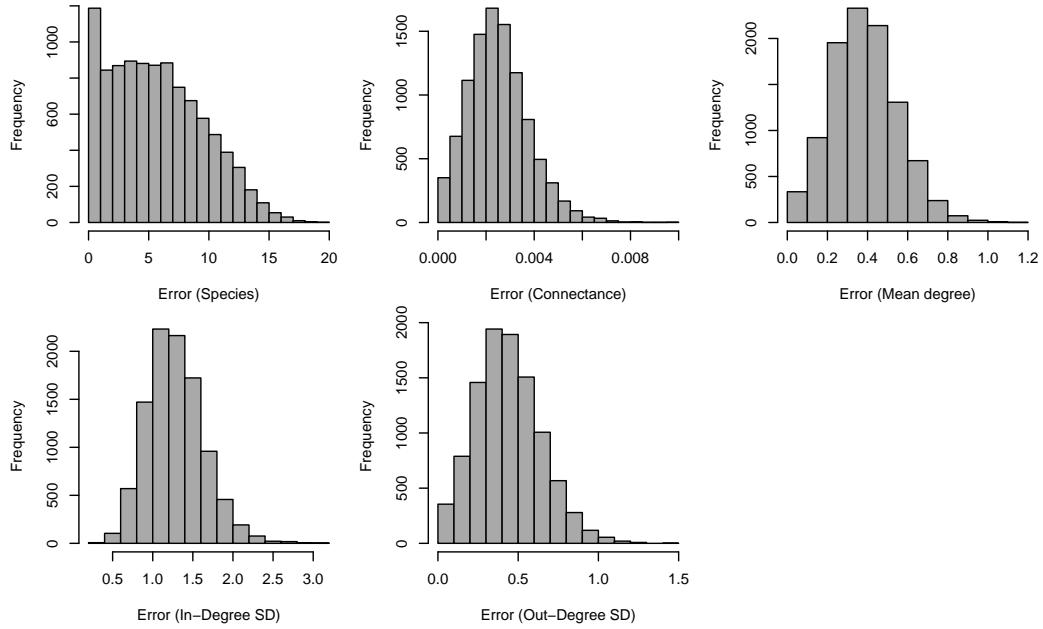


Figure S3. Error between structural measures of the assembly and Niche models. Error is measured as  $\sqrt{(m_i - m_j)^2}$ , where  $m_i$  and  $m_j$  are structural metrics for the assembly and Niche model, respectively. Only the trophic network of the assembly model was used to assess metrics.

