Supplementary Material

June 23, 2023

1 Datasets

The name of the 32 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 [23]:

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

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 measured in eq. 1. 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.
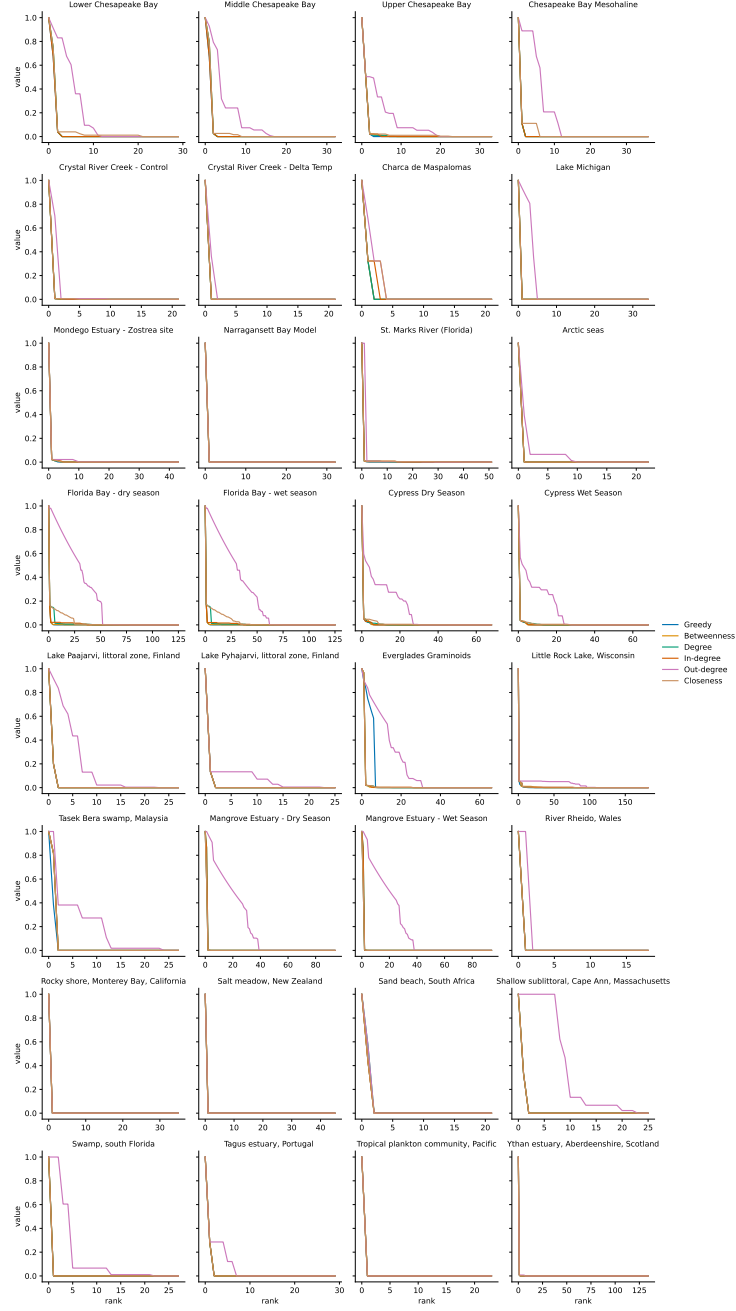
Figure 1: The effect of node deletion strategies (ON and LN) on network connectivity (eq. 1) is shown. Lines represent different deletion strategies (color-coded) and x-axis indicates node removal order. Y-axis shows connectivity ratio after node removal compared to original value.
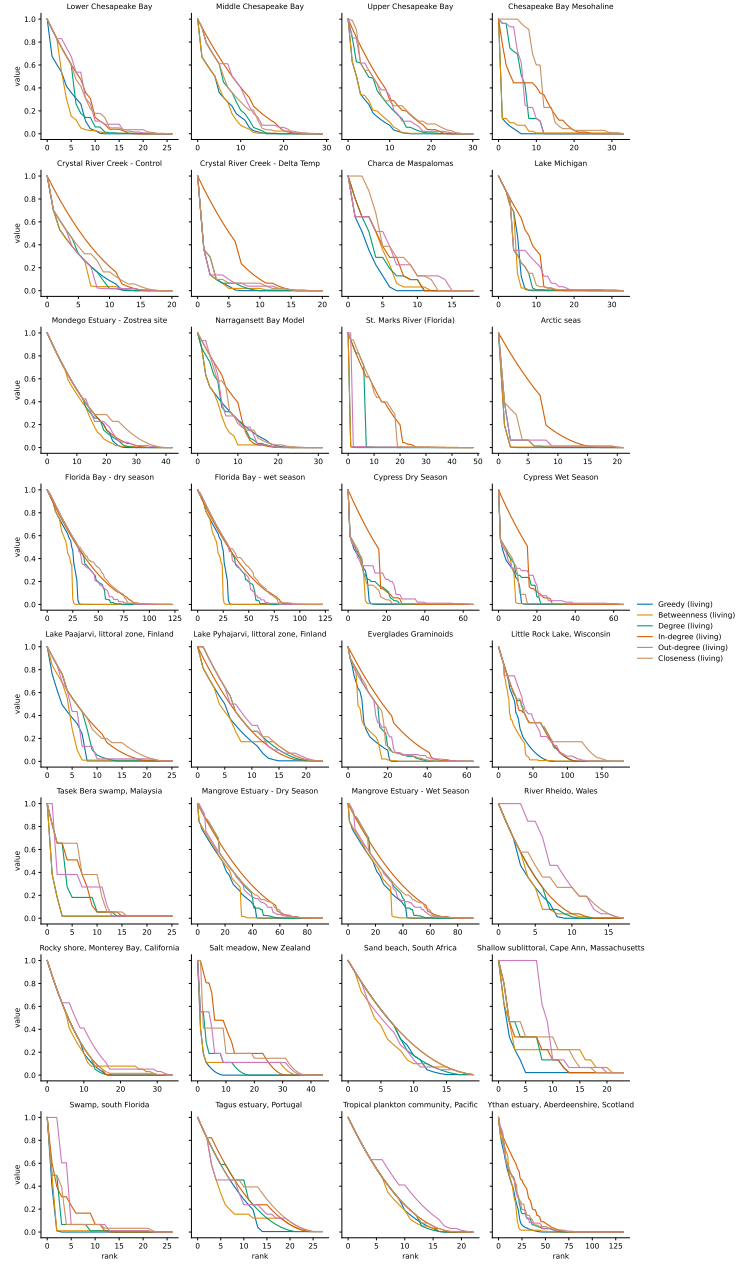
Figure 2: The effect of *LN* node deletion on network connectivity (eq. 1) is shown. Each line represents a different deletion strategy (color-coded), with x-axis indicating node removal order. Y-axis shows connectivity ratio after node removal compared to original value.

