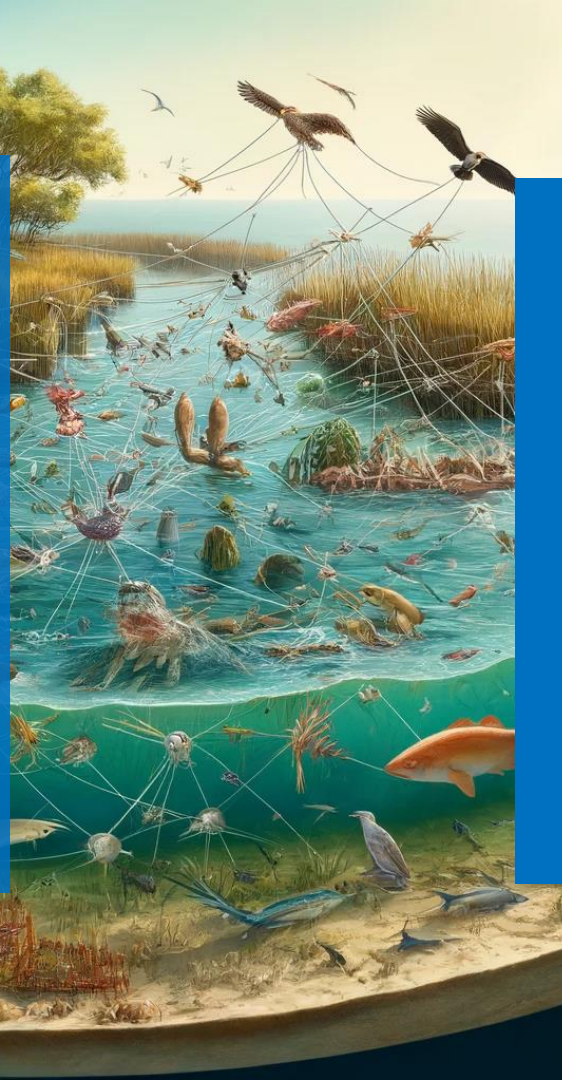# Analysis of food webs in Southern Florida

Project for the Web Mining Exam
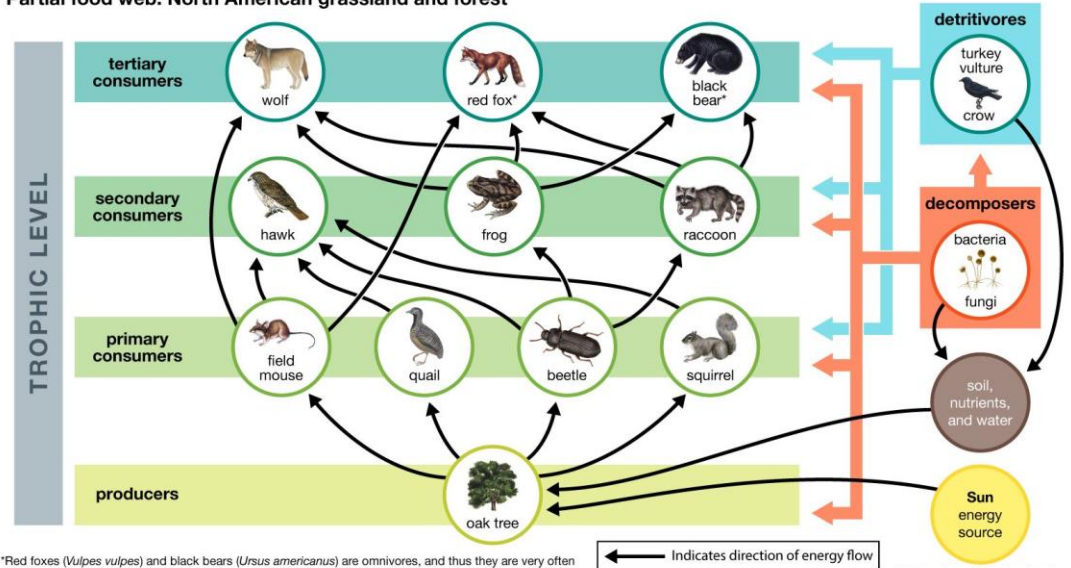By:
Alberto Biliotti
Jacopo Manetti

# What is a food web?

A food web is a type of directed network where each node represents a component of the ecosystem like an animal, a plant or a bacteria, while an edge represent  tranfers of material or energy between the components, mostly caused by predation or decomposition phenomena.

**Partial food web: North American grassland and forest**

TROPHIC LEVEL

| | |
|---|---|
| tertiary consumers | wolf, red fox*, black bear* |
| secondary consumers | hawk, frog, raccoon |
| primary consumers | field mouse, quail, beetle, squirrel |
| producers | oak tree |

detritivores: turkey vulture, crow

decomposers: bacteria, fungi

soil, nutrients, and water

Sun energy source

*Red foxes (*Vulpes vulpes*) and black bears (*Ursus americanus*) are omnivores, and thus they are very often considered to be secondary consumers. However, in this food web they function as tertiary consumers.

← Indicates direction of energy flow

© Encyclopædia Britannica, Inc.

# Networks Details

Our data consist of 4 different biomes:



Cypress wetlands

The Everglades graminoid marshes

The mangrove estuaries

The shallows of Florida Bay

For each biomes, there are 2 different networks: one for the dry season and one for the wet one.
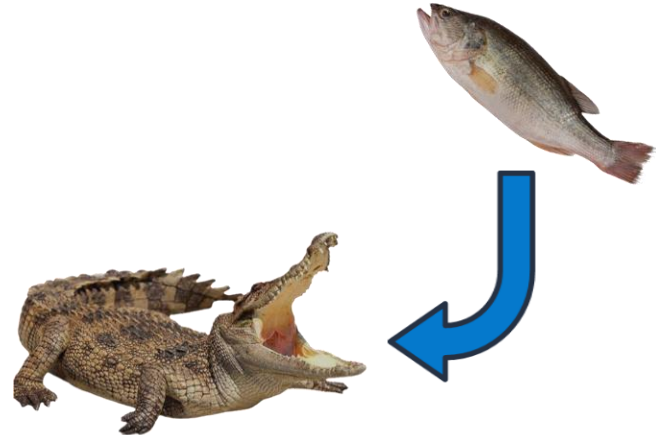
# Networks Details

## The Nodes

Each node has 3 attributes:

- Name of component
- Bio-masses: Indicates the quantity of organic matter present in each component of the ecosystem.
- Classification type:
  1. Living/producing compartment
  2. Other compartment
  3. Input
  4. Output
  5. Respiration

## The Edges

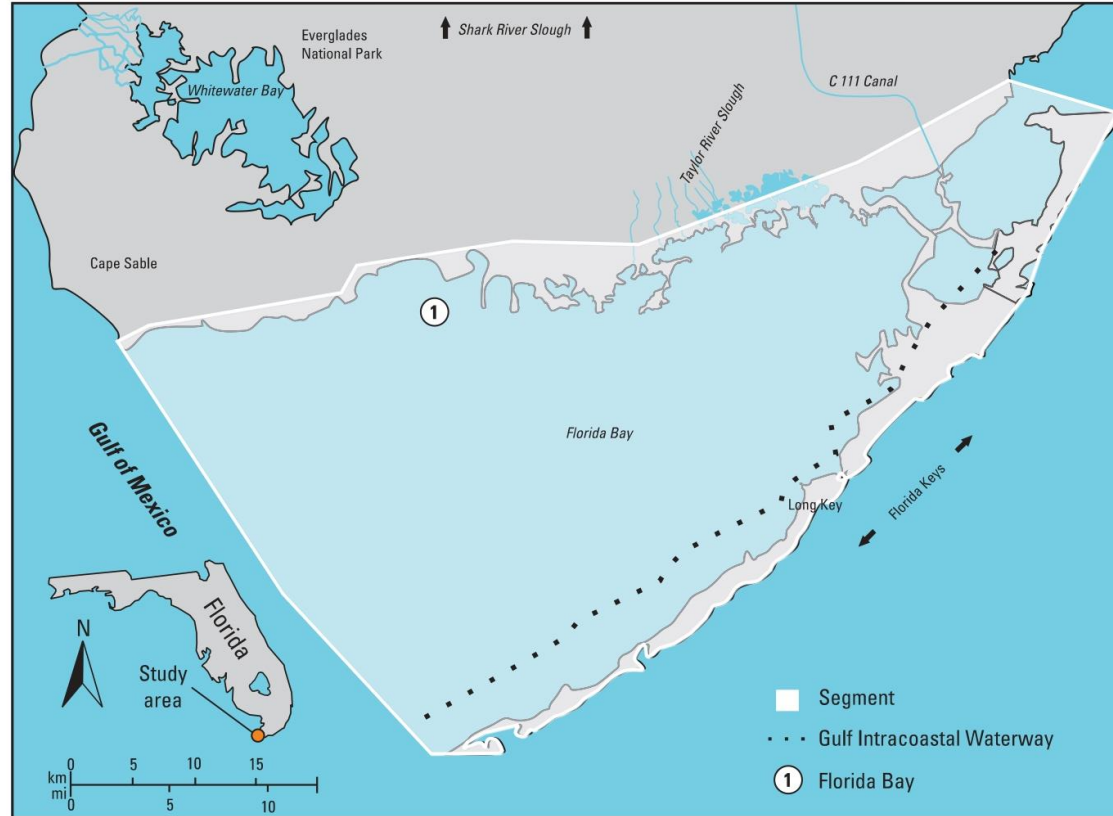There is an edge from $i$ to $j$ if compartment $j$ consumes carbon from compartment $i$