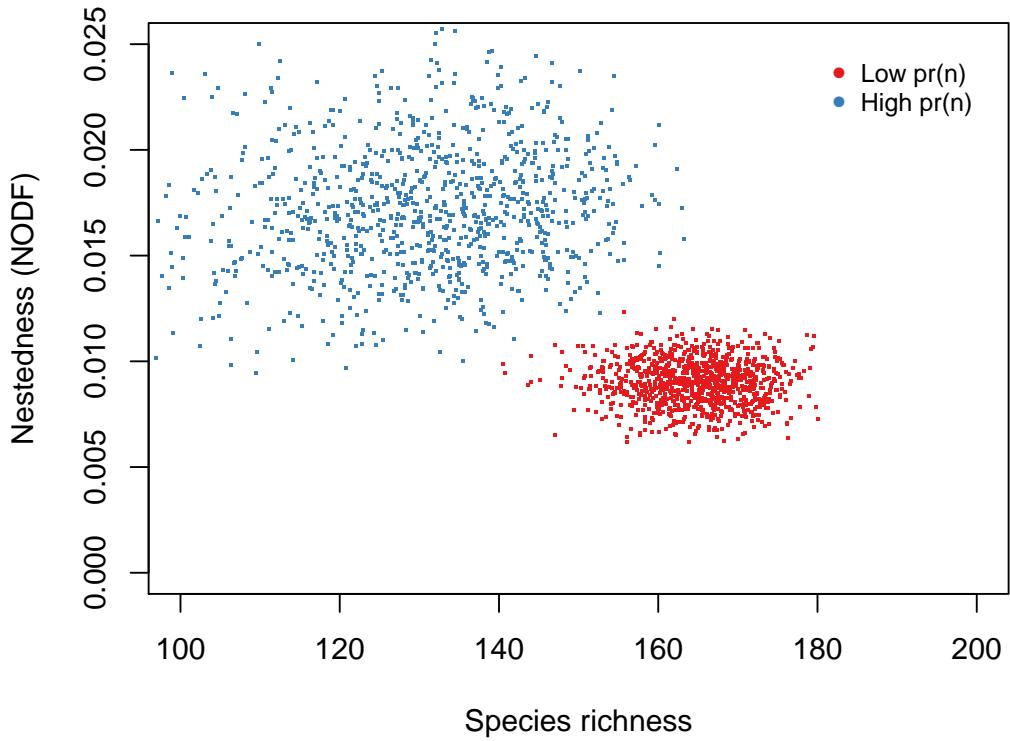


Figure S4. Nestedness (UNODF) as a function of steady state richness for 1000 replicated communities without service interactions ( $p_n = 0$ ) compared to those with a high frequency of service interactions ( $p_n = 0.002$ ). While higher frequencies of service interactions do lower steady state species richness (due to increasing secondary extinction rates), there is not a relationship between nestedness and species richness across replicates for a given service interaction frequency.

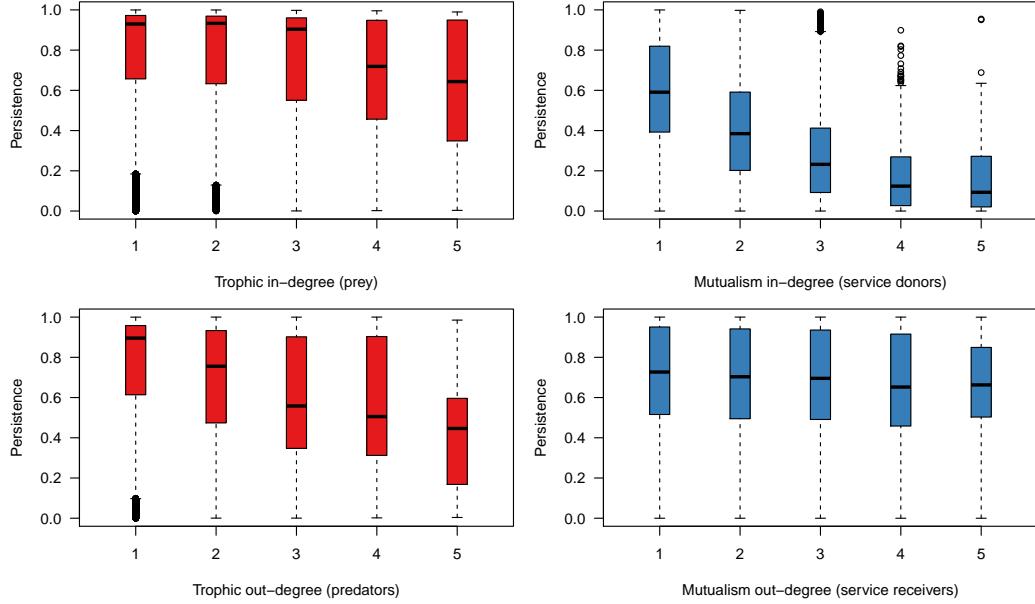


Figure S5. Persistence as a function of trophic and mutualism in/out-degree for communities with higher densities of service interactions ( $p_e = 0.01$ ;  $p_n = 0.002$ ). Left column: species-specific persistence as a function of trophic in-degree (the number of prey a species has; top) and out-degree (the number of predators a species has; bottom). Right column: species-specific persistence as a function of the mutualism in-degree (the number of service donors a species has; top) and out-degree (the number of service receivers a species has; bottom). As the trophic in- and out-degree of species increases, competition strength is lowered and persistence decreases. As the mutualism in-degree increases, so does the number of service donors that are needed for the receiving species to remain in the community. This introduces structural constraints that lowers persistence. Box edges define the 25th and 75th percentiles of simulation results, and whiskers represent  $1.5 \times$  the interquartile range.