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 eq. 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 (Blue), the in periphery (orange) 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   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 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.

6   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, respectively. 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).
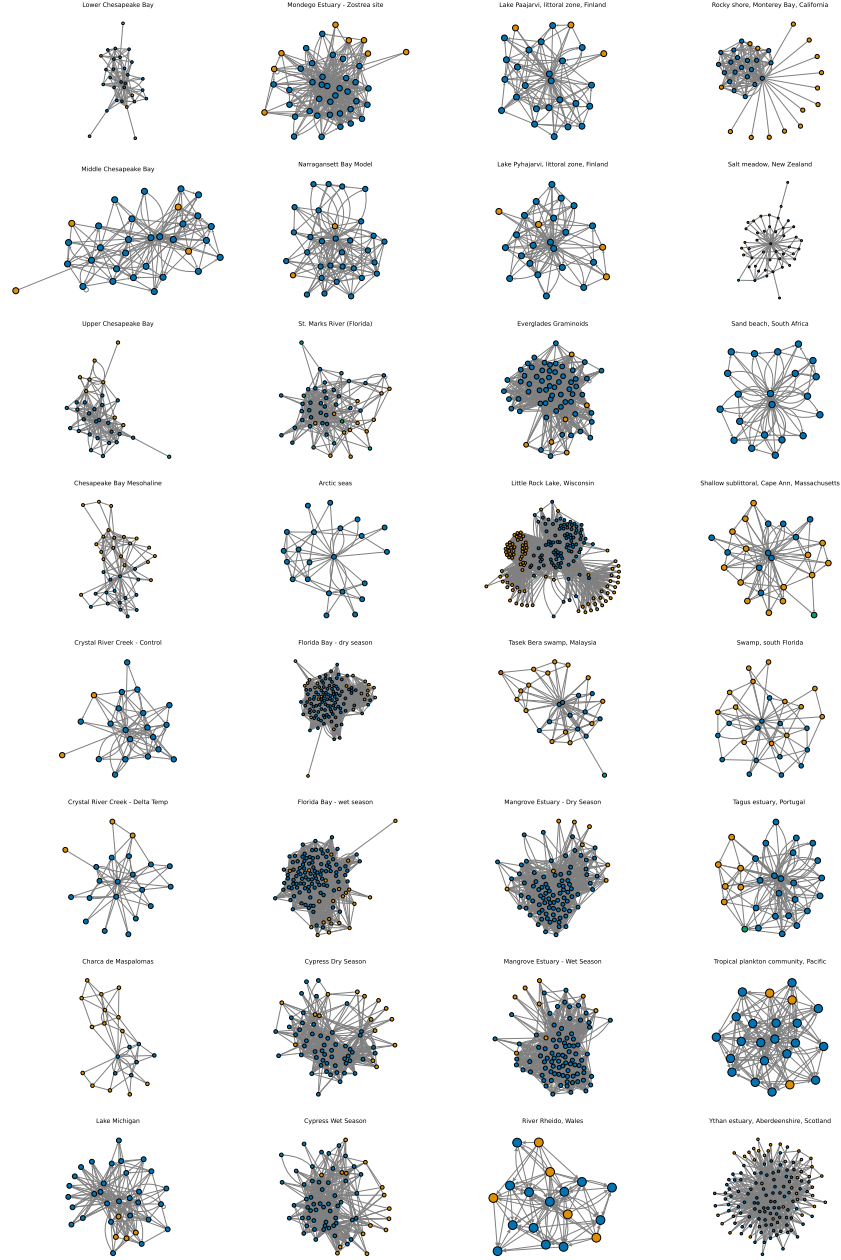
