

# Analysis of real and generated food webs and study on pollution spreading.

S. P. Lipani<sup>1</sup>, L. Negri<sup>1</sup>, M. Seppi<sup>1</sup>, N. Sinigaglia<sup>1</sup>

<sup>1</sup>Department of Physics and Astronomy "Galileo Galilei", University of Padova, Padova (Italy)

## Abstract

*Foodwebs are nowadays an interesting and challenging subject of study for network analysis. In this study real food webs are compared one to each other in terms of how authors collected data, in order to build a coherent dataset. Emergent patterns of these food webs are then quantified and three different generative models are used to try to reproduce them: cascade model, niche model and nested hierarchy model. A detailed comparison shows how niche model is the most suited of the three in accord with the data.*

*Finally, an original clustering algorithm based on the idea of pollution spread is proposed and studied over food webs generated with the niche model. This algorithm needs no parameters, is suited for directed networks and provides information on which species may be screened to understand if the considered food web is polluted.*

## 1. Introduction

A food web (FW) is the natural interconnection of food chains and a graphical representation of "what-eats-what" in an ecological community. In the history of ecological research the description of this interaction has been a priority, since it is the simplest one to describe [1]. Each FW can be represented as a directed graph, where the links illustrate the feeding pathways:

" $\mathbf{a} \rightarrow \mathbf{b}$ " means  $\mathbf{a}$  is eaten by  $\mathbf{b}$ .

One of the primary ways by which ecologists organize ecosystems is trophic structure: the partitioning of biomass between trophic levels [2]. Despite the challenge posed by the intricacy of these trophic networks, a way to capture some of their structural properties is with generative models. Many have been proposed [3], showing that the topology of FWs in nature is non-random. These models are of paramount importance since they provide a basis for understanding how a FW self-organises itself.

The first goal of this work is to study numerous properties in the topological structure of FWs such as: degree distribution, robustness, assortativity; and to identify which generative model is the most suited in accord with data. Three different models are considered: cascade model [4], niche

model [4] and nested hierarchy model [5].

An other important challenge is to understand how pollution would spread in a given ecosystem [6]. There are many ways a pollutant can do so, and one of these are through predations, which are described by a FW.

Finally, in this work an original clustering algorithm is defined following this rationale: "which species would suffer of the same pollution due to the food it eats?". The algorithm is then tested over real FWs and generated ones with the most suited method, with the goal of study how the minimum number of species that should be screened to understand if an ecosystem is polluted changes with respect to FW size and connectance.

## 2. Notations and definitions

$\mathbf{A}$  = Adjacency matrix

$A_{ij} = 1$  means that species  $i$  predate species  $j$ .

**Basal species** := species which predate no other species.

**Top species** := species which are not predated by other species, also called top predators.

## 3. Materials

This work is focused on FWs with "prey-predator" boolean linkages, where an active link  $A_{ij} = 1$  indicates a specific predation, without any quantitative information. To carry out the proposed study, many different FW datasets have been taken from web of life website <sup>1</sup>. For the purpose of the analysis, food networks should be in the form of FWs, representing a sort of ecological pyramid. However, during the collection of datasets, FWs named FW 017 in web of life website, coming from Porto Rico, Marshall islands, Madagascar, Hawaii, New Caledonia and Okinawa, were found to be bipartite graphs. It means that even if the study were carried on them, they would have showed just two levels of connections of basal and top predator species, and it wouldn't have been of interest for the main scope of the study. For this reason they were excluded from the rest of the work and only eight datasets are actually used. More details about them are reported in table 1.

<sup>1</sup><https://www.web-of-life.es/map.php>

Code	Type (location)	Reference
004	River (Brazil)	[7]
007	Seagrass (Florida, USA)	[2]
008	Marine (Caribbean)	[8]
010	Lake (Brazil)	[9]
011	Marine (Maine, USA)	[10]
012	Marine (North Carolina, USA)	[10]
015	Marine (New zeland)	[11]
016	Estuarine (Portugal)	[12]

Table 1: Details about used FWs:

**Code** = identificative number used during the work.

## 4. Methods

### 4.1. Network analysis

To compare different real and randomly generated networks, it is necessary to define some measurable metrics on which the study is performed. These chosen metrics and the visualization algorithm are presented in this section.

- **Parameter  $\gamma$  and scale free model**

The nodes degree distribution of many social and natural networks have been shown empirically to have a power-law behaviour  $p_k \propto k^{-\gamma}$  [13]. It usually emerges whenever there is a network that expresses the preferential attachment property: new nodes tend to connect to high degree ones. FWs might have this property, since new species may eat the most abundant or nutritious resources, already eaten by many other species. Whenever  $\gamma \in [2, 3]$  the network is considered scale-free. Since FWs are directed, the in and out degree distribution will be studied separately.

- **Trophic level**

$$Tr_i = \begin{cases} 1 + \frac{\sum_{j=1}^N A_{ij} Tr_j}{\sum_{j=1}^N A_{ij}} & \sum_{j=1}^N A_{ij} > 0 \\ 1 & \sum_{j=1}^N A_{ij} = 0 \end{cases} \quad (1)$$

The trophic level distribution is one of the most used and useful metric when studying FWs [2], since it takes advantage of the natural hierarchy of such networks and it is able to quantify it. Trophic levels are defined by the self consistency equation 1, where  $A$  is the adjacency matrix and  $Tr_i$  is the trophic level of species  $i$ . The equation assigns to each node the average trophic level of its preys increased by 1; assigning by definition 1 to basal species. A higher trophic level implies a higher ranking of a species as predator, quantifying how much it is high in the FW hierarchy.