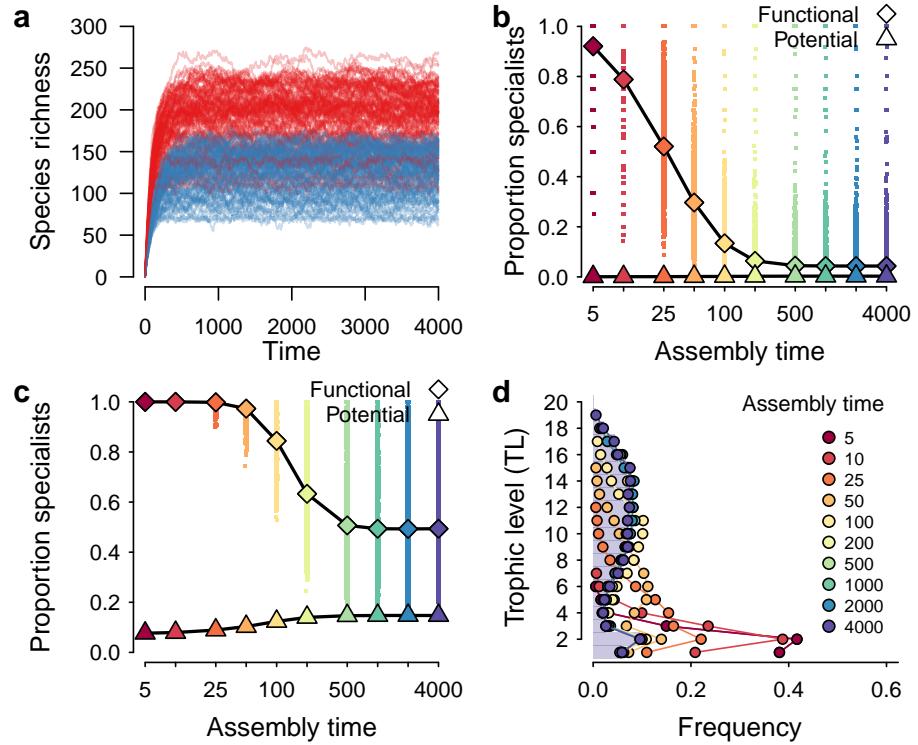


Figure S6. **a**, Assembling communities over time from a pool of 200 engineering species. Species richness is blue; modifier richness is red. Steady state species richness is reached by  $t = 250$ . **b**, The proportion of specialists as a function of assembly time, where a specialist is defined as a species with a generality index  $G_i < 1$  relative to the steady state link density.  $G_i$  is scaled to the steady state link density where links are direct trophic interactions between species. Diamonds represent functional (realized) trophic interactions; triangles represent potential trophic interactions. **c**, The proportion of specialists as a function of assembly time, where a specialist is defined as a species with a generality index  $G_i < 1$ . Here  $G_i$  is scaled to the steady state link density where links are composed of both direct trophic interactions between species and indirect trophic interactions between consumers and those species that produce modifiers as resources. Diamonds represent functional (realized) trophic interactions; triangles represent potential trophic interactions. **d**, The frequency distribution of trophic levels as a function of assembly time (iterations). Autotrophs occupy  $TL = 1$ . Measures were evaluated across  $10^4$  replicates; see Methods for parameter values.