Figure 3: A sketch for each network in the dataset to show the structure of the network: a large strongly connected giant component (blue) with a in periphery (orange) and out periphery (green)
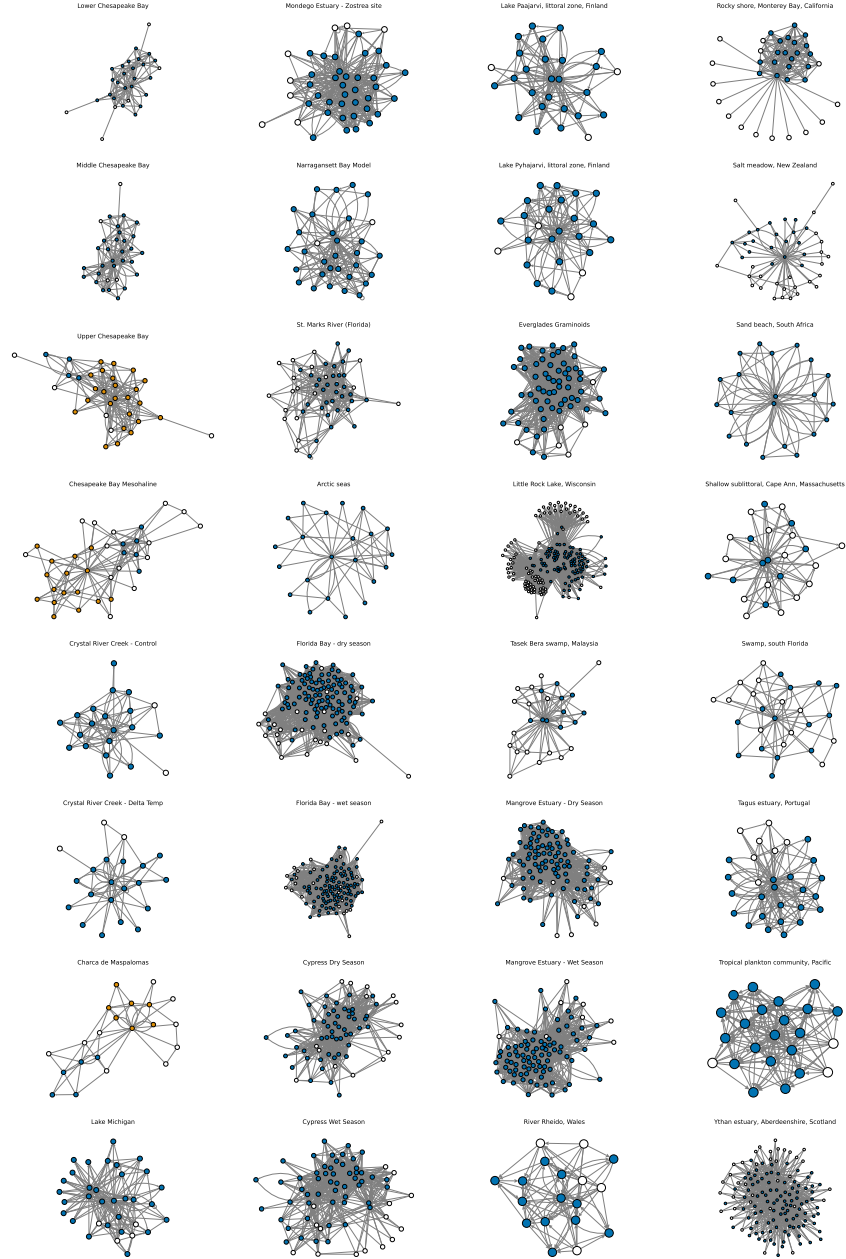
Figure 4: A sketch for each network in the dataset to show the different strongly connected component (colored nodes). Isolated note are colored in white.
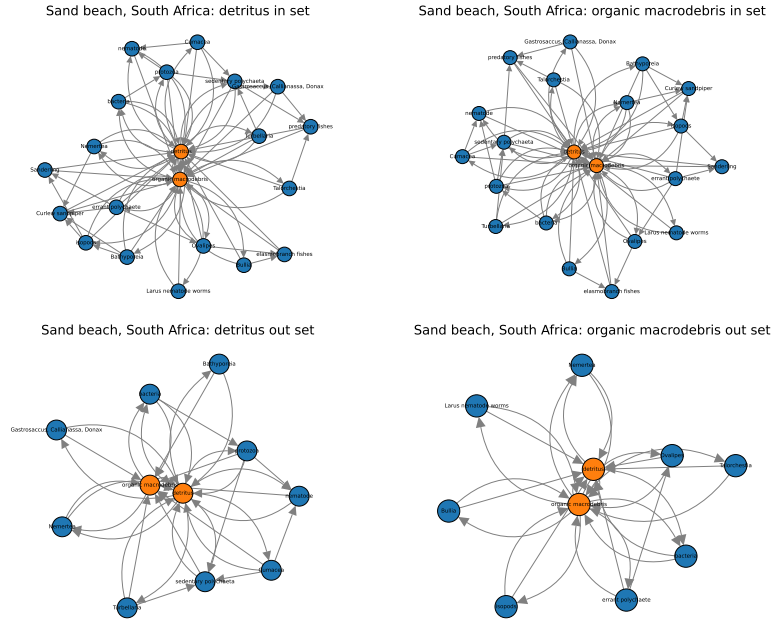
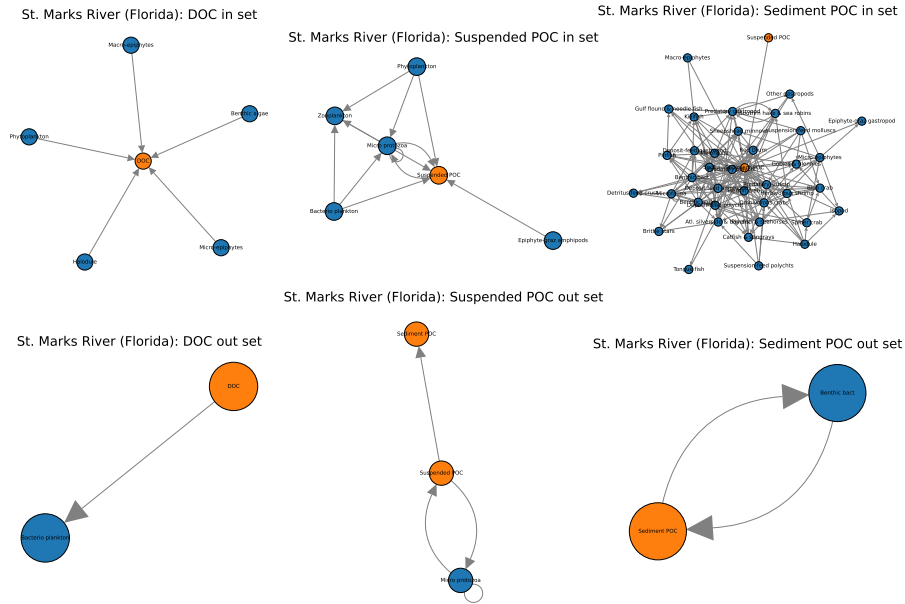Figure 5: Ego network of *ON* nodes for Sand beach, South Africa dataset.