- **Assortativity**

Assortativity measures how the average degree of the nearest neighbours  $\overline{k_{NN}^i}$  of a specific node  $i$  depends on the degree of the node  $k^i$ . In the case of a directed network, whether to use in or out degree needs to be specified. So, considering a predator  $i$ , the focus will be on the number of its preys ( $k_{in}^i$ ) and on the average number of predators of its preys ( $k_{NNout}^i$ ).

To quantify this modified metric assortativity the following equation is used:

$$\ln(\overline{k_{NNout}^i}) = \mu \ln(k_{in}^i)$$

Where  $\mu$  quantifies the assortativity: if it is positive the network is assortative, otherwise it is disassortative. Disassortativity may arise as an artifact from the degree distribution, for this reason  $\mu$  parameter will be compared to  $\mu_{RS}$  value, obtained by reshuffling randomly the FW edges; this in fact will maintain only the structural disassortativity.

- **PageRank**

To characterize the centrality of every node in a network one of the most widely used metrics is PageRank [14]. The basic idea behind this algorithm is that each node should have a higher "rank" if it has both a high degree and is connected with other high-ranking nodes. Due to FWs directed structure, the PageRank hubs value of node  $i$  will be strongly related to the number of total species which depend on species  $i$  for survival. In particular, the more a species is basal, the higher its PageRank hubs value will be.

- **Hierarchical robustness**

A way to measure the robustness of a network is with the well-known Molloy-Reed criterion [15], which is based on the assumption of random nodes removal and of having a scale free network. However, the hierarchical structure of FWs makes these two assumptions not appropriate. For this reason a variation of Molloy-Reed criterion is introduced: nodes are removed in an ordered way, following PageRank hubs value of each node in a decrescent way, until the giant component is smaller than the rest of the network.

Finally, hierarchical robustness is defined as the ratio between number of removed nodes and the FW number of nodes. If it is close to 0, it means that the FW breaks as soon as few basal species are removed.

- **Visualization**

In this work FWs are represented in a way that clearly highlight their hierarchical structure. The algorithm can be divided into three steps:

1. Basal and top nodes positions are fixed, creating two parallel horizontal lines;
2. With previous constraint, Fruchterman-Reingold algorithm [16] is run over FW adjacency matrix considering it as an undirected network. This gives the x coordinate of node final positions;
3. y coordinate of node final position is set to be equal to node squared trophic level.

## 4.2. Food webs generator

Two parameters are needed to create a FW: number of nodes ( $S$ ) and connectance ( $C$ ). Since it's uncommon that two species predate each other, in each simulation it is assumed that  $A_{ij} = 1$  must imply  $A_{ji} = 0$ .

- **Cascade model matrices**

The cascade model [4] assigns each species a random value drawn uniformly from the interval  $[0, 1]$  and each species has probability  $P = 2CS/(S - 1)$  of consuming species with values less than its own. This pecking order helps to explain species richness among trophic levels but underestimates interspecific trophic similarity and overestimates food-chain length and number in larger webs.

- **Niche model matrices**

The niche model similarly assigns each species a randomly drawn "niche value" ( $\nu_i$ ) extracted uniformly from the interval  $[0, 1]$ . Species  $i$  consumes all species falling in a range ( $r_i$ ) that is placed by uniformly drawing the centre of the range ( $c_i$ ) from  $[r_i/2, \nu_i]$  [4]. The single range adds a community-level contiguity of niche space to the cascade model by causing species with similar niche values to share consumers frequently within the community. The placement of the niche partially relaxes the cascade hierarchy by allowing up to half a consumer's range to include species with niche values higher than the consumer's value.

The species with the smallest  $\nu_i$  has  $r_i = 0$  so that every web has at least one basal species [17].

- **Nested hierarchy model matrices**

This modeling framework gives each species a niche value and a range, exactly as in the Niche Model. However, instead of establishing links directly to species within the range, first the number of prey to be consumed by each species is determined, in proportion to the range,  $k_{in}^i \propto r_i$ , so as to generate an expected number of links  $L$ , if the resulting number exceeded  $S-1$ , we arbitrarily reduced it to  $S-1$ . These links are then attributed in the following way. The species with lowest niche value has no prey, while the one with the highest has no predators (so there is always at least one basal species and one apex predator). Starting from the

species with second smallest niche value and going up in order of  $n$ , we take each species  $i$  and apply the rules in [5] to determine its  $k_{in}^i$  prey.

By creating groups of consumers, organization is determined by phylogenetic constraints. Links attributed to species free of consumers, and links distributed randomly, render the adaptation of consumers to new prey. In forcing consumers to form various trophic groups, this model escapes the one dimensional nature of former models, and better reflects hierarchies emanating from the phylogenetic structure of the community.

## 4.3. Clustering algorithm based on pollution

There are many ways a pollutant can spread on an ecosystem, and one of these is predation. The proposed clustering algorithm follows this very simple idea, which can be formalized in the question: "supposing each basal species is affected by a different pollutant, which species would suffer of the same pollution due to predation?"

With this setup all it is needed to do is to spread pollutants starting from basal species and following FW structure. Species affected by the same combination of pollutants will belong to the same cluster. Species with a unique combination of pollutants will create a cluster by themselves; for this reason they will be labeled as noise (cluster index = 0).

