Supplementary Material

May 15, 2023

1 Datasets

The name of the 30 graphs of the present study as well as their metrics are reported in Table 1. We have discarded the Input, Output and Respiration nodes, as well as the isolated nodes as they didn't contribute to the structure of the network analyzed in this study, but introduced outliers in the centrality measurements.

2 Measurements

2.1 Connectivity Value

We define [18]:

$$f(G) = \sum_{i=1}^{l} \binom{|C_i|}{2} \tag{1}$$

where C_i is the number of links i-th strongly connected component out of l strongly connected components available and S_G is the number of of species of the graph G (i.e. the number of vertices of the graph G).

3 Node removal strategies

Figures 1,2 show the change in connectivity value for each network under different deletion *strategies*. We explored various strategies to identify the smallest number of nodes that, if removed, would significantly decrease network connectivity. Additionally, we sought insights into which nodes play critical roles in maintaining network connectivity. We noticed that the nodes that represents detriti, sediments, et sim. (or other nodes, ON) afflict the connectivity value in a stronger way then the nodes representing living creatures (LN). Following this study and depending on these results and the aim of the study itself, we decided to use the greedy strategy applied only to LN of each graph.
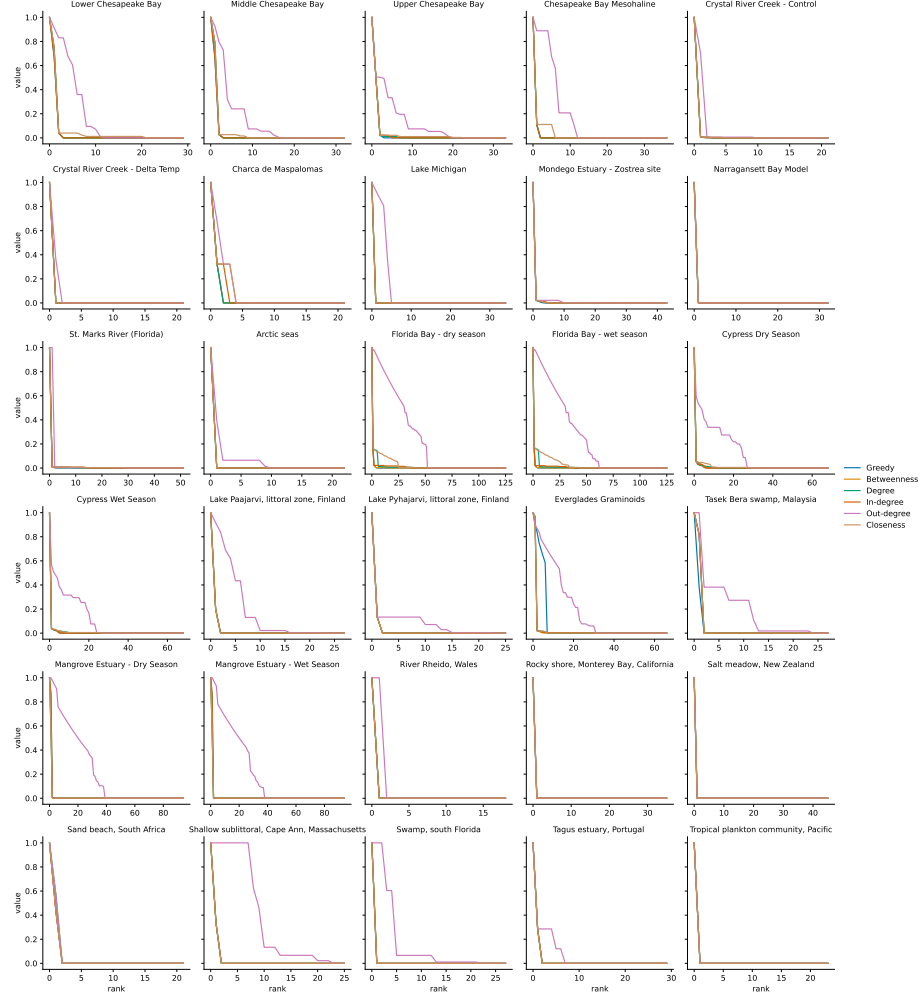
Figure 1: The effect of *ON* and *LN* node deletion on the connectivity value eq.1 of networks in the dataset is shown. Each line represents a different node deletion strategy (color-coded), and the x-axis indicates the order in which nodes were removed for each strategy. The y-axis shows the ratio of the current connectivity value of the network after node removal to its original value with all nodes present
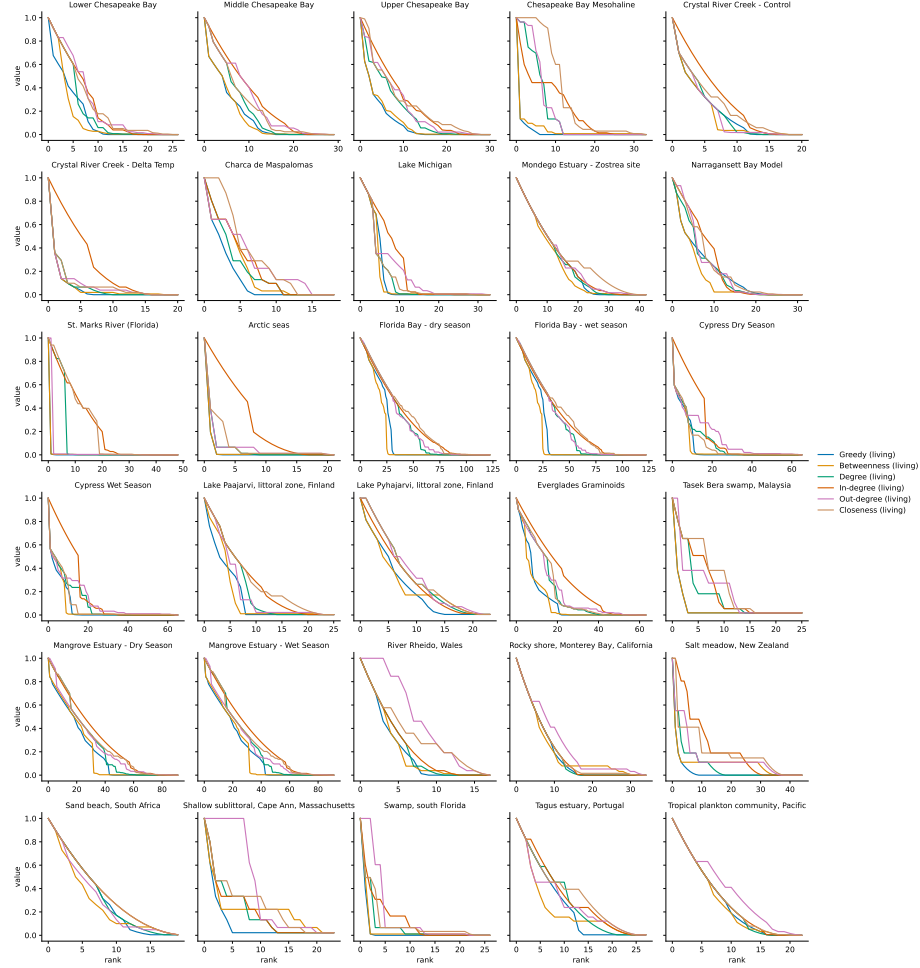
Figure 2: The effect of LN deletion on the connectivity value eq.1 of networks in the dataset is shown. Each line represents a different node deletion strategy (color-coded), and the x-axis indicates the order in which nodes were removed for each strategy. The y-axis shows the ratio of the current connectivity value of the network after node removal to its original value with all nodes present