# Choose the networks to analyze

We have chosen to analyze the two networks with the higher number of nodes, which are the food webs regarding the Florida bay for both the wet and dry season.

Both the networks are composed of the same 128 nodes.

# Networks preprocessing

We have removed from the networks all the special nodes that refer to compartments which are not organism, because we want to analyze the relationship between the organism in term of food chain and not their overall contribution to the total input/output of carbon in the ecosystem.
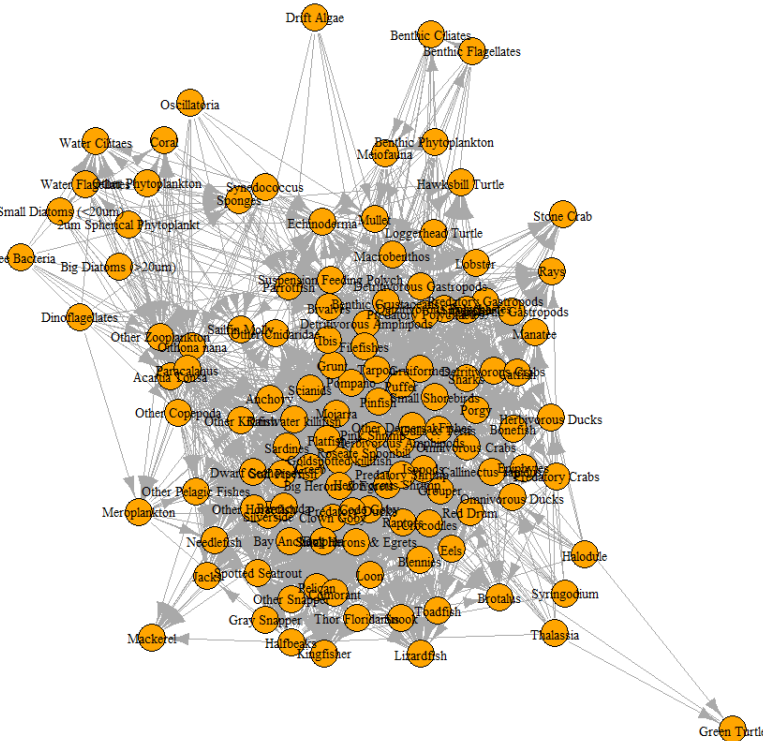
After this selection there are 121 nodes to analyze, the network regarding the dry season has 1799 edges, while the one for the wet season 1767.

Graph of the Florida Bay in the **wet** season

Graph of the Florida Bay in the **dry** season
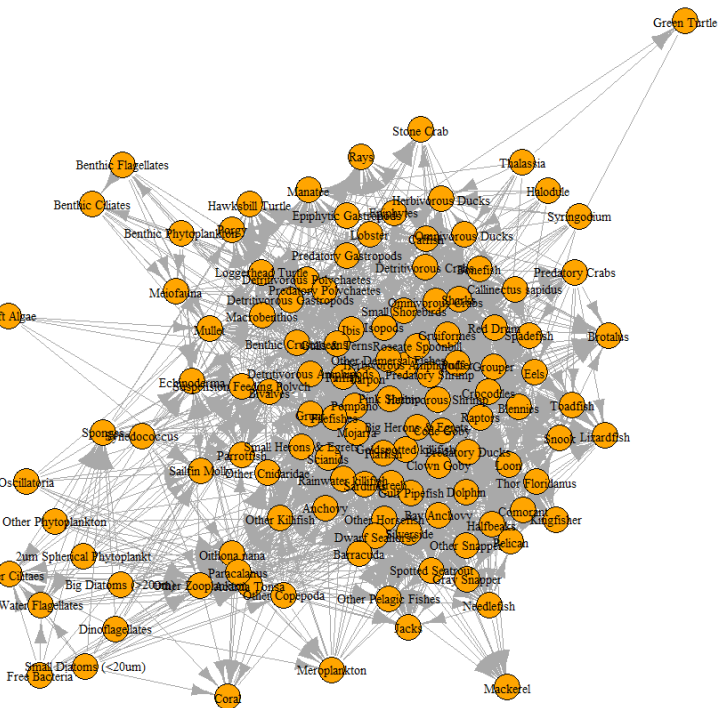
Giving the huge number of nodes and edges, it's hard to describe the networks by simply looking at their plots.

The only thing we can see is the presence of some apex predators like rays or mackerel and the presence of organism like bacteria or algae that don't consume any other organism

# Biomasses

```
       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
   0.000000   0.000192   0.002605   0.393831   0.020810  21.999550
```

```
 [1] "Thalassia"          "Epiphytes"          "Halodule"               "Drift Algae"
 [5] "Syringodium"        "Synedococcus"       "Meiofauna"              "Bivalves"
 [9] "Echinoderma"        "Benthic Flagellates" "Benthic Phytoplankton"  "Water Flagellates"
[13] "Sponges"            "Free Bacteria"      "Detritivorous Polychaetes" "Predatory Polychaetes"
[17] "Herbivorous Shrimp" "Isopods"            "2um Spherical Phytoplankt" "Pink Shrimp"
```

Here we can see that the great part of the organisms has a low biomass value, while just a few species have great biomasses values.

## Dry season

Here we can see that the things don't change much during the dry season, indeed we can only observe a little decrement in the overall biomass as expected

```
       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
   0.000000   0.000261   0.002486   0.398514   0.019020  22.998630
```

```
 [1] "Epiphytes"            "Thalassia"           "Drift Algae"          "Halodule"
 [5] "Syringodium"          "Meiofauna"           "Synedococcus"         "Bivalves"
 [9] "Echinoderma"          "Small Diatoms (<20um)" "Big Diatoms (>20um)"  "Benthic Phytoplankton"
[13] "Sponges"              "Free Bacteria"       "Benthic Flagellates"  "Detritivorous Polychaetes"
[17] "Predatory Polychaetes" "Herbivorous Shrimp" "Water Flagellates"    "Pink Shrimp"
```

As we can see, the species with the greater biomass values are mostly plants or algae, in fact the animals with the greatest biomasses are the meiofauna and the bivalves.

# Networks Statistics

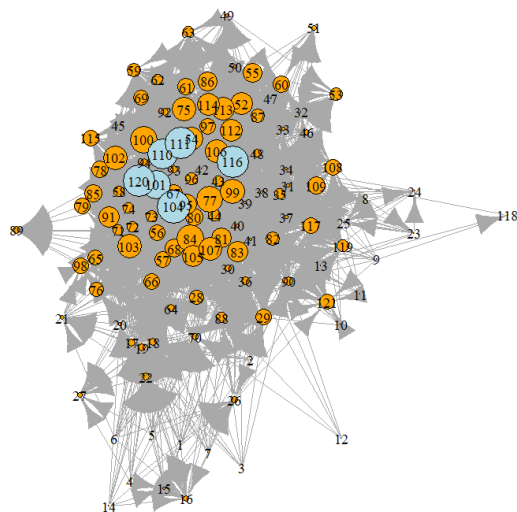| | Dry Season | Wet Season |
|---|---|---|
| Density | 0,12 | 0,12 |
| Reciprocity | 0,004 | 0,004 |
| Transitivity | 0,27 | 0,27 |
| Odds Ratio Transitivity | 2,63 | 2,62 |
| Biomass assortativity | 0,05 | 0,06 |

➢ We can see that the two networks have very similar statistics.

