Interpretation of graph theory metrics for species persistence in a metapopulation. The Gulf of Lion study case.

Institut Pythéas
Observatoire des Sciences de l'Univers
Aix-Marseille Université

COCONET

cnrs advancing the frontiers

Aix-Marseille université

mio Mediterranean Institute of Oceanography

Costa A., Doglioli A.M., Guizien K., Petrenko A.A.

METAPOPULATION MODEL

(see Guizien et al, 2014)

Species: soft-bottom polychaete
Site: Gulf of Lion

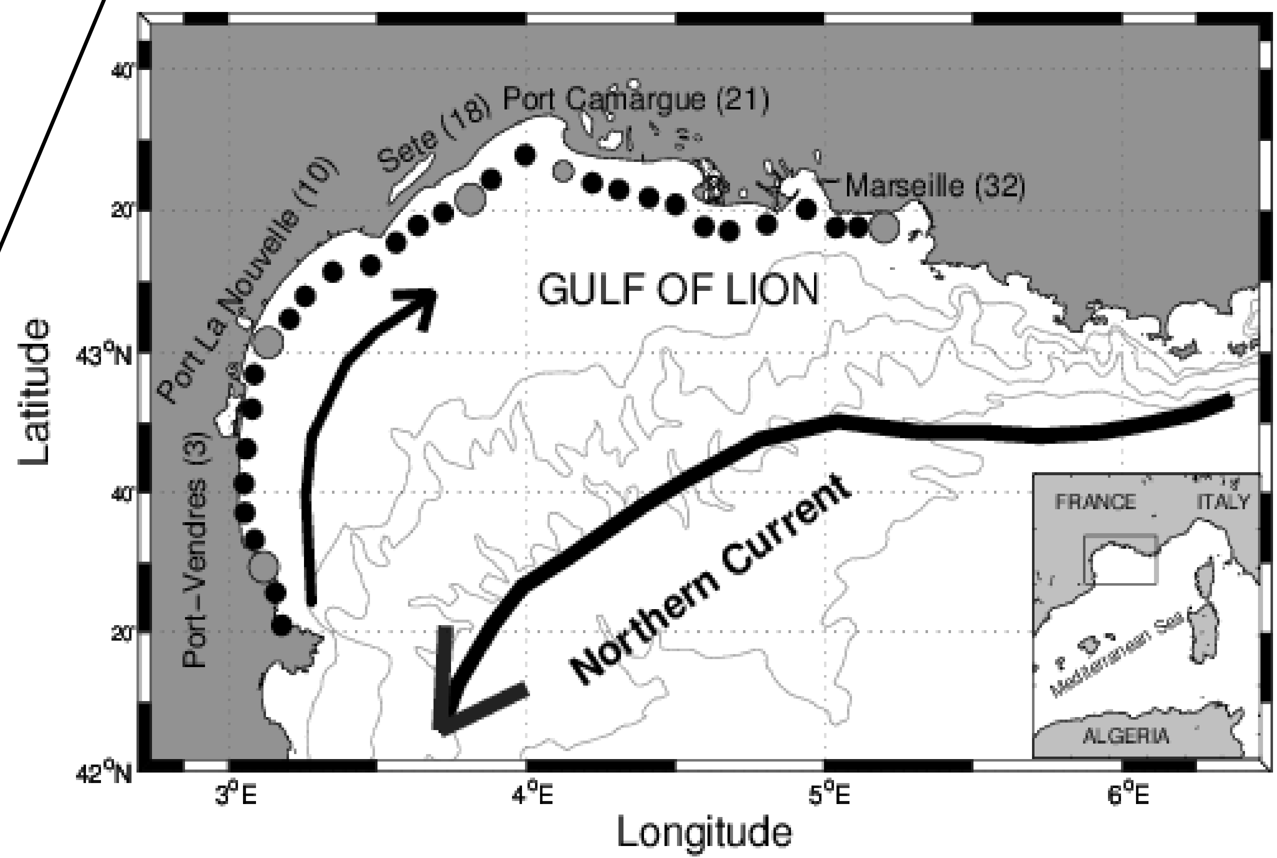
$$N_{t+1} = min\left(b_j C_{ij} a_i N_t + s_{jj} \delta_{ji} N_t, N_{max}\right)$$

Propagule production rate

Transfer probability

Recruitment success

Survivorship rate



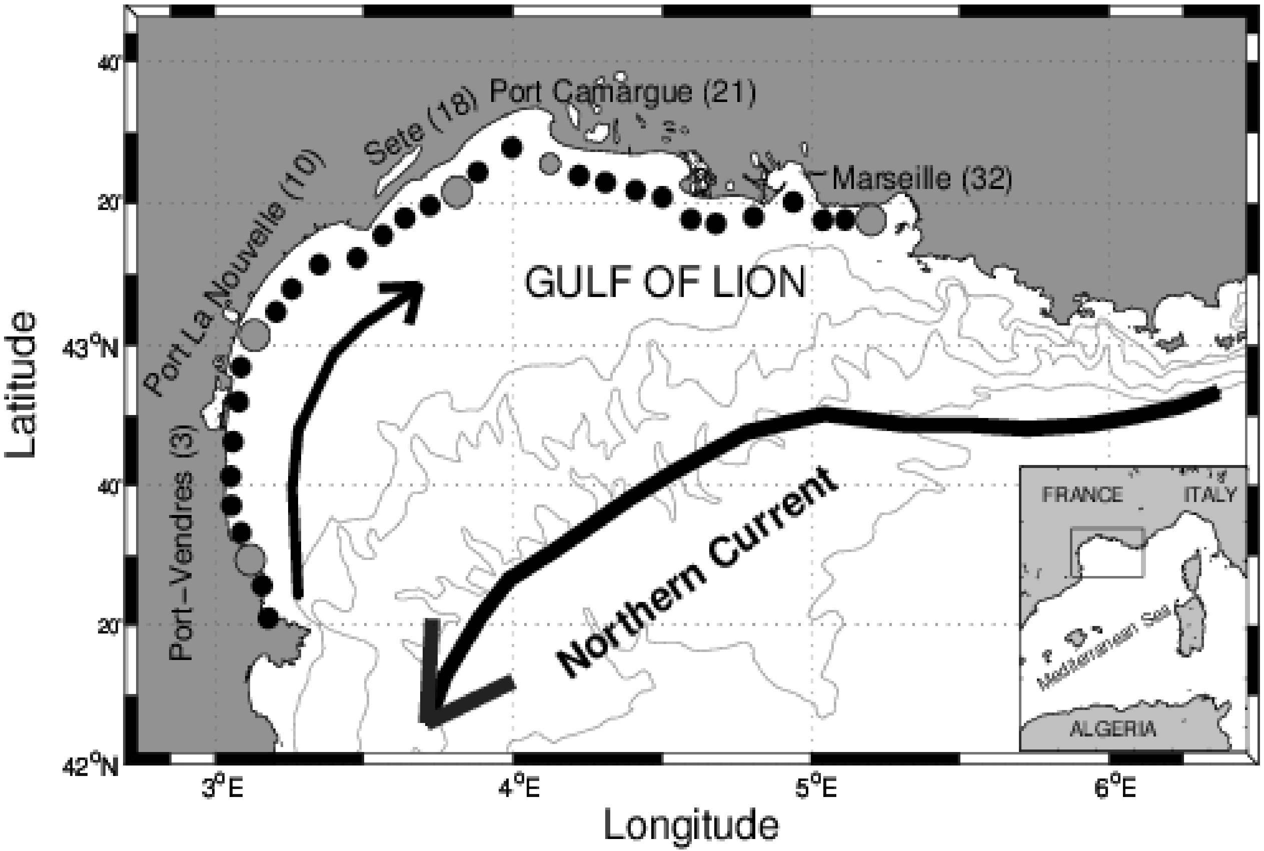
METAPOPULATION MODEL

(see Guizien et al, 2014)

Species: <u>soft-bottom polychaete</u>
Site: Gulf of Lion

$$N_{t+1} = min(b_j C_{ij} a_i N_t + s_{jj} \delta_{ji} N_t, N_{max})$$

Four sites analyzed.



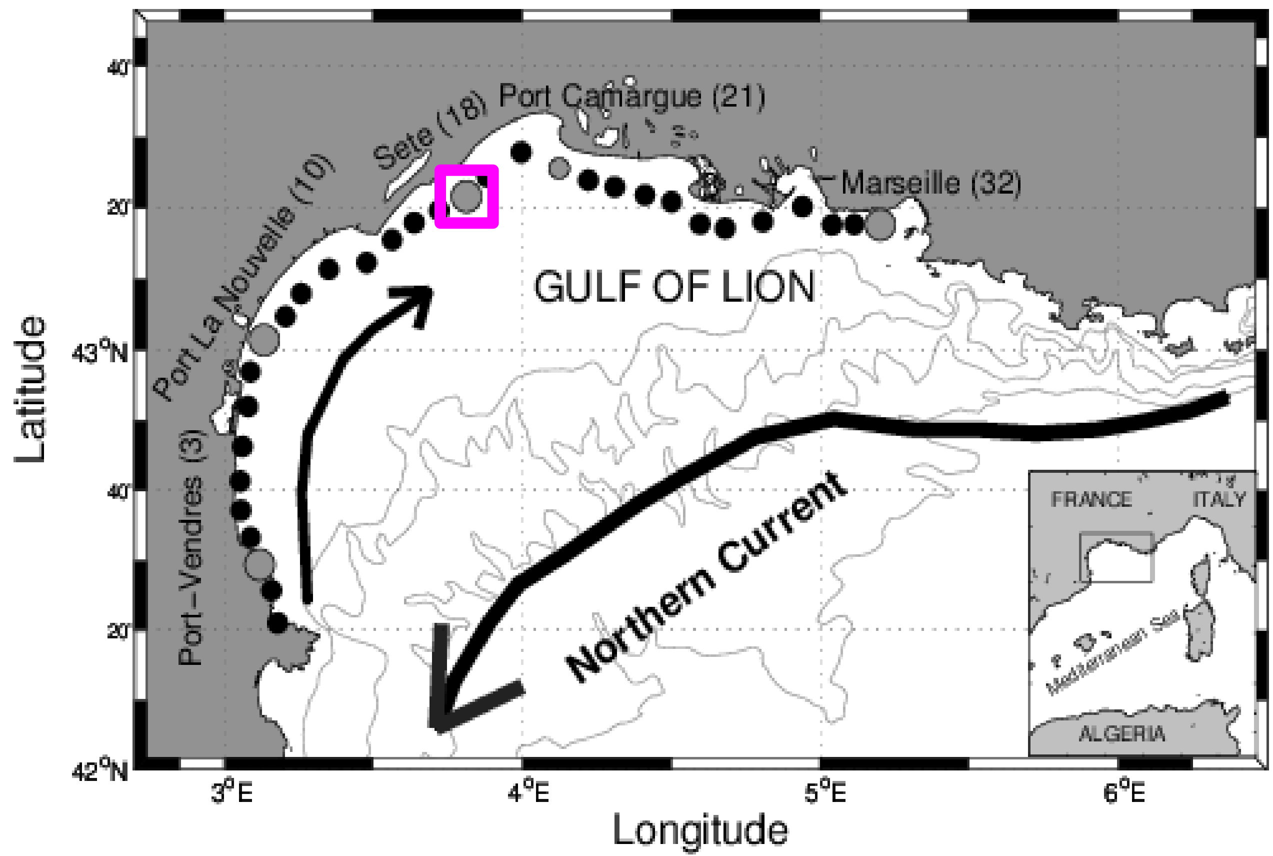
METAPOPULATION MODEL

(see Guizien et al, 2014)

Species: <u>soft-bottom polychaete</u>
Site: Gulf of Lion

$$N_{t+1} = min\left(b_j C_{ij} a_i N_t + s_{jj} \delta_{ji} N_t, N_{max}\right)$$

One essential node for persistence.



