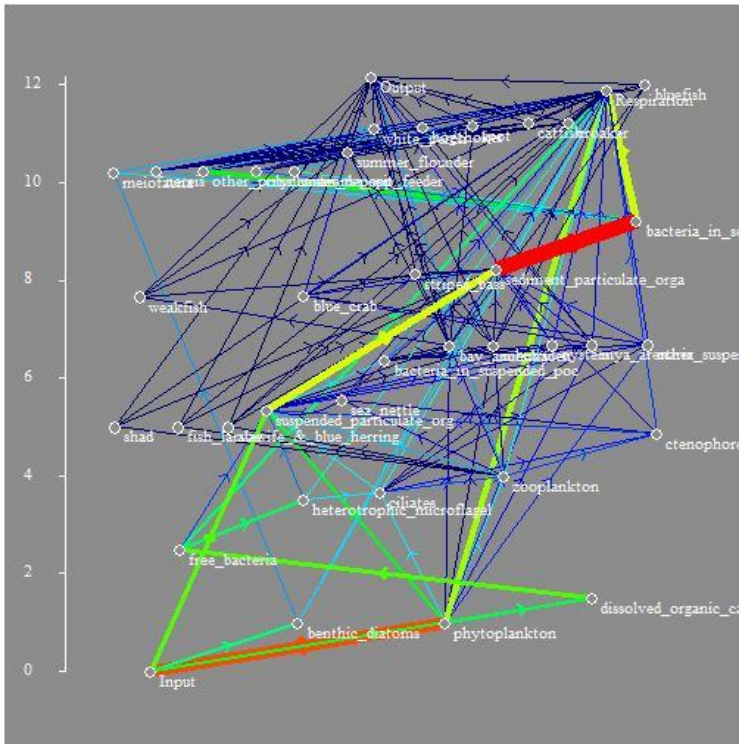
n_w program

ANALYSIS OF TROPHIC NETWORKS

FOOD WEBS



REFERENCE MANUAL

CONTENTS OF THE N_W REFERENCE MANUAL

Introduction
0. Running the program
1. Input files
2. User Interface
3. Graphics
4. Synthetic networks
5. Group detection
6. Network aggregation

Appendix A — Network descriptors

References
Technical notice

CONTENTS OF THE N_W DISTRIBUTION

nw.exe	n_w computer program
qtintf.dll	runtime library
nwref.pdf	reference manual

Example network files

Example.nw0	Example aquatic food web
Lazzaro_58.nw0	Food web from Lazzaro & al (2009)
Lazzaro_74.nw0	Food web from Gauzens & al (2013)
misc/Celegans_neural.gml	Neural network of the nematod *C. elegans*
(...)	

INSTALL

- Download self-extracting archive autonw.exe from website
- Double-click on autonw.exe to expand it inside a user chosen directory

RUN

Either
- Double-click on any network file *.nw0, open it with the n_w program nw.exe.
- Drag and drop any network file *.nw0 to the nw.exe icon to run the n_w program with the file as input.
- Open the n_w program by clicking on the executable file nw.exe. Then use  to open an input file.

WEB SITE

http://www.biologie.ens.fr/~legendre/n_w/n_w.html

AUTHORS

Stéphane Legendre & Benoît Gauzens

CONTRIBUTORS

Jacques Gignoux, Gérard Lacroix, Xavier Lazzaro, Elisa Thébault

INTRODUCTION

The n_w computer program is designed to study food webs, and more generally systems represented by directed graphs. Structural and functional properties of empirical and synthetic networks can be analyzed:

- Topological descriptors → structural properties of the network
- Graphical representations
- Network generation
- Network editing facilities
- Network partition into groups and modules → functional properties of the network
- Network aggregation into simplified networks → functional properties of the network

A food web, a generalization of the 3-levels food chain

Plant → Herbivores → Carnivores,

is a community of species related by trophic interactions (who eats whom) within an ecosystem (Pimm 2002). A food web comprising S species can be represented by a directed graph or network whose nodes are the species and whose arcs describe the feeding links. In the network representing the food web, there is an arrow joining each prey (resource) species to its predator (consumer) species.

Such a network can equivalently be represented by a (0,1)-matrix $A = (a_{ij})$ of size S, with the (0,1)-matrix of the network depending of the presence/absence of a feeding link between species i and species j. An entry 1 in position (i,j) means that species j is a predator of species i. Thus the 1's in row i of the matrix correspond to the indices j of the predator species of prey i, and the 1's in column j of the matrix correspond to the indices i of the prey species of predator j. Hence, rows correspond to prey, and columns correspond to predators.

We note that a node can represent a definite species (e.g., *Daphnia pulex*) or a set of species (e.g., Copepods, Bacteria), or even a compartment (e.g., detritus). In each case we shall talk of 'species'.

Beside the (0,1)-matrix A representing the network, a matrix $W = (w_{ij})$ ascribing non negative real values w_{ij} (weights) to trophic links can be introduced (A and W have the same nonzero entries). The matrix W represents biological values associated with trophic interactions, e.g., energy flux, probability of capture, consumption rate. Weighted descriptors can be associated with such a weighted network. The entries a_{ij} of the matrix A describe the links from resource to consumers in the food web. The entries w_{ij} of the matrix of weights W quantify energy fluxes associated with these links.

Though weighted networks can be studied with the n_w program, their study is not the purpose of the program. The EcoPath program (http://www.ecopath.org) is more appropriate for that.

The n_w program allows to consider several food webs simultaneously, empirical ones, which are specified in text files, or synthetic (random) ones, which are created using predefined models. The networks can be edited and displayed graphically. Results are stored in text and graphic files.

Though other directed networks than food webs can be studied with the n_w program, it is not designed for very large networks (no more than 2000 nodes). Analysis and visualization of large networks can be performed with the Pajek program (http://pajek.imfm.si/doku.php).

0. RUNNING THE PROGRAM

The n_w program uses input files describing the networks of interest. The input files are text files which can be edited using any text editor under Windows (e.g., the NotePad). The main format is the nw0 format, described in the next section.

To run the n_w program with the file *.nw0 as input, either

- Run the n_w program by clicking on the file nw.exe, and open the file using the  icon,
- Drag and drop the icon of the file onto the icon  of the program,
- Associate files with extensions nw0, (and also nw1, nw2) with the executable file nw.exe of the n_w program.

The n_w program opens the file and checks its consistency. For example, in the nw0 format, if the number S of species declared in the first line of the file does not match the number of subsequent lines (the lines of the matrix describing the network, see next section), an error is reported.

The processed file now appears in the EDIT window, and network descriptor values appear in the main window, the Net_Work window.

1. INPUT FILES

The input files of the n_w program are text files describing the networks to study. There are 3 main formats with the respective name extensions *.nw0, *.nw1, *.nw2.

In all these formats:

- The first line of the file indicates the number S of species,
- The data separator is the horizontal tab,
- A line beginning with { is a comment line and is not processed.

1.1. nw0 format

Species names followed by (0,1)-matrix rows describing their predators. The order of the species in columns must be identical to the order of the species in rows.

```
S
Species_1	0	0	0	1	...
Species_2	0	0	1	1	...
...
Species_S	0	1	0	0	...
```

Here Species_1 has Species_4 as predator.

There are S lines after the first one. For a weighted network, the matrix rows contain nonnegative real numbers, e.g.,

```
S
Species_1	0	0	0	0.5	...
Species_2	0	0	0.45	1.2	...
...
Species_S	0	0.90	0	0	...
```

The nw0 files are easily constructed from Excel files. By default, an input file in text format *.txt is considered to have the nw0 format.

Note that the species names are not case sensitive: ‘Daphnia’ and ‘daphnia’ is the same species. However, if the name is ‘Daphnia’ in the input file, it will be displayed as ‘Daphnia’.

1.2. nw1 format

List of prey-predator links. Each line contains the name of the prey species followed by the name of a predator species or nothing when there is no predator.

```
S
Prey_1		Predator_1a
...
Prey_1		Predator_1z
Prey_2		Predator_2a
...
```

```
Prey_2          Predator_2z
...
```

Here species Prey_1 has species Predator_1a, ..., Predator_1z as predators.

There are at least L lines after the first one, where L is the number of links. For a weighted network, the weight associated with the link is specified by a real number, e.g.,

```
S
Prey_1          Predator_1a     0.5
...
Prey_2          Predator_2a     0.45
Prey_2          Predator_2b     1.2
...
Prey_2          Predator_2z     0.90
...
```

1.3. nw2 format

The nw2 format is similar to the nw0 format, but with additional information. It allows to study food webs associated with experimental manipulations (see Lazzaro & al 2009). A 'cumulated' matrix represents all species present in the various experimental food webs. Each experimental food web is represented by an 'instantaneous' matrix, a submatrix of the cumulated matrix. The *.nw2 file describes such an instantaneous matrix. The beginning of the file corresponds to the cumulated matrix of size S in nw0 format, where S is the total number of species in the experiments. A number T ≤ S of additional lines are added at the bottom of the file to specify which species were observed in a given experiment:

```
species_1       a_1
species_2       a_2
...
species_T       a_T
```

These T additional lines specify the presence/absence of the species in the 'instantaneous' food web: a_i = 0, or a missing species in the list, means absence. When positive, the parameter a_i can be used to specify the biomass of species i, or any relevant biological information.

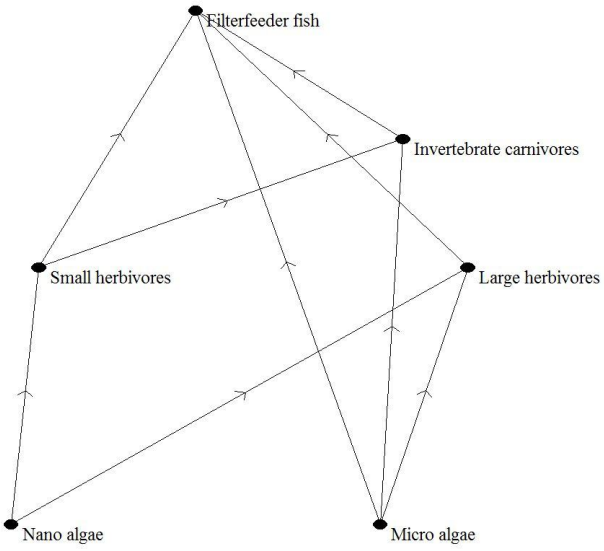
1.4. gml format

The gml (Graph Modelling Language) format is a convenient general purpose format to describe a network. It is shared among several applications (e.g., Pajek, Gephi and others, see http://en.wikipedia.org/wiki/Graph_Modelling_Language). In the n_w program, network files can be opened and saved in this format.