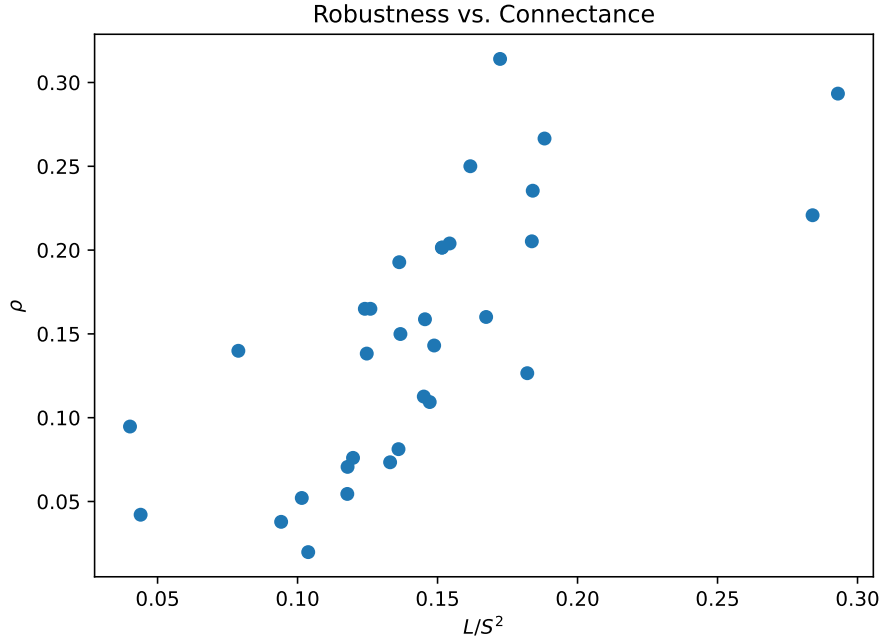Figure 6: Ego network of *ON* nodes for St. Marks River (Florida) dataset.