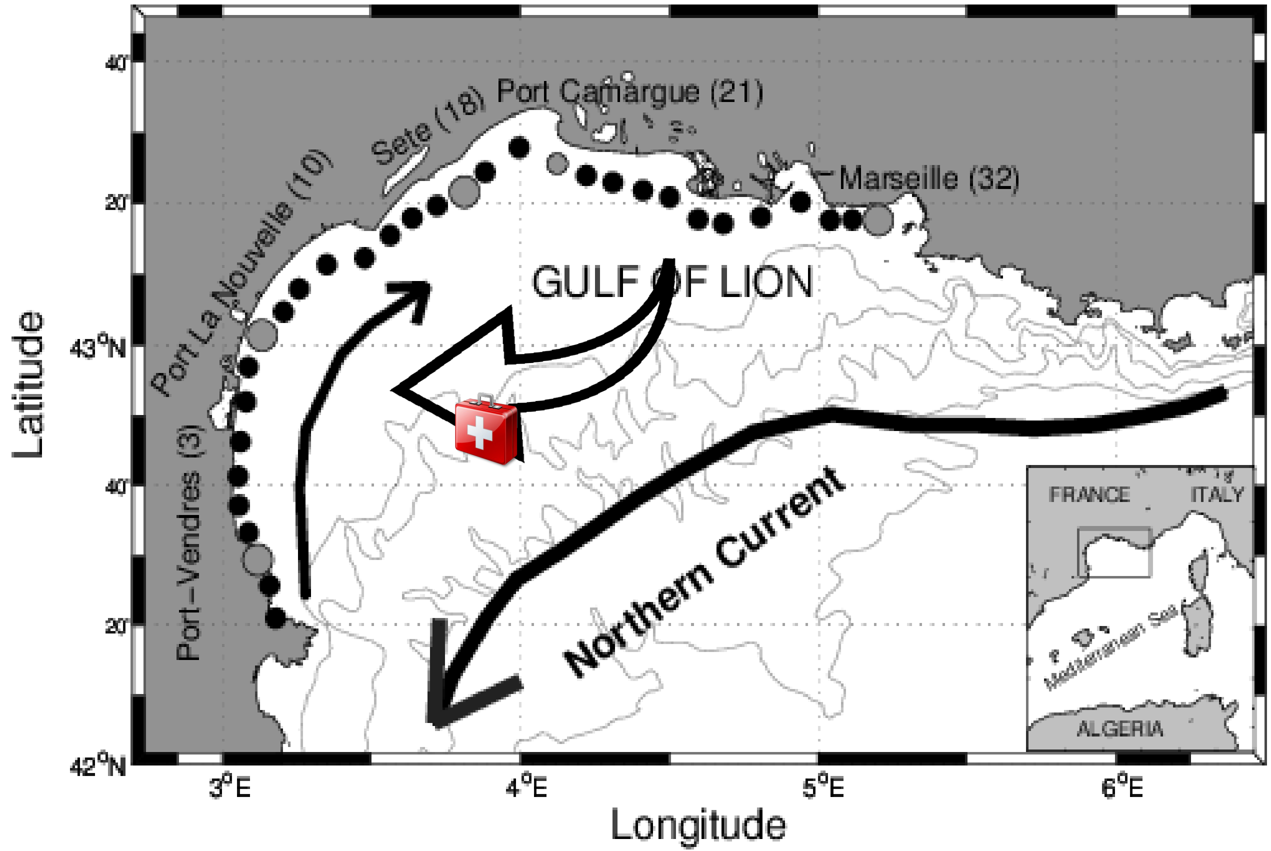
METAPOPULATION MODEL

(see Guizien et al, 2014)

Species: soft-bottom polychaete
Site: Gulf of Lion

$$N_{t+1} = min\left(b_j C_{ij} a_i N_t + s_{jj} \delta_{ji} N_t, N_{max}\right)$$

Presence of a rescue mechanism.



METAPOPULATION MODEL

(see Guizien et al, 2014)

Species: <u>soft-bottom polychaete</u>
Site: Gulf of Lion

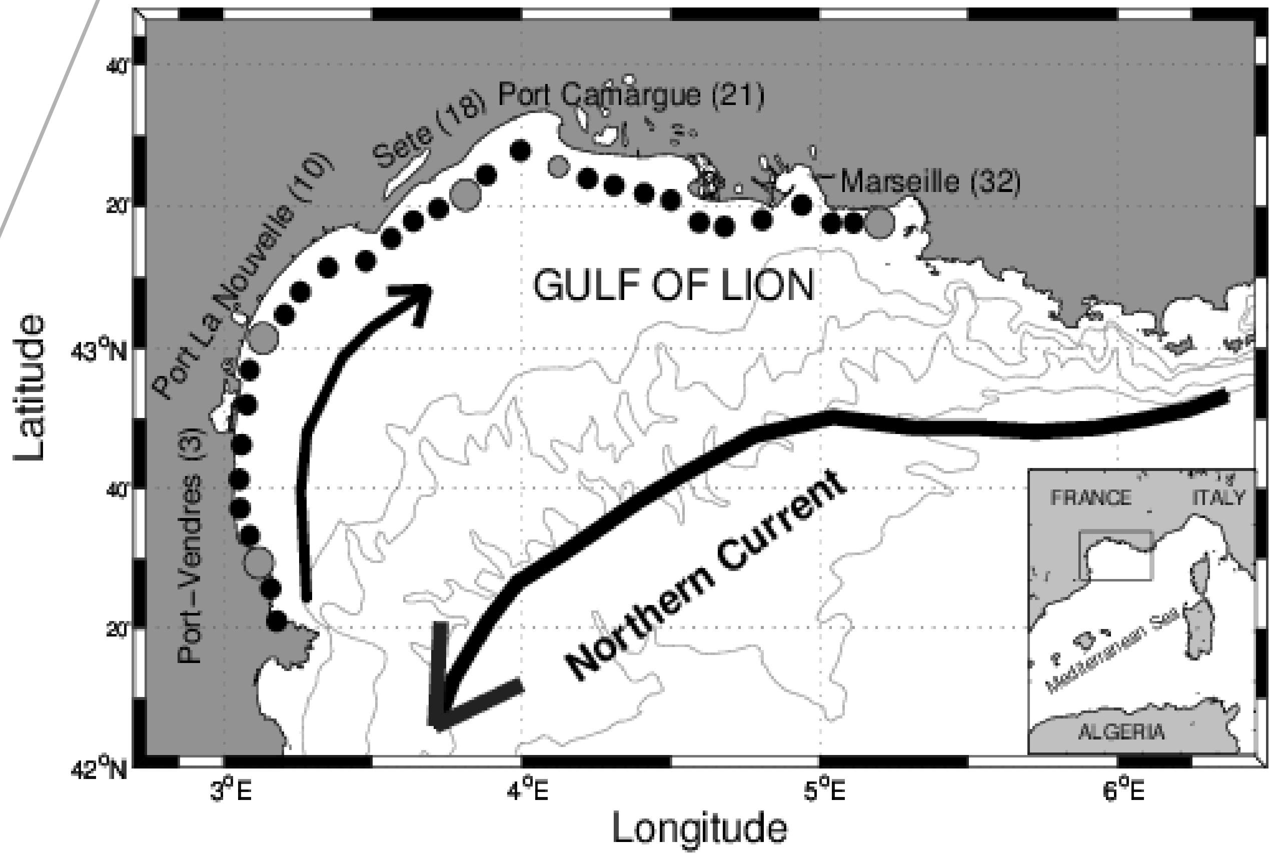
$$N_{t+1} = min(b_j C_{ij} a_i N_t + s_{jj} \delta_{ji} N_t, N_{max})$$

Propagule production rate

Transfer probability

Recruitment success

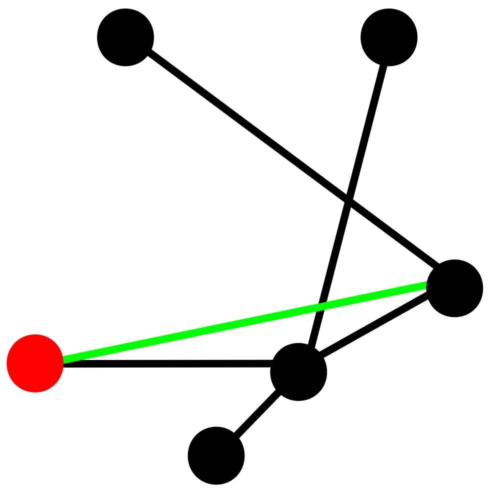
Survivorship rate



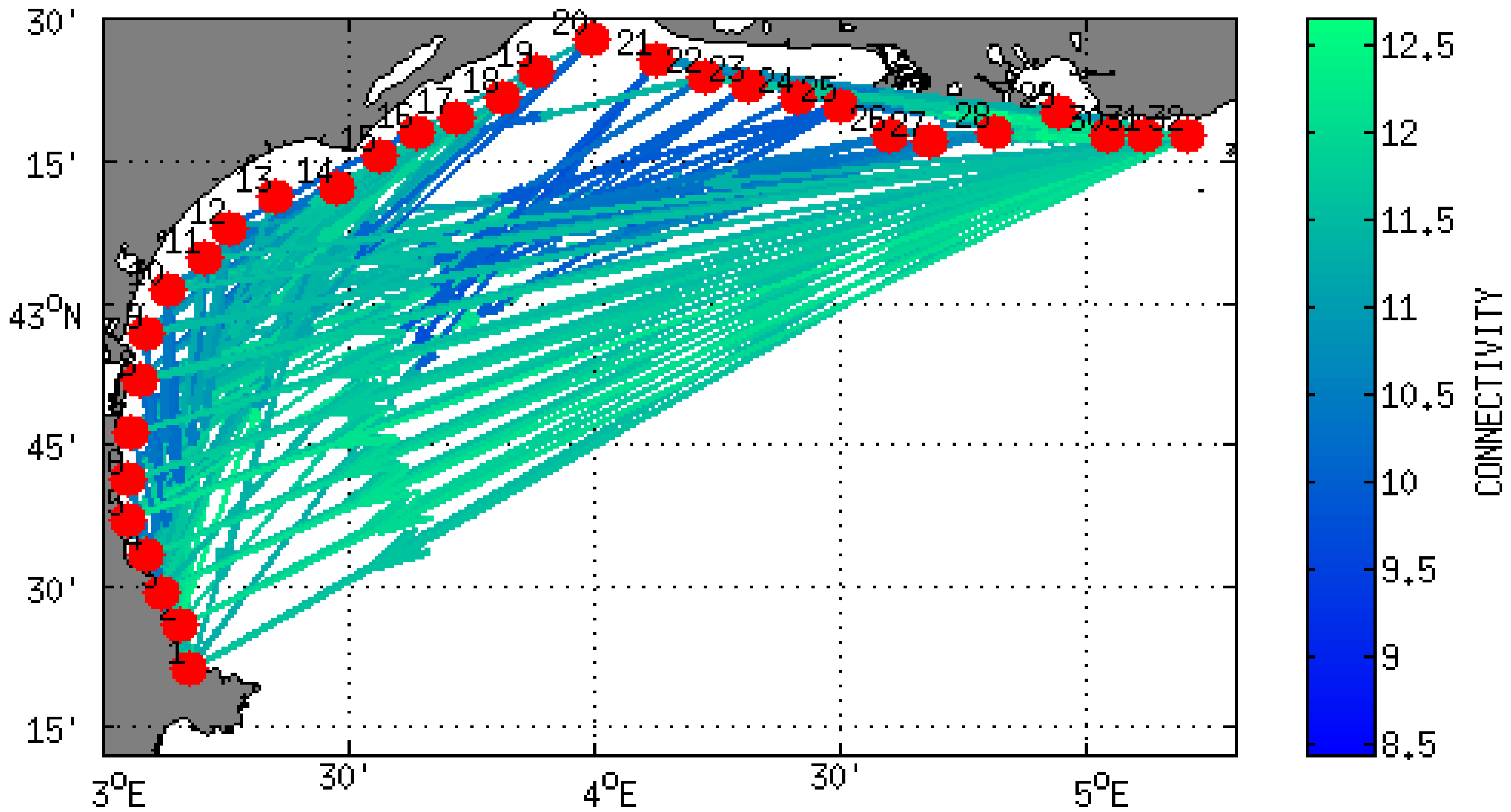
GRAPH THEORY

Graph: representation of pairwise relations between nodes (**sites**).

Relations are represented by links (C_{ij}).



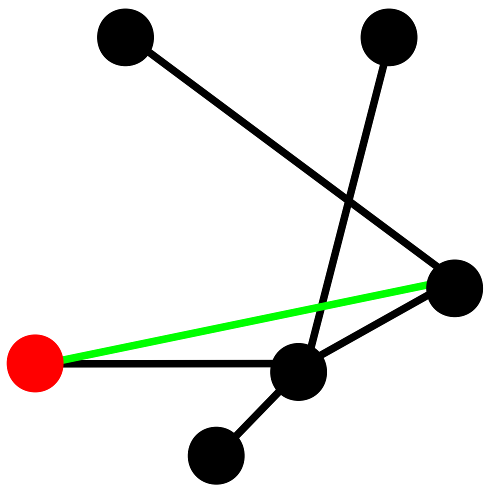
WHAT IT LOOKS LIKE



GRAPH THEORY

Graph: representation of pairwise relations between nodes (**sites**).

Relations are represented by links (C_{ij}).



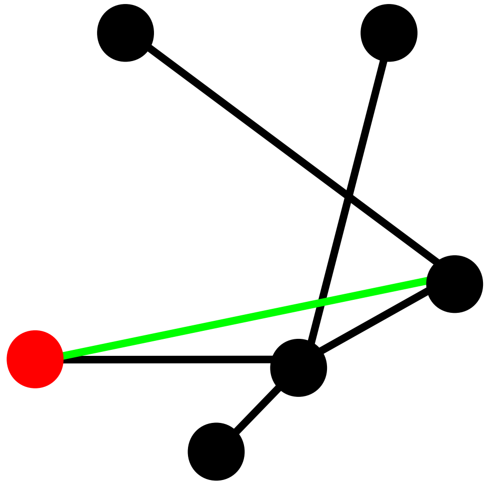
New distance between nodes:

$$d_{i,j} = \log\left(\frac{1}{C_{i,j}}\right)$$

GRAPH THEORY

Graph: representation of_pairwise relations between nodes (**sites**).

Relations are represented by links (C_{ij}).