➢ The networks reciprocity is very low, as we would expect in a food chain.

➢ The transitivity instead is high, especially if compared with the density using the odds ratio. Possible presence of ecological niches.

➢ Finally, the biomass of the organisms do not seem to have a relevant effect on the food chains.

Wet season In-degree centrality

Dry season In-degree centrality
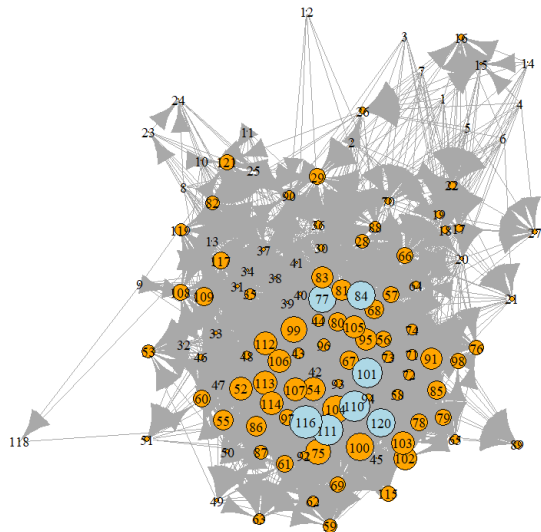
| [1] | "Crocodiles" | "Predatory Ducks" | "Raptors" | "Greeb" |
|---|---|---|---|---|
| [5] | "Scianids" | "Dolphin" | "Pompano" | "Loon" |
| [9] | "Big Herons & Egrets" | "Other Demersal Fishes" | "Grouper" | "Tarpon" |
| [13] | "Small Shorebirds" | "Sharks" | "Pelican" | "Comorant" |
| [17] | "Ibis" | "Roseate Spoonbill" | "Gruiformes" | "Gulls & Terns" |
| [21] | "Small Herons & Egrets" | "Pinfish" | "Barracuda" | "Flatfish" |

| [1] | "Big Herons & Egrets" | "Raptors" | "Crocodiles" | "Dolphin" |
|---|---|---|---|---|
| [5] | "Predatory Ducks" | "Greeb" | "Scianids" | "Loon" |
| [9] | "Pompano" | "Other Demersal Fishes" | "Grouper" | "Pelican" |
| [13] | "Comorant" | "Tarpon" | "Ibis" | "Roseate Spoonbill" |
| [17] | "Small Shorebirds" | "Sharks" | "Gruiformes" | "Gulls & Terns" |
| [21] | "Small Herons & Egrets" | "Pinfish" | "Barracuda" | "Flatfish" |

These nodes are typically top predators or charismatic species with a significant impact on the food web structure. The central species by in-degree remain similar across seasons, but their ranking changes slightly. For instance, "Big Herons & Egrets" drop from the first position in the dry season to the ninth in the wet season, suggesting seasonal variations in predation or prey availability.
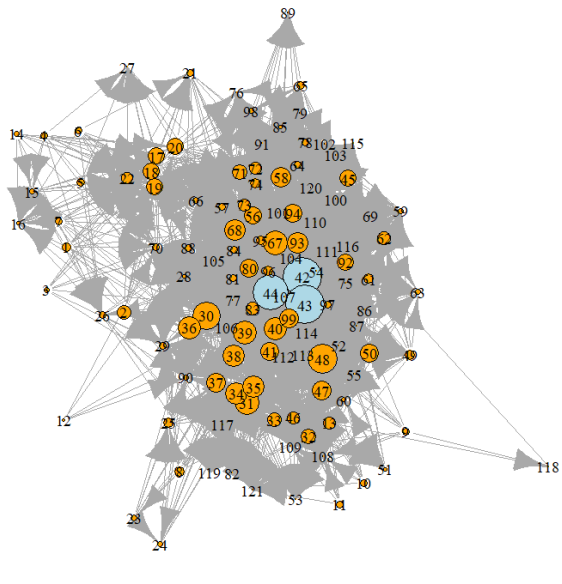
These nodes are often small invertebrates that are suitable for consumption by many other species, reflecting their importance as crucial prey in the food web. The species with the highest out-degree remain the same across both seasons, indicating that small invertebrates like shrimp and crabs maintain high vulnerability to predation year-round.
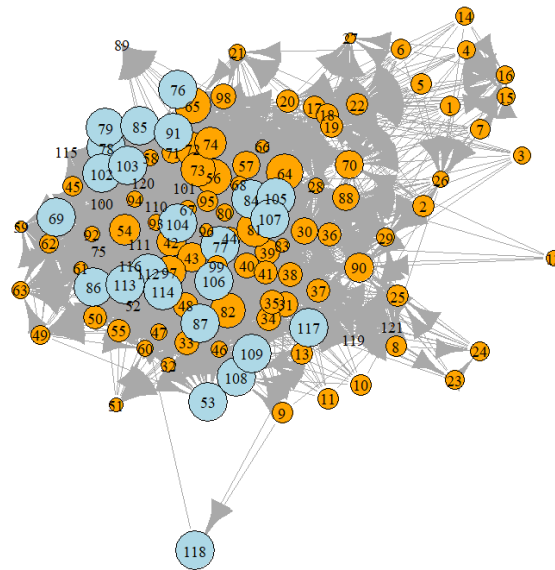
Wet season Closeness centrality

Dry season Closeness centrality

| [1] | "Rays" | "Snook" | "Jacks" | "Pompano" |
|---|---|---|---|---|
| [5] | "Other Snapper" | "Gray Snapper" | "Scianids" | "Spotted Seatrout" |
| [9] | "Red Drum" | "Spadefish" | "Barracuda" | "Pelican" |
| [13] | "Comorant" | "Big Herons & Egrets" | "Small Herons & Egrets" | "Ibis" |
| [17] | "Roseate Spoonbill" | "Herbivorous Ducks" | "Omnivorous Ducks" | "Gruiformes" |
| [21] | "Small Shorebirds" | "Gulls & Terns" | "Loggerhead Turtle" | "Green Turtle" |

| [1] | "Rays" | "Snook" | "Jacks" | "Pompano" |
|---|---|---|---|---|
| [5] | "Other Snapper" | "Gray Snapper" | "Scianids" | "Spotted Seatrout" |
| [9] | "Red Drum" | "Spadefish" | "Barracuda" | "Pelican" |
| [13] | "Comorant" | "Big Herons & Egrets" | "Small Herons & Egrets" | "Ibis" |
| [17] | "Roseate Spoonbill" | "Herbivorous Ducks" | "Omnivorous Ducks" | "Gruiformes" |
| [21] | "Small Shorebirds" | "Gulls & Terns" | "Loggerhead Turtle" | "Green Turtle" |

These nodes are often medium-sized predatory fish that play crucial roles in trophic dynamics, with rapid influence on many other species. The species with high closeness centrality do not vary between seasons, reflecting their stable and central position in the food web.
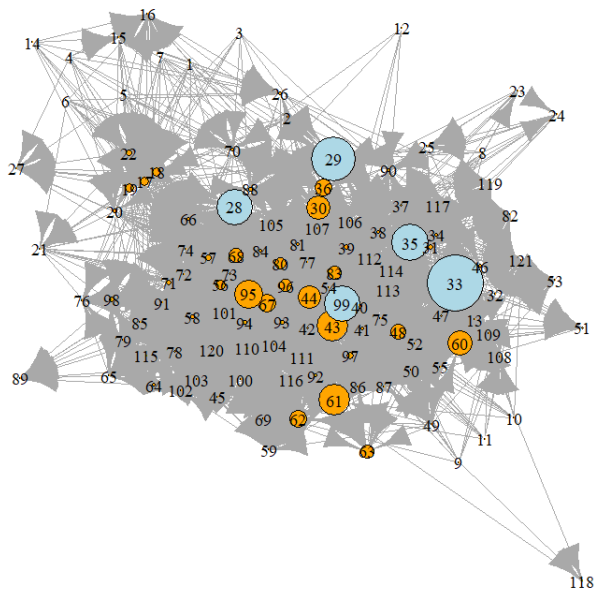
## Wet season Betweenness centrality