Algorithm performed, the following quantity can be defined:  $N_{screening} :=$  minimum number of species to screen in order to have full knowledge of pollutants presence. This is an interesting observable since it is able to quantify the minimum effort needed to understand if any pollutant is present in a FW. An easy routine can be build to find  $N_{screening}$ : it always screens a species from the cluster which allow to discover the bigger number of new pollutants. At each step the discovered pollutants are updated and the routine stops when all pollutants have been discovered.

## 5. Results

### 5.1. Limits of this study

Before any analysis, it should be pointed out its limitations. Different networks are often produced by different authors, with different research methods that can influence the overall variables that are studied. This dishomogeneity in methods and not in actual natural behaviours can lead to erroneous correlation of observables. This will be taken in account when comparing random models in section 5.3.

A second limitation arise from the small size of most FWs; this may leads to spurious correlations and noisy observables, some of which the calculation implies that they are dealing with large networks. Every obtained result must be taken with a grain of salt.

	FW 004	FW 007	FW 008	FW 010	FW 011	FW 012	FW 015	FW 016
#Species	32	48	293	39	105	71	107	37
Connectance	0.14	0.1	0.04	0.16	0.03	0.03	0.08	0.18
fraction of top	0.06	0.19	0.15	0.08	0.31	0.37	0.12	0.0
fraction of basal	0.13	0.13	0.17	0.10	0.46	0.54	0.47	0.16
$\gamma_{out}$	$2.6 \pm 0.2$	$2.7 \pm 0.3$	$2.60 \pm 0.04$	$2.75 \pm 0.07$	$2.38 \pm 0.07$	$1.8 \pm 0.1$	$3.0 \pm 0.2$	$3.0 \pm 0.3$
$\gamma_{in}$	$2.8 \pm 0.2$	$2.7 \pm 0.2$	$2.14 \pm 0.04$	$2.7 \pm 0.2$	$3.6 \pm 0.2$	$2.6 \pm 0.2$	$2.8 \pm 0.2$	$2.1 \pm 0.2$
$\mu - \mu_{RS}$	-0.37	-0.03	-0.18	-0.13	-0.10	-0.02	0.35	-0.01
robustness	0.88	0.67	0.69	0.64	0.43	0.21	0.72	0.95

Table 2: Object of study observables of datasets where error on gamma is the result of a linear fit.

## 5.2. Real food webs analysis

In order to distinguish and analyse emerging patterns on food webs and comparing them with random models of matrices generation it's useful to first compare the observables discussed in the section 4.1 between the real food webs. Results of these computations are displayed in table 2. For the purpose of seeing if a comprehensive analysis of these networks makes sense it should be pointed out if the latter are homogeneous for comparison or not.

The chosen FWs vary a lot in the number of species, from 32 to 293. It's important to notice that the number of species does not identify the structure of the food web except for its complexity. Similarly connectance depends both on the abundance of species inside the network and the number of links, therefore it is not alone an indicator of the variety of the species diets. Connectance ranges in  $[0.03, 0.18]$  and from observations seems to be inversely proportional to the number of species.

There exists a big difference in basal species fraction, in particular two groups seems to form: some FWs as 008 and 016 have a basal species fraction around 0.15, while FWs 011, 012 and 015 around 0.45. This can be visualized in figure 5. The analysis of the species defined in each food web highlights how these differences arise from how authors of each paper define and aggregate different species. For, instance in FW 012 there are four species of *Achnanthes* that are a kind of microalgae while in FW 016 just *green algae* are mentioned. Concerning the number of top predators there can be seen that just FW 016 presents anomalies. There are indeed zero top predators as it is shown in table 2 since high trophic level species eat each other.

About the  $\gamma_{out}$  parameter, the first 4 Food Webs have similar values, around 2.6; these values would indicate that the network exhibits the scale free property. However the last 4 vary wildly between 1.8 and 3.0, meaning that maybe these degree distributions belong to another regime. The  $\gamma_{in}$  parameter has some wild variations between all of the FWs, from 2.1 to 3.6. This means that FWs may not have the scale-free property in out degree distribution.

When comparing assortativities, as it is possible to see

from figure 4, checking the Random-Shuffled version is necessary, since structural disassortativity is common in such small networks. Most networks seems to have a disassortative behaviour, with  $\mu \sim -0.1$ , but some others seem to be neutral (FW 012, FW 016) or even assortative (FW 015). This would indicate that predators with a very differentiate diet prefer preys that have less competition.

Robustness ranges in  $[0.2, 0.9]$  meaning there is a structural dependence that makes different measured data with a different robustness. Furthermore the FWs 011, 012, 015 show an outlier behavior and eventually FWs 007, 008 and 010 display a similar behaviour. As to visualize the dependance between robustness and complexity of a food web it could be proper to display how robustness tends to increase with respect to connectance. As visible in figure 1, the hierarchical robustness is likely to increase when links are more frequent.

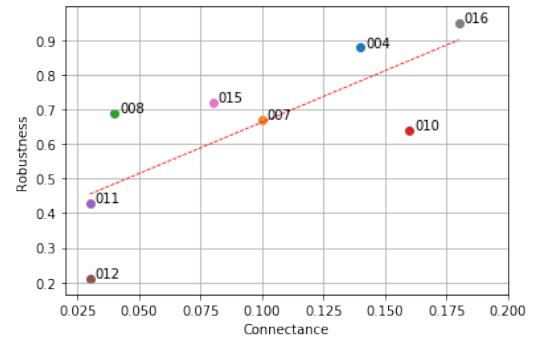


Figure 1: Comparison between Hierarchical robustness and connectance for each analyzed FW.