A file opened in any of the formats nw0, nw1, txt, gml can be saved in any other of these formats.

Example of nw0 file: Lazzaro_Small_Filter.nw0

```
{ Reduced food web from
{ Lazzaro X, G Lacroix, B Gauzens, J Gignoux & S Legendre. 2009.
{ Predator foraging behaviour drives food-web topological structure.
{ Journal of Animal Ecology 78:1307-1317.
{ Top predator = Filter feeder fish                                    comment lines
{
6                                                                      number of species
Filterfeeder fish        0   0   0   0   0   0                         top predator
Invertebrate carnivores  1   0   0   0   0   0
Small herbivores         1   1   0   0   0   0
Large herbivores         1   0   0   0   0   0
Nano algae               0   0   1   1   0   0                         basal species
Micro algae              1   1   0   1   0   0                         basal species
```



1.5. Editing an input file within the program

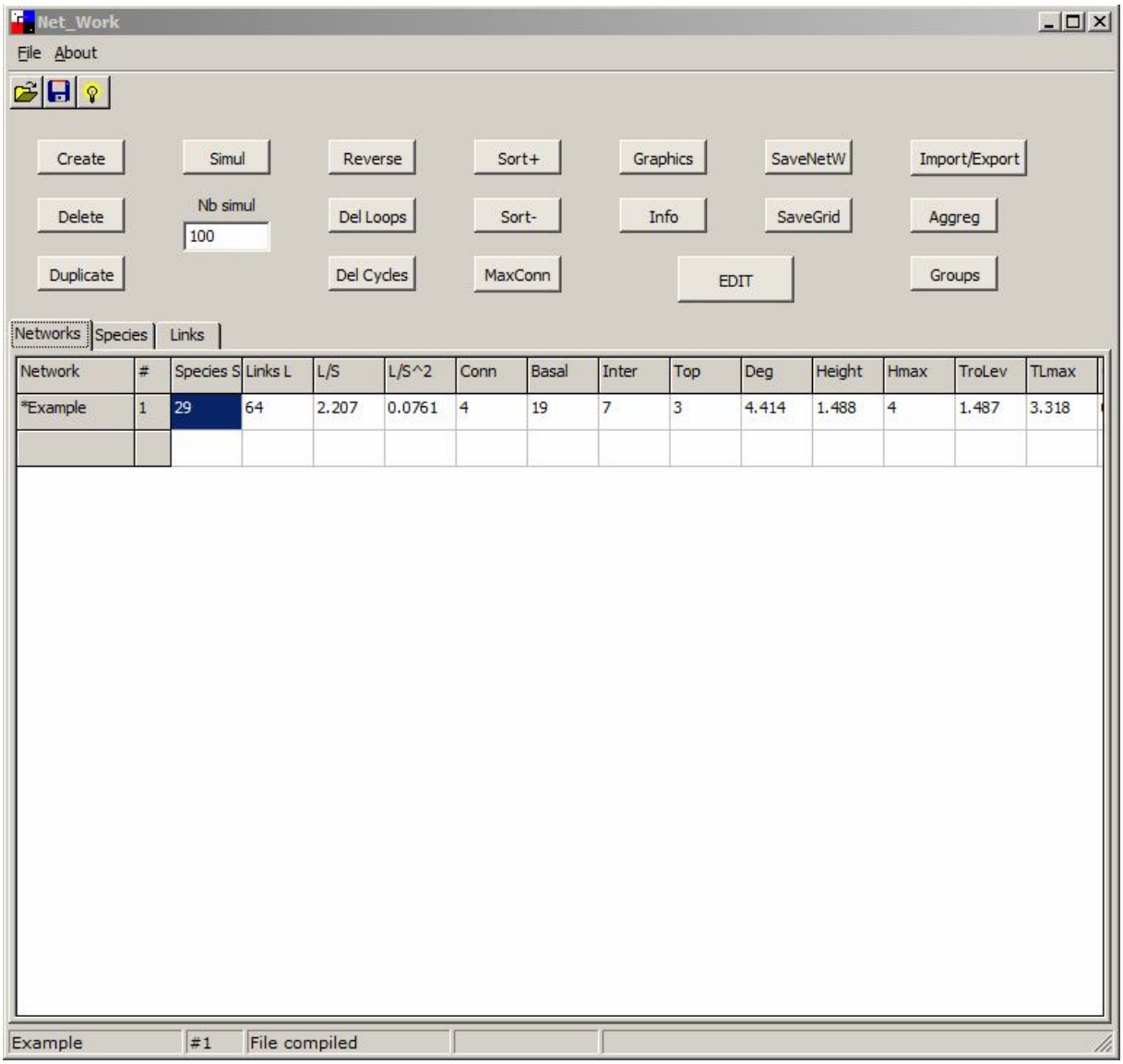
Any input file can be edited, and saved, from the EDIT window. You have to click the compile button  to process the file.

2. USER INTERFACE

Let us run the n_w program on the input file Example.nw0 (click on file nw.exe in the n_w directory, then click on to open the file). There are now 4 windows on the screen. The main n_w window (Net_Work) is on the right and 3 other windows appear on the left: the GRAPHICS windows (grey background), the EDIT window where the input file is displayed, and the PAD window where processing information is given.

2.1. The main window

The main n_w window allows to perform commands, and displays information about the networks under study.



The upper part of the main window contains command buttons (Create, Simul, Delete, ...) and the lower part is constituted of a multigrid where descriptor values of the networks, and their

constituent species, and their links are displayed. There are 3 layers in the multigrid that are described below: Networks, Species and Links. These layers are accessed by clicking the respective tab.

2.1a. Networks grid

The Networks grid displays descriptor values of the networks under study. The signification of these descriptors is summarized below, and explained in more detail in Appendix A.

A network is selected by clicking the grey cell containing its name. The selected network is highlighted by an * in front of its name.

Network	#	Species S	Links L	L/S	L/S^2	Conn	Basal	Inter	Top	Deg	Height	Hmax
*Example	1	29	64	2.207	0.0761	4	19	7	3	4.414	1.488	4

- Network = name of the network
- # = number attributed to the network by the n_w program
- Species S = number of species = size of the matrix
- Links L = number of links = number of 1's in the matrix
- L/S = linking intensity
- L/S^2 = connectance = number of links divided by the total number of possible links
- Conn = number of connected components in the underlying undirected network
- Basal = number of basal species, i.e., species having no prey
- Inter = number of intermediate species, i.e., species having prey and predators
- Top = number of top species, i.e., species having no predators and at least one prey
- Deg = average degree = average number of prey and predator per species
- Height = average height of the food web
- Hmax = maximum height of the food web

Selecting a white cell in the Networks grid and moving the cursor using the → arrow gives access to further descriptor values, as detailed below.

Network	#	TroLev	TLmax	OI	PathLen	Loops	Cycles	CycLen	Diameter	Radius	CharLen	Entropy	ScalEnt
*Example	1	1.487	3.318	0.3088	1.092	0	0	0	3	0	2.086	1.696	

- Network = name of the network
- # = number attributed to the network by the n_w program
- TroLev = average trophic level of the food web
- TLmax = maximum trophic level in the food web
- OI = omnivory index
- PathLen = average path length in the network
- Loops = number of self-loops = number of cannibalistic species
- Cycles = lower bound for the number of directed cycles (zero when the directed network is acyclic)
- CycLen = estimation of average directed cycle length
- Diameter = diameter of the undirected network
- Radius = radius of the undirected network
- CharLen = characteristic path length of the undirected network
- Entropy = entropy H
- ScalEnt = scaled entropy = H/ln(S)

Note on the computation of trophic height

Trophic heights and trophic levels cannot be computed in networks without basal species. For food webs having loops or cycles, all paths in the network are searched exhaustively to compute trophic heights. The algorithm has exponential complexity and is time consuming. In the n_w program, the computation of heights is stopped as soon as the computation time exceeds 6 seconds (indicated by '*** path search - time out' in the PAD window). It is possible to use the commands Del Loops and Del Cycles to create a network without loops and without cycles.

2.1b. Species grid

The Species grid displays in columns species-specific descriptor values associated with the species constituting the network selected in the Networks grid. The signification of these descriptors is summarized below, and explained in more detail in Appendix A.

*Species	#	Conn	Mod	AicG	TroG	In	Out	Deg	Pos	Hmin	Hmean	Hmax	TroLev
*Dceped	1	1	0	0	0	12	0	12	T	1	2.328	4	2.742
NaupCy	2	1	0	0	0	4	3	7	I	1	1.25	2	2.25
Mesoc	3	1	0	0	0	6	0	6	T	2	2.462	4	3.318
NaupCa	4	1	0	0	0	4	3	7	I	1	1.25	2	2.25
Daphn	5	1	0	0	0	8	2	10	I	1	1.125	2	2.125
Simoc	6	1	0	0	0	11	2	13	I	1	1	1	2
Aspl	7	1	0	0	0	4	2	6	I	2	2.176	3	3.156
Brach	8	1	0	0	0	8	3	11	I	1	1.125	2	2.125
Chiro	9	1	0	0	0	6	0	6	T	1	1.167	2	2.167
...													

- Species = name of the species
- # = index of the species in the row of the food web matrix
- Conn = index of connected component the species belongs to
- Mod = index of module (when module detection has been run, see Section 5)
- AicG = index of AIC group (when AIC group detection has been run, see Section 5)
- TroG = index of trophic group (when trophic group detection has been run, see Section 5)
- In = in-degree = number of prey
- Out = out-degree = number of predators
- Deg = degree = in-degree + out-degree = number of prey and predators
- Pos = trophic position (B = basal, I = intermediate, T = top)
- Hmin = minimum height from basal species
- Hmean = average height from basal species
- Hmax = maximum height from basal species
- TroLev = trophic level of the species
- OI = omnivory index of the species
- Clust = clustering coefficient
- Eccen = eccentricity
- Rank = importance rank as defined by Allesina & Pascual (2009b)
- CycTime = cycle time

2.1c. Links grid

The Links grid displays the (0,1)-matrix associated with the network selected in the Networks grid. The matrix can be modified when the EDIT button is ON, thus allowing to modify the links of the network. Select the cell to be modified, type the new value, then RETURN. Links can also be created or deleted using the Create and Delete commands.