Figure 7: Robustness as defined in the article (y-axis) and connectance [10] (x-axis) of each graph in the dataset.

7   The datasets and their robustness

Table 2 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. The Pearson correlation coefficient of these two quantities is $r = 0.73$.

8   Representation of three-species motifs

We investigate food-web 'motifs' which are n-species connected subgraphs found within the food web. We focus on 5 single-link motifs ($S_1 - S_5$) and 8 double-link motifs ($D_1 - D_8$) as defined in [28] (fig. 8). Each of these motifs has a particular biological meaning, for example competition, omnivory and food chain [4, 28]. For each of our datasets we consider the number of these motifs. To understand whether the motif is over- or under-represented in the network, we compared the result to a random network generated with the Erdős -Rényi model with the same number of vertices and edges. Interestingly the motif S_1, which represents a Food Chain, is under-represented in with the Erdős-Rényi.
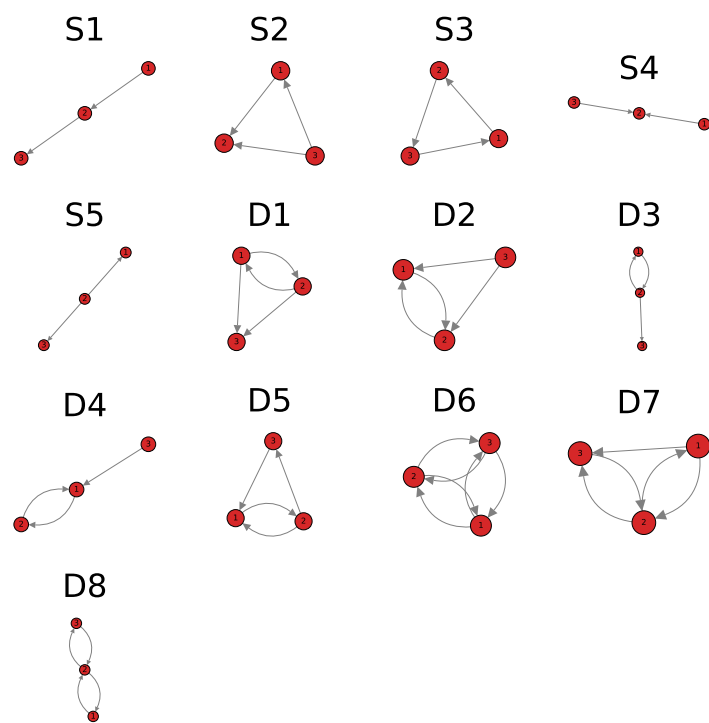
Figure 8: The set of 13 motifs compared in the study.

8.1 Motif representation in real Networks versus Erdős-Rényi graphs

In table 3 we schematically report the result of the comparison of the motifs between the real networks and an ensamble of 50 random Erdős-Rényi graphs. The arrows represents the value of the z-score defined as:[19, 28, 4]

$$z_i = \frac{N_{i,real} - \langle N_i \rangle_{ER}}{\sigma_i} \tag{2}$$

where $N_{i,real}$ is the number of the i-th motif, $\langle N_i \rangle_{ER}$ is the average number of the i motif on the Erdős-Rényi ensemble and σ_i it's standard deviation. In table 3 we denoted with ↑ a value of $z_i \geq 2$ (over-representation); with ↓ a value of $z_i \leq 2$ (under representation) and with – a value $-2 < z_i < 2$. In figure 9 we reported the profile score ($\mathcal{P}_i = z_i / \sqrt{\sum_j z_j^2}$ [20]) of the motifs for each graph in the dataset[1].

Notice that the motif S_2 is over-represented in most of the networks, S_3 and D_3 are under-represented, confirming the results in [28]. In certain cases, the outcomes regarding motifs are inconclusive. This means that in approximately half of the networks, the motif appears to be under-represented, while in the remaining half, it is over-represented. It is valuable to explore the biological characteristics of the network, if any, that are associated with the representation of the motif.