## 5.3. Models comparison

During the analysis of top and basal species abundance it had been noticed how there are differences between papers authors choices: basal species are sometimes collapsed together and sometimes not. This difference influences many network properties, so when comparing these parameters it would be proper to consider only one of these groups. Since

Observables	Real	Cascade	Niche	Nested
fraction of top	$0.10 \pm 0.07$	$0.02 \pm 0.01$	$0.01 \pm 0.01$	$0.02 \pm 0.01$
fraction of basal	$0.14 \pm 0.02$	$0.75 \pm 0.11$	$0.20 \pm 0.08$	$0.21 \pm 0.05$
$\gamma_{out}$	$2.91 \pm 0.05$	$4.48 \pm 0.45$	$4.73 \pm 0.53$	$4.40 \pm 0.41$
$\gamma_{in}$	$2.66 \pm 0.03$	$2.50 \pm 0.16$	$2.92 \pm 0.05$	$2.74 \pm 0.08$
$\mu - \mu_{RS}$	$-0.14 \pm 0.12$	0	$0.01 \pm 0.16$	$0.04 \pm 0.03$
robustness	$0.76 \pm 0.12$	1	$0.96 \pm 0.06$	$0.80 \pm 0.11$

Table 3: Mean value of observables over the five real FWs and over the five FW generated with the random models with same S and C. Gamma errors have been estimated by the propagation of errors formula since a linear fit error corresponds to each of them for each FW; then the mean propagates the components errors. Other observables do not have any intrinsic error therefore the above given estimation is the result of the std.deviation of different FW values.

collapsed basals FWs are five out of eight, only FWs 004, 007, 008, 010, 016 will be used in the models comparison. Given a real FW, a new one is generated from each model, using as parameters the number of species and connectance of the real FW. In this way a dataset of five FWs will be obtained for each model. In table 3 observables average are reported for real and simulated data, and in figure 7 FWs generated from FW 007, are shown. All the three models predict a lower number of top predators than real networks. It happens because they have been thought in order to ensure in-link complexity. On the other hand, the fraction of basals does not depend on the in degree, and niche and nested models are able to reproduce that behaviour.

$\gamma_{out}$  is not well predicted by random models, it shows a steep slope that cut off the the big hubs while the real FWs present a wider behavior in the degree distribution. On the other hand  $\gamma_{in}$  is well depicted by the random models.

While real networks have a more or less disassortative behavior, this is not captured by any of the random models, predicting slightly assortative or neutral behavior. The only model that predicts in some cases disassortativity is the niche one. The cascade model exhibits the peculiarity of having exactly a neutral action.

Concerning robustness, cascade model is built as an upper triangle matrix with increasing number of interactions for each species. When nodes are removed by following importance order, cascade remains robust since no fundamental food chain is broken; indeed cascade robustness is equal to 1. Other models get close to the real network robustness with a bit of overestimation, again probably due to the inter-linkage property of these models. Eventually trophic level distribution is quite heterogeneous for real data as it can be seen from figures 6 (a). It means that species are distributed in different trophic levels without any particular underlying pattern. However, as visible from figure 7, random models are built for reproducing some particular behaviors:

Cascade model creates an ordered hierarchy where top species predates almost every other species: a single predator is on the top of a ordered ecological pyramid.

Niche model organizes the network in slightly interacting groups, called niches, trying to reproduces the variety of connections inside the same niche as in a real FW.

Nested model is similar to niche one in predicting trophic levels but instead of having independent niches it tends to make them interact. It implies that many species could end up by having the same diet. This can be seen from figure 6 (d), FW 008 simulation. Indeed lower and higher trophic levels are not much populated. The biggest group of species is detected in middle-end trophic levels. It happens because as the number of preys for each predator increases then it is easier for a new species to mimic the diet of the previous variegated ones, since there are would be already some in the adjacency matrix. The largest diets start to show in middle trophic levels and as species are "put" there the more others are attracted to have the same diet and enlarge that particular trophic level.

Overall niche model behaves pretty well together with nested in predicting observables as depicted in table 3, but the niche model produces trophic level distribution more similar to the expected ones, as visible from 6 (c). This makes Niche model the best performing of the three.

## 5.4. Pollution Clustering

In figure 8 are reported results of pollution clustering algorithm applied over FW 007 and 016. In both cases a big cluster is present, due to the high connectance of the two FWs. In fact, in both cases that cluster is the one made of species which are sensible to all possible pollutants.

It is also interesting to notice how many small clusters (2 or 3 species) are present, and how they all tend to be spatially compact. This fact suggests that the structure obtained using the visualization algorithm is strongly related to the hierarchical structure induced by prey-predator links.

FW code	#Species	connectance	$N_{\text{screening}}$
004	32	0.14	1
007	49	0.1	1
008	293	0.04	1
010	39	0.16	1
011	105	0.03	6
012	71	0.03	10
015	107	0.08	1
016	37	0.18	1

Table 4: Node number, connectance and  $N_{\text{screening}}$  for each considered FW in the section 5.2. It is clear how connectance, more than nodes number, plays a very important role on determining the number of species to screen.

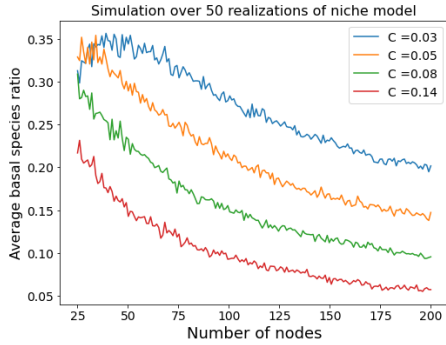


Figure 2: Average basal species ratio with respect to number of nodes and connectance  $C$ , computed performing a Monte Carlo over niche model generated FWs.