Matrix	*Dceped	NaupCy	Mesoc	NaupCa	Daphn	Simoc	Aspl	...
*Dceped	0	0	0	0	0	0	0	
NaupCy	1	0	1	0	0	0	1	
Mesoc	0	0	0	0	0	0	0	
NaupCa	1	0	1	0	0	0	1	
Daphn	1	0	1	0	0	0	0	
Simoc	1	0	1	0	0	0	0	
Aspl	1	0	1	0	0	0	0	
...								

Click on the Matrix cell in the upper left corner to access circularly to different options:

- Matrix. The (0,1)-matrix A = (a_{ij}) of the network describing the presence/absence of a feeding link. The matrix A can be modified when the EDIT button is ON.
- Val. The matrix of weights W = (w_{ij}) of the network describing quantitative values associated with feeding links when such values are present (otherwise W is identical to A). For an aggregated network (see Section 6), the entry w_{ab} of matrix W contains the number of links joining the species in the original network that have been aggregated to form the groups a and b. The same holds for a network issued from group detection (see Section 5).
- LenMin. The matrix whose entry (i,j) gives the minimum (oriented) path length from species i to species j.
- LenMean. The matrix whose entry (i,j) gives the average (oriented) path length from species i to species j.
- LenMax. The matrix whose entry (i,j) gives the maximum (oriented) path length from species i to species j.
- NbPath. The matrix whose entry (i,j) gives the total number of (oriented) paths going from species i to species j.

 To edit the matrix of weights

Under the Links grid, with the upper left corner in the Matrix option, click the EDIT button to turn it ON. Enter the weight value of the link (e.g., replace 1 by 1.5) in the (0,1)-matrix. The edited network is computed, and the matrix entry turns to 0 or 1. You may check, by clicking in the upper left corner to go to the Val option, that the link has the desired weight.

2.2. Commands

Clicking on a grey cell of the grids allows for selection. The item selected is indicated by *. Click on a command button to apply the corresponding operation to the selected item.

Examples

Let us assume that you have opened the file Example.nw0, an aquatic food web comprising S = 29 species and L = 64 links. The network is selected in the Networks grid (indicated by a * in front of the name of the network).

- Click Duplicate → a copy of the selected network is created, and now selected.
- Click Delete → the newly created network is deleted.
- Click Graphics → a graphical representation of the network (see section 4) appears in the GRAPHICS window.
- Click Info → the INFO window appears displaying all descriptor values of the network (some descriptors do not appear in the Networks grid). Moreover, the species constituting the network can be travelled using the arborescence on the left panel of the INFO window, and all descriptor values associated with these species are displayed.
- Click SaveNetW → a dialog box appears, allowing to save the file in any of the nw0, nw1, gml or txt formats.

Now select the Species grid.

By default the first species 'Dceped' is selected (indicated by a * in front of the name of the species), and the first column Species is selected.

- Click Sort+ → a copy of the selected network is created with the species names sorted in alphabetical order (you see a second network Example#3 in the Networks grid, that is now the selected network).
- Select column Deg (a * appears), click Sort- → a copy of the Example#3 network (Example#4) is created with the species sorted according to decreasing degrees, that is in decreasing order of the total number of prey and predators per species. The first species for this order is 'Simoc'.
- Select column Deg, click Graphics → The degree distribution in the network is displayed in the GRAPHICS window (Fig. 3).
- Select column Hmean, click Graphics → The distribution of the species heights is displayed in the GRAPHICS window.
- Click Delete → a new network is created (Example#5) with the selected species ('Simoc') deleted. If the EDIT mode is ON, the species is deleted in the actual network.
- Select the Networks grid. You see that the created network has S = 28 species and L = 51 links.
- Select the Species grid again. Click SaveGrid → a dialog box appears, allowing to save the grid in text format.
- Select the Networks grid. Select the original network Example#1.

Now select the Links grid.

The (0,1)-matrix of the network is displayed. The name 'Dceped' is selected in row.

- Select the name 'Simoc' in column. Click the EDIT button: the EDIT mode is now ON. In the matrix select the entry 0 at the intersection of the row 'Dceped' and the column 'Simoc'. The 0 is highlighted in dark blue. Type 1, then RETURN. A new link joining 'Dceped' to 'Simoc' has been created (Note: this link is not biologically relevant). You see in the Networks grid that the network now has L = 65 links, and that its height has increased. You may click the GRAPHICS button to observe the new network.

EDIT mode

The EDIT mode is used to modify the network, via the matrix entries in the Links grid. Click the EDIT button, then you get EDIT ON and the EDIT mode is activated. Click on EDIT again to deactivate. When the EDIT mode is ON, the selected network is modified under operations like Delete, Create, Reverse, etc... When the EDIT mode is OFF, a new network is created when these operations are performed.

In the Links grid, select a species name in row, select a species name in column. Type 1 or 0 in the corresponding (0,1)-matrix cell. Type RETURN. The network now has the corresponding link if the value was 0 and is now 1, and has the link deleted if the value was 1 and is now 0. If a value distinct from 0 or 1 is entered, this affects the weighted matrix W (click on Matrix in the upper left corner to access to Val where the matrix W is displayed). In this case, the network is considered to be weighted. However, the (0,1)-matrix A still describes the presence/absence of a link.

The Import/Export command

The Import/Export command allows to compute and store the descriptor values of a set of networks that are in the same directory and have the same n_w format (nw0, nw1, gml, txt)). The Networks grid in the main window been selected, click the Import/Export button to open the IMPORT/EXPORT window. Click the IMPORT button. Open the directory where are the files you want to process. Select the file format (nw0 is the default). Select any file of the desired format, say the nw0 format, in the directory and click Open in the Open window. The n_w program opens all *.nw0 files in the chosen directory and creates as many networks. These created networks appear in the Networks grid of the main window.

To export the descriptor values of these networks, select the desired descriptors in the IMPORT/EXPORT window where all availablee descriptors are shown. Click EXPORT. A Save as window opens. Type the name of the output text file. The n_w program computes the desired descriptors and saves their values, separated by horizontal tabs, in the output file. The output file can easily be processed using Excel.

The command EXPORT can be used at any time to export the descriptor values of the set of networks that are currently present in the Network grid of the main window.

Commands summary

BUTTON\GRID	Networks grid	Species grid	Links grid
Create	Create synthetic network	Create new species	
Delete	Delete selected network	Delete selected species and associated links	
Duplicate	Duplicate selected network	Duplicate selected species and associated links	
Simul	Run Monte Carlo simulation of selected network		
Nb simul	Number of networks generated in the simulation		
Reverse	Reverse all links, i.e., prey become predators and vice versa		
Del Loops	Delete loops in the selected network		
Del Cycles	Delete cycles in the selected network		
Sort+		Sort species in increasing order of the parameter of the selected column	
Sort-		Sort species in decreasing order of the parameter of the selected column	
MaxConn	Extract maximal connected component of selected network		
Graphics	Graphical representation of selected network	Graphical representation of species distributions according to parameter of the selected column	Graphical representation of selected network
Info	Display descriptors of selected network and species		
SaveNetW	Save selected network to file		
SaveGrid	Save to file data in Networks grid	Save to file data in Species grid	Save to file data in Links grid
EDIT		Edit species names	Edit network matrix
Import/Export	Import a set of n_w files in a selected directory. Export their descriptor values to text file		
Aggreg	Run network aggregation methods (biological info, trophic similarity)		
Groups	Run group detection methods (modules, AIC groups, trophic groups)		

3. GRAPHICS

Under the Networks grid of the main window, click the Graphics button to get a graphical representation of the selected network. In the GRAPHICS window, the food web is represented according to the trophic heights of the species when this information is available (Fig. 1). Clicking in the Height cell of the status bar at the bottom of the GRAPHICS window allows to circle through various representations of the food web (Fig. 2).

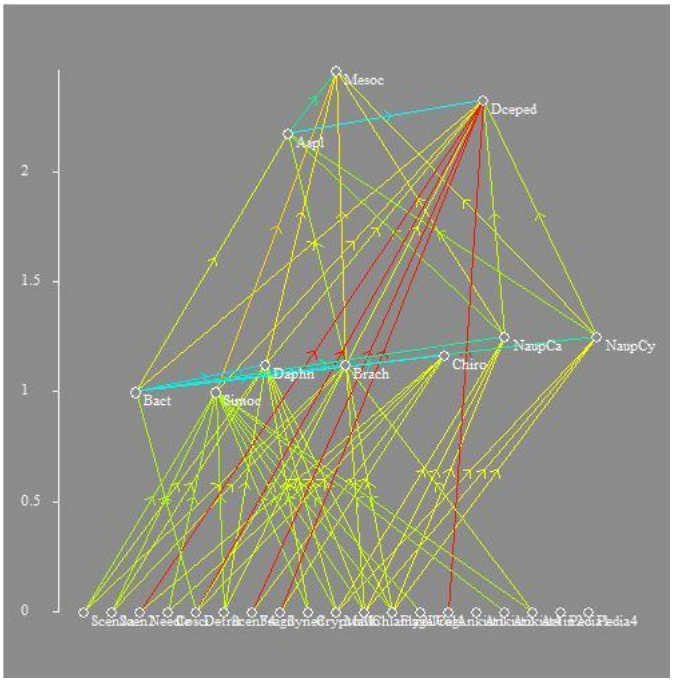


Figure 1. Representation of the species of the Example food web according to trophic height. Basal species (height = 0) at the bottom, intermediate species with height in the range 1-1.2, an intermediate species at height 2.2 (Aspl), and two top species (Dceped, height = 2.2, and Mesoc, height = 2.3).