New distance between nodes:

$$d_{i,j} = \log\left(\frac{1}{C_{i,j}}\right)$$

Shortest path

=

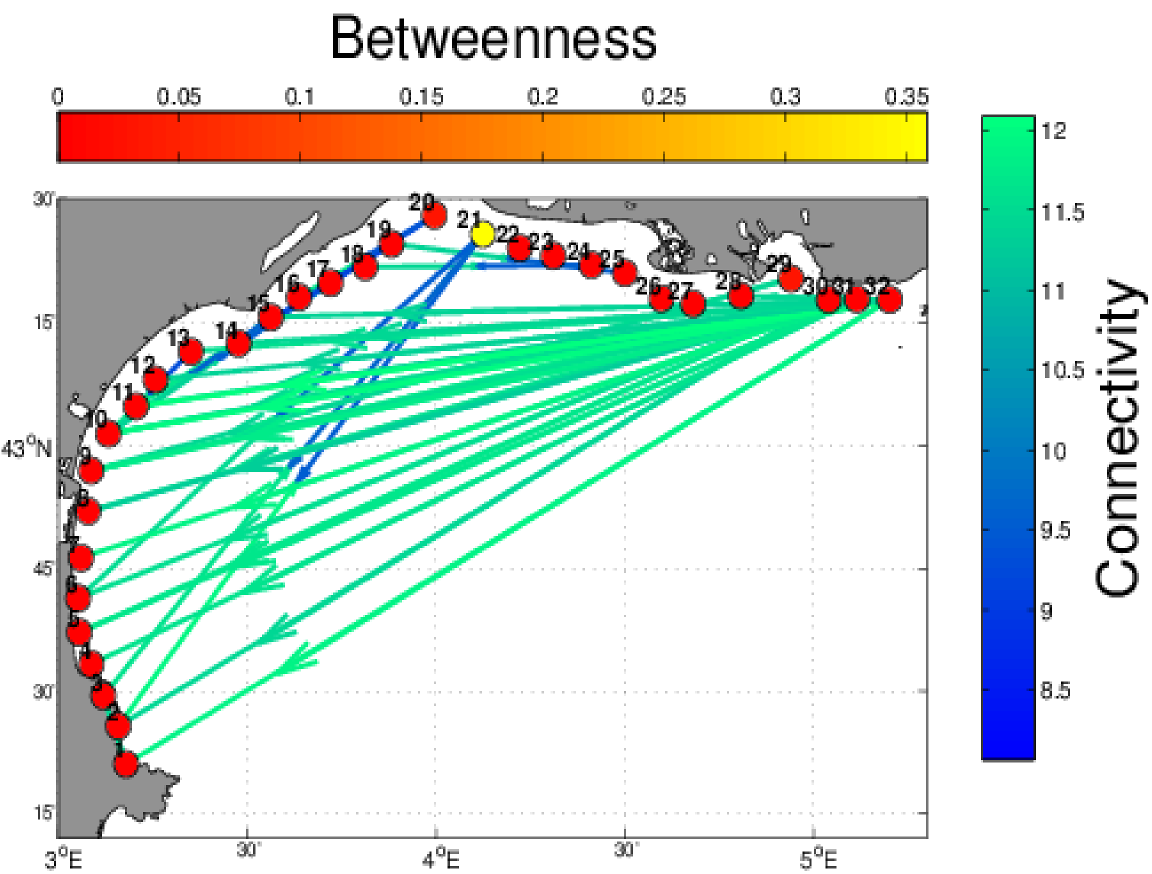
most probable

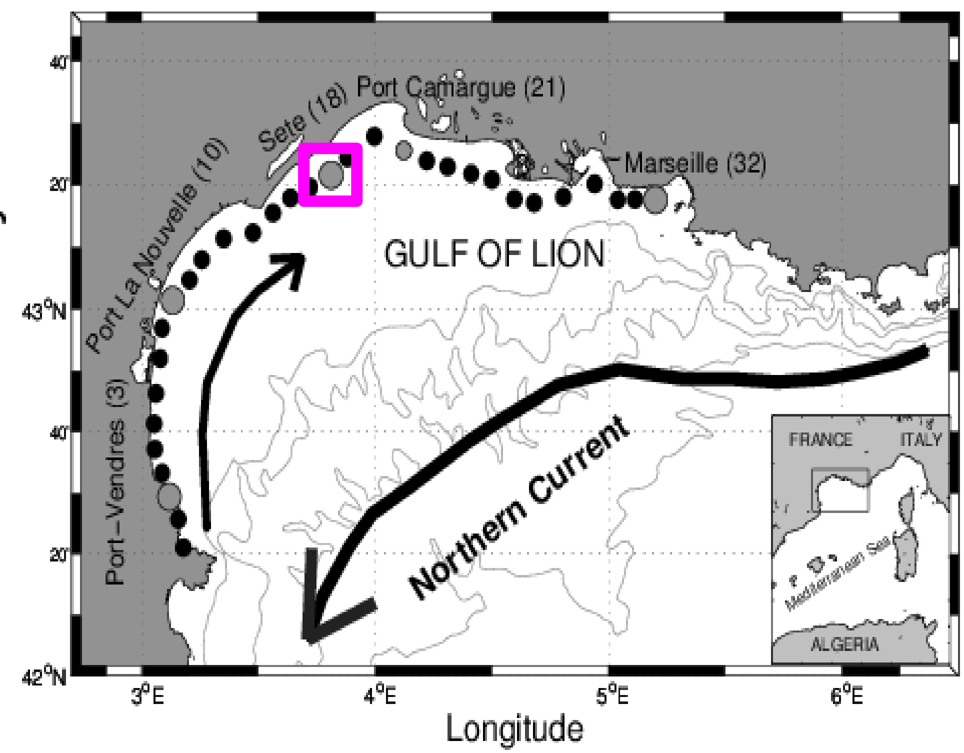
Betweenness: the number of shortest paths that pass through a node

Betweenness: the number of shortest paths that pass through a node

Gateways → Persistence

Betweenness: the number of shortest paths that pass through a node





Gateways $\not\rightarrow$ Persistence

Thus we introduce a new measure.

BRIDGING CENTRALITY

Thus we introduce a new measure.

BRIDGING CENTRALITY

Nodes with:

High basin-scale flux of propagule

Lying the border of clusters

Thus we introduce a new measure.

BRIDGING CENTRALITY

Nodes with:

High basin-scale flux of propagule ⟹ **High betweenness**

Lying the border of clusters ⟹ **Topological factor**

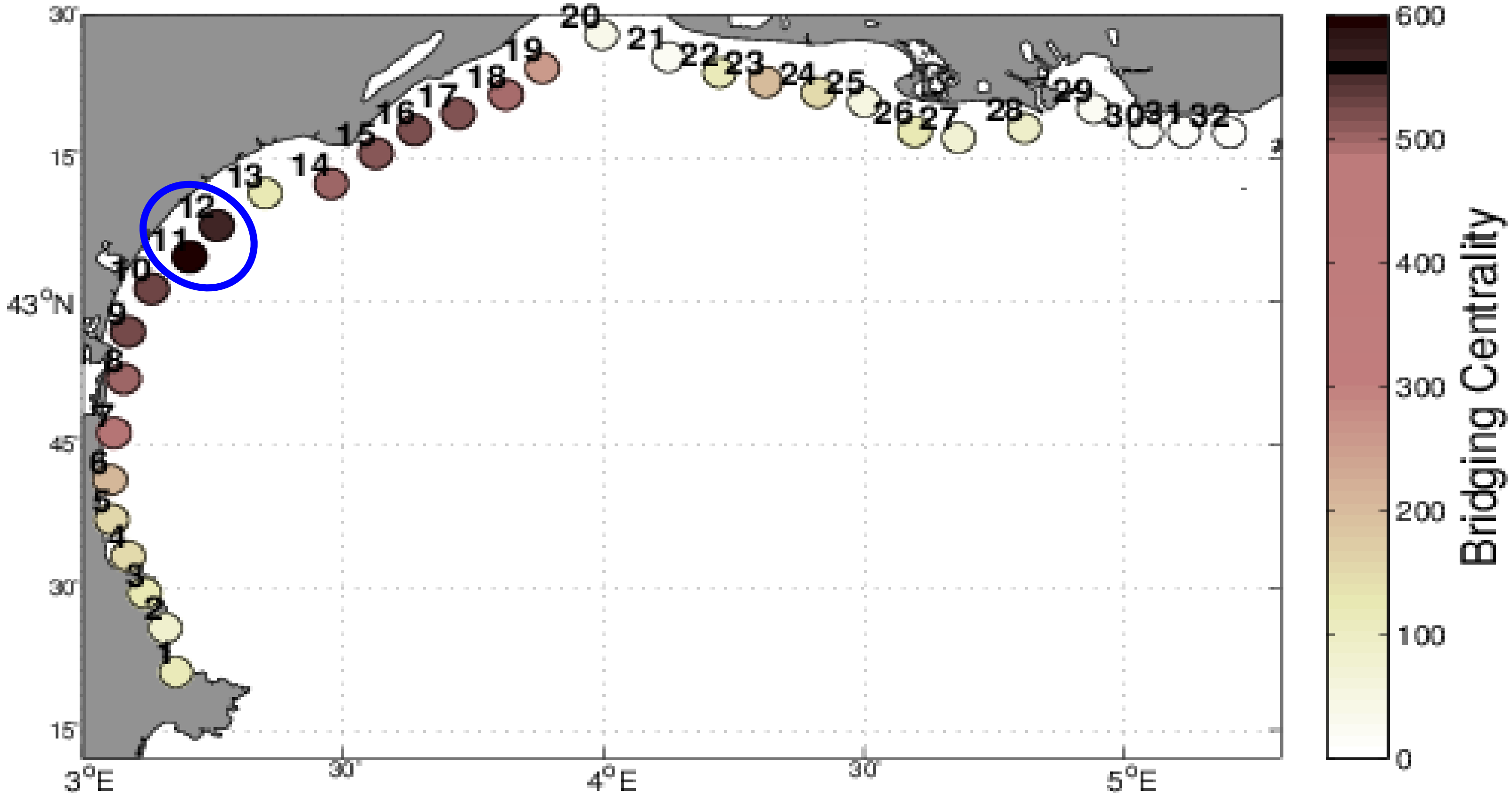
Thus we introduce a new measure.

BRIDGING CENTRALITY

Nodes with:

High basin-scale flux of propagule ⟹ **High betweenness**

Lying the border of clusters ⟹ **Topological factor**



High-bridging centrality nodes ensure the network's integrity

Thus we look for clusters.

Thus we look for clusters.

Cluster of nodes = Community

Cluster = Statistically surprising disposition of links

Thus we look for clusters.

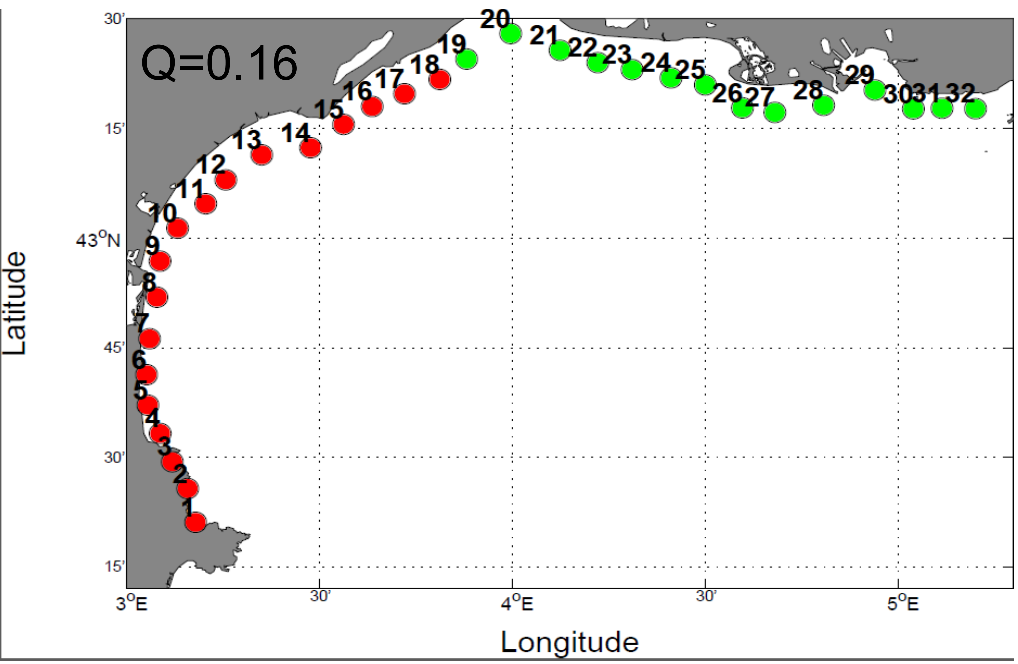
Cluster of nodes = Community

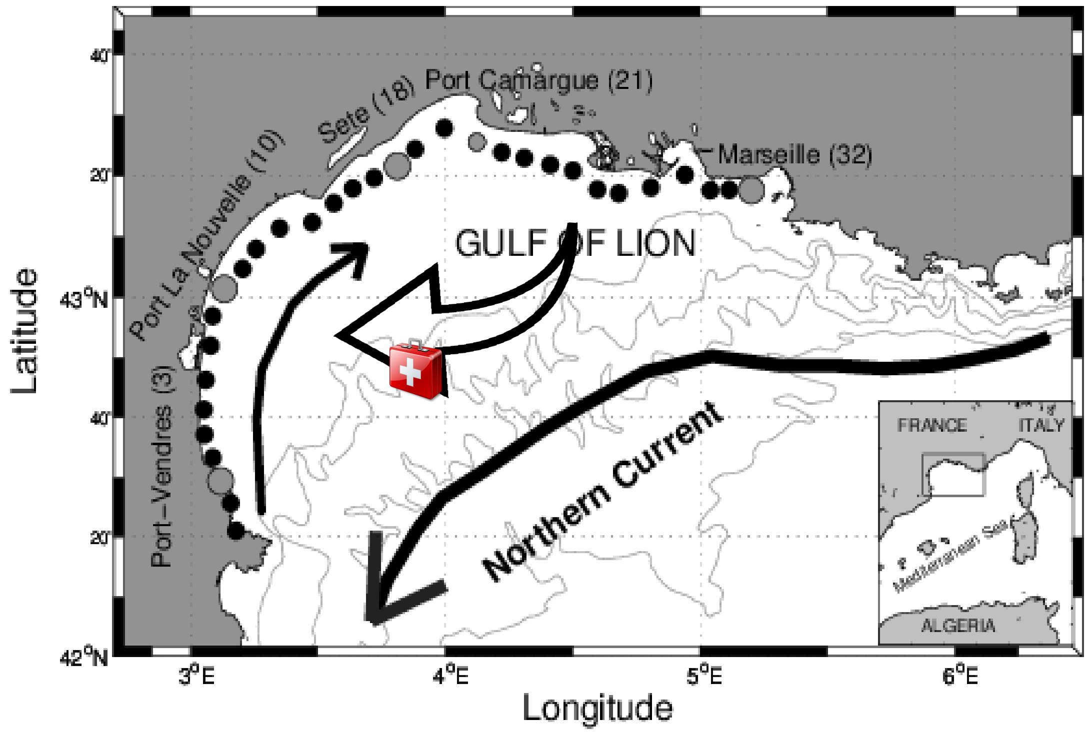
Cluster → Maximization of modularity Q

Thus we look for clusters.