A way to approach  $N_{\text{screening}}$  study is to analyze the distribution of the number of pollutants per cluster: it is shown for FWs 008, 011 and 016 in figure 9. As it is visible, for FWs 008 and FWs 016 there is exactly one cluster which has a number of pollutants equals to the number of basal species: this implies that each species in that cluster is sensitive to all pollutants. Which in turn implies that  $N_{\text{screening}}$  is equal to 1 for these two FWs, which is not true for FW 011. In table 4 it is reported  $N_{\text{screening}}$  value for each considered FW in the section 5.2.

Finally, niche model is used to generate FWs with different number of nodes and connectance, with the goal of reproduce trends visible in table 4.

In graph 3 it is reported how  $N_{\text{screening}}$  changes with respect to network size and connectance, while in graph 2 it is reported how fraction of basal species changes with respect to the same parameters. Comparing the two it is clear how  $N_{\text{screening}}$  is directly proportional to the fraction of basal species, which is inversely proportional to FW size and connectance. Moreover, data reported in tabel 4 can be used to

test one more time the validity of niche model. Considering for example FW 007 (connectance = 0.1 , number of nodes = 49) and checking graph 3, it is clear how predicted  $N_{\text{screening}}$  from niche model is 1, which is the observed one. The same reasoning can be applied to each FW, showing how niche model predictions successfully capture the general trend, even if basal species of real FWs are organized in different ways depending on authors.

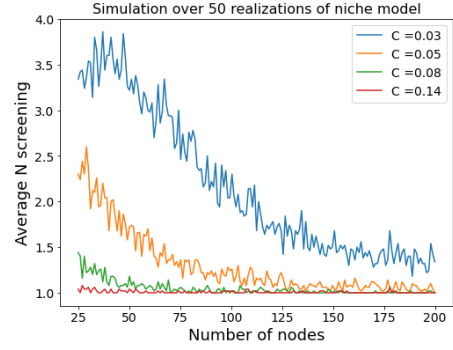


Figure 3: Average  $N_{\text{screening}}$  with respect to number of nodes and connectance  $C$ , computed performing a Monte carlo over niche model generated FWs. If connectance is too high  $N_{\text{screening}}$  tends to be 1; same as FW becomes bigger.

## 6. Conclusions

In this work, starting from real food webs dataset, there is an attempt to define relevant summary statistics to be able to describe them. These observables are used to analyze the real data, highlighting how authors choices may have a big impact on a food web structure and on generality of obtained results. Three different generative models are then compared exploiting these summary statistics:

Cascade model predicts perfect ecological pyramid shapes, which are in contrast with observations;

Niche model is able to capture the biggest number of food webs observables and up to some extent to create realistic trophic structures;

Nested model provides similar results to niche model, but is not able to reproduce properly the trophic structure observed in real data.

Finally, a clustering algorithm is defined following the intuition of how pollutants would spread in a food web. This algorithm, which needs no parameters and provides a way to efficiently organize species screening to detect pollution, is then tested over niche model, highlighting how this model predictions are close to what is observed in real data.

Future works may include more studies about which is the most correct agreement to choose when speaking on grouping basal species, and improvements on the pollution clustering, which is based on the limiting assumption that only basal species can be polluted.

## 7. Contributions

**Conceptualization:** S. P. Lipani, L. Negri, M. Seppi, N. Sinigaglia.

**Writing:** S. P. Lipani, L. Negri, M. Seppi, N. Sinigaglia.

**Data curation and materials:** S. P. Lipani, L. Negri.

### About coding:

**Food web generators:** S. P. Lipani.

**Degree distribution, Assortativity:** L. Negri.

**Trophic level:** M. Seppi.

**Robustness, Visualization:** N. Sinigaglia.

**Real food web analysis :** S. P. Lipani, L. Negri.

**Models comparison:** S. P. Lipani, M. Seppi

**Pollution clustering:** M. Seppi, N. Sinigaglia.

## References

- [1] Angelo Antonio Agostinho Ronaldo Angelini. Food web model of the upper paran river floodplain: description and aggregation effects. 2004.
- [2] Joseph J. Luczkovich Robert R. Christian. Organizing and understanding a winter’s seagrass foodweb network through effective trophic levels. 1998.
- [3] Mercedes Pascual Stefano Allesina, David Alonso. A general model for food web structure. 2008.
- [4] Neo D. Martinez Richard J. Williams. Simple rules yield complex food webs. 2000.
- [5] Marie-France Cattin et al. Phylogenetic constraints and adaptation explain food web structure. 2004.
- [6] Wenjing Guo et al. Persistent organic pollutants in food: Contamination sources, health effects and detection methods. 2019.
- [7] Angelo Antonio Agostinho Ronaldo Angelini and Luiz Carlos Gomes. Modeling energy flow in a large neotropical reservoir: a tool do evaluate fishing and stability. 2006.
- [8] Enric Sala Jordi Bascompte, Carlos J. Melin. Interaction strength combinations and the overfishing of a marine food web. 2005.
- [9] R. Angelini et al. Aquatic food webs of the oxbow lakes in the pantanal: A new site for fisheries guaranteed by alternated control? 2013.
- [10] C. R. Townsend R. M. Thompson. Impacts on stream food webs of native and exotic forest: an intercontinental comparison. 2003.
- [11] C. R. Townsend R. M. Thompson. Is resolution the solution?: the effect of taxonomic resolution on the calculated properties of three streamfood webs. 2000.
- [12] Alexandra Baeta et al. Modelling the effects of eutrophication, mitigation measures and an extreme flood event on estuarine benthic food webs. 2011.
- [13] Albert Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. 1999.
- [14] Page et al. The pagerank citation ranking bringing order to the web. 1999.
- [15] Molloy M. and Reed B. Random structures and algorithms. 1995.
- [16] Fruchterman et al. Graph drawing by force-directed placement. 1991.
- [17] Drew Purves Richard J. Williams, Ananthi Anandanadesan. The probabilistic niche model reveals the niche structure and role of body size in a complex food web. 2010.