## Dry season Betweenness centrality

Wet season legend:
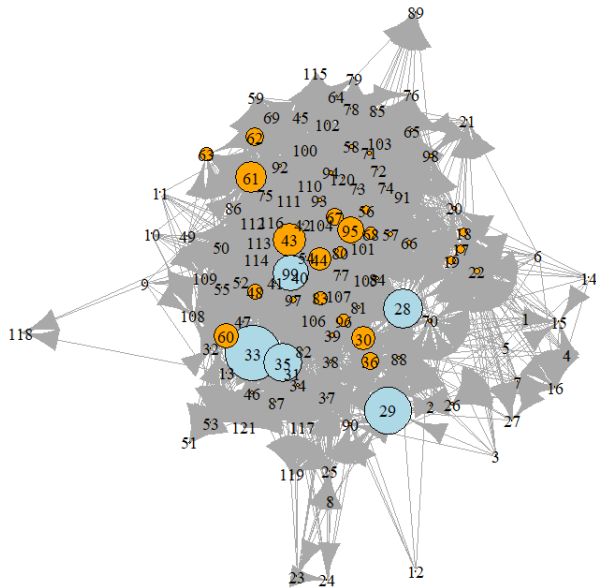
| | | | |
|---|---|---|---|
| [1] "Predatory Gastropods" | "Echinoderma" | "Predatory Polychaetes" | "Other Cnidaridae" |
| [5] "Other Demersal Fishes" | "Predatory Shrimp" | "Eels" | "Flatfish" |
| [9] "Catfish" | "Bivalves" | "Pink Shrimp" | "Goldspotted killifish" |
| [13] "Toadfish" | "Suspension Feeding Polych" | "Omnivorous Crabs" | "Rainwater killifish" |
| [17] "Pinfish" | "Brotalus" | "Filefishes" | "Mojarra" |
| [21] "Sardines" | "Acartia Tonsa" | "Oithona nana" | "Paracalanus" |

Dry season legend:

| | | | |
|---|---|---|---|
| [1] "Predatory Gastropods" | "Echinoderma" | "Other Cnidaridae" | "Predatory Polychaetes" |
| [5] "Other Demersal Fishes" | "Predatory Shrimp" | "Eels" | "Flatfish" |
| [9] "Catfish" | "Bivalves" | "Pink Shrimp" | "Toadfish" |
| [13] "Goldspotted killifish" | "Suspension Feeding Polych" | "Omnivorous Crabs" | "Pinfish" |
| [17] "Brotalus" | "Rainwater killifish" | "Filefishes" | "Mojarra" |
| [21] "Sardines" | "Acartia Tonsa" | "Oithona nana" | "Paracalanus" |

These nodes represent species that connect different trophic compartments, playing a key role in the stability and cohesion of the network. Key species by betweenness centrality remain constant across seasons, suggesting that their role as connectors within the food web is critical year-round.
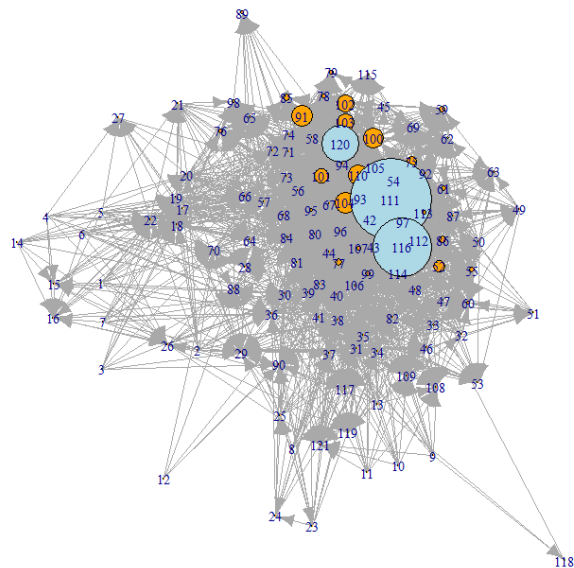
Wet season Eigenvector centrality

Dry season Eigenvector centrality
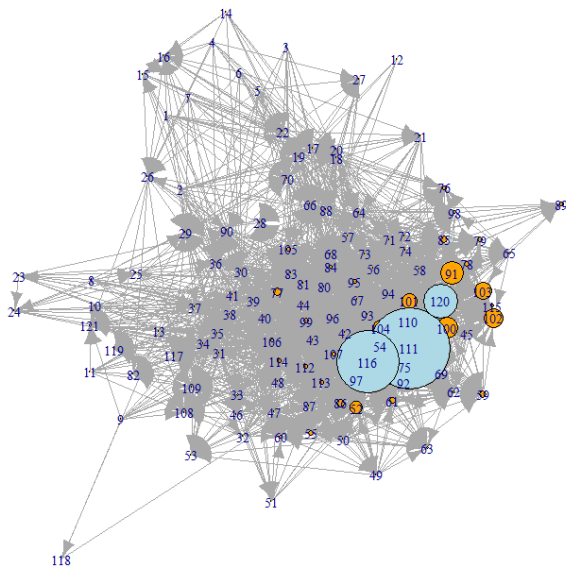
| [1] "Raptors" | "Crocodiles" | "Dolphin" | "Barracuda" |
|---|---|---|---|
| [5] "Loon" | "Predatory Ducks" | "Pelican" | "Comorant" |
| [9] "Big Herons & Egrets" | "Greeb" | "Tarpon" | "Sharks" |
| [13] "Grouper" | "Red Drum" | "Pompano" | "Spotted Seatrout" |
| [17] "Lizardfish" | "Eels" | "Bonefish" | "Other Demersal Fishes" |
| [21] "Other Snapper" | "Gray Snapper" | "Roseate Spoonbill" | "Small Herons & Egrets" |

| [1] "Raptors" | "Crocodiles" | "Dolphin" | "Big Herons & Egrets" |
|---|---|---|---|
| [5] "Barracuda" | "Loon" | "Predatory Ducks" | "Pelican" |
| [9] "Comorant" | "Greeb" | "Tarpon" | "Sharks" |
| [13] "Grouper" | "Red Drum" | "Pompano" | "Spotted Seatrout" |
| [17] "Lizardfish" | "Eels" | "Bonefish" | "Other Demersal Fishes" |
| [21] "Gray Snapper" | "Other Snapper" | "Roseate Spoonbill" | "Small Herons & Egrets" |

These nodes represent top predators with broad influence networks, often at the apex of food chains. Species with high eigenvector centrality are very similar across seasons, with slight variations in their ranking.
This reflects that the most influential species tend to maintain their status regardless of the season.

# Centrality Indices

|  | Dry Season | Wet Season |
|---|---|---|
| In-degree | 0,22 | 0,21 |
| Out-degree | 0,39 | 0,38 |
| Betweenness | 0,03 | 0,03 |

➢ We can not derive the Closeness Centrality index since some nodes are not reachable from other nodes, so the distance between them results to be infinite.

➢ Overall, we see that the networks are heavily centralized only if we consider Eigenvector centrality, while there is a little bit less centralization considering Out degree closeness.

➢ There seems to be no centralization at all if we consider In degree or Closeness centrality.