Cluster of nodes = Community

Cluster → Maximization of modularity Q





Graph theory detects rescue effects.

CONCLUSIONS

1) We propose a new consistent node-to-node metric.

2) Betweenness is not correlated with persistence.

3) Bridging centrality is a good indicator for persistence.

4) Modularity can identify communities and rescue mechanisms.

5) Methodology independent from demographic parameters.

THANK YOU

FOR YOUR

KIND ATTENTION

BIBLIOGRAPHY

Graph theory:

Costa A., Guizien K., Doglioli A.M., Petrenko A.A., *submitted to Limnology and Oceanography*

Metapopulation model:

Guizien K., Belharet M., Moritz C., Guarini J., *Diversity and Distributions*, 2014.

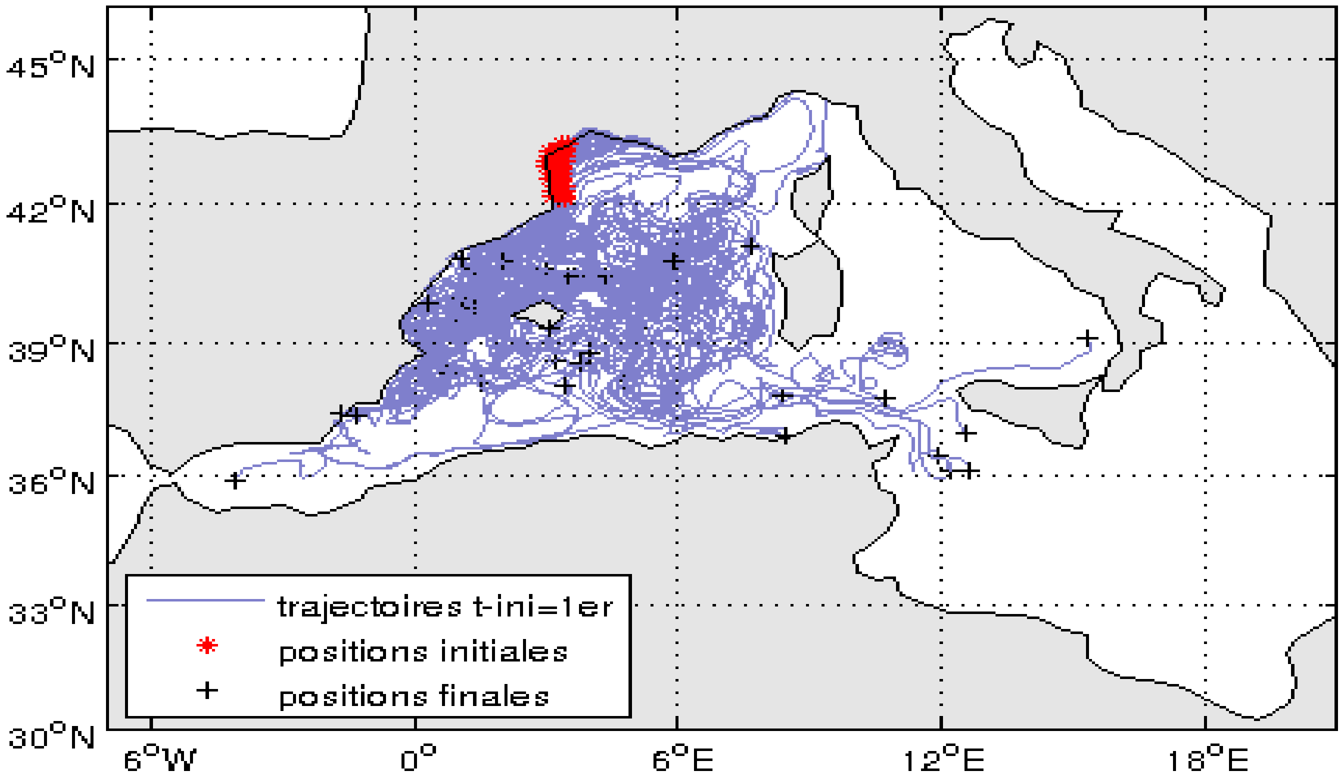
Modularity:

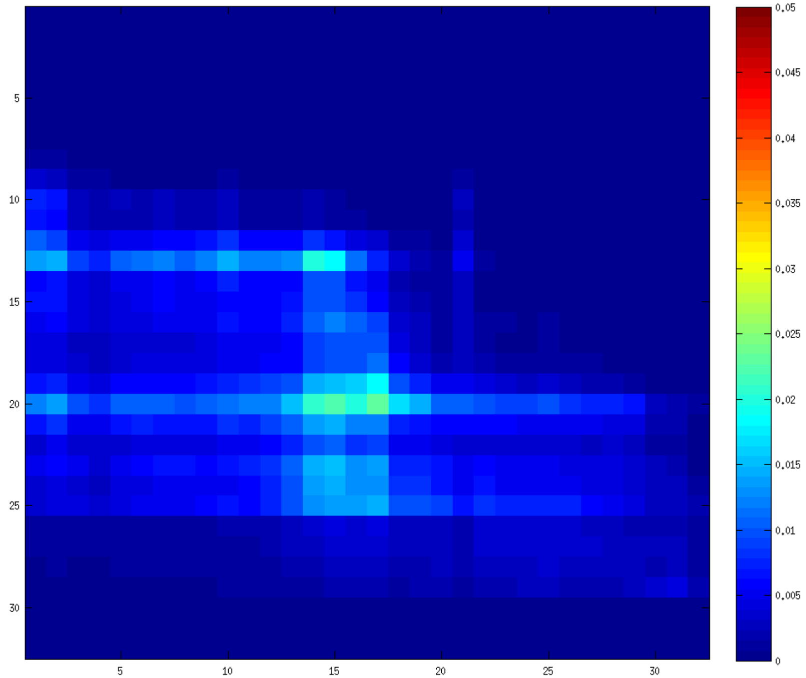
Newman M., Girwan M., *Physical Review E 69*, 2004.

Bridging centrality:

Hwang, W., Taehyong, K., Murali, R., Aidong, Z., *Proceedings KDD '08*, 2008.

EXTRA SLIDES





Graph Theory

enables us to

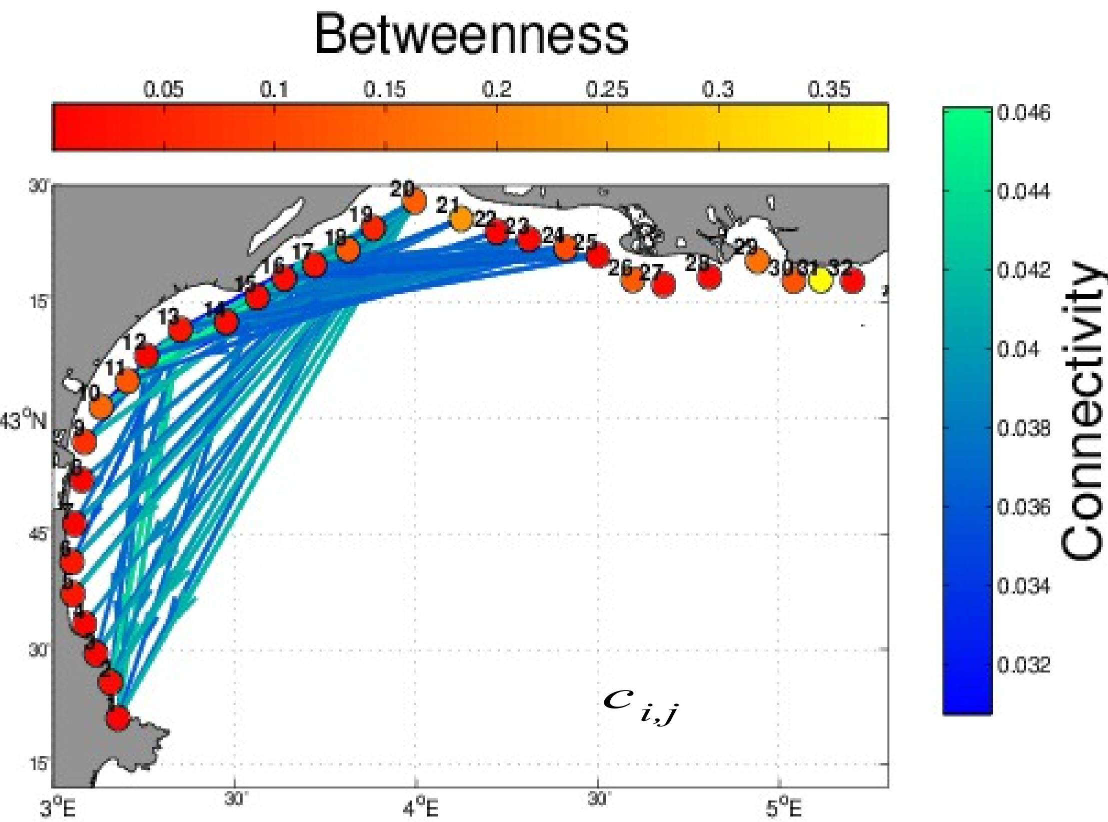
derive more information about

crucial points of the system.

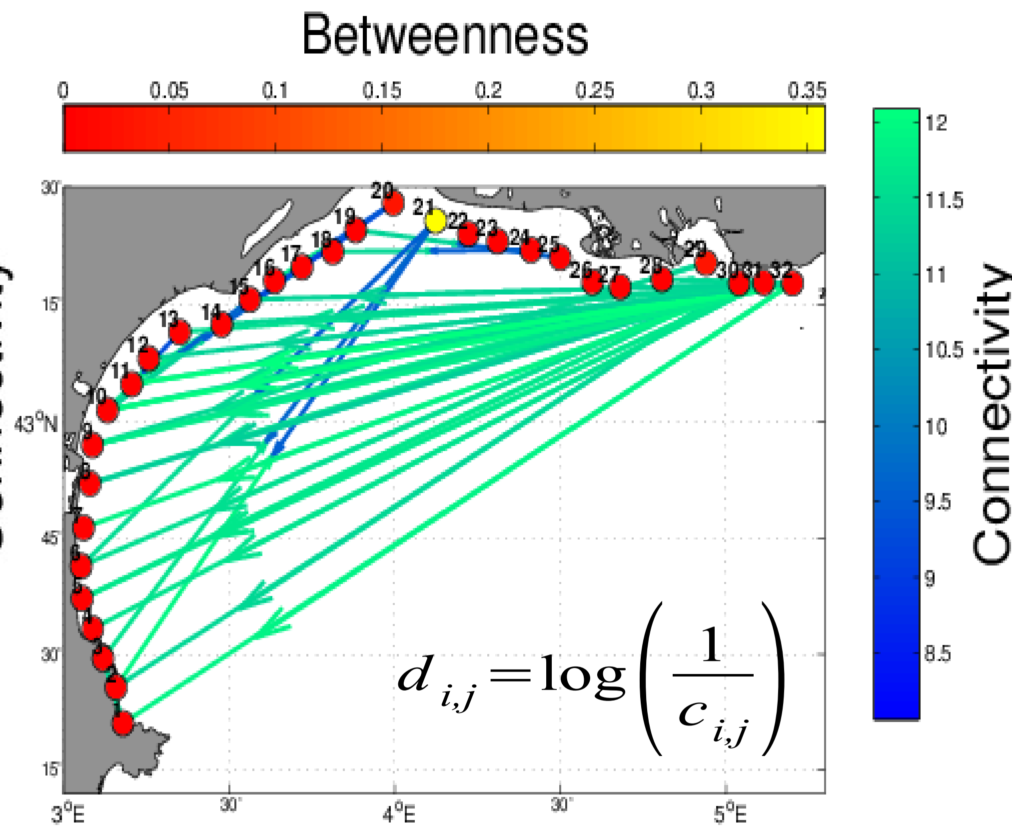
Graph theory for **key site selection:**

pay great attention on the metrics !!!