## 8. Appendix

Graphs that need more than one column to be visualised are shown in this section.

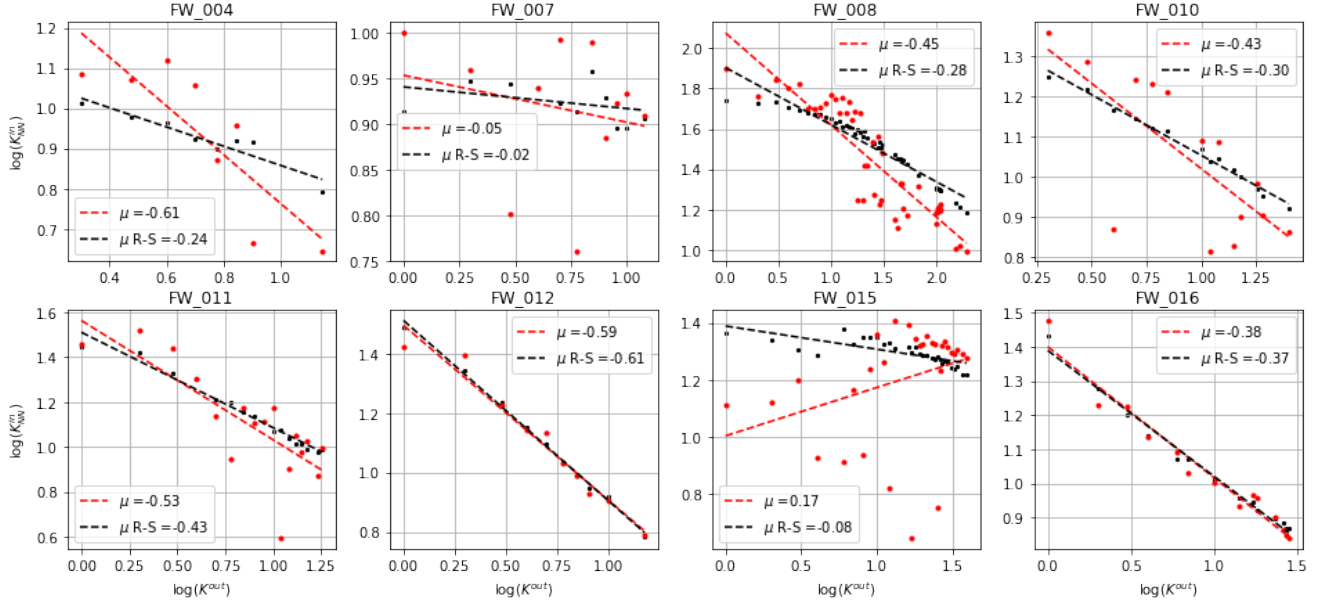


Figure 4: Assortativity for real and Random-Shuffled (RS) version for each network. Most of the FWs exhibit a neutral or disassortative behaviour, with the exception of FW 015, which has a particular internal structure.

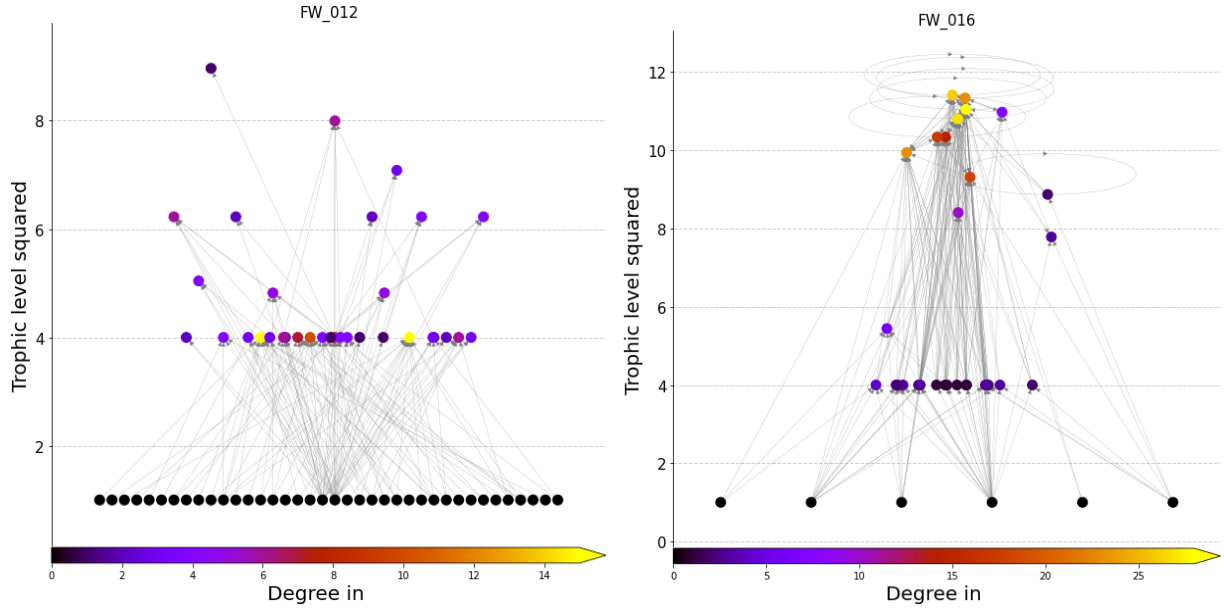


Figure 5: FWs 012 and 016 network visualization. It is important to notice how the number of basal species has consequences on the rest of the network: FW 012 looks more flattened than FW 016, as visible from trophic levels, similarly to a bipartite network. Another consequence is the high average in degree of trophic level 2 in FW 012.