# Simple Random Graph Model

We start modelling our networks with the homogenous simple random graph model.

```
Call:
ergm(formula = rete ~ edges)

Maximum Likelihood Results:

       Estimate Std. Error MCMC % z value Pr(>|z|)
edges  -1.97648    0.02538      0  -77.87   <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 20129  on 14520  degrees of freedom
 Residual Deviance: 10753  on 14519  degrees of freedom

AIC: 10755  BIC: 10763  (Smaller is better. MC Std. Err. = 0)
```

```
Call:
ergm(formula = rete_d ~ edges)

Maximum Likelihood Results:

       Estimate Std. Error MCMC % z value Pr(>|z|)
edges  -1.95602    0.02519      0  -77.66   <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 20129  on 14520  degrees of freedom
 Residual Deviance: 10879  on 14519  degrees of freedom

AIC: 10881  BIC: 10889  (Smaller is better. MC Std. Err. = 0)
```
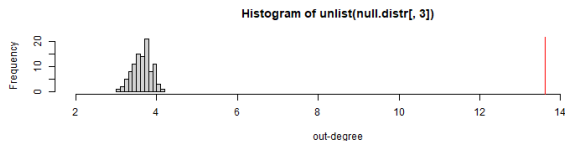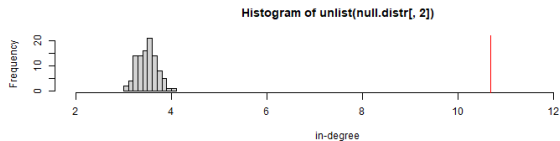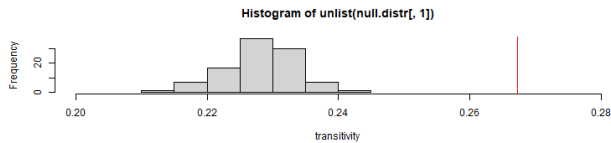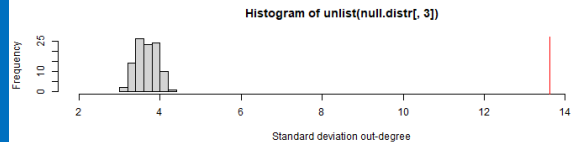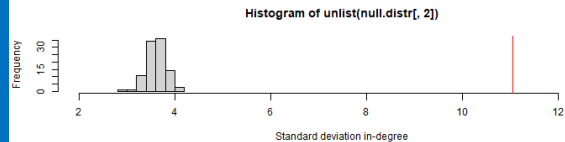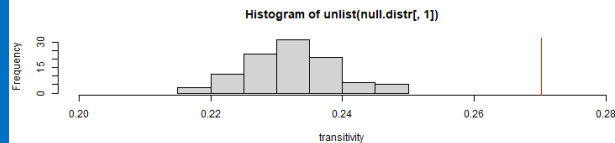
# Simple Random Graph Model

Obviously, this approach does not seem to achieve great results, as we can see from the distributions of the  simulated networks statistics, compared to our networks ones:

# SRG with biomass attribute

Now, we try add the biomass nodal attributes at the Simple Random Graph Model specification for our networks:

```
Call:
ergm(formula = rete_d ~ edges + nodecov("bio.masses...baydry.dat") +
    absdiff("bio.masses...baydry.dat"), control = control.ergm(seed = 1))

Maximum Likelihood Results:

                                  Estimate Std. Error MCMC % z value Pr(>|z|)
edges                             -1.92693    0.02584      0 -74.560   <1e-04 ***
nodecov.bio.masses...baydry.dat   -0.32072    0.34794      0  -0.922    0.357
absdiff.bio.masses...baydry.dat    0.27979    0.34944      0   0.801    0.423
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null Deviance: 20129  on 14520  degrees of freedom
 Residual Deviance: 10854  on 14517  degrees of freedom

AIC: 10860  BIC: 10883  (Smaller is better. MC Std. Err. = 0)
```

**Wet season**

```
Call:
ergm(formula = rete_w ~ edges + nodecov("bio.masses...baywet.dat") +
    absdiff("bio.masses...baywet.dat"), control = control.ergm(seed = 1))

Maximum Likelihood Results:

                                  Estimate Std. Error MCMC % z value Pr(>|z|)
edges                             -1.94299    0.02601      0 -74.709   <1e-04 ***
nodecov.bio.masses...baywet.dat   -0.28830    0.27146      0  -1.062    0.288
absdiff.bio.masses...baywet.dat    0.23819    0.27310      0   0.872    0.383
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null Deviance: 20129  on 14520  degrees of freedom
 Residual Deviance: 10722  on 14517  degrees of freedom

AIC: 10728  BIC: 10751  (Smaller is better. MC Std. Err. = 0)
```

We can see that the parameters for both networks relating to main effects and to homophily are non-significantly different from zero, so we can say that the biomasses of the organism has no significant effect on the network.

# Non-Homogeneous SRG

Now, we try add sender and receiver effect to the SRG model specification accounting for the in and the out degree of the networks' nodes:

```
     Null Deviance: 16186  on 11676  degrees of freedom
 Residual Deviance:  7651  on 11435  degrees of freedom

 AIC: 8083  BIC: 9673  (Smaller is better. MC Std. Err. = 0)

 Warning: The following terms have infinite coefficient estimates:
  receiver2 receiver3 receiver4 receiver5 receiver6 receiver7 receiver8 receiver9 receiver10 receiver11 receiver12 receiver13 receiver14 sender52 sender75 sender89 sender100 sender101 sender110 sender111 sender115 sender116 sender119 sender120 sender121
```

## Wet season

```
     Null Deviance: 16186  on 11676  degrees of freedom
 Residual Deviance:  7651  on 11435  degrees of freedom

AIC: 8083  BIC: 9673  (Smaller is better. MC Std. Err. = 0)

Warning: The following terms have infinite coefficient estimates:
  receiver2 receiver3 receiver4 receiver5 receiver6 receiver7 receiver8 receiver9 receiver10 receiver11 receiver12 receiver13 receiver14 sender52 sender75 sender89 sender100 sender101 sender110 sender111 sender115 sender116 sender119 sender120 sender121
```

Unfortunately, in the food webs there are some organisms that do not eat any other organism and some others that are not eat by any other one. This led to the presence of some nodes that have in or out degree equal to zero.

This make impossible to estimate the relative parameters, in fact we can see that some parameters are equal to minus infinite.

Giving the fact that we can't insert the receiver and sender effects due to the minus infinite parameter problem, we only consider the mutual relations.

```
Call:
ergm(formula = rete_d ~ edges + mutual, control = control.ergm(seed = 1))

Monte Carlo Maximum Likelihood Results:

        Estimate Std. Error MCMC % z value Pr(>|z|)
edges  -1.80788    0.02939      0 -61.504   <1e-04 ***
mutual -3.59036    0.50954      0  -7.046   <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null Deviance: 20129  on 14520  degrees of freedom
 Residual Deviance: 10661  on 14518  degrees of freedom

AIC: 10665   BIC: 10680   (Smaller is better. MC Std. Err. = 2.981)
```

**Wet season**

```
Call:
ergm(formula = rete_w ~ edges + mutual, control = control.ergm(seed = 1))

Monte Carlo Maximum Likelihood Results:

        Estimate Std. Error MCMC % z value Pr(>|z|)
edges  -1.83004    0.02898      0  -63.16   <1e-04 ***
mutual -3.53978    0.50353      0   -7.03   <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null Deviance: 20129  on 14520  degrees of freedom
 Residual Deviance: 10548  on 14518  degrees of freedom

AIC: 10552   BIC: 10567   (Smaller is better. MC Std. Err. = 3)
```

The parameter associated to the mutual relations is negative, as expected in a food web, and significatively different from zero.
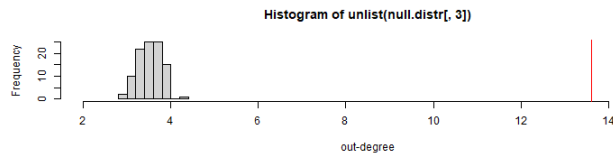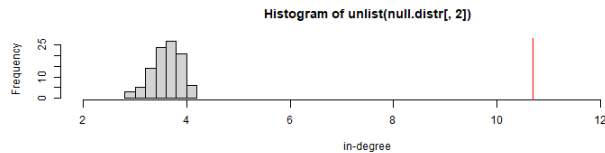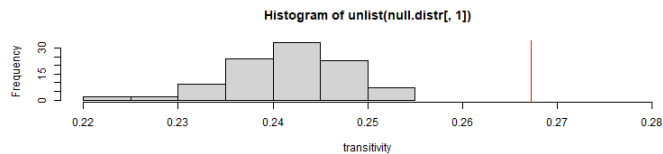
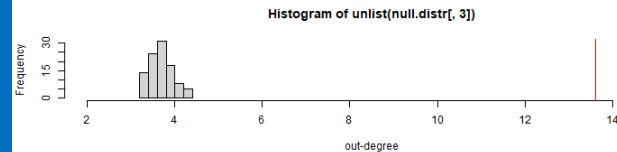The AIC and BIC do not decrease too much with respect to the previous models.
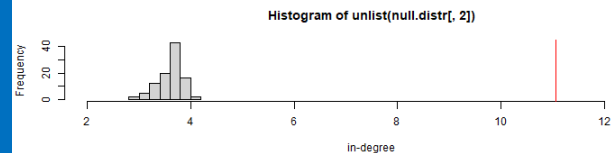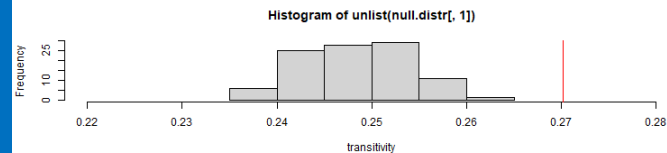
# p1 Model

This model does not seem to achieve great results too, as we can see from the distributions of the simulated networks statistics, compared to our networks ones:

# Markov Graph Model

We tried to include in the model the 2 in and out stars together with the triangle statistics, but both models could not be estimated due to degeneracy issues. Even trying to remove some of these statistics or adding the alternate stars statistic does not help to mitigate these issues.