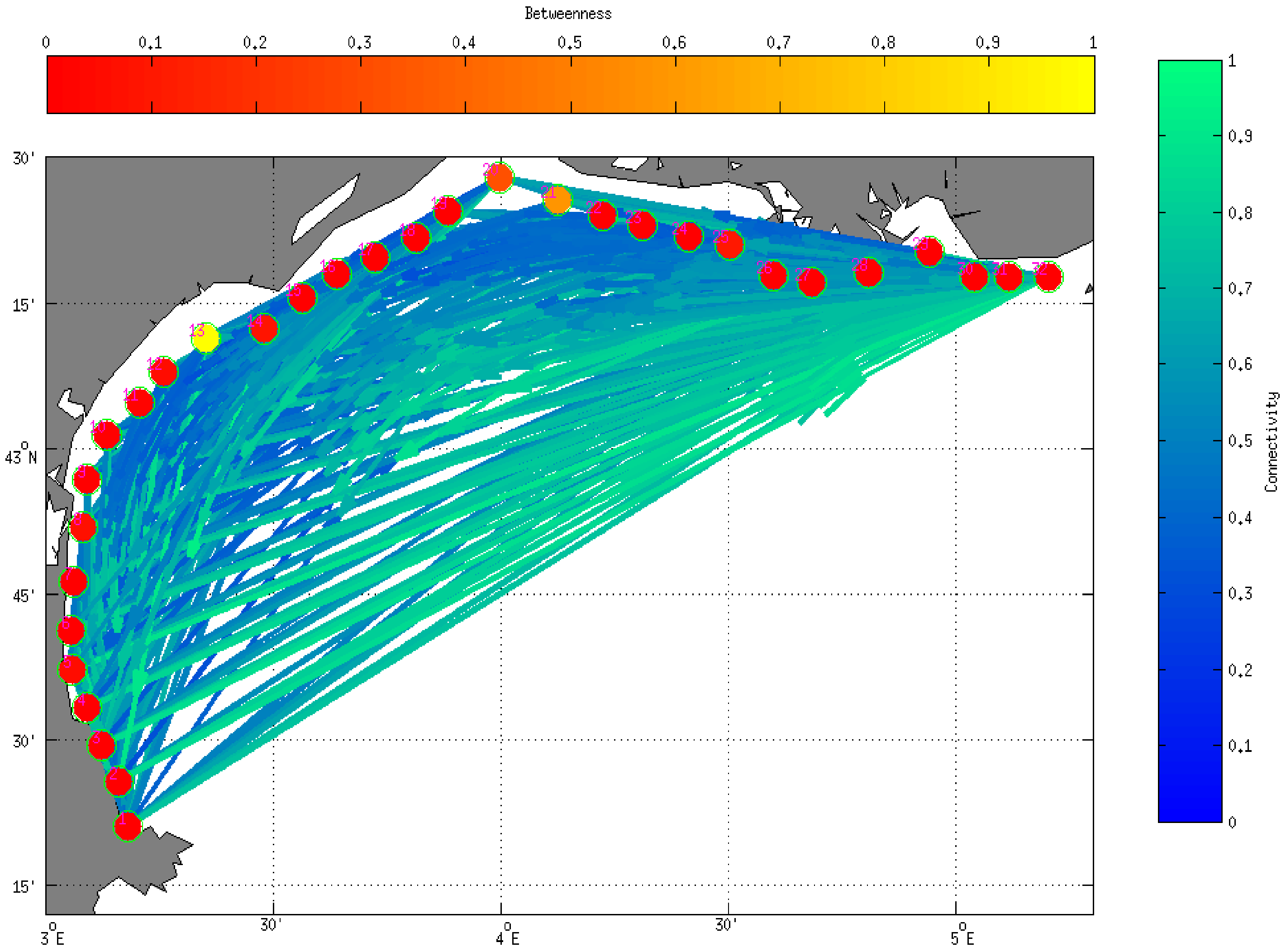
Betweenness: identifies gateways (i.e. nodes through which most of the propagules flow, e.g. soft-bottom polychaete in the Gulf of Lion)



Andrello et al., Low connectivity between Mediterranean Marine Protected Areas: a biophhysical modeling approach for the dusky grouper: *Epinephelus Marginatus*, *PlosOne*, 2013

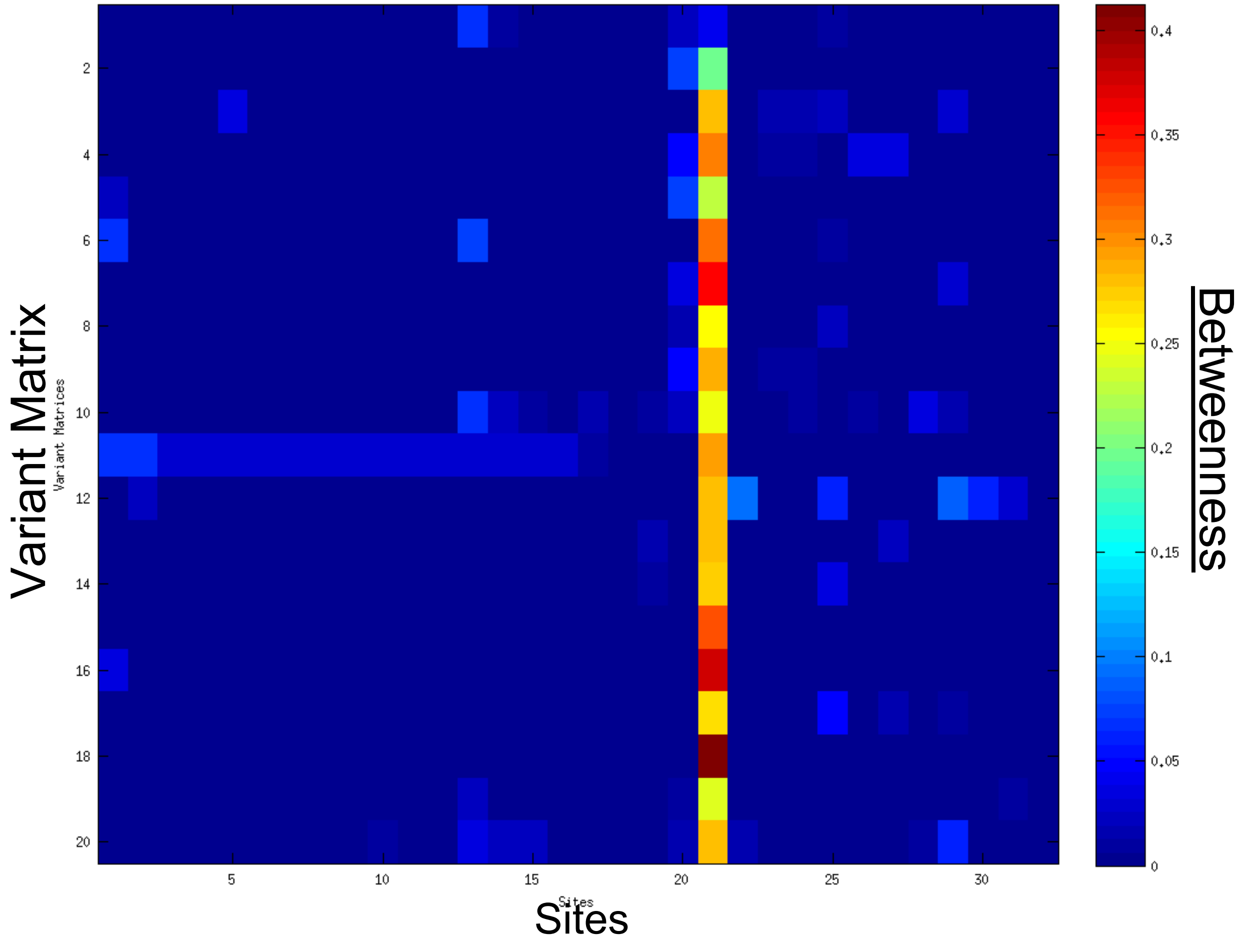


Costa et al., Estimation of connectivity in marine biological networks: Graph Theory versus Metapopulation Models. The Gulf of Lion study case, in preparation

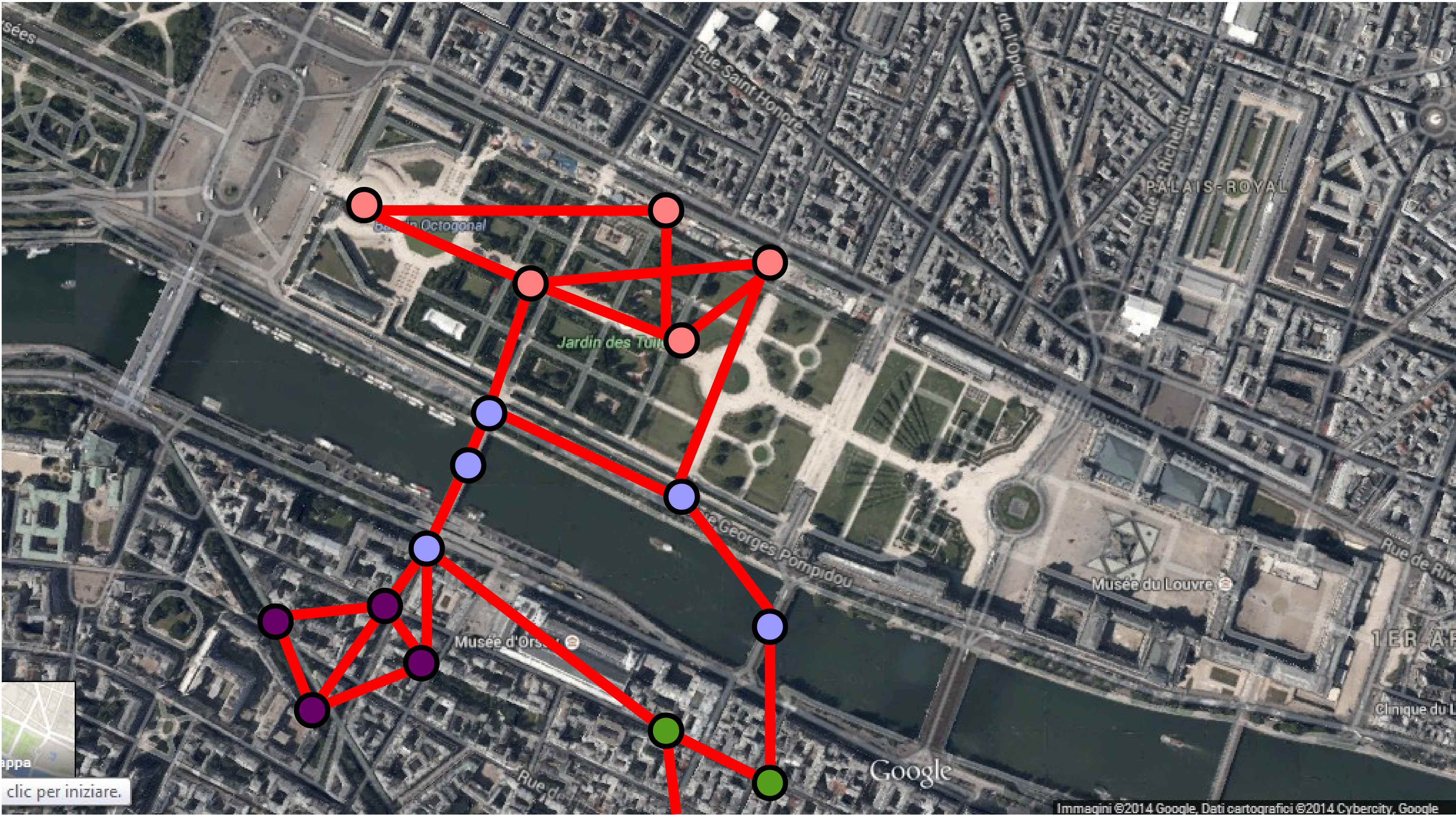


Betweenness
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Connectivity
30'
15'
43°N
45'
30'
15'
3°E
30'
4°E
30'
5°E