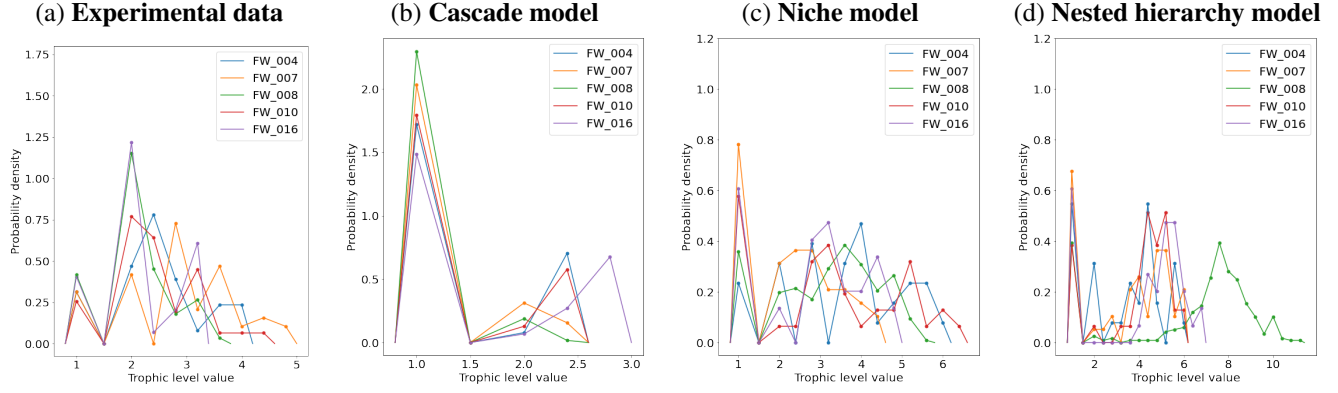


Figure 6: Trophic level distributions of real and simulated FWs. Each FW is simulated with same number of species and connectance of the original one. Histograms bins have width 0.4 and are represented by dots; the first bin is [1,1.4].

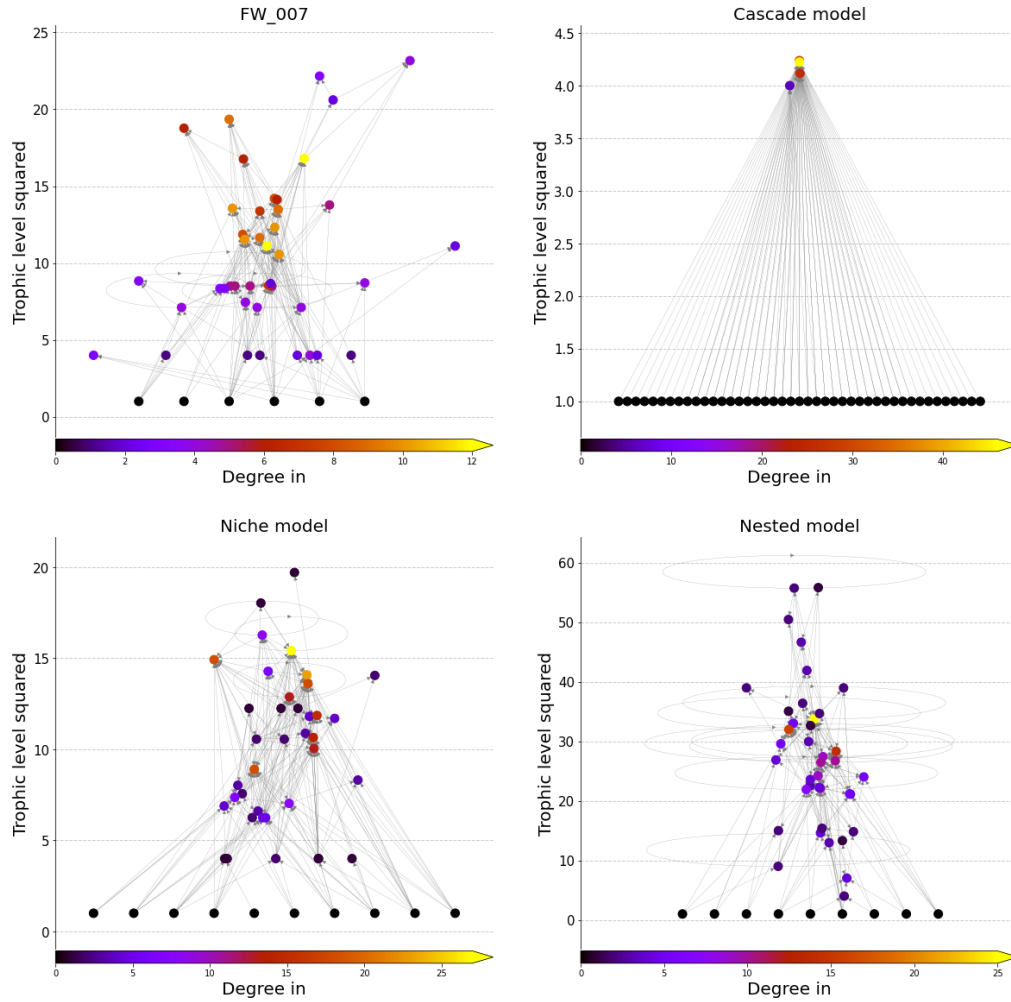


Figure 7: Visualization of FW 007 and random networks generated with same number of species and connectance. The unnatural "triangle shape" of cascade model is clearly noticeable, while both niche and nested model are able to capture most of the features. A difference to point out is the presence in FW 007 of many top predators, which is not found in the models.