---

[1]The complete comparison between the motifs measured on the Erdős-Rényi graphs and the real graph is available at `https://github.com/davidetorre92/AquaNet/tree/6da811d9b01024103d4c6c97d1325dff030a04ba/Motif/Motif\%20representation\%20in\%20real\%20networks\%20vs.\%20Erdos\%20Renyi`
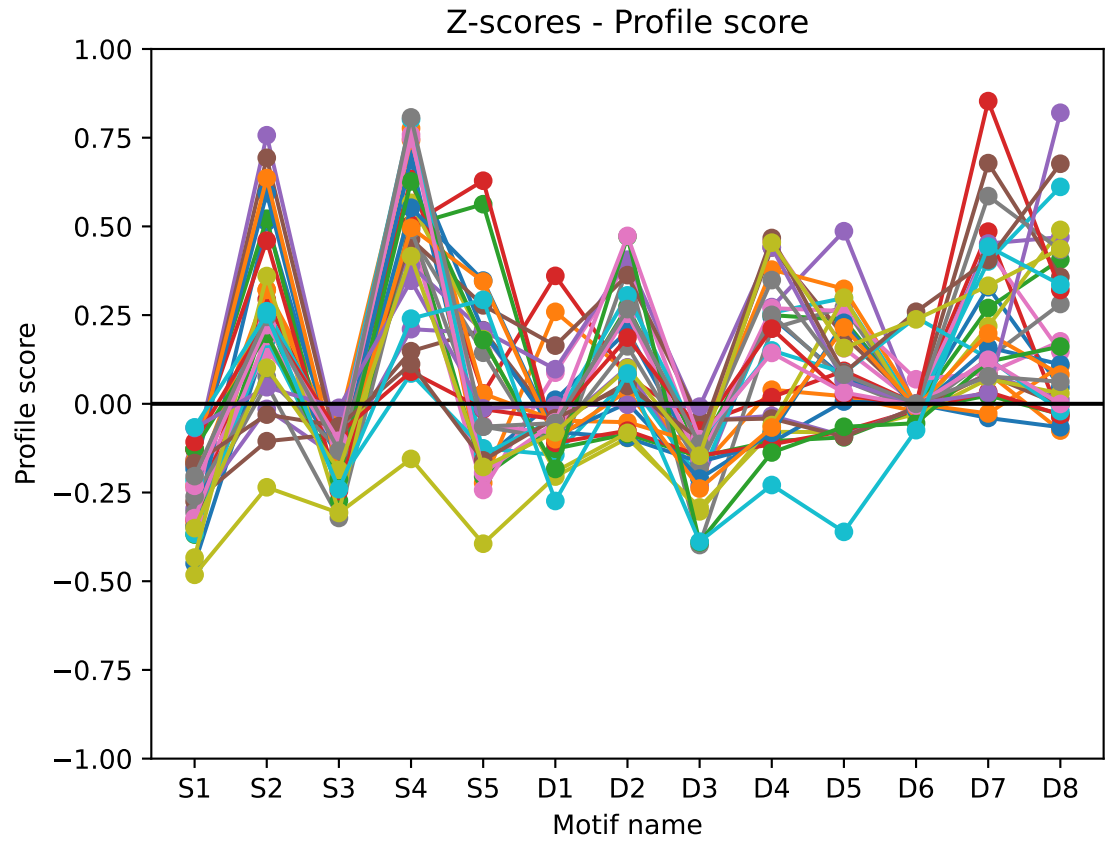
Figure 9: Profile score $\mathcal{P}_i = z_i / \sqrt{\sum_j z_j^2}$ [20] for each motif, for each dataset.