3

3.1   Deletion by centrality measurement

To identify for each graph the sequence of nodes that result in the least amount of reduction in the connectivity value (as defined in Equation 1) upon their removal, we have defined an algorithm based on the centrality measurements as follows:

1. Evaluate the centrality measurements (either degree, out-degree, in-degree, betweenness or closeness) of the nodes of the graph. Let $C(v_1), \ldots, C(S_G)$ be the sequence of measure for each node in the graph G for the different *centrality measurements*.

2. Find the node that with the maximum value:

   $v = argmax_{v \in V}\big(C(v)\big)$

   In case of multiple nodes with the same value, we pick the one with the lowest index in the graph.

3. Remove the vertex from the graph G:

   $G \leftarrow G \setminus \{v\}$

4. Loop back to step 2 until G has no more nodes.

Where V is either the set of all the nodes of the graph G (LN and OC) or the set of nodes representing living creatures (LN) in the graph G.

3.2   Greedy strategy

To identify for each graph the sequence of nodes that result in the least amount of reduction in the connectivity value (as defined in Equation 1) upon their removal, we have defined an algorithm, named *greedy strategy*, as follows:

1. Evaluate the connectivity value of the graph;

2. For each node in the graph, evaluate the node that minimizes the connectivity value the most:

   $v = argmin_{v \in V}\big(f(G) \setminus \{v\})\big)$

   In case of multiple nodes minimizing the connectivity value with the same amount, we pick the one with the lowest index in the graph.

3. Remove the vertex from the graph G:

   $G \leftarrow G \setminus \{v\}$

4. Loop back to step 2 until G has no more nodes.

   Where V is either the set of all the nodes of the graph G (LN and OC) or the set of nodes representing living creatures (LN) in the graph G.

4   Network Structure

Figure 3 shows the structure of the networks. Each color reflects which node belong to which structure, either the largest giant strongly connected component (violet), the in periphery (blue) and the out periphery (green). As reported, the majority of nodes in these food webs belong to the the largest strongly connected component, with a mean value of 0.79 of all the nodes of the network, and a standard deviation of 0.16. The incident set and outgoing set components are much smaller, with mean values of 0.20 and 0.01, respectively. The nodes outgoing category are either birds (e.g. St. Marks River) or jellyfish (in the case of the Lower and Upper Chesapeake Bay in Summer). Other marginal nodes are measured. Some nodes can be reached from the incident structure but do not belong to the core. These nodes have been filled with yellow and are present in three networks: Lower Chesapeake Bay (Rotifers); Parana, Brazil (Rhaphiodon Vulpinus) and Tasek Bera swamp, Malaysia (swallows). These nodes are predators of the species in the in periphery. In this dataset - with the only exception of the food web of Parana, Brazil - there is at least one ON node belonging to the largest strongly connected component.

5   Generality of nodes in the peripheries

We define [9]:

$$Gen_i = \frac{1}{L_G/S_G} \sum_j^{S_G} a_{ji} \qquad (2)$$

As the *species' normalized prey counts*[26] of the node i. If this quantity is greater than 1, then the species represented by node i has more prey than the average of the other species in the network, indicating that this species exhibits a *generalist* behavior in the ecosystem represented by G. We report in Table 2 the count of the nodes divided by their structure and the number of generalist nodes. As mentioned in the main corpus, in 90% of the datasets, all the nodes that exhibit a higher generalist behavior ($Gen_v > 1$) belong exclusively to the core, with exception of Chesapeake Bay Mesohaline, Charca de Maspalomas, Salt meadow New Zealand.