**Dry season**

```
Iteration 11 of at most 60:
Errore in ergm.MCMLE(init, s, s.obs, control = control, verbose = verbose,  :
  Unconstrained MCMC sampling did not mix at all. Optimization cannot continue.
In aggiunta: Messaggio di avvertimento:
In ergm_MCMC_sample(s, control, theta = mcmc.init, verbose = max(verbose -  :
  Unable to reach target effective size in iterations alotted.
```

**Wet season**

```
Errore in ergm.MCMLE(init, s, s.obs, control = control, verbose = verbose,  :
  Unconstrained MCMC sampling did not mix at all. Optimization cannot continue.
In aggiunta: Messaggio di avvertimento:
In ergm_MCMC_sample(s, control, theta = mcmc.init, verbose = max(verbose -  :
  Unable to reach target effective size in iterations alotted.
```

# Social circuit Model

We insert the alternating k-triangles and the k-2-paths statistics in the model. Both parameters associated with those statistics result to be very significantly different from zero, while the mutual parameter seems to be moderately significantly different from zero for both networks.

```
Call:
ergm(formula = rete_d ~ edges + mutual + gwesp(decay = 1, fixed = T) +
    gwdsp(decay = 1, fixed = T), control = control.ergm(seed = 1))

Monte Carlo Maximum Likelihood Results:

                   Estimate Std. Error MCMC % z value Pr(>|z|)
edges             -0.677162   0.108433      0  -6.245 < 1e-04 ***
mutual            -1.265164   0.519848      0  -2.434 0.01494 *
gwesp.OTP.fixed.1  0.141897   0.045442      0   3.123 0.00179 **
gwdsp.OTP.fixed.1 -0.160738   0.005508      0 -29.181 < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null Deviance: 20129  on 14520  degrees of freedom
 Residual Deviance:  8554  on 14516  degrees of freedom

AIC: 8562   BIC: 8593   (Smaller is better. MC Std. Err. = 4.97)
```
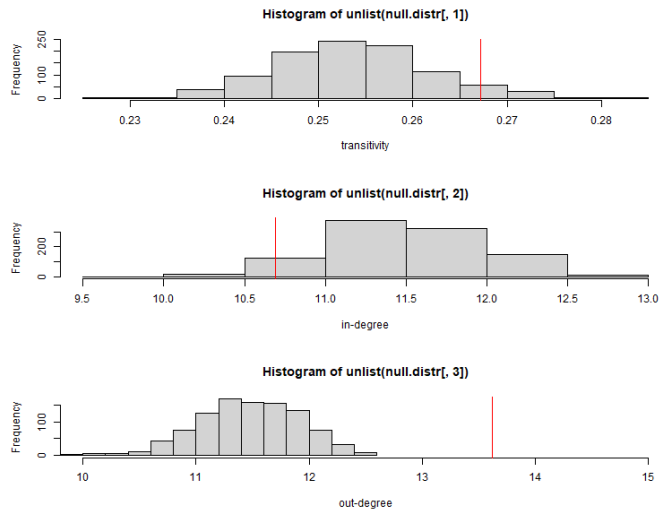
**Wet season**

```
Call:
ergm(formula = rete_w ~ edges + mutual + gwesp(decay = 1, fixed = T) +
    gwdsp(decay = 1, fixed = T), control = control.ergm(seed = 1))

Monte Carlo Maximum Likelihood Results:

                   Estimate Std. Error MCMC % z value Pr(>|z|)
edges             -0.834507   0.109293      0  -7.636 <1e-04 ***
mutual            -1.326966   0.522297      0  -2.541 0.0111 *
gwesp.OTP.fixed.1  0.201855   0.045018      0   4.484 <1e-04 ***
gwdsp.OTP.fixed.1 -0.162527   0.005489      0 -29.608 <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null Deviance: 20129  on 14520  degrees of freedom
 Residual Deviance:  8533  on 14516  degrees of freedom

AIC: 8541   BIC: 8572   (Smaller is better. MC Std. Err. = 6.905)
```
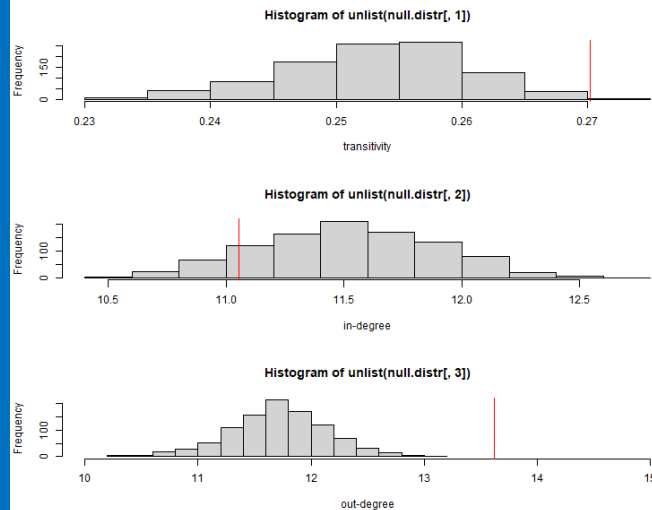
# Social circuit Model

This model seems to achieve better results than the ones we have seen before: indeed, we can see that the in degree for both networks and the transitivity for the wet network seems to be well

Histogram of unlist(null.distr[, 1])

Histogram of unlist(null.distr[, 2])

Histogram of unlist(null.distr[, 3])

**Wet season**



Histogram of unlist(null.distr[, 1])

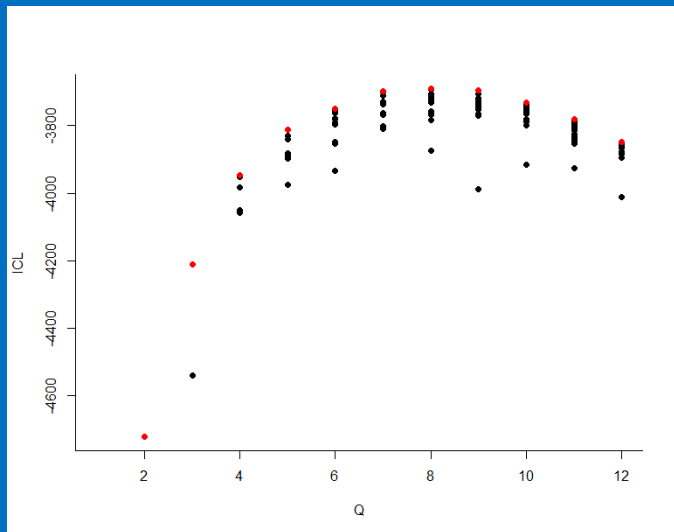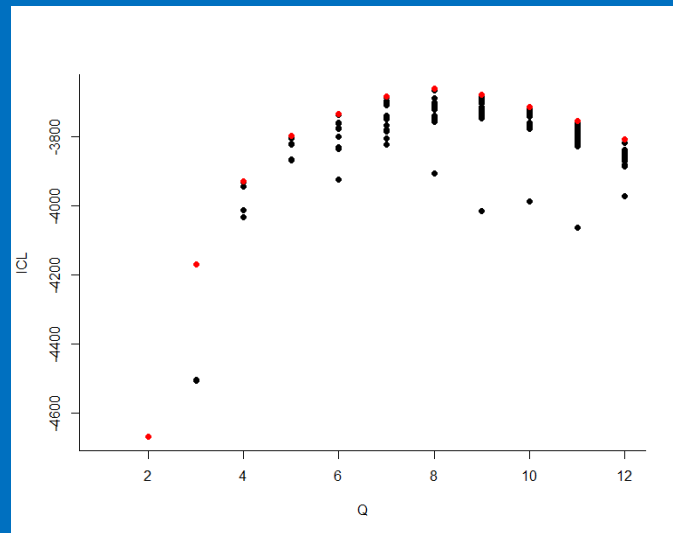Histogram of unlist(null.distr[, 2])

Histogram of unlist(null.distr[, 3])

# Stochastic Block Model

Finally, we want to see how the stochastic block model perform on our data.