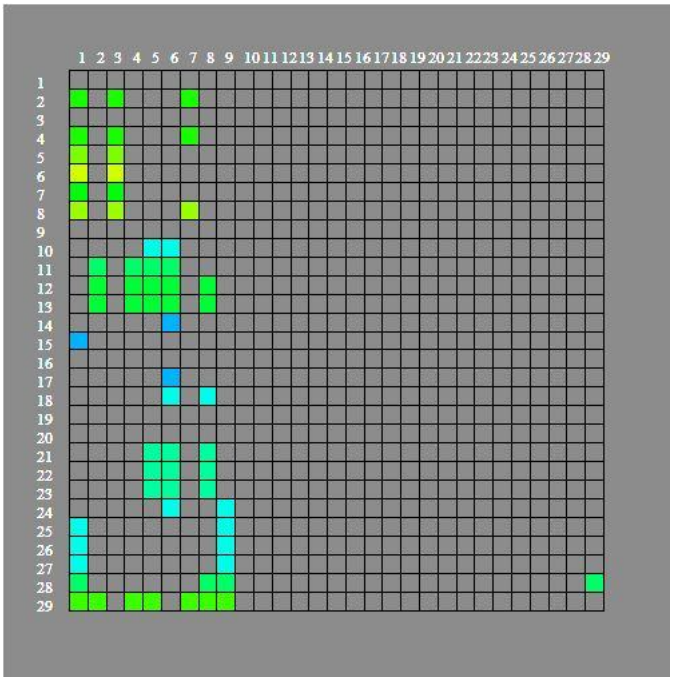


Figure 2. Presence-absence (0,1)-matrix associated with the 29 species of the Example food web.

There are several graphics windows (10 in all), allowing to display representation of networks simultaneously. The windows are labelled GRAPHICS<1>, GRAPHICS<2>, ...

Under the Species grid of the main window, select a descriptor column, for example Deg which gives the number of prey and predators. Click the Graphics button to get a graphical representation of the selected descriptor distribution (Fig. 3).

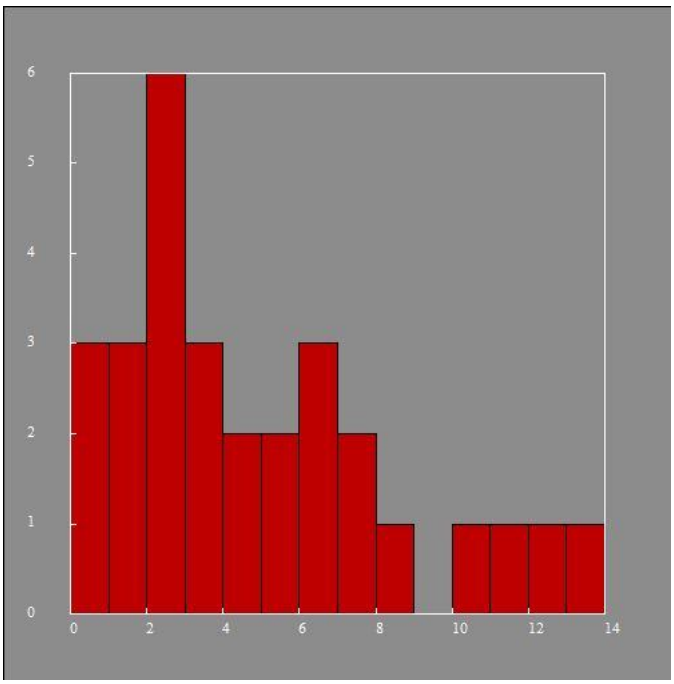


Figure 3. Degree distribution of the Example food web. There are 6 species of degree 2, and 3 isolated species (degree 0).

- Click on the Clear option to clear graphics.
- Graphics can be parameterized using the Settings option which opens the GRAPHICS SETTINGS window.
- Graphics can be saved in a bitmap file *.bmp using the Save option. The bitmap format *.bmp can easily be converted into the *.jpg format which is much lighter in terms of size (e.g., using the Paint program).

4. SYNTHETIC NETWORKS

The Networks grid been selected, click the Create button in the main window to open the CREATE window allowing to construct predefined types of networks. This feature is provided mainly for exploratory purpose.

At the top left of the CREATE window the number S of species in the network to create can be edited (the default is 30, otherwise it is the number of specie of the network currently selected).

Note on the random generator seed

The random generator seed (default 1) can be modified in the edit panel at the upper right of the CREATE window. This allows for different trajectories of the random generator, hence for different values of the random functions called by the program. Changing the seed provides different realizations of the random processes. It does not in any case changes the statistical significance or characteristics of the results.

In the CREATE window, the networks described below can be created using the corresponding radio buttons.

4.1. Random networks

- Degree Model — parameter <selected network> → A random network with the same number of nodes and the same node degrees as the selected network. The index # of the selected network is specified next to the radio button. The created network is obtained by performing a random walk in the set of (0,1)-matrices representing all networks with the same node degrees. Each step in the walk consists in selecting 2 random row indices i_1, i_2 and 2 random column indices j_1, j_2 such that the corresponding submatrix as the form $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ or $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and swapping the 0's and 1's in the submatrix. This procedure changes the links without changing the node degrees.
- BIT Model (Basal-Intermediate-Top model) — parameter <selected network> → A random food web with the same connectances in the basal, intermediate and top compartments as the selected network, whose index # is specified next to the radio button. More precisely, the matrix of the food web is decomposed into 9 blocks BxB, BxI, BxT, IxB, IxI, IxT, TxB, TxI, TxT corresponding to the B = Basal, I = Intermediate, T = Top species. The matrix of the created network is constructed so that the 4 blocks BxI, BxT, IxI, IxT have in average the same number of 1's as the original matrix. The other 5 blocks are 0, by definition, and are left unchanged. In particular, the created food web has the same number of basal, intermediate, and top species as the original one (possibly a very slight difference).
- Niche Model — parameters <number of species S, connectance C = L/S^2> → A random network aiming at mimicking a food web by using one dimensional niches (Williams & Martinez 2000).

- Random Network — parameters <number of species S, probability p> → A Erdös-Rényi random graph in which any 2 nodes are connected with constant probability p. Thus, the average number of links is $L = CS^2$ where $p = C$ is the connectance. The degree distribution is binomial with parameter p.
- Random Uniform Network — parameters <number of species S, number of links L>. Another Erdös-Rényi random graph where the graph is drawn uniformly among all graphs with the same number S of nodes and L of links (Batagelj & Brandes 2005).
- Random SmallWorld Network — parameters <number of species S, probability p, mean degree d> → A random 'small world' network constructed according to the model of Watts & Strogatz (1998). In a small world network, the average distance between any 2 nodes (the characteristic path length) is short, scaling as ln(S), but the clustering coefficient is large. In a random graph, the characteristic path length also scales as ln(S), but the clustering coefficient is small. In a scale-free network, whose degree distribution follows a power law, and which is a particular small world network, the characteristic path length scales as ln(ln(S)). Many real world networks exhibit the small world property.
- Random SubNetwork — parameter <selected network> → A network whose nodes are a random proper subset of the set of species of the selected network. You need to specify the number S' of species of the subnetwork at the top of the CREATE window. The size S' of the subnetwork must be strictly less that the size S of the selected network.
- Random Tree — parameters <probability p, in-degree d, number of levels N> → A random tree with N levels and average in-degree p*d. There is only one top predator at the apex of the tree. When p = 1, the tree is not random. In this case, each non basal species has exactly d prey, and the number of species is $1 + d + d^2 + \ldots + d^{N-1}$. When p < 1, each non basal species has in average p*d prey, and at most d.

4.2. Other networks

- Void Network — parameter <number of species S> → A (non random) network containing S nodes without links. This is a way to construct a user-defined network, using the EDIT command and progressively adding links in the Links grid.
- Rooted — parameter <selected network> → When the selected network has basal species (species without preys), a special node called '_Root_' is added to the network with links describing the circulation of matter (Note: there is no dissipation). The root node has outgoing links pointing to all basal species (primary producers) and ingoing links are added from all species to the root node (see Entropy, Rank and CycTime in Appendix A). The matrix of the created network is irreducible in general. If the original network has S species and L links, the rooted network has S+1 nodes and L + B + S links, where B is the number of basal species in the original network. For a weighted network, the links to and from the Root are arbitrarily set to 1.

- Compose — parameters <2 selected networks> → A network that is the union of 2 networks of indices #a and #b (default a = b = # of selected network). The number of nodes in the created network is $S = S_a + S_b$, and the number of links is $L = L_a + L_b$. No links are created joining the component networks which are therefore disjoint, but links can be added between the two, using EDIT options. This is a way to construct new networks from old ones.

4.1. Random network simulation

Clicking the Simul button in the main window generates a user-defined number Nb simul (the default is 100) of realizations of the selected network, and computes average descriptor values on the set of networks generated. At the end of this Monte Carlo procedure, whose steps appear in the PAD window, the average descriptor values are displayed in the main window below the descriptor values of the selected network.

For an empirical (non random) network from an input file, the realizations are all identical, so that the average descriptor values produced by the simulation procedure and displayed below the network in the Networks grid match exactly the original values.

Examples

- Open the file Example.nw0. Click the Create button to open the CREATE window. In Nb of Species at the top of the window, replace 29 (the number of species of the food web) by 20. Select the radio button Random SubNetwork. Type OK in the CREATE window. A random subnetwork #1_Sub is created and selected. Now click the Simul button: 100 subnetworks of size 20 are generated. The average descriptor values of the random subnetworks are displayed.
- Select the Example network. Click the Create button. Select the BIT Model (Basal, Intermediate, Top model). Click OK. The random network #1_BIT is created and selected. Click the Simul button: 100 networks of size 29 that are realizations of the BIT model for the Example food web are generated.

Network	#	Species S	Links L	L/S	L/S^2	Conn	Basal	Inter	Top	Deg	Height	Hmax
Example	1	29	64	2.207	0.0761	4	19	7	3	4.414	1.488	4
#1_Sub	2	20	37	1.85	0.0925	3	13	5	2	3.7	1.254	3
>Simul 100		20	29.63	1.482	0.07408	4.2	13.44	4.34	2.22	2.963	1.402	3
*#1_BIT	3	29	62	2.138	0.07372	4	19	7	3	4.276	1.823	5
>Simul 100		29	61.58	2.123	0.07322	2.66	19.01	6.91	3.08	4.247	2.092	5.18

5. GROUP DETECTION

Species in ecological networks — more generally nodes in real world networks — can often be grouped according to some criterion, for example when they play similar functional roles. In a network, a module is a set of nodes that have more links between themselves than with other nodes. Modularity detection has been applied to networks pertaining to various domains (e.g., Newman 2006, Olesen & al 2007, Guimerà & al 2007). Modules then reflect functional units that are partially autonomous.