6   The strongly connected components of the graphs

Figure 4 shows a sketch of the networks of the dataset in which each node is colored according to the strongly connected component (SCC) to which it belongs. The isoleted nodes (*i.e.* the nodes that do not belong to any SCC) are colored in white. We can see that usually each network exhibit a large SCC (that is categorized as the core) and more rarely has multiple SCC. The network that exhibit the latter behaviour are Charca de Maspalomas, Upper Chesapeake Bay, Chesapeake Bay Mesohaline. For Chesapeake Bay Mesohaline
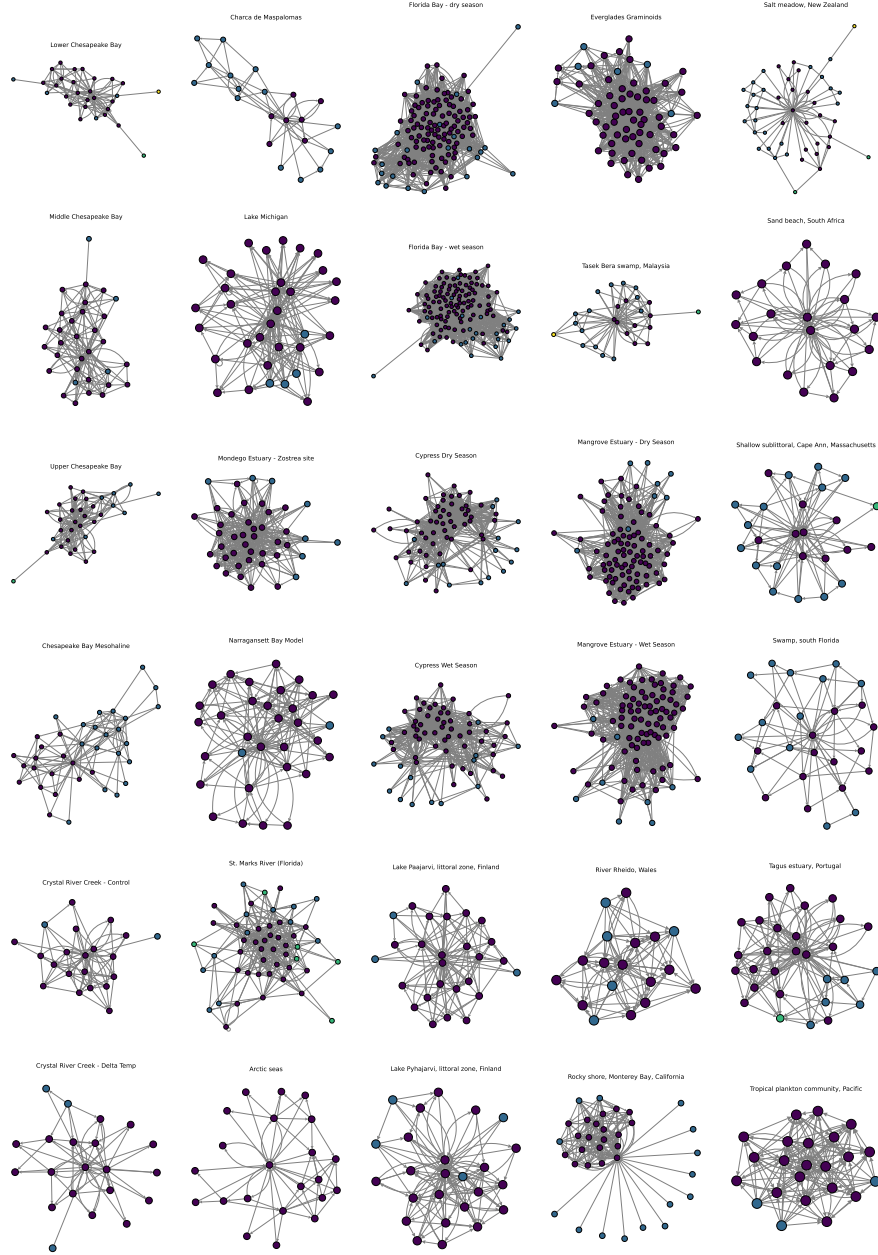
Figure 3: A sketch for each network in the dataset to show the structure of the network: a large strongly connected giant component (violet) with a in periphery (blue) and out periphery (green)

and Charca de Maspalomas we measured that the number of generalist in the In-periphery is higher than usual. This observation may imply that the structure of these networks is characterized by the coexistence of multiple cores in the same ecosystem, as opposed to a structure that is divided into distinct in-periphery and core components described before.

7 Difference between most robust network and least robust network

Figure 5, 6 show the ego network of the ON of the most robust network and the least robust network. The main difference occurs in the number of *out* edges as the number of species feeding from these nodes is larger in the case of the most robust network. This suggest that a key difference regarding the robustness of a food web may be found in the number of nodes that feed from the lower part of foodweb (detriti, sediments, POC).

8 The datasets and their robustness

Table 3 reports the dataset, their first three nodes of the *greedy* sequence (sec. 3.2). Fig 7 shows a scatter plot of the robustness and graph connettance of the dataset. It can be seen that the robustness increases with the connettance. The Pearson correlation coefficient of these two quantities is $r = 0.73$.