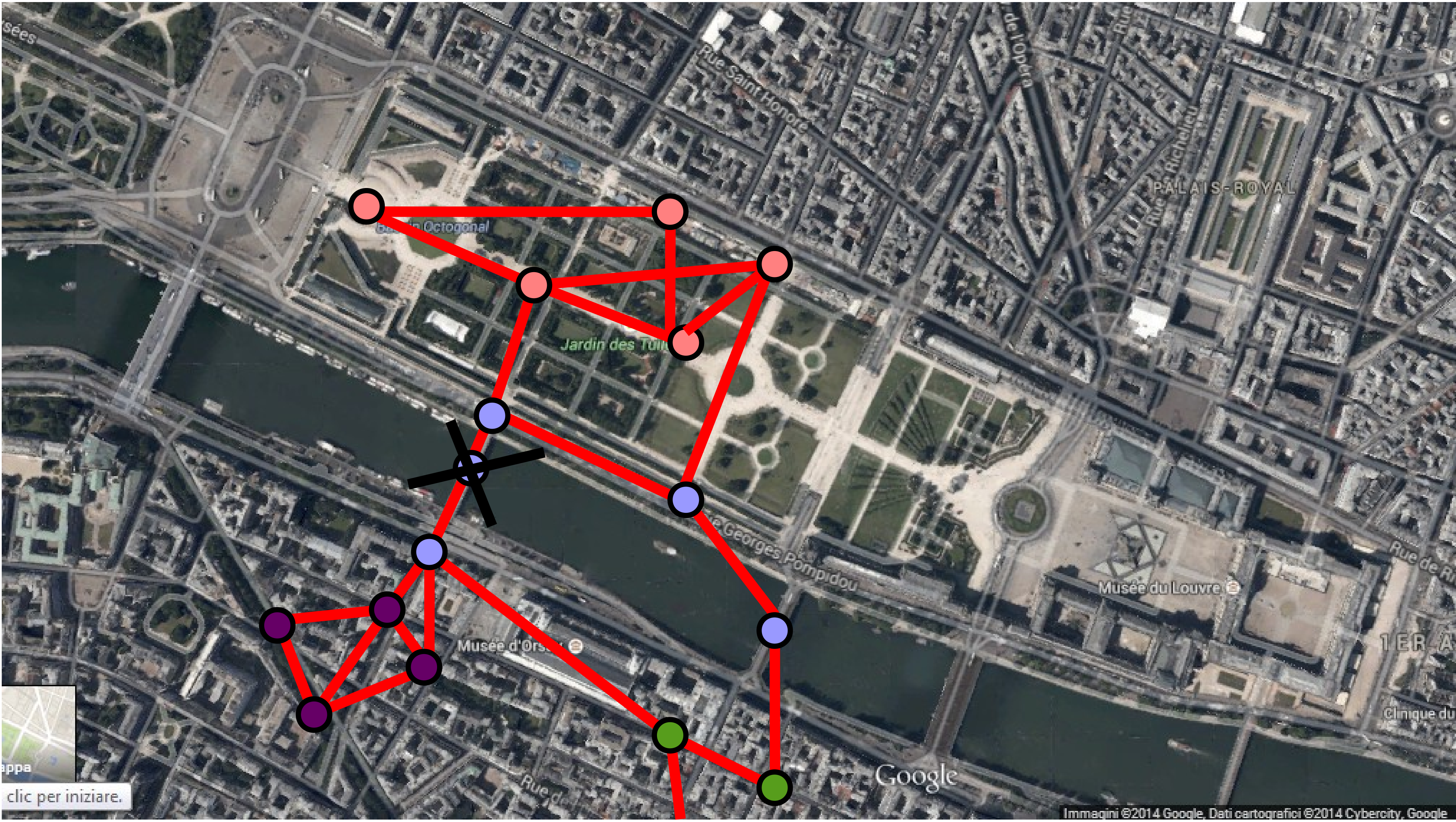
First measure: betweenness

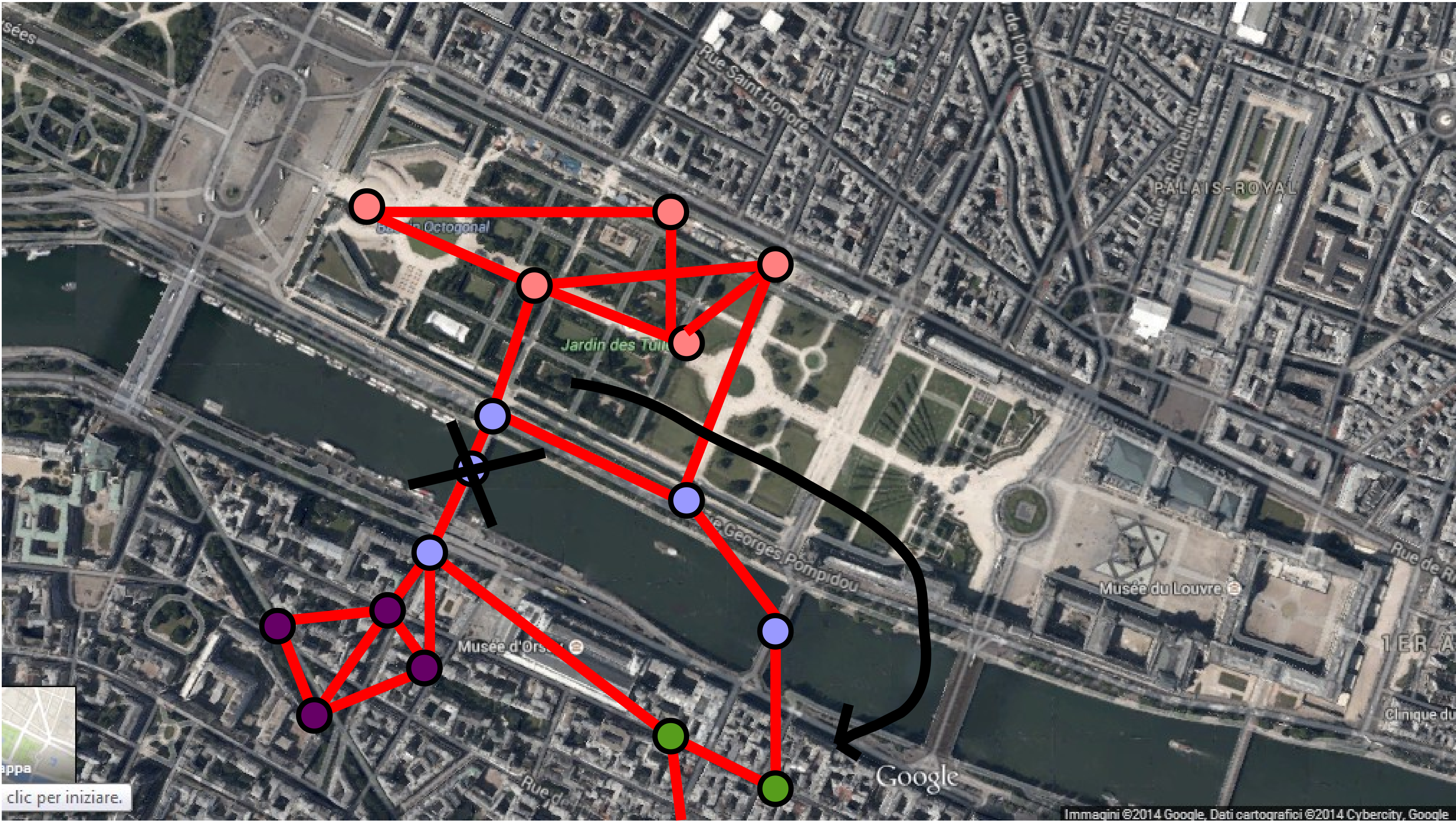


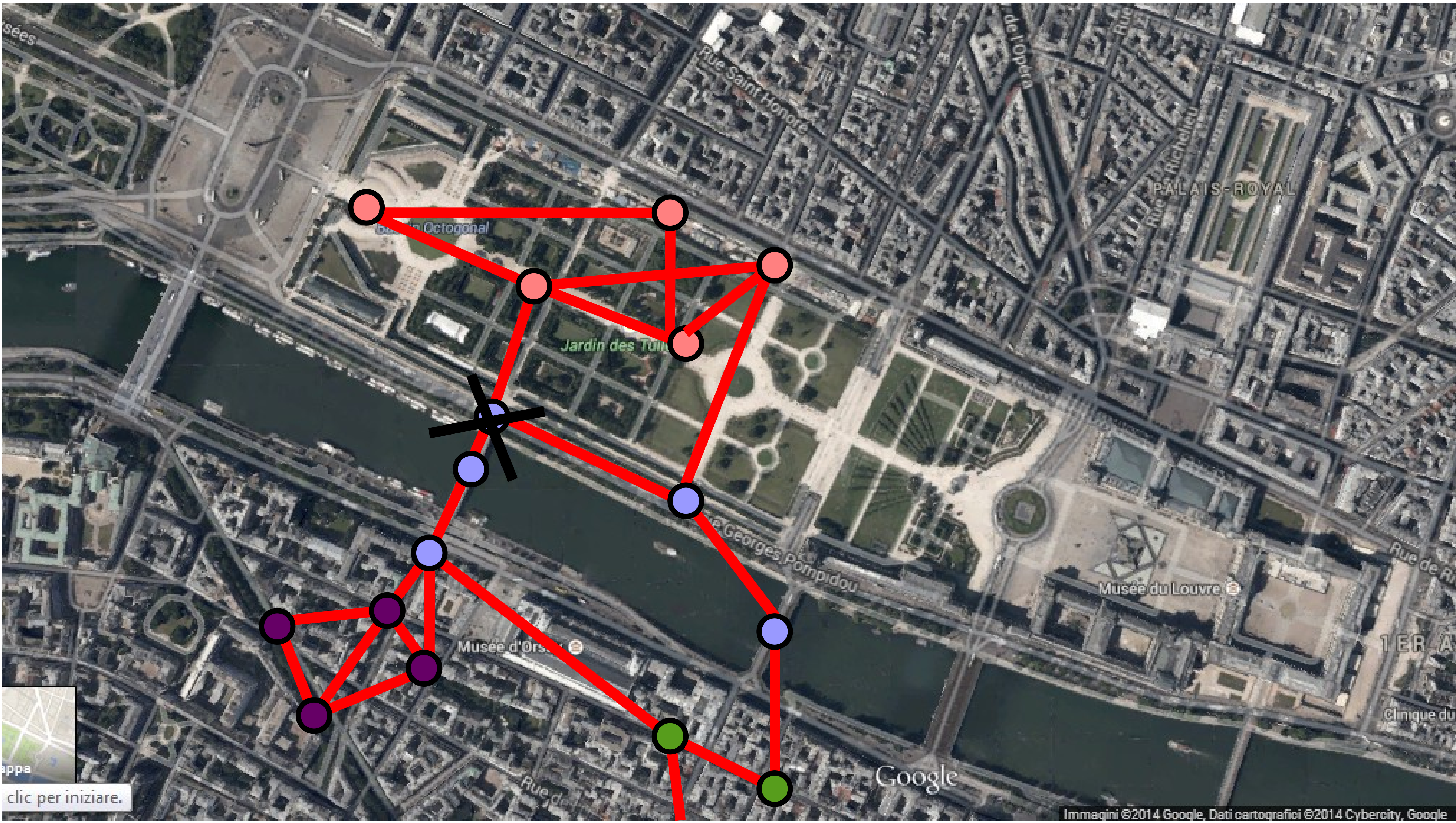
There is **no correspondence between high betweenness and important sites for persistence**.

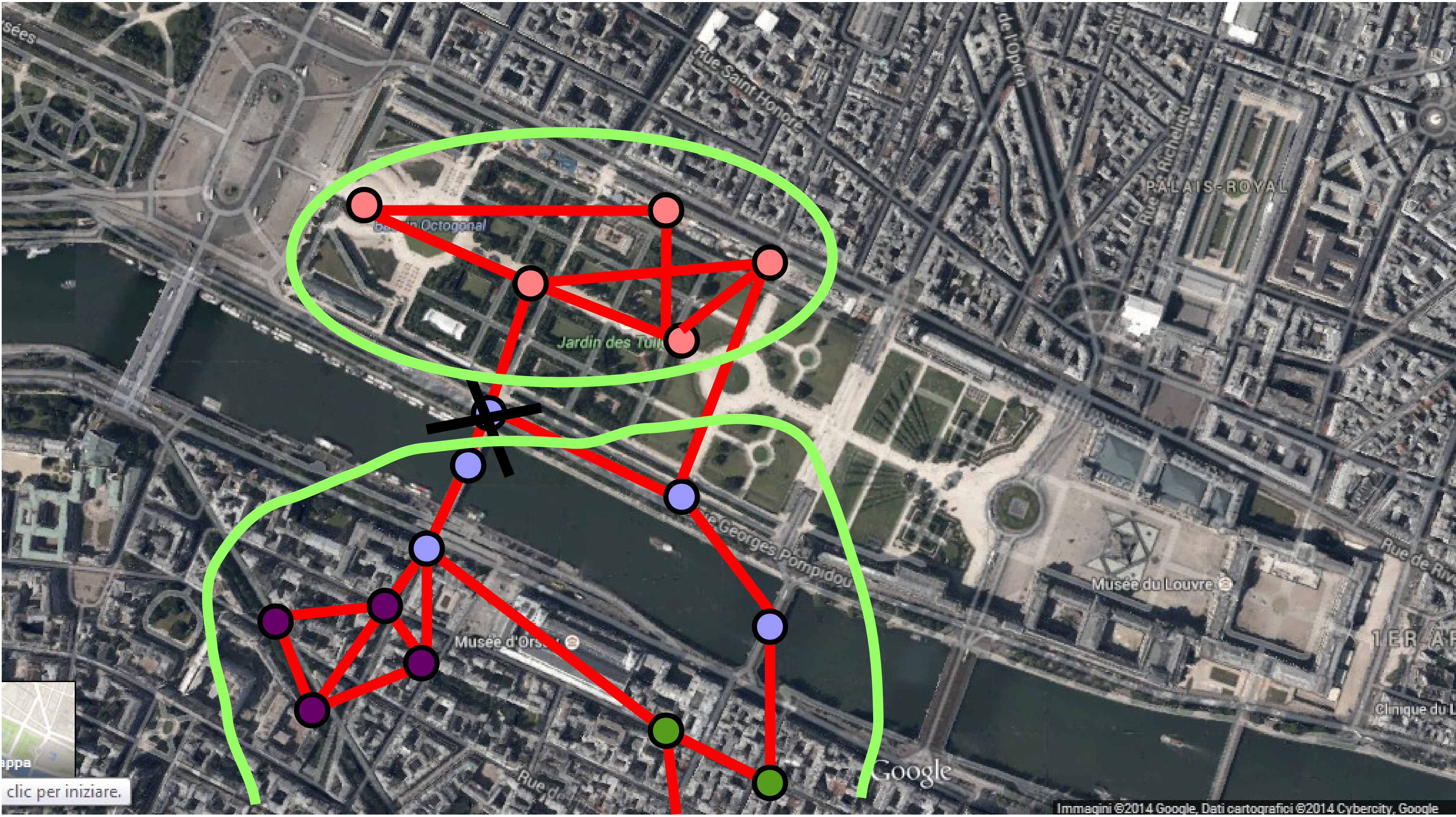


Rue Saint Honoré
PALAIS-ROYAL
Musée du Louvre
Google
clic per iniziare.
Immagini ©2014 Google, Dati cartografici ©2014 Cybercity, Google