Species can be grouped according to other criteria, some of which are handled by the n_w program. Click the Groups button to run the group detection methods in the GROUPS window. The network can be partitioned into 3 types of groups — modules, AIC groups, trophic groups — that are described below. In all cases, a simulated annealing algorithm is used (Guimerà & Nunes Amaral 2005). The algorithm tries to find a global optimum (a peak of highest 'energy') while avoiding to be trapped in local optima. In fact, it is not guaranteed to find the optimal partition because group detection is a combinatorial problem that is computationally hard (NP), and for which an optimal solution can not been found in reasonable time. However, the partition found by the simulated annealing algorithm is close to optimal. for a large network, the completion of the algorithm can take a long time.

Group detection cannot be run for a network containing isolated species. This is easily remedied using the MaxConn button, which allows to extract the largest connected component of the selected network, hence discarding isolated species.

5.1. Modules

The partitioning of a network into modules is performed by optimizing a modularity index. The modularity index M(E) of a given partition E of the network into modules — a particular arrangement of the species into disjoint groups — is defined as the difference between the within-groups link density and its expectation in a random (Erdös-Rényi) graph:

$$M(E) = \sum_{s=1}^{|E|} \left(\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right).$$

Here |E| is the number of elements in the partition (the number of modules), l_s is the number of links between nodes in the module of index s, L is the total number of links, and d_s is the degree of the species belonging to module s. The parameter l_s/L is the fraction of links inside module s (within-group link density), and $(d_s/2L)^2$ is an approximation of this expected quantity by chance alone.

The n_w program finds the partition with the best modularity index M(E) using the simulated annealing algorithm.

The decomposition of a network into modules leads to a (0,1)-matrix whose diagonal blocks are denser than off-diagonal blocks (Fig. 4).

Example

- Click  in the main window to open the file Example.nw0. The network Example#1 has S = 29 species and L = 64 links.
- Click the MaxConn button. This extracts the largest connected component of the network and creates a new network Example#2 that has S = 26 species and L = 64 links.
- Click the Groups button to open the GROUPS window. Select the Modules radio button and click the Run button. Information about the simulated annealing procedure appears in the PAD window while the algorithm is running. At the end of the procedure, 3 modules are found, of respective sizes 8, 12, 6. The final energy is E = 0.281372 at step 176 of the algorithm.
- Click the Graphics button. The network is represented according to height in the GRAPHICS<2> window. Click the Settings option  to open the GRAPHICS SETTINGS<2> window. Select the Modules radio button in the Groups panel. Click OK. The GRAPHICS<2> window now displays to which module the species belong, indicated by the module index and a specific color (red green blue). Within-module links are of the specific module color.
- Notice that a new network has been created, called #2_mod (modules of network #2). This network represents the network of modules in the Example#2 network: each node corresponds to a module in the original network. Hence, the #2_mod network has 3 nodes. You may select this network and display it graphically using the Graphics button. The thickness of the links corresponds to the number of between-modules links in the original network.
- Select the Example#2 network. Select the Species grid. Select the Mod column, which displays the module index of the species (selection is materialized by an *). Click the Sort+ button. This sorts the species according to increasing module index, and creates a new network with the species sorted as requested.
- Select the Networks grid. You see the newly created network Example#4 selected. Click the Graphics button. Click the Settings option  to open the GRAPHICS SETTINGS<4> window, select Modules in the Groups panel, click OK. Click the height option in the status bar, at the bottom of the GRAPHICS<4> window. Repeat clicking to get to the Circle representation where you can appreciate the within-module links. Click again to get to the Matrix representation to appreciate the 3 diagonal blocks corresponding to modules in the network matrix (Fig. 4).

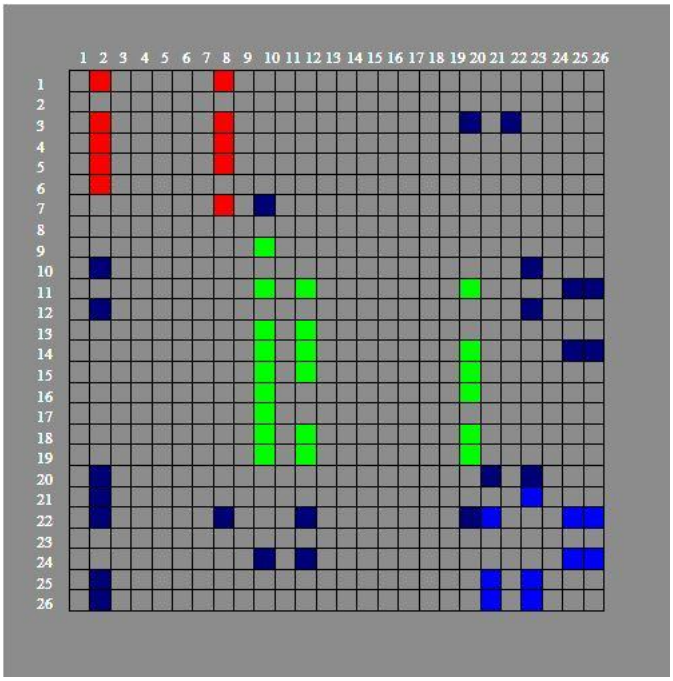


Figure 4. Module partitioning of the Example food web. The 3 modules correspond to red, green blue diagonal blocks in the food web matrix.

5.2. AIC groups

In this method proposed by Allesina & Pascual (2009a), the groups obtained are such that a best compromise is found between the reduction of the network into fewer 'species' — nodes representing groups when the network is partitioned into groups — and the information content of the network, as given by its structure.

Let us consider a given network N(S,L) with S nodes and L links. If the network is random (Erdös-Rényi), each edge is drawn with probability p according to the Bernoulli distribution. The probability to reproduce the network is

$$\mathbf{P}\left(N(S,L)\middle|p\right) = p^{L}(1-p)^{S^2-L}.$$

This probability can also be read as the likelihood of the parameter p given the network N(S,L):

$$\mathbf{L}\left(p\middle|N(S,L)\right) = p^{L}(1-p)^{S^2-L}.$$

This probability (or likelihood) is maximized when $p = L/S^2$ (the connectance).

When the network is partitioned into k groups, one considers the probability

$$\mathbf{P}\left(N(S,L)\middle|\vec{p}\right) = \prod_{i=1}^{k}\prod_{i=1}^{k} p_{ij}^{L_{ij}}(1-p_{ij})^{S_iS_j-L_{ij}}.$$

Here p_{ij} is the probability that a node in group i is joined to a node in group j, $\vec{p}$ is the vector containing all the probabilities p_{ij}, L_{ij} is the number of edges connecting nodes belonging to group i to nodes belonging to group j, and S_i is the number of nodes in group i. To maximize the probability P, it is sufficient to set $p_{ij} = L_{ij}/(S_iS_j)$ for all combinations of groups.

The AIC criterion (Akaike 1974) is now used to select the best model and partition a given network N(S,L) into groups accordingly. The parameters for each model are a vector $\vec{p}$ specifying the assignment of each species to a group, and k^2 probabilities, where k is the number of groups. The AIC for any partition is

$$AIC\left(N(S,L)\middle|\vec{p},k\right) = 2k^2 + 2S - 2\left(\sum_{i,j=1}^{k} L_{ij}\ln(p_{ij}) + (S_iS_j - L_{ij})\ln(1-p_{ij})\right).$$

Minimizing the AIC yields the group arrangement that maximizes the likelihood of reproducing the original network while keeping the number of groups, and therefore the number of free parameters, low.

An example of application of the method is given in Baskerville & al (2011).

5.3. Trophic groups

The notion of trophic group has been used by ecologists since a long time (Elton 1927), but has never been given a formal definition. A trophic group is defined as a set of species interacting with other well-defined sets of species, and trophic groups are usually determined by biological expertise.

In Gauzens & al (2015), a trophic group is defined using the notion of trophic similarity. The trophic similarity T(i,j) of two species i and j is their number of common prey and predators divided by their total number of prey and predators (Martinez 1991):

$$T(i,j) = \frac{\left|P_i \cap P_j\right| + \left|p_i \cap p_j\right|}{\left|P_i \cup P_j\right| + \left|p_i \cup p_j\right|}.$$

Here P_i and p_i represent respectively the set of predators and prey of species i, $\left|P_i \cap P_j\right|$ is the cardinal of the intersection of P_i and P_j, i.e., the number of predators common to species i and j, and $\left|P_i \cup P_j\right|$ is the total number of predators of species i and j.

Trophic groups detection is performed similarly to modules detection using a trophic similarity index which compares the observed trophic similarity between all pairs of species in the same group to the expected trophic similarity in a random network. For a given partition E of the network into trophic groups, the index is

$$G(E) = \sum_{s=1}^{|E|} \frac{1}{|g|} \sum_{\substack{i,j \in g \\ i<j}} \left(T(i,j) - T_r(i,j)\right).$$

Here $|g|$ is the number of nodes in group g, $|E|$ is the number of groups in the partition E, $T_r(i,j)$ is the trophic similarity of species i and j expected by chance (the trophic similarity of nodes i and j in a random network having S nodes.

5.4. Parameters of the simulated annealing algorithm

The parameters of the simulated annealing algorithm can be modified, but this not recommended:

- Initial temperature T_i (default 0.995). Increasing this parameter gives more opportunity to the algorithm for exploration. The system cools down later, avoiding being stuck on a non optimal peak.
- Final temperature T_f (default 1 E-13). This positive parameter is a halt criterion when the expected halt criterion is not met. Should not be changed.
- Cooling rate C (default 0.95). This multiplicative parameter, $0 < C < 1$, determines the rate of cooling according to temperature change $T' = CT$ at each phase of the algorithm: temperature T decreases exponentially from T_i down to T_f with rate C. The algorithm usually terminates before T_f is reached. Lowering C leads to a faster temperature decrease.
- Epsilon ε (default 1 E-10).

6. NETWORK AGGREGATION

The species represented by the nodes of the trophic network are the actors of the trophic interactions. However, these 'species' are not taxonomic species in general. Indeed, some nodes can represent large groups, like Bacteria, while some species can be represented by different nodes to account for different ontogenic stages where individuals do not feed on the same prey.