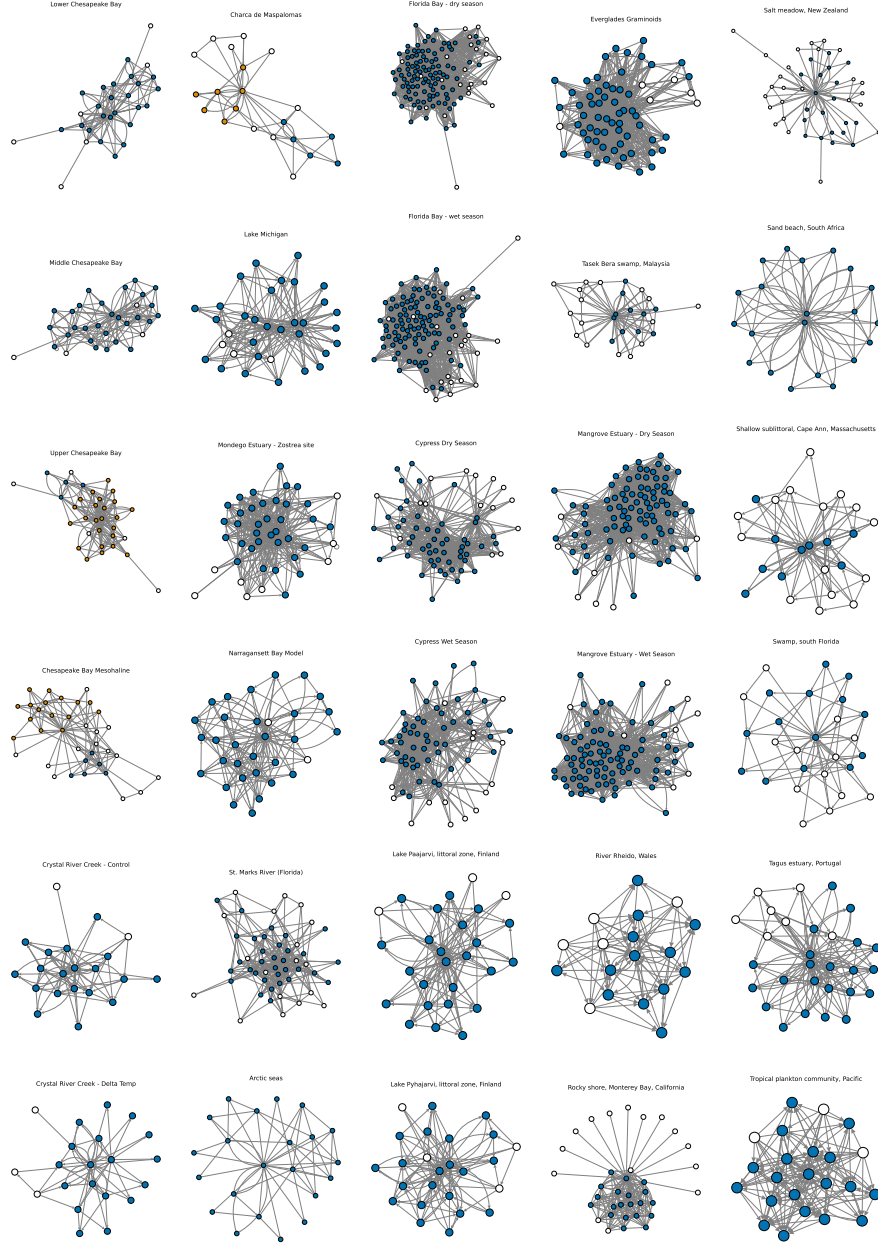Figure 4: A sketch for each network in the dataset to show the different strongly connected component (colored nodes). Isolated note are colored in white.
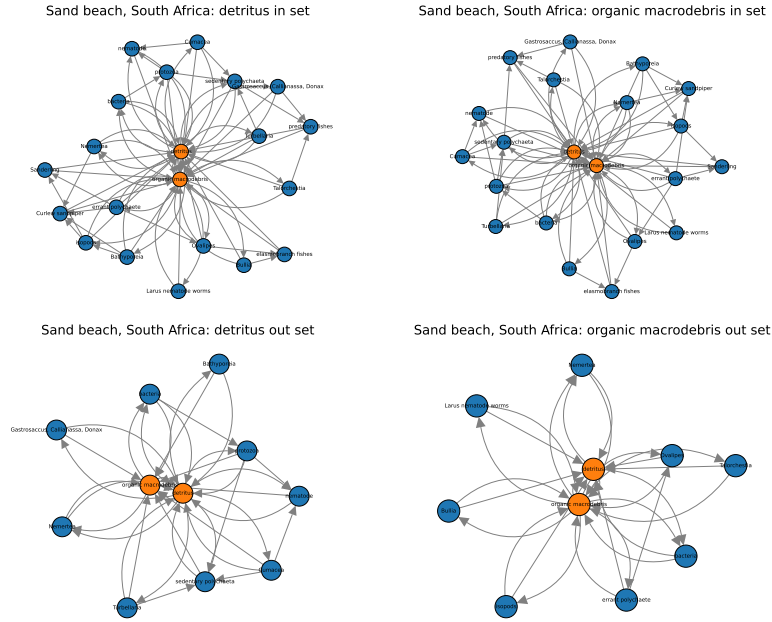
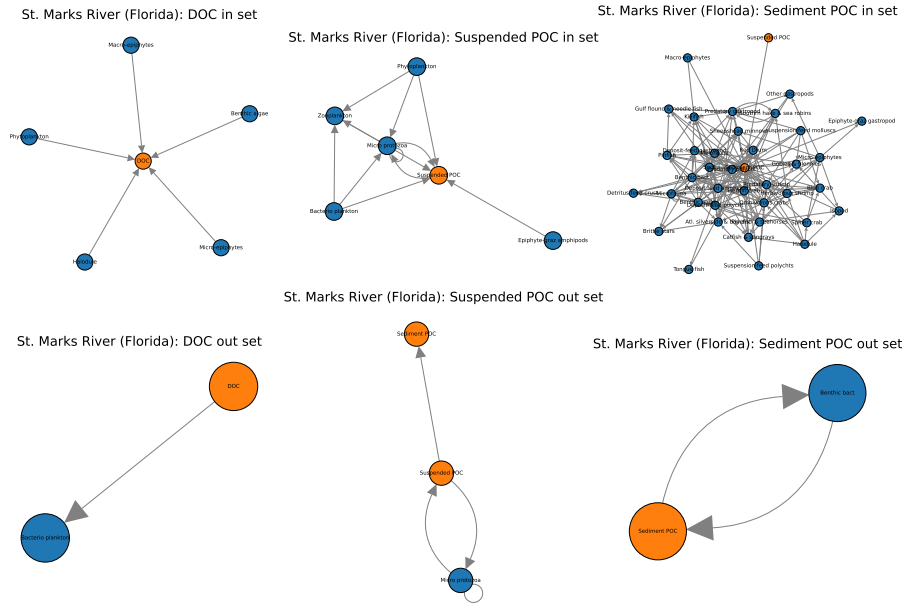Figure 5: Ego network of *ON* nodes for Sand beach, South Africa dataset.