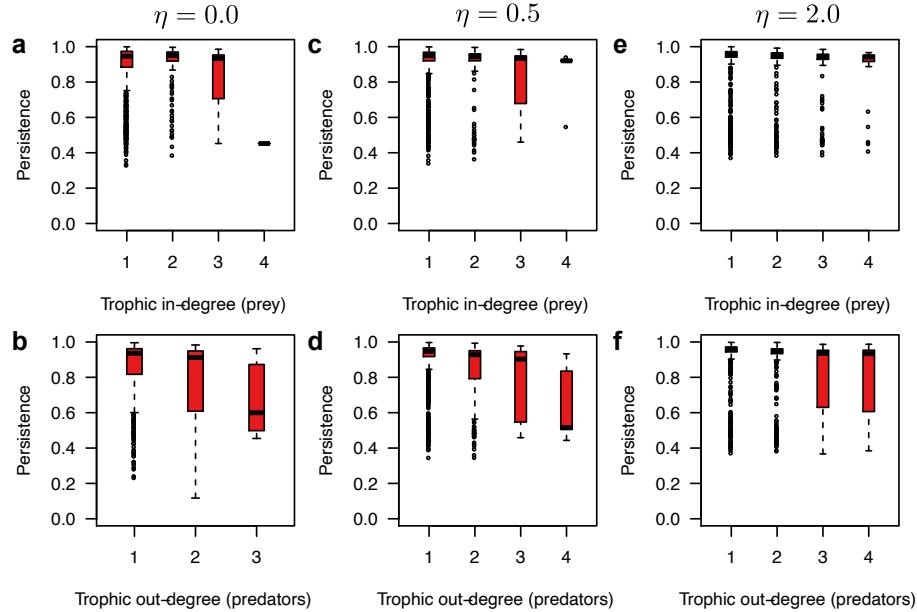


Figure S7. Species-specific persistence as a function of **a**, trophic in-degree (number of resources a species has; top) and **b**, out-degree (number of consumers that eat the species; bottom) when there are no engineers in the community. Species-specific persistence as a function of **c**, trophic in-degree (number of resources a species has; top) and **d**, out-degree (number of consumers that eat the species; bottom) when engineers are rare ( $\eta = 0.5$ ). The notion that having a small number of engineers and modifiers in the community increases rates of primary extinction (Fig. ??a) by stabilizing consumers at the expense of their prey is supported by *i*) increased persistence of generalist consumers, and *ii*) the presence of species with larger number of predators. Species-specific persistence as a function of **e**, trophic in-degree (number of resources a species has; top) and **f**, out-degree (number of consumers that eat the species; bottom) when engineers are common ( $\eta = 2.0$ ). The notion that a large number of engineers and modifiers in the community decrease rates of primary extinction (Fig. ??a) due to expanding niche space (diffusing the effects of competitive exclusion) is supported by the lack of correlation between trophic in/out-degree and persistence. Box edges define the 25th and 75th percentiles of simulation results, and whiskers represent  $1.5 \times$  the interquartile range.