Graph name	Habitat Type	S	L	$LSCC$	$f(G)$	L/S^2
Artic Seas [9, 6]	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 [31]	Estuary/Marsh	21	81	0.9	171	0.18
Crystal River Creek - Delta Temp [31]	Tidal creek	21	60	0.86	153	0.14
Cypress Dry Season [29]	Swamp	68	554	0.78	1378	0.6
Cypress Wet Season [29]	Swamp	68	545	0.78	1378	0.12
Everglades Graminoids [30]	Marsh	66	793	0.91	1770	0.18
Florida Bay - dry season [29]	Bay	125	1969	0.82	5253	0.13
Florida Bay - wet season [29]	Bay	125	1938	0.82	5253	0.12
Lake Michigan [17]	Freshwater lake	34	172	0.88	435	0.15
Lake Paajarvi, littoral zone, Finland [27, 6]	Lake	27	122	0.89	276	0.17
Lake Pyhajarvi, littoral zone, Finland [6]	Lake	25	115	0.84	210	0.18
Little Rock Lake, Wisconsin [18]	Lake	182	2612	0.53	4560	0.08
Lower Chesapeake Bay [13]	Estuary	29	115	0.79	253	0.14
Mangrove Estuary - Dry Season [30]	Estuary	94	1339	0.91	3655	0.15
Mangrove Estuary - Wet Season [30]	Estuary	94	1340	0.91	3655	0.15
Middle Chesapeake Bay [13]	Estuary	32	149	0.88	378	0.15
Mondego Estuary - Zostrea site [22]	Estuary	43	348	0.81	595	0.19
Narragansett Bay Model [21]	Estuary	32	158	0.94	435	0.15
River Rheido, Wales [16, 6]	Estuary	18	92	0.72	78	0.28
Rocky shore, Monterey Bay, California [12, 6]	Shore	35	167	0.57	190	0.14
Salt Meadow, New Zealand [24, 6]	Wetland	45	89	0.44	190	0.19
Sand beach, South Africa [5, 6]	Shore	21	76	1	210	0.17
Shallow sublittoral, Cape Ann, Massachusetts [8, 6]	Costal	25	92	0.4	45	0.15
St. Marks River (Florida) [2]	Seagrass	51	270	0.65	528	0.1
Swamp, South Florida [14, 6]	Swap	27	74	0.52	91	0.1
Tagus estuary, Portugal [26, 6]	Estuary	29	136	0.76	231	0.16
Tasek Bera swamp, Malaysia [11, 6]	Swamp	27	97	0.41	55	0.13
Tropical plankton community, Pacific [25, 6]	Marine	23	155	0.87	190	0.29
Upper Chesapeake Bay [13]	Estuary	33	158	0.73	282	0.15
Ythan estuary, Aberdeenshire, Scotland [15]	Estuary	134	721	0.74	4851	0.04

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*[7] of the network.

Graph name	$[v_1^G, v_2^G, v_3^G]$	ρ_G
Sand beach, South Africa	['Gastrosaccus, Callianassa, Donax', 'bacteria', 'Cumacea']	0.314
Tropical plankton community, Pacific	['bacteria', 'Appendicularia large', 'Acartia large']	0.293
Mondego Estuary - Zostrea site	['Hydrobia ulvae', 'Melita palmata', 'Ampithoe ferox']	0.267
Tagus estuary, Portugal	['Ulva, Iteromoplla', 'Chelon labrosus, Liza ramada, L. aurata', 'Crassostrea angulata']	0.250
Lake Pyhajarvi, littoral zone, Finland	['Cyclopoida', 'Keratella, Kellicottia', 'Codonella, Vorticella']	0.235
River Rheido, Wales	[' Rhithrogena', ' Baetis', ' Chironomidae']	0.221
Crystal River Creek - Control	['benthic invertebrates', 'mullet', 'zooplankton']	0.205
Narragansett Bay Model	['Mesozooplankton', 'Microzooplankto', 'Shrimp(Pal,Crg)']	0.204
Mangrove Estuary - Wet Season	['INSCT', 'MICR. H2O', 'EPIFN']	0.201
Mangrove Estuary - Dry Season	['INSCT', 'MICR. H2O', 'EPIFN']	0.201
Rocky shore, Monterey Bay, California	['other phytoplankton', 'Suidasia sp.', 'Tegula funebralis']	0.193
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']	0.160
Middle Chesapeake Bay	['Free Bacteria', 'Particle Attached Bacteria', 'Meroplankton']	0.159
Lower Chesapeake Bay	['Free Bacteria', 'Particle Attached Bacteria', 'Meroplankton']	0.150
Lake Michigan	['Rotifers', 'Cyclopoids', 'Cladocerans']	0.143
Little Rock Lake, Wisconsin	['Yellow perch (Perca flavescens)', 'Yellow perch (Perca flavescens)', 'Largemouth bass (Micropterus salmoides)']	0.140
Charca de Maspalomas	['Benthic Deposit Feeders', 'Pelagic Bacteria', 'Microzooplankton']	0.138
Everglades Graminoids	['Mesoinverts', 'Terrestrial Inverts', 'Fishing spider']	0.127
Upper Chesapeake Bay	['Free Bacteria', 'Suspension Feeding Benthos', 'Bay anchovy']	0.113
Shallow sublittoral, Cape Ann, Massachusetts	['annelids', 'Pomatomus, Poronatus', 'isopods, Gammarus, Caprella']	0.109
Ythan estuary, Aberdeenshire, Scotland	['Flounder Platichthyes fiesus', 'Corophium volutator', 'Ragworm Nereis diversicolor']	0.095
Crystal River Creek - Delta Temp	['benthic invertebrates', 'bay anchovy', 'zooplankton']	0.081
Cypress Dry Season	['Ter. Invertebrates', 'Aquatic Invertebrates', 'Mink']	0.076
Tasek Bera swamp, Malaysia	['bacteria, fungi', 'detriuvorous invertebrates', 'detritivorous fishes']	0.073
Cypress Wet Season	['Terrst. I', 'White ibis', 'Aquatic I']	0.071
Arctic seas	['phytoplankton', 'benthonic invertebrates', 'bacteria']	0.055
Swamp, south Florida	['insect larvae', 'crayfish', 'amphipods']	0.052
Salt meadow, New Zealand	['amphipods', 'collembola', 'mites']	0.042
Chesapeake Bay Mesohaline	['bacteria in sediment poc', 'bacteria in suspended poc', 'ciliates']	0.038
St. Marks River (Florida)	['Benthic bact', 'Micro protozoa', 'Predatory gastropod']	0.020