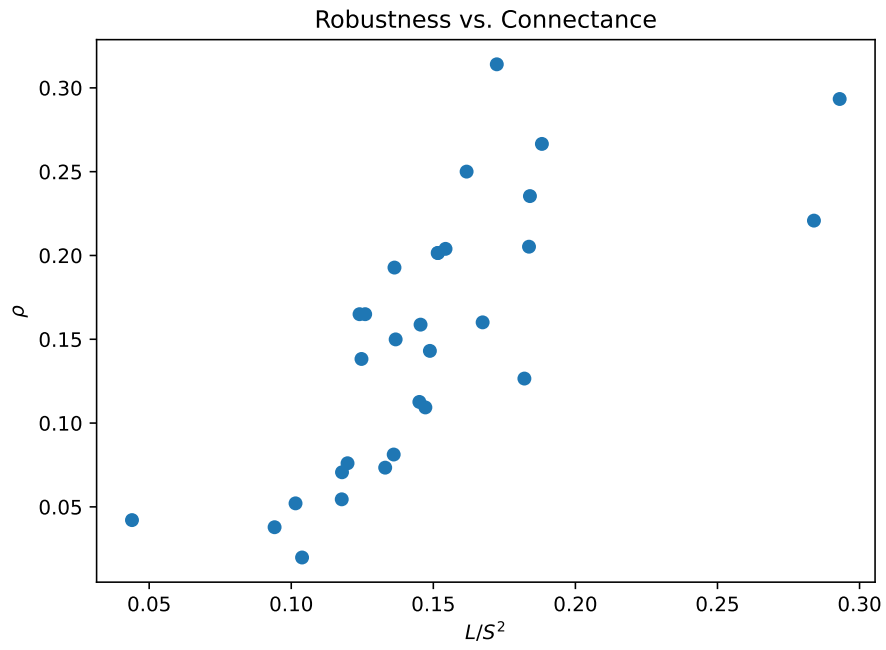Figure 6: Ego network of *ON* nodes for St. Marks River (Florida) dataset.

Figure 7: Robustness as defined in the article (y-axis) and connectance [9] (x-axis) of each graph in the dataset.

Table 1: Dataset main metrics summary. S is the number of species (i.e. the nodes) in the network, L is its number of interactions (i.e. the edges), $LSCC$ the fraction of the nodes belonging to the largest strongly connected component, $f(G)$ is the connectivity value in eq. 1, L/S^2 the *connectance*[6] of the network.