Furthermore, aggregating well-resolved food webs — lumping their nodes into fewer nodes — allows to simplify them and to focus on functional properties (Gauzens & al 2013).

The n_w program provides several methods to aggregate trophic networks, either from an auxiliary file describing expertise knowledge for lumping species into groups, or according to several criteria: trophic similarity, biological information (e.g., body size), or random. In the last 3 cases, hierarchical clustering is performed. At each step of the process, 2 nodes that are closest according to the criterion are lumped into a single node. The clustering is performed from the original un-aggregated network until only 1 node remains.

6.1. From file

This option is made for aggregating species according to biological expertise. The nodes to be aggregated are described in an auxiliary text file whose format is the following:

group_1
> species_1a
> species_1b
...
group_2
> species_2a
> species_2b
...

Group_i is the name of the aggregated node lumping species_ia, species_ib, etc... of the original network. As for all n_w files, items are separated by horizontal tabs.

The species listed in the auxiliary file should match exactly those of network to be aggregated.

Example

- Click  in the main window to open the file aggreg/Lazzaro_74.nw0 in the aggreg directory, a 74-nodes aquatic food web.
- Click the Aggreg button to open the AGGREGATION window.
- Click the From file radio button, then the Select file button to open the auxiliary file Laz74_agg17.txt.

- Click the Run button. This aggregates the Lazzaro_74 food web into the 17 groups described by the auxiliary file, creating the 17-nodes network #1_Agg_file (i.e., the network #1 has been aggregated according to file information).
- The 74-nodes original network is selected in the main window. Click the Info button to display information about the network. The 17 groups are listed at the bottom of the INFO window.
- Select the 17-nodes aggregated network in the main window. Click the Graphics button to display the aggregated network (Fig. 5).

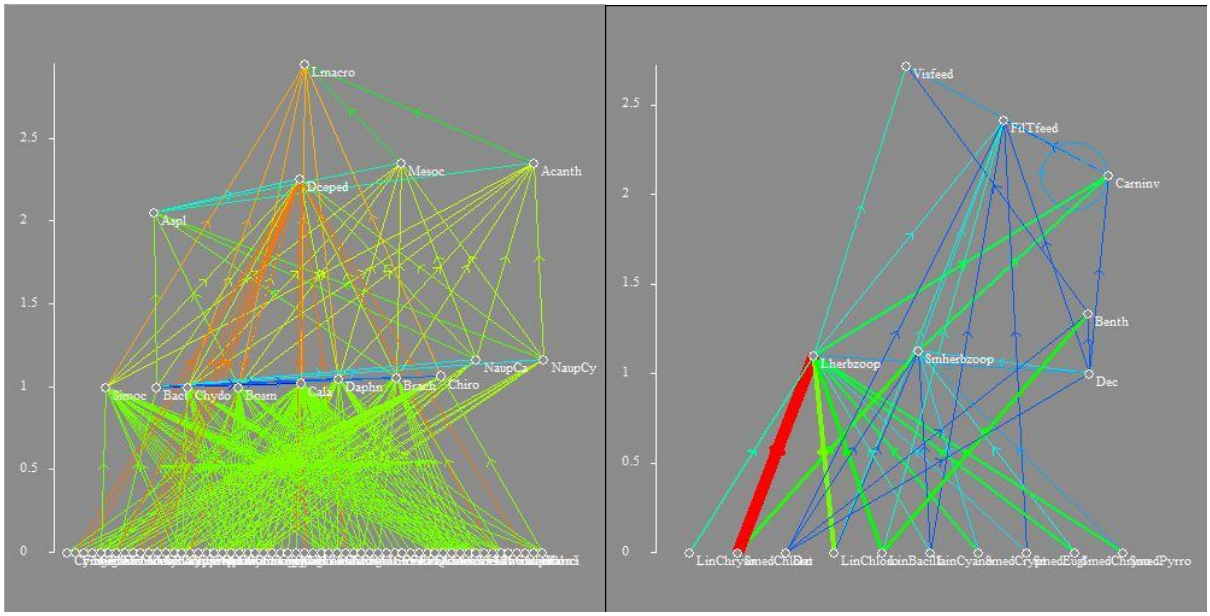


Figure 5. The Lazzaro_74 aquatic food web. Left: fully resolved 74-nodes network. Right: 17-nodes network, aggregated according to biological expertise; the thickness and color of the links correspond to the number of links from the original 74-nodes network that have been lumped by the aggregation process.

6.2. Trophic similarity

Under this criterion, species with the closest trophic similarity are lumped at each level of the aggregation process. The trophic similarity T(i,j) of two species (nodes) i and j is their number of common prey (in-degree) and predators (out-degree) divided by their total number of prey and predators (total degree) (see section 5.3).

At a given aggregation level, the trophic similarities T(i,j) of all pair of nodes (i,j) are computed. Two nodes a and b such that T(a,b) is maximal among all the T(i,j)'s are chosen as the nodes to be lumped at the next aggregation level. If the name of species a is say 'Namea' and the name of species b is 'Nameb', the name of the species lumped from a and b is 'Namea.Nameb'.

Example

- Click  in the main window to open the file aggreg/Lazzaro_74.nw0.
- Click the Aggreg button to open the AGGREGATION window.
- Click the Trophic similarity radio button.

- Click the Run button. This aggregates the Lazzaro_74 food web into groups according to trophic similarity from 74 nodes down to 1 node. Thus 73 networks are created, constituting the whole aggregation gradient. They appear in the main window, and all information about them is available. It is possible to restrict the aggregation gradient by specifying the lower bound for the number of nodes. For example if the lower bound is 60, the aggregation gradient will comprise all networks from 74 nodes down to 60 nodes.
- You may want to use the Import/Export command to export in a text file selected descriptor values of the networks along the aggregation gradient.

6.3. Biological criterion

This criterion of aggregation uses biological characteristics of the species in the network, like size, behavior... The biological characteristics are described in an auxiliary text file. The auxiliary file should reside in the same directory as the file describing the food web and bear the same name, but with the extension 'bio'.

For example, for the network file aggreg/Lazzaro_74.nw0, the file describing the biological characteristics of the species is called Lazzaro_74.bio. The format of the auxiliary file Lazzaro_74.bio begins:

```
species Dceped
size    14500000
behaviour       0
species Lmacro
size    750000000
behaviour       1
species NaupCy
size    203.2569389
behaviour       0
species NaupCa
size    280
behaviour       0
...
```

Here, for each species, there are 2 biological criteria, size and behaviour, and values associated with these criteria: body mass for size, and a variable taking values 0, 1, 2 for the behaviour criterion.

The *.bio file lists exactly all species of the corresponding network file, and associates a fixed number of criteria with these species, with empirical values for each criterion. Use horizontal tabs as separators. When present, the *.bio file is in fact compiled at the same time as the network file (just after).

Under the biological aggregation criterion, species with the smallest difference in biological characteristics are lumped at each level of the aggregation process. More precisely, at a given aggregation level, the Euclidian distances

$$D(i, j) = \sqrt{(B_i - B_j)^2}$$

are computed for each pair (i,j) of species. Here B_i and B_j are the biological characteristics of species (nodes) i and j, for a given criterion (e.g., size). Two nodes a and b such that D(a,b) is minimal among all the D(i,j)'s are chosen as the nodes to be lumped at the next aggregation level. When several criteria are used (e.g., size and behaviour), the distance is computed as

$$D(i, j) = \sqrt{\sum_k (B_{ik} - B_{jk})^2}$$

where the sum is over the different criteria k. When species a with characteristic B_a is lumped with species b with characteristic B_b, the lumped species has characteristic $(B_a + B_b)/2$. Note that all criteria are defined as real numbers, even if they code categories like behavior.

Example

- Click  in the main window to open the file aggreg/Lazzaro_74.nw0.
- Click the Aggreg button to open the AGGREGATION window.
- Click the Biological criteria radio button, then the Select criteria button to open a small window listing the criteria (described in the Lazzaro_74.bio text file), here size and behaviour.
- Select both criteria. Click OK.
- Click the Run button. This aggregates the Lazzaro_74 food web into groups according to both size proximity and behaviour proximity, from 74 nodes down to 1 node.

6.2. Random

Under this criterion, randomly chosen species are aggregated at each level of the aggregation process. This can be considered a null model of species aggregation.

APPENDIX A — NETWORK DESCRIPTORS

A directed graph is a mathematical object constituted of a set of nodes or vertices and a set of arcs. Arcs are oriented edges, or arcs, joining nodes. In ecology, food webs are represented by directed graphs. In this case, nodes are called species and arcs representing trophic interactions are called links. A link joining a species to itself (cannibalistic species) is called a loop.

In a directed graph, a (directed) path is a sequence of arcs joining successive nodes. The path length is the number of arcs in the path. A closed path, a path whose end node coincide with the starting node, is called a circuit or a cycle. The cycle length is the number of arcs in the cycle. We mainly consider simple cycles where except for the start node which is identical to the end node, nodes appear only once in the closed path.

A directed acyclic graph or DAG is a directed graph containing no cycle. Food webs can often be represented by DAGs.

For a given directed graph, one can consider the underlying undirected graph: its edges are the arcs of the directed graph with the orientation disregarded. In the undirected graph, the length d(i,j) of the shortest (undirected) path between any two nodes i and j defines a distance.

The main network descriptors are listed below.

S — Number of nodes, or species.

L — Number of arcs, or links.

L/S — Linking intensity.

L/S^2 — Connectance.

Number of links divided by the total number S^2 of possible links (corresponding to all entries of the (0,1)-matrix representing the directed graph set to 1). When cannibalistic species are excluded, the connectance is computed as L/S(S-1). Connectance represents the probability of 2 randomly chosen nodes in the network been connected. In a food web, the number L of links tend to scale allometrically with the number S of species, $L \propto S^a$. It has been hypothesized that food webs have constant exponent a equal to 2 (constant connectance hypothesis, Dunne 2006, Riede & al 2010).

Conn — Number of connected components in the underlying undirected graph.

In a connected undirected graph, there is an undirected path from any node to any other node. Otherwise, the graph can be partitioned into a finite number of connected components that are disjoint.

Basal — Number of basal species = number of species having no prey.

Inter — Number of intermediate species = number of species having prey and predators.

Top — Number of top species = number of species having no predator and ta least one prey.

Deg — Degree of a node, average degree of a network.