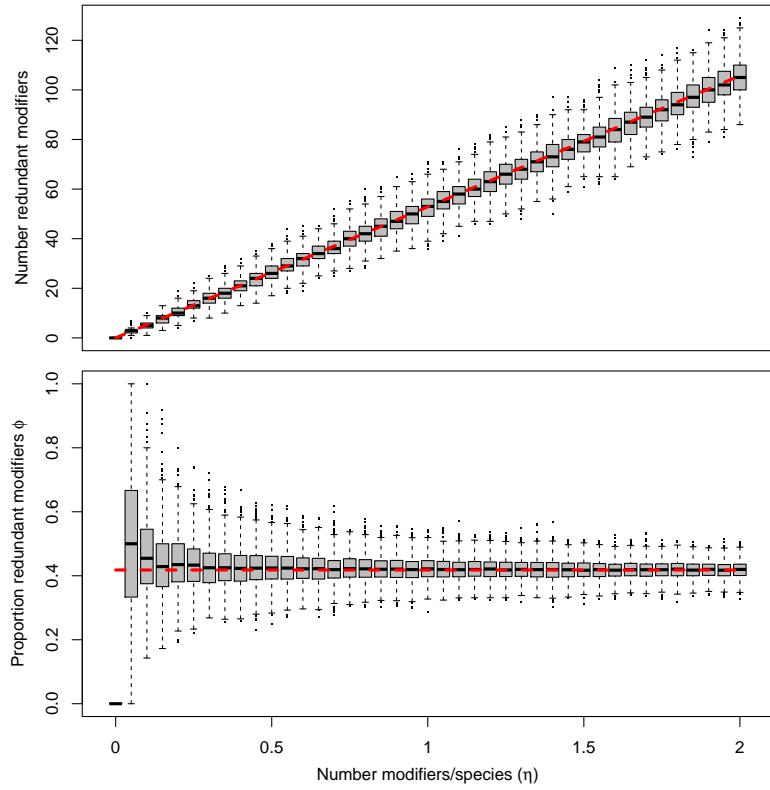


Figure S8. **a**, Number of redundant modifiers in the source pool as a function of the expected number of modifiers made per species  $\eta$ . The red dashed line shows the analytical expectation (Equation (S3)). **b**, Proportion of redundant modifiers  $\phi$  versus the total number of modifiers in the source pool as a function of the expected number of modifiers made per species  $\eta$ . The red dashed line shows the analytical expectation of  $\phi \approx 0.418$  (Equation (S4)). Box edges define the 25th and 75th percentiles of simulation results, and whiskers represent  $1.5 \times$  the interquartile range.

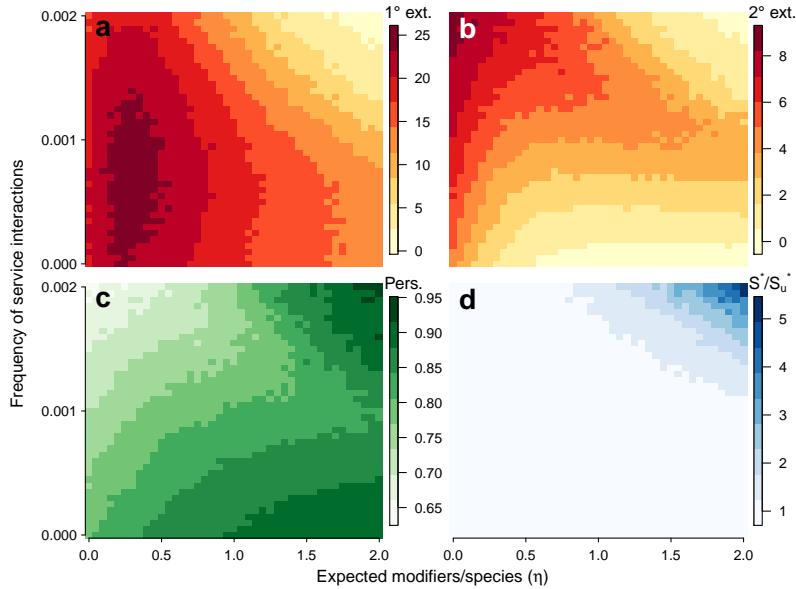


Figure S9. Measures of community stability as a function of the frequency of service interactions and number of modifiers per species, where each modifier is uniquely made by an engineer. **a**, Mean rates of primary extinction, where primary extinctions occur from competitive exclusion of consumers over shared resources. **b**, Mean rates of secondary extinction, which cascade from primary extinctions. **c**, Mean species persistence, defined as the percent simulation time the community is occupied by a given species, averaged across all species that successfully colonize. **d**, The ratio  $S^*/S_u^*$ , where  $S_u^*$  denotes steady states for systems where all engineered modifiers are unique to each engineer, and  $S^*$  denote steady states for systems with redundant engineering. Higher values of  $S^*/S_u^*$  mean that systems with redundant engineers have higher richness at the steady state than those without redundancies. Primary and secondary extinction rates were evaluated at the community level, whereas persistence was determined for each species and averaged across the community. Each measure reports the expectation taken across 50 replicates. See Methods and Supplementary Appendix 2 for parameter values.