Graph name	Habitat Type	S	L	$LSCC$	$f(G)$	L/S^2
Artic Seas [8, 5]	Marine	22	56	0.45	45	0.12
Charca de Maspalomas [1]	Coastal Lagoon	21	55	0.33	31	0.12
Chesapeake Bay Mesohaline [3]	Estuary	36	122	0.44	135	0.09
Crystal River Creek - Control [25]	Estuary/Marsh	21	81	0.9	171	0.18
Crystal River Creek - Delta Temp [25]	Tidal creek	21	60	0.86	153	0.14
Cypress Dry Season [23]	Swamp	68	554	0.78	1378	0.6
Cypress Wet Season [23]	Swamp	68	545	0.78	1378	0.12
Everglades Graminoids [24]	Marsh	66	793	0.91	1770	0.18
Florida Bay - dry season [23]	Bay	125	1969	0.82	5253	0.13
Florida Bay - wet season [23]	Bay	125	1938	0.82	5253	0.12
Lake Michigan [15]	Freshwater lake	34	172	0.88	435	0.15
Lake Paajarvi, littoral zone, Finland [22, 5]	Lake	27	122	0.89	276	0.17
Lake Pyhajarvi, littoral zone, Finland [5]	Lake	25	115	0.84	210	0.18
Lower Chesapeake Bay [12]	Estuary	29	115	0.79	253	0.14
Mangrove Estuary - Dry Season [24]	Estuary	94	1339	0.91	3655	0.15
Mangrove Estuary - Wet Season [24]	Estuary	94	1340	0.91	3655	0.15
Middle Chesapeake Bay [12]	Estuary	32	149	0.88	378	0.15
Mondego Estuary - Zostrea site [17]	Estuary	43	348	0.81	595	0.19
Narragansett Bay Model [16]	Estuary	32	158	0.94	435	0.15
River Rheido, Wales [14, 5]	Estuary	18	92	0.72	78	0.28
Rocky shore, Monterey Bay, California [11, 5]	Shore	35	167	0.57	190	0.14
Salt Meadow, New Zealand [19, 5]	Wetland	45	89	0.44	190	0.19
Sand beach, South Africa [4, 5]	Shore	21	76	1	210	0.17
Shallow sublittoral, Cape Ann, Massachusetts [7, 5]	Costal	25	92	0.4	45	0.15
St. Marks River (Florida) [2]	Seagrass	51	270	0.65	528	0.1
Swamp, South Florida [13, 5]	Swap	27	74	0.52	91	0.1
Tagus estuary, Portugal [21, 5]	Estuary	29	136	0.76	231	0.16
Tasek Bera swamp, Malaysia [10, 5]	Swamp	27	97	0.41	55	0.13
Tropical plankton community, Pacific [20, 5]	Marine	23	155	0.87	190	0.29
Upper Chesapeake Bay [12]	Estuary	33	158	0.73	282	0.15

Graph name	Core	In	Core Gen	In Gen
Arctic seas	22	0	3	0
Charca de Maspalomas	7	14	4	4
Chesapeake Bay Mesohaline	16	20	6	9
Crystal River Creek - Control	19	2	7	0
Crystal River Creek - Delta Temp	18	3	7	0
Cypress Dry Season	53	15	24	0
Cypress Wet Season	53	15	25	0
Everglades Graminoids	60	6	21	0
Florida Bay - dry season	103	22	57	0
Florida Bay - wet season	103	22	56	0
Lake Michigan	30	4	10	0
Lake Paajarvi, littoral zone, Finland	24	3	7	0
Lake Pyhajarvi, littoral zone, Finland	21	4	8	0
Lower Chesapeake Bay	23	4	14	0
Mangrove Estuary - Dry Season	86	8	35	0
Mangrove Estuary - Wet Season	86	8	34	0
Middle Chesapeake Bay	28	4	11	0
Mondego Estuary - Zostrea site	35	8	16	0
Narragansett Bay Model	30	2	12	0
River Rheido, Wales	13	5	6	0
Rocky shore, Monterey Bay, California	20	15	15	0
Salt meadow, New Zealand	20	22	5	5
Sand beach, South Africa	21	0	4	0
Shallow sublittoral, Cape Ann, Massachusetts	10	14	5	0
St. Marks River (Florida)	33	12	15	0
Swamp, south Florida	14	13	7	0
Tagus estuary, Portugal	22	6	7	0
Tasek Bera swamp, Malaysia	11	14	5	0
Tropical plankton community, Pacific	20	3	11	0
Upper Chesapeake Bay	24	8	11	1