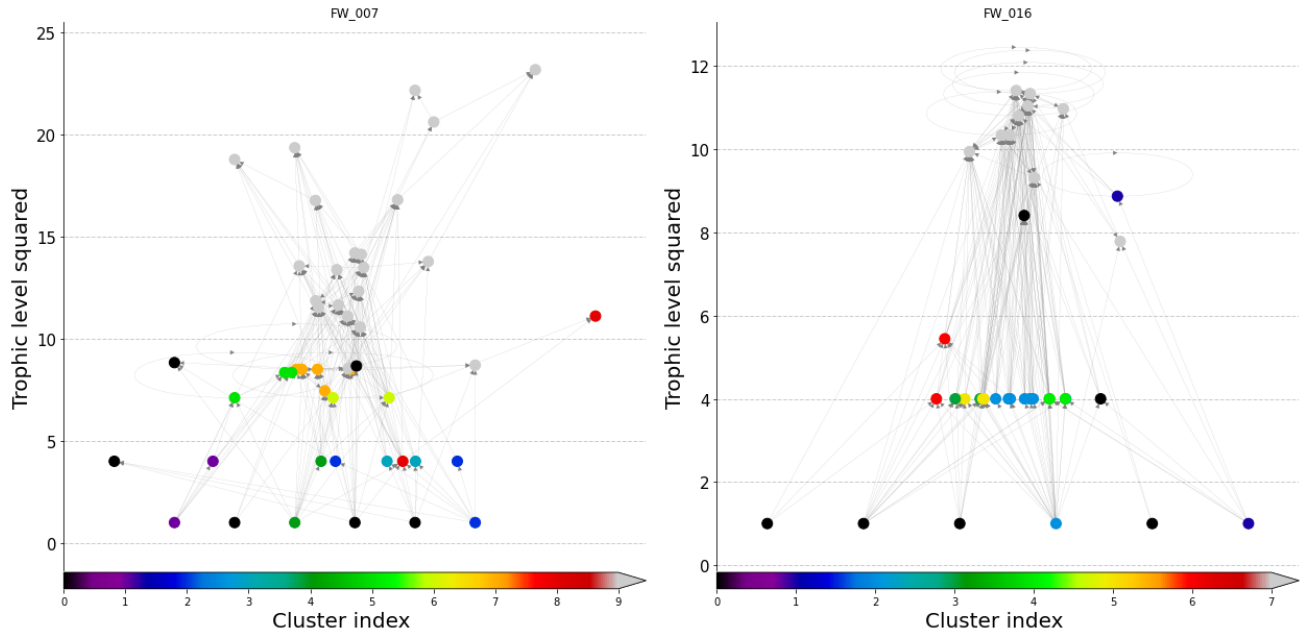


Figure 8: Results of pollution clustering over FWs 007 and 016. The nodes labeled as 0 are "1-element clusters", which can be identified as noise for the sake of visualization, but still are separated clusters. In both cases the big grey cluster is the one made of species which are sensitive to all possible pollutants

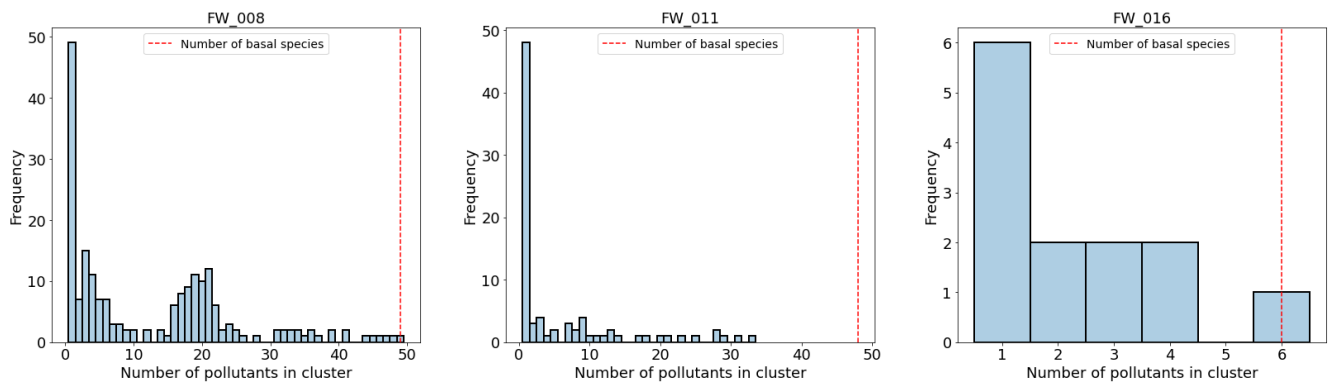


Figure 9: Clusters frequency with respect to the number of pollutants contained in each of them.

If there is at least one cluster with number of pollutants equals to number of basal species, it means that each species in that cluster is sensitive to all pollutants.