The degree of a node is the number of ingoing arcs plus the number of outgoing arcs. In a food web, the degree of a species is its number of prey plus its number of predators. For the whole network, the average degree is the average number of arcs entering and leaving the nodes, i.e., the average number of prey and predator per species for a food web.

In — In-degree d^{in} of a node.

The number of ingoing arcs, i.e., the number of prey of a species.

Out — Out-degree d^{out} of a node.

The number of outgoing arcs, i.e., the number of predators of a species.

In a directed graph with S nodes we have the degree-sum formula:

$$\sum_{i=1}^{S} d^{in}(i) = \sum_{i=1}^{S} d^{out}(i) = L .$$

Height — Trophic height of a species, of a food web.

Basal species have height 0. The height H(j) of a non basal species j is the average path length of all directed paths leading from all basal species to this species. The height of a given species can be thought of as the average energy reaching the species from primary producers in the food web. The height of the whole food web is the average of the heights of its constituent non basal species.

Hmax — Maximum height of a species, of a food web.

The maximum height of a non basal species is the maximum path length from basal species to this species. The maximum height of the whole food web is the maximum of the maximum heights of its non basal constituent species. It is the length of the longest path in the directed network.

Hmin — Minimum height of a species.

The minimum height of a non basal species is the minimum path length from basal species to this species.

TroLev — Trophic level of a species, of the food web.

This descriptor plays a role similar to height. Basal species have trophic level 1. The trophic level TL(j) of a non basal species j is defined as 1 plus the average of the trophic levels TL(i) of its prey:

$$TL(j) = 1 + \frac{1}{N_j} \sum_{i \in In(j)} TL(i) ,$$

where $N_j = |In(j)|$ is the number of prey of species j (the in-degree). The trophic level of the whole food web is the average of the trophic levels of its constituent species.

TLmax — Maximum trophic level of the food web.

Maximum value of the trophic levels of the constituent species of the food web.

Pos — Trophic position of a species.

A gross classification of species into B = Basal (no prey), I = Intermediate (having prey and predators) and T = Top (no predator, at least one prey), reflecting the classical plant $\rightarrow$ herbivore $\rightarrow$ carnivore tri-trophic food chain.

OI — Omnivory index of a species, of the food web.

A species is said omnivorous when it feeds at different trophic heights. The omnivory index of a non basal species is the standard deviation of the trophic heights of its prey. The omnivory index of the whole food web is the average of the omnivory indices of all non basal species. A related definition of the omnivory index uses trophic levels instead of trophic heights.

PathLen — Average path length.

Average of path lengths of all (directed) paths in the network.

Loops — Number of loops.

Number of links joining a species to itself (self-loops). In a food web, number of cannibalistic species.

Cycles — Number of (simple) cycles.

The n_w program does not compute the total number of cycles (which could be very large in some networks) but gives a lower bound. Food webs usually have few or no cycles.

CycLen — Average (simple) cycle length.

The n_w program computes an estimation of the average length of simple cycles, based on the subset of the set of simple cycles that have been detected.

Eccen — Eccentricity of a node, of the undirected graph.

For any focal node in the undirected graph, consider the set of shortest (undirected) paths from that node to any other node. The eccentricity of the focal node is the maximum length of these shortest paths. The eccentricity of the whole graph is the average of the eccentricities of its constituent nodes.

Diameter — Diameter of the undirected graph.

The diameter of the undirected graph is the maximal eccentricity of its constituent nodes.

Radius — Radius of the undirected graph.

The radius of the undirected graph is the minimal eccentricity of its constituent nodes. The centre of the graph is the set of nodes whose eccentricity is equal to the radius. This notion is used by the n_w program to provide a graphical representation of the network (see Section 4).

CharLen — Characteristic path length.

Average of lengths of shortest (undirected) paths between any two nodes in the network.

Clust — Clustering coefficient of a node, of the graph.

For a given node i, we look for all the nodes that can be reached from i by a directed link, i.e., that are predators of i. The number q_i of links joining these k_i neighbour nodes is determined. Then the clustering coefficient of node i is computed as

$$C(i) = \frac{q_i}{k_i^2}.$$

Here k_i^2 is the maximum number of edges joining the k_i neighbour nodes of i that could be realized, taking self-loops into account. The clustering coefficient of the whole graph is the average of the clustering coefficients of its constituent nodes.

Assort — Network assortativity.

Computed as $r = \frac{A - C}{B - C}$ where

$$A = \frac{1}{L}\sum_{(i,j)\in\Lambda} d_i d_j \ , \ B = \frac{1}{2L}\sum_{(i,j)\in\Lambda}\left(d_i^2 + d_j^2\right), \ C = \left[\frac{1}{2L}\sum_{(i,j)\in\Lambda}\left(d_i + d_j\right)\right]^2 ,$$

Λ is the set of links (i.e., sums are performed over nodes i, j that are connected), and d_i is the total degree of node i (Newman 2002). Network assortativity r lies in the range [-1,1], measuring a form of correlation between the degrees of the nodes, hence their propensity to attach to nodes of similar degree.

Prim

Index of primitivity of the network.

A directed graph is irreducible when there exists a directed path from any node to any other (the graph is strongly connected). An irreducible network necessarily has a cycle through any node. Indeed, consider a node i. The node i has a path to any other node j, and there is a path from j to i, hence a cycle through i. A food web is not irreducible because it has basal species, and no path leads to a basal species. A reducible network has primitivity index 0. When a network is irreducible it may be periodic (a rare case in fact) or aperiodic. Aperiodicity means that the cycle lengths of the network are relatively prime. The index of primitivity is the greatest common divisor of the cycle lengths, hence is 1 when the network is aperiodic and > 1 when it is periodic. A network that is irreducible and aperiodic is primitive. Primitivity ensures in particular that the (0,1)-matrix associated with the network has a dominant eigenvalue, a single real eigenvalue $\lambda > 0$ that is largest in modulus than all other eigenvalues. To summarize:

Prim = 0 $\rightarrow$ reducible (not primitive)

Prim = 1 $\rightarrow$ irreducible and aperiodic $\rightarrow$ primitive

Prim > 1 $\rightarrow$ irreducible but periodic $\rightarrow$ not primitive

Entropy

Network entropy.

Network entropy is a measure of network complexity. More precisely, entropy measures the diversity of pathways of energy flow between species in the network. The computation of the entropy H of an irreducible directed graph is described in Demetrius and Manke (2005). In fact, for a (0,1)-matrix, $H = \ln(\lambda)$ where λ is the dominant eigenvalue of A. To make a food web irreducible, we use the method proposed by Allesina & Pascual (2009b) that accounts for the circulation of energy in the network: a root node is attached to the food web with outgoing links pointing to all basal species (primary producers), and ingoing links are added from all species to the root node. Entropy is also a quantitative measure of network robustness (Demetrius and Manke 2005), that is the capacity of the network to remain functional in the face of random perturbations.

When the network has no basal species (is not a food web), entropy is computed directly from the the (0,1)-matrix of the network, which is checked for irreducibility. For a weighted network, a 'weighted entropy' is also computed from the quantitative matrix W using the same procedure. In this case, the added links emanating from the root node and going to it are arbitrarily weighted by 1.

For a general directed network, even if it has 'basal' nodes (nodes without in-links), the rooted network is not always irreducible. This is however most always the case for a food web.

ScalEnt Scaled network entropy.

Network entropy H increases with the number S of constituent nodes. Maximal entropy occurs when all matrix entries are 1, in which case H = ln(S). To disentangle complexity coming from an increase in S and from the very structure of the network, network complexity can be measured using scaled entropy, $H^* = \frac{\ln(\lambda)}{\ln(S)}$. Thus, scaled entropy measures complexity independently of the number of nodes.

Rank Importance rank of a species.

This notion has been introduced by Allesina & Pascual (2009b) by analogy with the PageRank algorithm of the Google search engine. A species is important if it points to (has links toward) many other species further up in the food web. Indeed, its extinction can lead to many co-extinctions. As for PageRank, the importance rank is computed by finding the right eigenvector of a conveniently constructed Markov chain matrix. The computation of the rank requires a primitive matrix. For a food web this is achieved as for the computation of entropy by adding a root node connected to all basal species and to which all species are connected. The resulting graph is irreducible. The n_w program further checks that the rooted graph so constructed is acyclic which ensures its primitivity. Then, in the rooted network matrix, each entry is normalized by the column sum, i.e., each entry a_{ij} is replaced by $\frac{a_{ij}}{\sum_j a_{ij}}$ (each arc (i,j) is weighted by the inverse of the in-degree of j). When the network has no basal species (is not a food web), the rank is computed directly from the normalized matrix, which is checked for primitivity.

CycTime Cycling time of a species.

As for the computation of entropy and rank, the food web is made irreducible by adding a root node, and the rooted graph so constructed is checked for primitivity. An irreducible Markov chain is associated with this network. The cycle time of species i is

the inverse of the probability π_i of first return to i where i is the state of the Markov chain associated with species i. When the network has no basal species (is not a food web), cycle times are computed directly from the Markov chain associated with the network matrix, which is checked for primitivity.

GenTime Generation time of the food web.

Computed as the cycle time of the root node when the food web is rooted as previously explained.

Weighted descriptors

Some network descriptors can be generalized to the case of a weighted network, where weights w_{ij} are associated with the links (i,j). A different approach than the one used here is described in Bersier (2002).

WGen Weighted generalism.

The generalism of a non basal species is the number of its prey, i.e., the in-degree. We define the weighted generalism of species i using the Simpson's index. Let

$$v_{ki} = \frac{w_{ki}}{\sum_{k=1}^{p} w_{ki}}$$

be the normalized in-flux to species i where p is the number of its prey. Then the weighted generalism of species i is computed as

$$\text{Gen(i)} = 1 - \sum_{k=1}^{p} v_{ki}^{2} .$$

WVul Weighted vulnerability.

The vulnerability of a non basal species is number of its predators (consumers) divided by the number of its prey, i.e., the out-degree divided by the in-degree. The weighted vulnerability of the non basal species i is

$$\frac{\sum_{j=1}^{P} w_{ij}}{\sum_{k=1}^{p} w_{ki}}$$

where P is the number of its predators, and p the number of its prey.

WHeight Weighted height of a species, of the food web.