Table 2: Number of generalist vertex for each graph and periphery. Core is the number of vertex in the core, In is the number of vertex in the In-periphery, Core Gen is the number of vertex in the Core with $Gen_v > 1$ (eq. 2) and In Gen is the number of vertex in the In-periphery with $Gen_v > 1$

Graph name	$[v_1^G, v_2^G, v_3^G]$	ρ_G
Sand beach, South Africa	['Gastrosaccus, Callianassa, Donax', 'bacteria', 'Cumacea']	0.3141
Tropical plankton community, Pacific	['bacteria', 'Appendicularia large', 'Acartia large']	0.2934
Mondego Estuary - Zostrea site	['Hydrobia ulvae', 'Melita palmata', 'Ampithoe ferox']	0.2666
Tagus estuary, Portugal	['Ulva, Iteromoplla', 'Chelon labrosus, Liza ramada, L. aurata', 'Crassostrea angulata']	0.25
Lake Pyhajarvi, littoral zone, Finland	['Cyclopoida', 'Keratella, Kellicottia', 'Codonella, Vorticella']	0.2354
River Rheido, Wales	[' Rhithrogena', ' Baetis', ' Chironomidae']	0.2208
Crystal River Creek - Control	['benthic invertebrates', 'mullet', 'zooplankton']	0.2052
Narragansett Bay Model	['Mesozooplankton', 'Microzooplankto', 'Shrimp(Pal,Crg)']	0.204
Mangrove Estuary - Wet Season	['INSCT', 'MICR. H2O', 'EPIFN']	0.2015
Mangrove Estuary - Dry Season	['INSCT', 'MICR. H2O', 'EPIFN']	0.2015
Rocky shore, Monterey Bay, California	['other phytoplankton', 'Suidasia sp.', 'Tegula funebralis']	0.1928
Florida Bay - dry season	['Water Flagellates', 'Water Cilitaes', 'Acartia Tonsa']	0.165
Florida Bay - wet season	['Water Flagellates', 'Water Cilitaes', 'Acartia Tonsa']	0.165
Lake Paajarvi, littoral zone, Finland	['bacterioplankton', 'Copepoda', 'Mollusca']	0.1601
Middle Chesapeake Bay	['Free Bacteria', 'Particle Attached Bacteria', 'Meroplankton']	0.1587
Lower Chesapeake Bay	['Free Bacteria', 'Particle Attached Bacteria', 'Meroplankton']	0.1499
Lake Michigan	['Rotifers', 'Cyclopoids', 'Cladocerans']	0.1431
Charca de Maspalomas	['Benthic Deposit Feeders', 'Pelagic Bacteria', 'Microzooplankton']	0.1382
Everglades Graminoids	['Mesoinverts', 'Terrestrial Inverts', 'Fishing spider']	0.1266
Upper Chesapeake Bay	['Free Bacteria', 'Suspension Feeding Benthos', 'Bay anchovy']	0.1126
Shallow sublittoral, Cape Ann, Massachusetts	['annelids', 'Pomatomus, Poronatus', 'isopods, Gammarus, Caprella']	0.1093
Crystal River Creek - Delta Temp	['benthic invertebrates', 'bay anchovy', 'zooplankton']	0.0812
Cypress Dry Season	['Ter. Invertebrates', 'Aquatic Invertebrates', 'Mink']	0.076
Tasek Bera swamp, Malaysia	['bacteria, fungi', 'detriuvorous invertebrates', 'detritivorous fishes']	0.0734
Cypress Wet Season	['Terrst. I', 'White ibis', 'Aquatic I']	0.0706
Arctic seas	['phytoplankton', 'benthonic invertebrates', 'bacteria']	0.0545
Swamp, south Florida	['insect larvae', 'crayfish', 'amphipods']	0.0521
Salt meadow, New Zealand	['amphipods', 'collembola', 'mites']	0.0421
Chesapeake Bay Mesohaline	['bacteria in sediment poc', 'bacteria in suspended poc', 'ciliates']	0.0379
St. Marks River (Florida)	['Benthic bact', 'Micro protozoa', 'Predatory gastropod']	0.0198

Table 3: The network of dataset with the first 3 nodes of the sequence constructed with the *greedy* strategy (section 3.2)