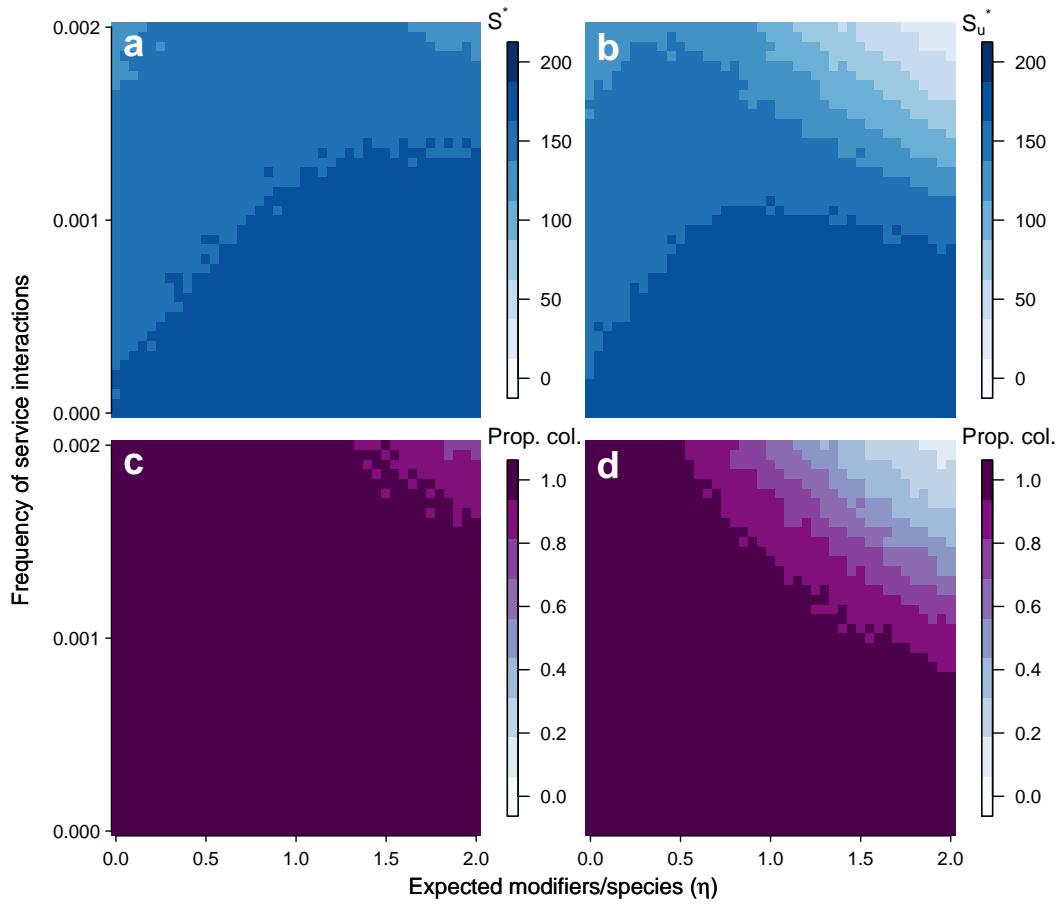


Figure S10. **a**, Steady state community richness with redundant engineering. **b**, Steady state community richness without redundant engineering. **c**, Proportion of species in the source pool that colonize the community at least once throughout the simulation (with redundant engineering). **d**, Proportion of species in the source pool that colonize the community at least once throughout the simulation (without redundant engineering).