## Wet season





We can see, from the ICL resulting from fitting a model with different number of blocks, that the better choice is 8 blocks for both networks

# Stochastic Block Model

Here we can see the results of parameters estimation for both networks and the proportion of organisms that have been assigned to each block.

|       | [,1]  | [,2]  | [,3]  | [,4]  | [,5]  | [,6]  | [,7]  | [,8]  |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| [1,]  | 0.111 | 0.003 | 0.471 | 0.002 | 0.488 | 0.236 | 0.859 | 0.329 |
| [2,]  | 0.103 | 0.040 | 0.679 | 0.001 | 0.387 | 0.283 | 0.074 | 0.004 |
| [3,]  | 0.006 | 0.002 | 0.012 | 0.001 | 0.338 | 0.009 | 0.571 | 0.682 |
| [4,]  | 0.230 | 0.303 | 0.068 | 0.053 | 0.011 | 0.210 | 0.030 | 0.005 |
| [5,]  | 0.008 | 0.001 | 0.014 | 0.000 | 0.068 | 0.006 | 0.029 | 0.508 |
| [6,]  | 0.010 | 0.001 | 0.014 | 0.000 | 0.141 | 0.042 | 0.072 | 0.313 |
| [7,]  | 0.001 | 0.000 | 0.001 | 0.000 | 0.001 | 0.001 | 0.002 | 0.073 |
| [8,]  | 0.000 | 0.000 | 0.001 | 0.000 | 0.001 | 0.001 | 0.001 | 0.056 |

**Wet season**

|       | [,1]  | [,2]  | [,3]  | [,4]  | [,5]  | [,6]  | [,7]  | [,8]  |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| [1,]  | 0.110 | 0.468 | 0.003 | 0.002 | 0.496 | 0.317 | 0.899 | 0.322 |
| [2,]  | 0.007 | 0.015 | 0.002 | 0.001 | 0.345 | 0.021 | 0.606 | 0.720 |
| [3,]  | 0.103 | 0.649 | 0.039 | 0.001 | 0.340 | 0.438 | 0.079 | 0.008 |
| [4,]  | 0.229 | 0.060 | 0.301 | 0.053 | 0.009 | 0.169 | 0.032 | 0.005 |
| [5,]  | 0.009 | 0.011 | 0.001 | 0.000 | 0.082 | 0.001 | 0.027 | 0.480 |
| [6,]  | 0.007 | 0.021 | 0.001 | 0.000 | 0.162 | 0.033 | 0.072 | 0.230 |
| [7,]  | 0.001 | 0.001 | 0.000 | 0.000 | 0.001 | 0.001 | 0.002 | 0.064 |
| [8,]  | 0.000 | 0.001 | 0.000 | 0.000 | 0.001 | 0.001 | 0.001 | 0.051 |

`[1] 0.10765347 0.09934553 0.14847048 0.16421137 0.17299043 0.07527693 0.11588193 0.11616985`

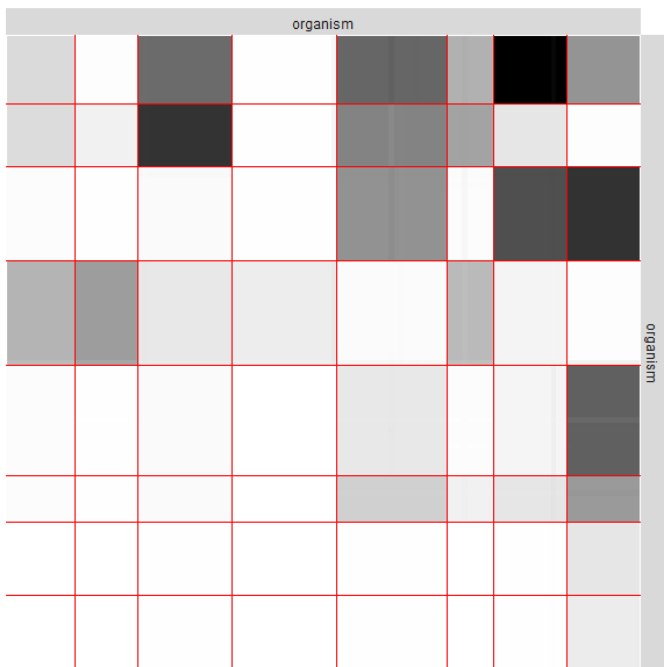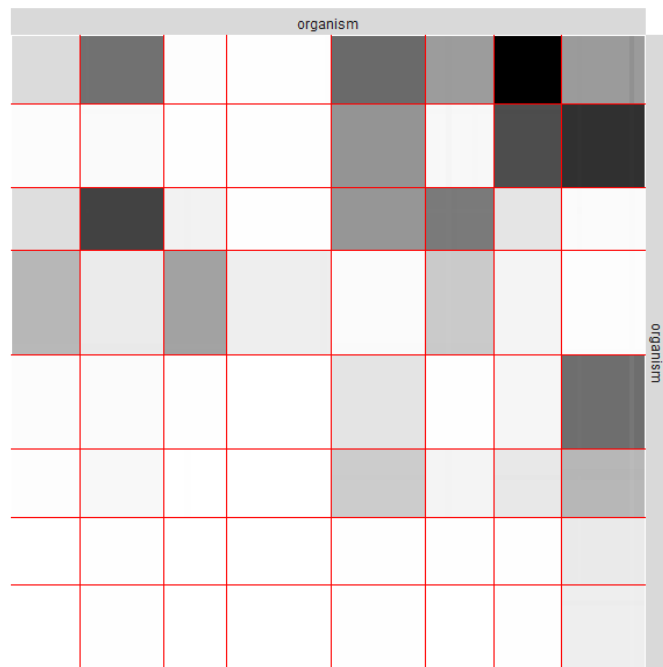`[1] 0.10757037 0.13224396 0.09939813 0.16489748 0.14891548 0.10733794 0.10809115 0.13154549`
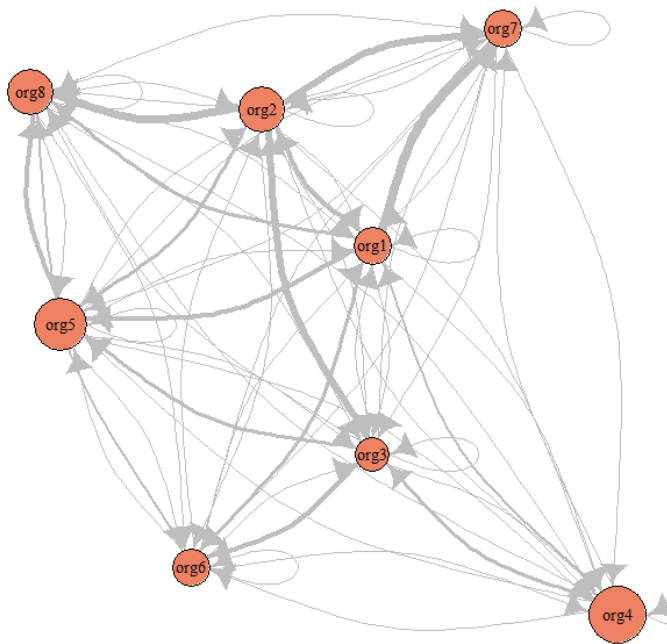
Wet season Data