The length of a path from species i to species j is the number l(i,j) of links to go from i to j following this path. The weighted height WH(j) of a given species j is computed in the same way as the height H(j):

$$WH(j) = \frac{\sum_{i \in B} \sum_{allpath(i,j)} l(i,j) W(i,j)}{\sum_{i \in B} \sum_{allpath(i,j)} W(i,j)} .$$

Here B is the set of basal species, and W(i,j) is the product of weights along the path (i,j) of length l = l(i,j):

$$W(i,j) = w_{ik_1} w_{ik_2} \ldots w_{ik_l} .$$

When the network is not weighted, i.e., the weights are 0/1, we recover the definition of ordinary height (W(i,j) is replaced by 1). The weighted height has the form of an expectation.

WOI — Weighted omnivory index of a species, of the food web.
Computed as the omnivory index, but using weighted heights.

WTroLev — Weighted trophic level of a species, of the food web.
Computed similarly to the unweighted case. Basal species have weighted trophic level 1. The weighted trophic level WTL(j) of a non basal species j is defined as 1 plus the weighted average of the weighted trophic levels TL(i) of its prey:

$$WTL(j) = 1 + \frac{1}{A_j} \sum_{i \in In(j)} w_{ij} \, WTL(i) ,$$

where $A_j = \sum_{i \in In(j)} w_{ij}$. The weighted trophic level of the whole food web is the average of the weighted trophic levels of its constituent species.

WTLmax — Maximum weighted trophic level of the food web.
Maximum value of the weighted trophic levels of the constituent species of the food web.

WEnt — Entropy of the weighted network. Computed similarly to the binary case.

WGenTime — Generation time of the weighted network. Computed similarly to the binary case.

WCycTime — Cycle time of the species of the weighted network. Computed similarly to the binary case.

APPENDIX B — EXAMPLE FILES

References are given as comments within the files.

Directory	File name	S	L	Description
n_w	Benguela.nw0	29	203	Aquatic food web
	BridgeBrookLake.nw0	75	553	Aquatic food web
	CarribeanReef.nw0	249	3313	Marine food web
	ChesapeakeBay.nw0	33	72	Marine food web
	Example.nw0	29	64	Example aquatic food web
	Creteil.nw0	67	64	Aquatic food web
	Lazzaro_74.nw0	74	229	Aquatic food web
	Lazzaro_small_Filter.nw0	6	9	Reduced aquatic food web
	Nepenthes.nw0	19	36	Food web in pitcher plant *Nepenthes*
	(...)			
n_w/aggreg	Lazzaro_74.bio			Biological information
	Laz74_aggreg17.txt			File for aggregation 74 → 17
	(...)			
n_w/import	day_14_BG_B10-4.nw2	26	27	Mesocosm aquatic food web
	day_14_BG_B10-9.nw2	23	29	Mesocosm aquatic food web
	(...)			
n_w/InteractionWeb	AkatoreA.nw0	85	227	Terrestrial food web
	AkatoreB.nw0	58	117	Terrestrial food web
	(...)			
n_w/misc	Celegans_metabolic.nw1	453	2040	Metabolic network of *C. elegans*
	Celegans_neural.nw1	297	2345	Neural network of *C. elegans*
	Celegans_neural_nocyc.nw1	297	1424	Network *C. elegans* , cycles removed
	Drosophila.nw1	14	17	Life cycle of *Drosophila melanogaster*
	Prison.nw0	67	182	Social network
	USairlines.nw1	332	2126	Transportation network
	(...)			
n_w/weighted	Carpinteria.nw0	128	2290	Terrestrial food web + parasitic links
	ConeSpring.nw0	8	17	Terrestrial food web
	LovinkHoeve.nw0	19	57	Terrestrial food web
	(...)			
n_w/weighted/paj	Baydry.nw0	128	2137	Weighted aquatic food web
	Chesapeake.nw0	39	177	Weighted aquatic food web
	(...)			

REFERENCES

- Akaike H. 1974. A new look at the statistical model identification. IEEE Transaction on Automatic Control 19:716–723.
- Allesina S & M Pascual. 2009a. Food web models: a plea for groups. Ecology Letters 12:652–62. http://www.ncbi.nlm.nih.gov/pubmed/19453619
- Allesina S & M Pascual. 2009b. Googling food webs: can an eigenvector measure species' importance for coextinctions? PLoS Computational Biology 5:e1000494. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2725316&tool=pmcentrez&rendertype=abstract
- Baskerville EB, AP Dobson, T Bedford, S Allesina, TM Anderson & M Pascual. 2011. Spatial Guilds in the Serengeti Food Web Revealed by a Bayesian Group Model (L.A. Meyers, Ed.). PLoS Computational Biology 7 e1002321. http://dx.plos.org/10.1371/journal.pcbi.1002321
- Batagelj V & U Brandes. Efficient generation of large random networks. Physica Review E 71, 036113. DOI: 10.1103/PhysRevE.71.036113
- Bersier LF, C Banašek-Richter & MR Cattin. 2000. Quantitative descriptors of food-web matrices. Ecology 83:2394–2407.
- Demetrius L & T Manke. 2005. Robustness and network evolution – an entropic principle. Physica A 346:682-696.
- Dunne JA. 2006. The network structure of food webs. In Ecological networks: linking structure to dynamics in food webs (M Pascual & JA. Dunne eds), Oxford University Press), pp. 27–86.
- Dunne JA, RJ Williams & ND Martinez. 2002. Food-web structure and network theory: The role of connectance and size. Proceedings of the National Academy of Sciences USA 99:12917–12922. http://www.pnas.org/cgi/content/abstract/99/20/12917
- Elton C. 1927. Animal ecology. Prentice-Hall International, London. http://www.cabdirect.org/abstracts/19632204195.html
- Gauzens B, S Legendre, X Lazzaro & G Lacroix. 2013. Food-web aggregation, methodological and functional issues. Oikos 122:1606–1615.
- Gauzens B, E Thébault, G Lacroix & S Legendre. 2015. Trophic groups and modules: two levels of group detection in food webs. Interface Focus.
- Girvan M & MEJ Newman. 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences USA 99:7821–7826. http://www.pnas.org/cgi/content/abstract/99/12/7821
- Guimerà R & LA Nunes Amaral. 2005. Functional cartography of complex metabolic networks. Nature 433:895–900. http://dx.doi.org/10.1038/nature03288
- Guimerà R, M Sales-Pardo & L Amaral. 2007. Module identification in bipartite and directed networks. Physical Review E 76. http://arxiv.org/abs/physics/0701151

- Guimerà R, DB Stouffer, M Sales-Pardo, EA Leicht, MEJ Newman & L Nunes Amaral. 2010. Origin of compartmentalization in food webs. Ecology 91:2941–51. http://www.ncbi.nlm.nih.gov/pubmed/21058554
- Hulot FD, G Lacroix, F Lescher-Moutoué & M Loreau. 2000. Functional diversity governs ecosystem response to nutrient enrichment. Nature 405:340–344. http://www.ncbi.nlm.nih.gov/pubmed/10830961
- Lazzaro X, G Lacroix, B Gauzens, J Gignoux & S Legendre. 2009. Predator foraging behaviour drives food-web topological structure. Journal of Animal Ecology 78:1307–1317.
- Martinez ND. 1991. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. Ecological Monographs 61:367–392. http://www.jstor.org/stable/2937047
- Newman MEJ. 2002. Assortative mixing in networks. Physical Review Letter 89:208701. http://arxiv.org/abs/cond-mat/0205405/
- Newman MEJ. 2006. Modularity and community structure in networks. Proceedings of the National Academy of Sciences of the United States of America 103:8577–8582. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1482622&tool=pmcentrez&rendertype=abstract
- Olesen JM, J Bascompte, YL Dupont & P Jordano. 2007. The modularity of pollination networks. Proceedings of the National Academy of Sciences USA 104:19891–19896. http://www.pnas.org/cgi/content/abstract/104/50/19891
- Pimm SL. 2002. Food Webs. The University of Chicago Press.
- Pimm SL, JH Lawton & JE Cohen. 1991. Food web patterns and their consequences. Nature 350:669–674.
- Riede JO, BC Rall, C Banasek-Richter, SA Navarrete, EA Wieters, MC Emmerson, U Jacob & U Brose. 2010. Scaling of food-web properties with diversity and complexity across ecosystems. Advances in Ecological Research 42:139–170.
- Rooney N, MS McCann & JC Moore. 2008. A landscape theory for food web architecture. Ecology Letters 11:867–881.
- Stouffer DB, J Camacho, R Guimerà, CA Ng & LA Nunes Amaral. 2005. Quantitative patterns in the structure of model and empirical food webs. Ecology 86:1301–1311**.**
- Strogatz SH. 2001. Exploring complex networks. Nature 410:268–276.
- Thébault E & C Fontaine. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329:853–856. http://www.ncbi.nlm.nih.gov/pubmed/20705861
- Ulanowicz RE. Quantitative methods for ecological network analysis. Computational Biology and Chemistry 28:321–339.
- Williams RJ & ND Martinez. 2000. Simple rules yield complex food webs. Nature 404:180–183.
- Watts DJ & SH Strogatz. 1998. Collective dynamics of 'small-world' networks. Nature 393:440–442. http://www.ncbi.nlm.nih.gov/pubmed/9623998

TECHNICAL NOTICE

Computer	PC
System	Windows
Minimal memory required	2 G
Programming language	Object Pascal – Borland Delphi 6
Source code size	~ 16000 lines
Exec code size (nw.exe)	~ 840 K
	+ 4000 K runtime library qtintf.dll

PROGRAM BOUNDS

maximum number of networks	300
maximum number S of species per network	2000
maximum number of links	1000000

AUTHORS

Stéphane Legendre & Benoît Gauzens
Laboratoire d'écologie
Ecole Normale Supérieure
46 rue d'Ulm – 75005 Paris
Email: legendre@ens.fr
Web site: http://www.biologie.ens.fr/~legendre

ACKNOWLEDGEMENTS

InteractionWeb Data Base by Vasquez & al http://www.nceas.ucsb.edu/interactionweb, Pajek website http://vlado.fmf.uni-lj.si/pub/networks/data/bio/foodweb/foodweb.htm, and all persons who made network datasets available on Internet.
R. Guimerà kindly shared his simulated annealing algorithm with Elisa Thébault. Our algorithm is based on his via Elisa's implementation.