Rue Saint Honoré
PALAIS-ROYAL
Musée du Louvre
Google
clic per iniziare.
Immagini ©2014 Google, Dati cartografici ©2014 Cybercity, Google


Rue Saint Honoré
PALAIS-ROYAL
Musée du Louvre
Google
clic per iniziare.
Immagini ©2014 Google, Dati cartografici ©2014 Cybercity, Google

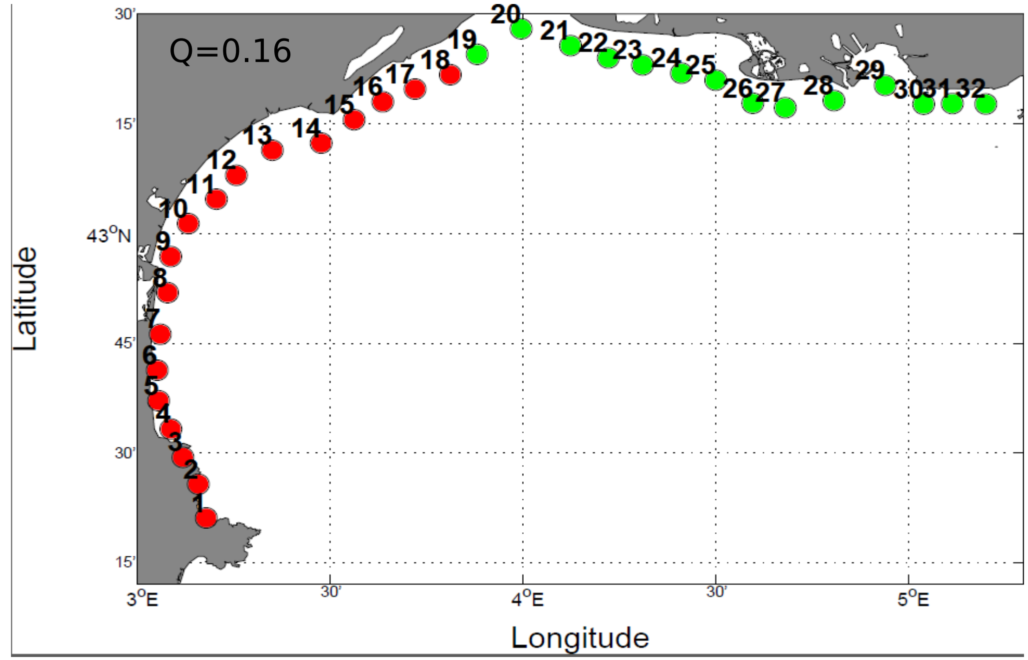

Rue Saint Honoré
PALAIS-ROYAL
Georges Pompidou
Musée du Louvre
Google
clic per iniziare.
Immagini ©2014 Google, Dati cartografici ©2014 Cybercity, Google



More isoleted components !!!

Second measure: modularity

Q is positive



Congruent

with what emerges from

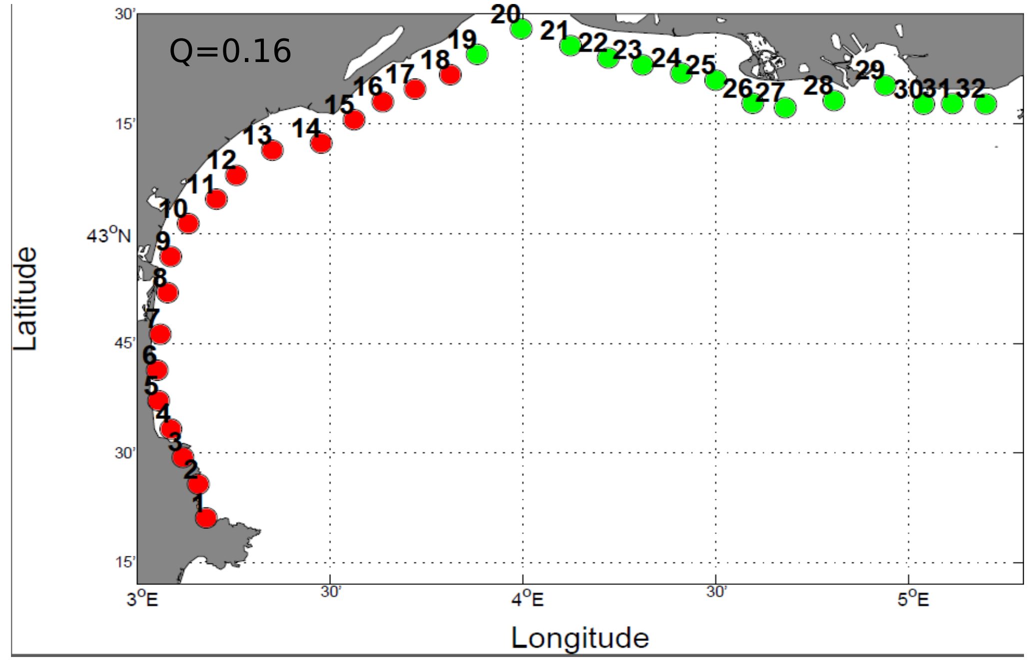
the metapopulation model analysis.

Second measure: modularity

Q is positive

⇓

There are distinct communitites



Congruent

with what emerges from

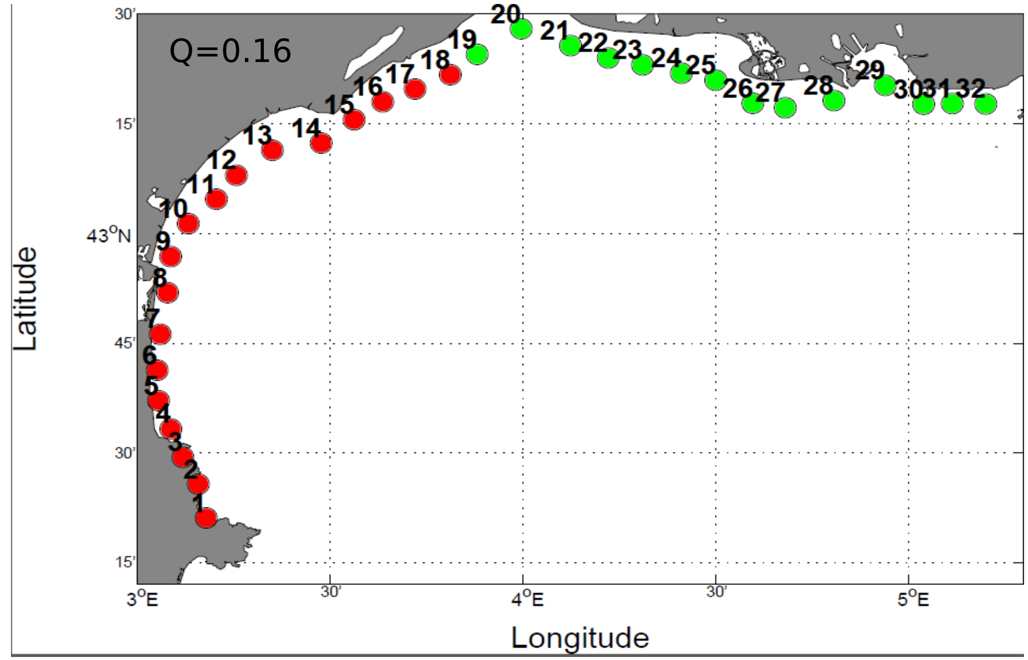
the metapopulation model analysis.

Second measure: modularity

Q is positive

⇓

There are distinct communitites



Congruent

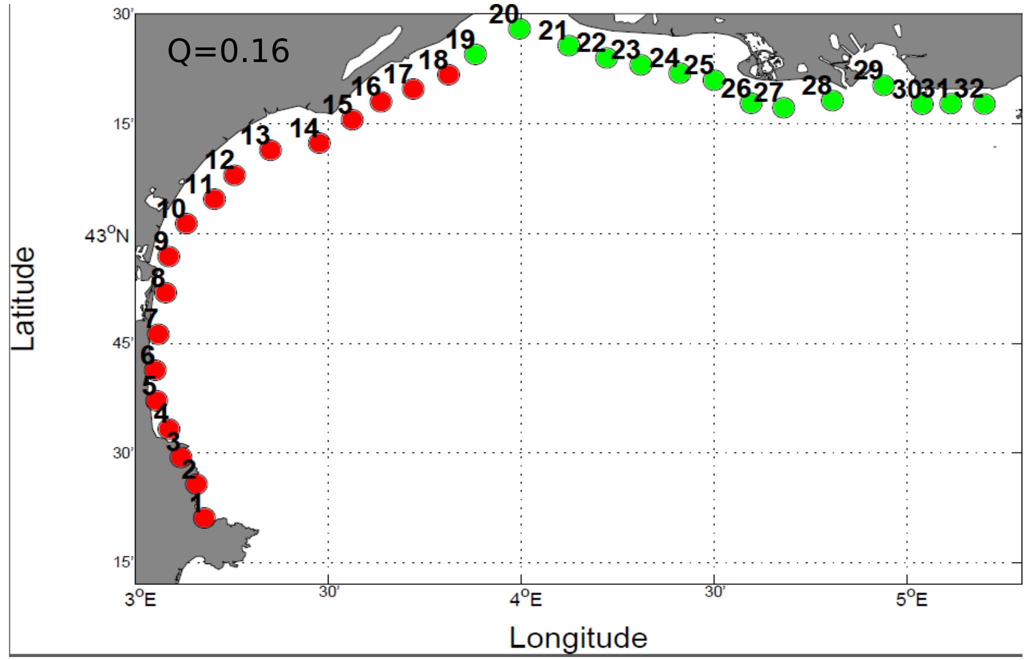
with what emerges from

the metapopulation model analysis.

But what about the rescue effect?