Table 2: The network of dataset with the first 3 nodes of the sequence constructed with the *greedy* strategy (section 3.2)

14

	S1	S2	S3	S4	S5	D1	D2	D3	D4	D5	D6	D7	D8
Lower Chesapeake Bay	↓	↑	↓	↑	↓	↓	↑	↓	↑	↓	↓	↑	↓
Middle Chesapeake Bay	↓	↑	↓	↑	↓	↑	↓	↓	↓	↓	↓	↓	↓
Upper Chesapeake Bay	↓	↑	↓	↑	↓	↓	↑	↓	↑	↑	↓	↑	↑
Chesapeake Bay Mesohaline	↓	↑	↓	↑	↓	↑	↓	↓	↓	↓	–	↑	↓
Crystal River Creek - Control	↓	↓	↓	↑	↑	↓	↓	↓	↑	↑	↓	↑	↑
Crystal River Creek - Delta Temp	↓	↓	↓	↑	↑	↓	↓	↓	↑	↑	–	↑	↑
Charca de Maspalomas	↓	↓	↓	↑	↓	↓	↑	↓	↑	↑	–	↑	↓
Lake Michigan	↓	↓	↓	↑	↓	↓	↓	↓	↑	↑	↓	↓	↑
Mondego Estuary - Zostrea site	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑
Narragansett Bay Model	↓	↑	↓	↓	↓	↓	↓	↓	↑	↑	↓	↑	↑
St. Marks River (Florida)	↓	↑	↓	↑	↑	↓	↓	↓	↓	↓	–	↓	↓
Arctic seas	↓	↑	↓	↑	↓	↓	↓	↓	↑	↑	–	↓	↓
Florida Bay - dry season	↓	↑	↓	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓
Florida Bay - wet season	↓	↑	↓	↑	↑	↓	↓	↓	↓	↓	↓	↑	↓
Cypress Dry Season	↓	↑	↓	↑	↑	↑	↑	↓	↓	↓	↓	↓	↓
Cypress Wet Season	↓	↑	↓	↑	↑	↑	↑	↓	↓	↓	↓	↑	↓
Lake Paajarvi, littoral zone, Finland	↓	↑	↓	↑	↓	↓	↑	↓	↑	↑	↓	↑	↑
Lake Pyhajarvi, littoral zone, Finland	↓	↑	↓	↑	↓	↓	↑	↓	↑	↓	↓	↑	↑
Everglades Graminoids	↓	↑	↓	↑	↑	↓	↓	↓	↓	↑	↓	↑	↓
Tasek Bera swamp, Malaysia	↓	↑	↓	↑	↓	↓	↑	↓	↑	↑	↑	↑	↓
Mangrove Estuary - Dry Season	↓	↑	↓	↑	↑	↓	↓	↓	↓	↓	↓	↑	↑
Mangrove Estuary - Wet Season	↓	↑	↓	↑	↑	↓	↓	↓	↓	↓	↓	↑	↑
River Rheido, Wales	↓	↑	↓	↑	↑	↓	↑	↓	↓	↓	↓	↓	↑
Rocky shore, Monterey Bay, California	↓	↑	↓	↑	↓	↓	↑	↓	↑	↓	↓	↑	↑
Salt meadow, New Zealand	↓	↑	↓	↑	↓	↓	↓	↓	↑	↑	–	↑	↑
Sand beach, South Africa	↓	↓	↓	↑	↓	↓	↓	↓	↑	↑	↑	↑	↑
Shallow sublittoral, Cape Ann, Massachusetts	↓	↑	↓	↑	↓	↓	↑	↓	↑	↓	↓	↑	↓
Swamp, south Florida	↓	↑	↓	↑	↓	↓	↑	↓	↑	↓	–	↓	↓
Tagus estuary, Portugal	↓	↑	↓	↑	↓	↓	↑	↓	↑	↑	↑	↑	↑
Tropical plankton community, Pacific	↓	↑	↓	↑	↑	↓	↓	↓	↓	↓	↓	↑	↑

Table 3: Representation of real network Motif with Erdős-Rényi graph representation. ↑ represents a value of the z score ≥ 2. ↓ represents a value of the z score $\leq$ -2. – represents $2 < z_i < 2$.