Dry season Data

Stochastic Block Model
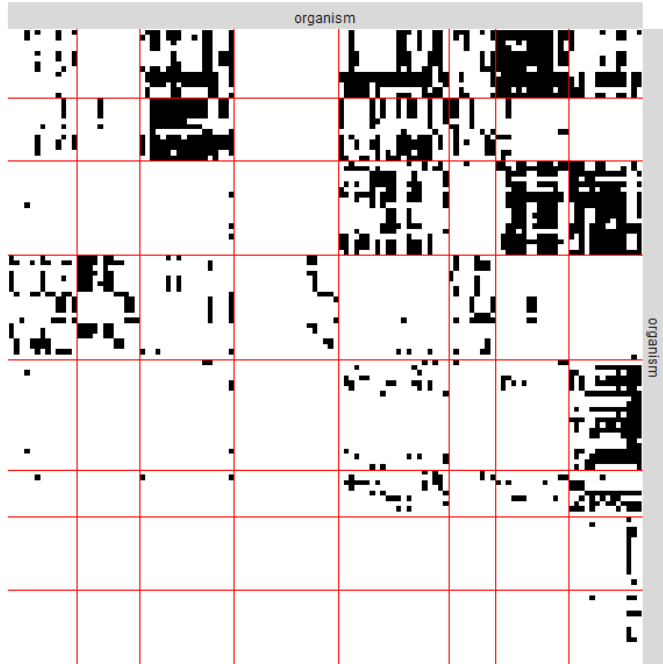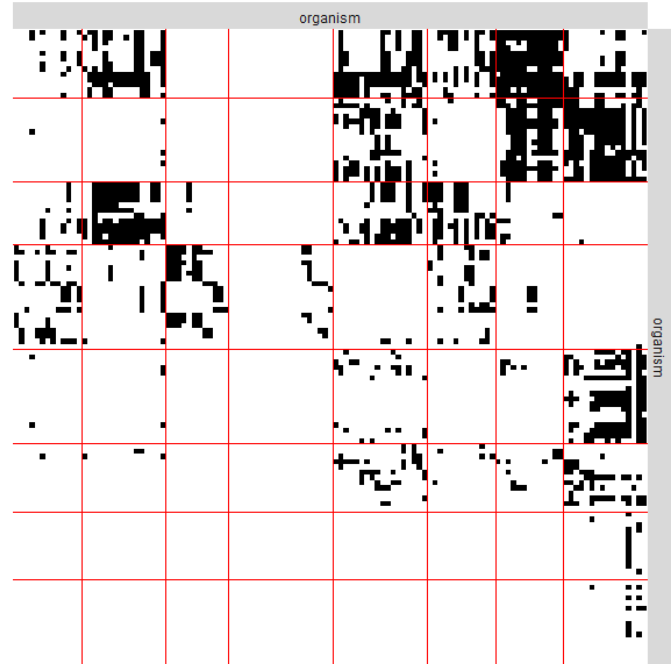
Wet season Expected

Dry season Expected

Stochastic Block Model

Wet season meso

Dry season meso

Stochastic Block Model
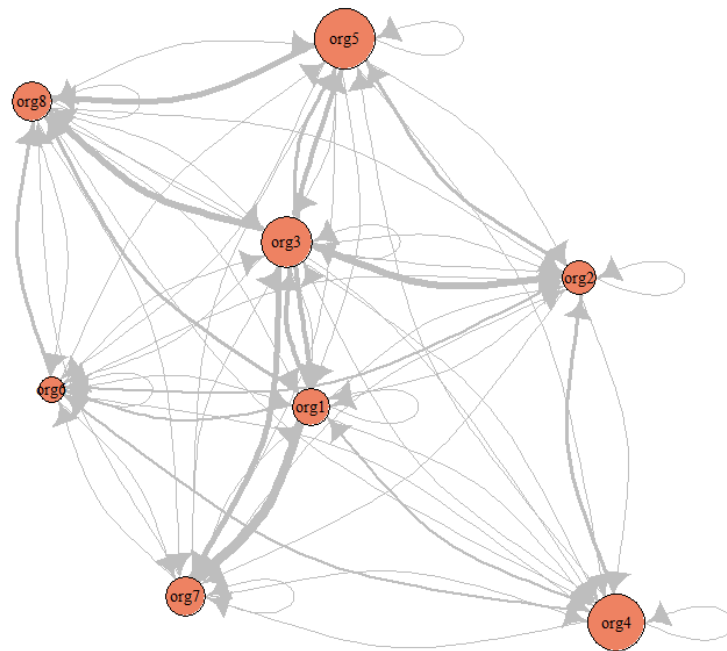
Wet season Block
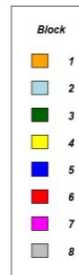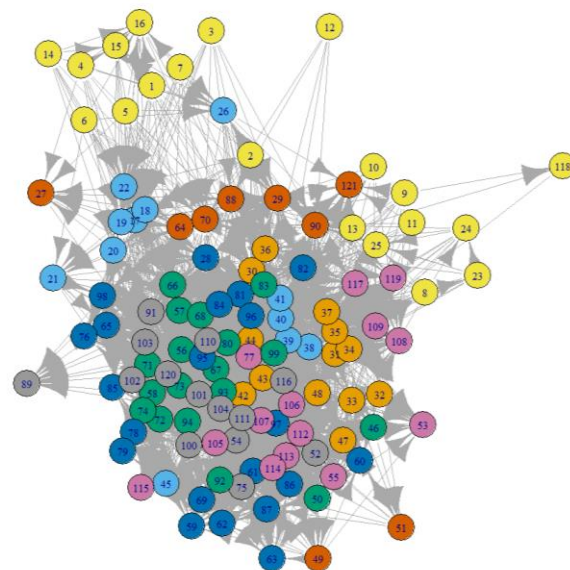
Dry season Block

Stochastic Block Model
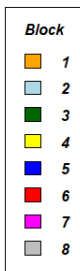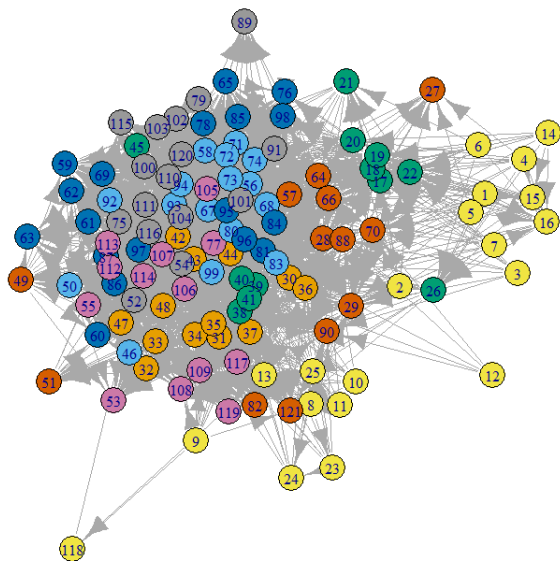
# Block Composition

➢ Every block contains organism that behave similarly.

➢ We have noticed that the groups are mostly made by similar organism.

➢ For example, the fourth group contains mostly microorganism, like bacteria, algae or plankton (except the turtle).

➢ These organism rarely eat any other organism in the network, and overall tend to have less (in or out) edges in the graph than the other nodes.

**Fourth Group (yellow)**

```
[1]  "2um Spherical Phytoplankt"  "Synedococcus"        "Oscillatoria"         "Small Diatoms (<20um)"
[5]  "Big Diatoms (>20um)"        "Dinoflagellates"     "Other Phytoplankton"  "Benthic Phytoplankton"
[9]  "Thalassia"                  "Halodule"            "Syringodium"          "Drift Algae"
[13] "Epiphytes"                  "Free Bacteria"       "Water Flagellates"    "Water Cilitaes"
[17] "Benthic Flagellates"        "Benthic Ciliates"    "Meiofauna"            "Green Turtle"
```

Another example is group eight that contains nodes which eat a lot of other organism.
This group is composed by great predator, like crocodiles, herons, raptors, sharks, barracudas or pelicans.

```
[1]  "Sharks"              "Tarpon"           "Grouper"    "Gray Snapper"  "Mackerel"
[6]  "Barracuda"           "Loon"             "Greeb"      "Pelican"       "Comorant"
[11] "Big Herons & Egrets" "Predatory Ducks"  "Raptors"    "Kingfisher"    "Crocodiles"
[16] "Dolphin"
```

**Dry season**

**Wet season**

```
[1]  "Sharks"           "Tarpon"    "Grouper"      "Mackerel"   "Barracuda"
[6]  "Loon"             "Greeb"     "Pelican"      "Comorant"   "Big Herons & Egrets"
[11] "Predatory Ducks"  "Raptors"   "Crocodiles"   "Dolphin"
```

# Conclusions

- ❖ We have observed that the network for the wet and the dry season are not so different.
- ❖ The two networks do not seem to be very centralized, except if we consider the out degree.
- ❖ The best model we were able to fit is the social circuit model, but it was not able to fully capture all the networks' characteristics.
- ❖ The SBM model divides the nodes into 8 blocks that contains organism which have  both similar predatory habits and are predate similarly.
- ❖ The biomass nodal attribute seems to have no relevant effect on the predation's habits of the organism.
- ❖ Finally, we have shown that the are no organism which is more important than others, but each organism have its own role and part in the food chain and in the whole ecosystem.

# THANKS FOR THE ATTENTION!

Project for the Web Mining Exam
By:
Alberto Biliotti
Jacopo Manetti