Second measure: modularity

Q is positive **but low**

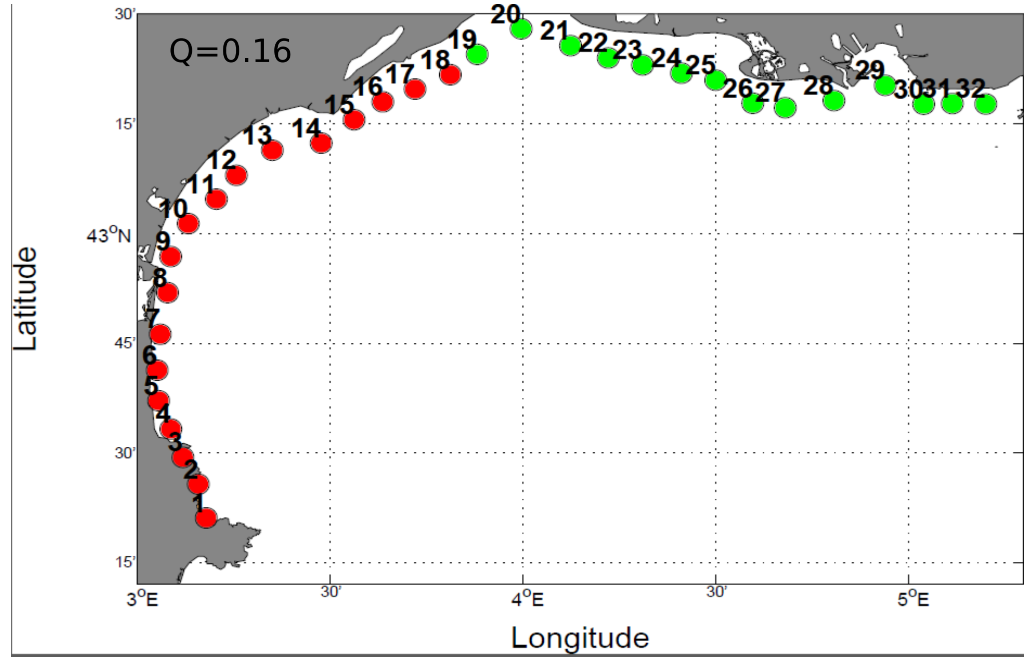


Second measure: modularity

Q is positive **but low**

⇓

high exchange



Second measure: modularity

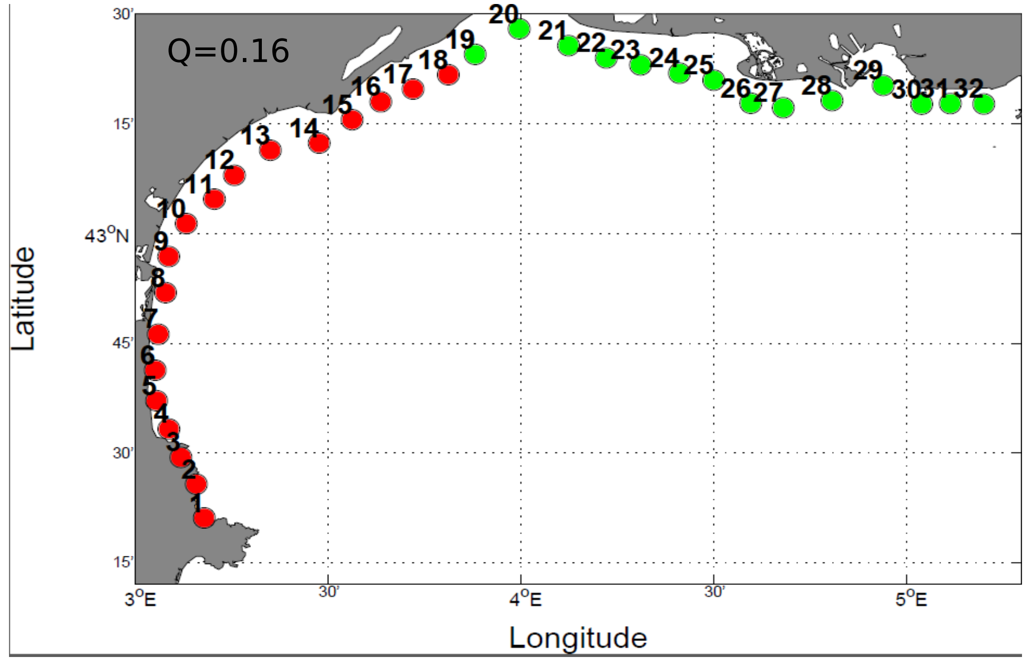
Q is positive **but low**

⇓

high exchange

⇓

rescue effect is possible



Second measure: modularity

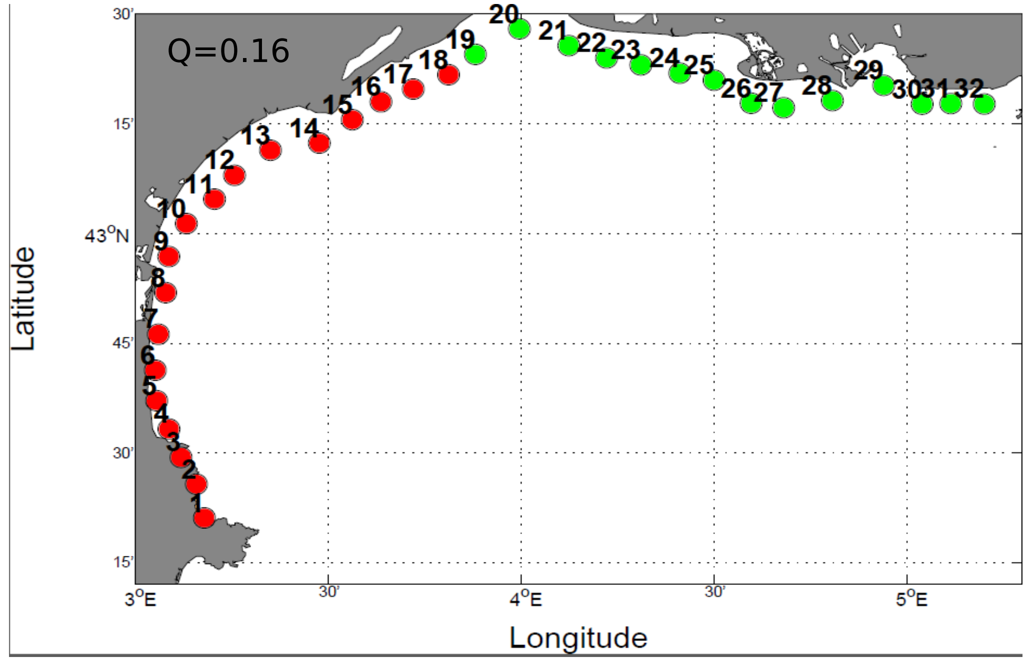
Q is positive **but low**

⇩

high exchange

⇩

rescue effect is possible



Graph analysis reflects well
the organization of polychate population in the gulf
resulting from Guizien et al.(2014)

Third measure: minimum cycles

We search autosufficient sites

using

a simplified metapopulation model

without density dependent factors:

$$N_{t+1} = \left(b_j C_{ij} a_i\right) \cdot N_t$$

Hasting (2006) demonstrates that

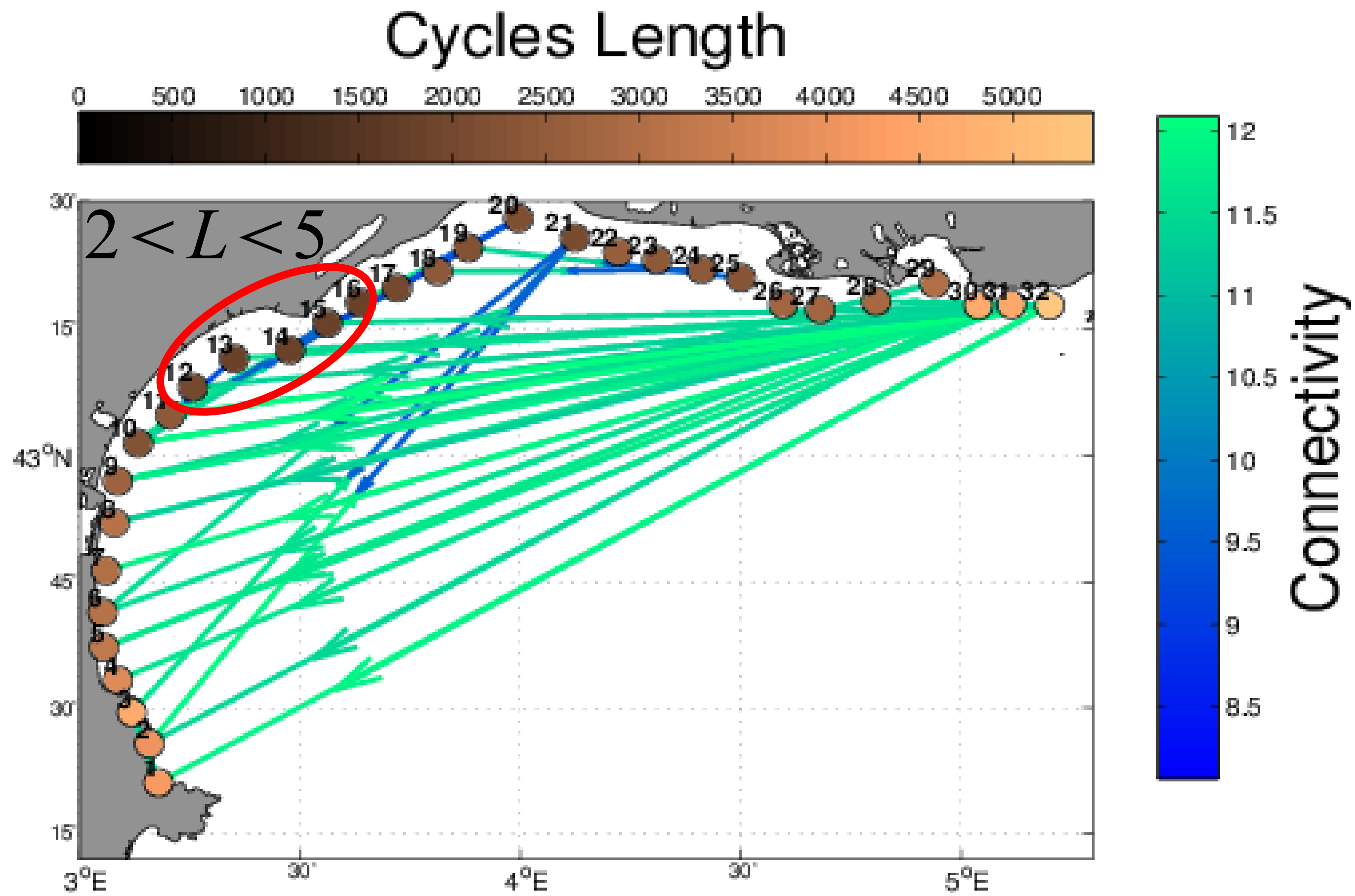
self sufficient sites have **cycles** starting from it.

The more the cycle is probable,

the more the site is self-sufficient:

it is a source of *larvae*.

Third measure: minimum cycles



Nodes **13-16** composes the majority of the cycles.

Core of well connected nodes in the centre of the GoL.

In the meatapop. model nework collapses when nodes around Sète (node 18) are removed.

Clustering coefficient:

$$C(v_i) = E / (V*(V-1) / 2)$$

V = # of neighbors

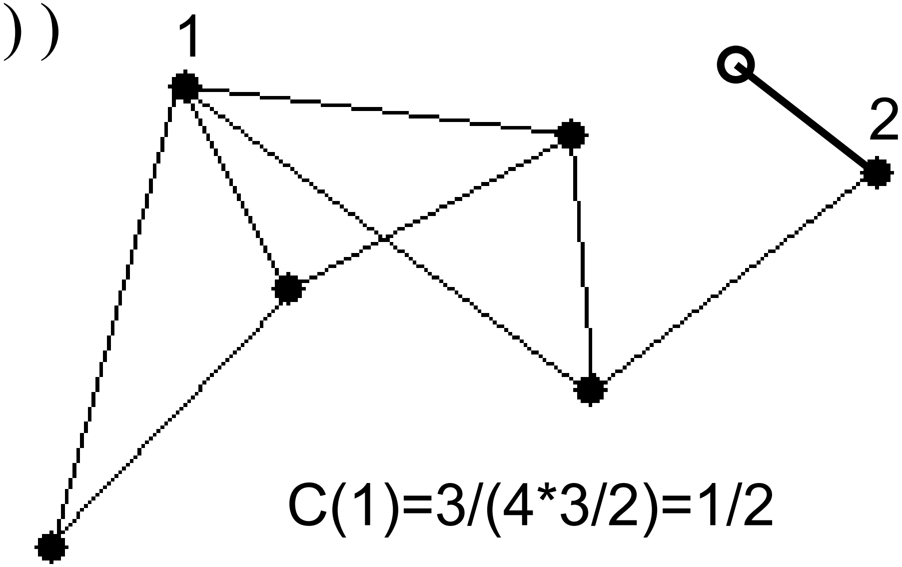
E = # of connections between V

cluster: a set of nodes that are “close” to each other, with respect to a certain measure of distance or similarity

A possible measure is the number of neighbors of a nodes that are also neighbors between themself (as Facebook ;-))

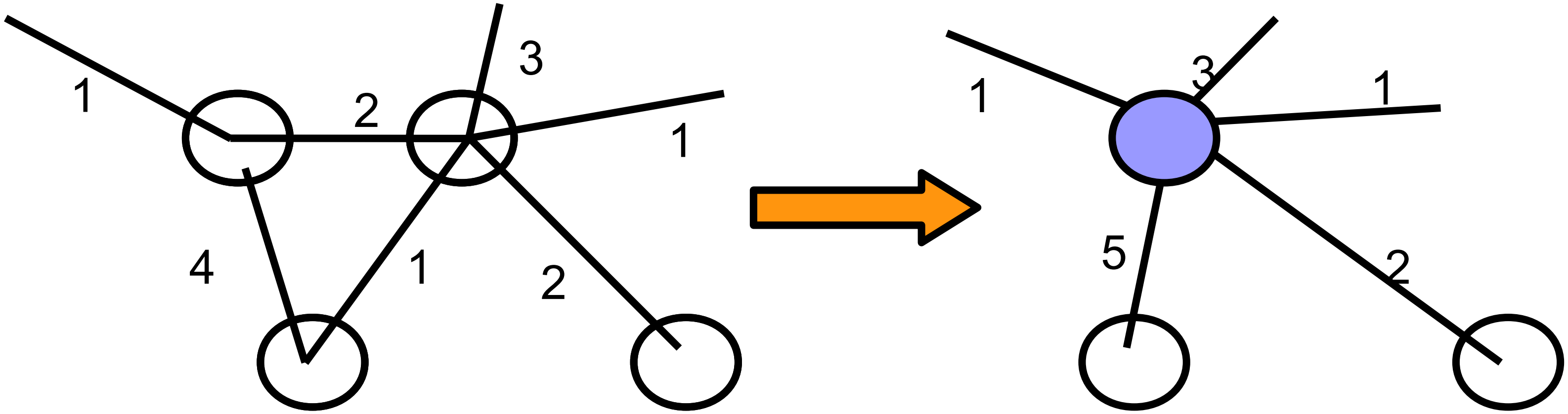
↓

We can identify the more well connected areas of sea, that is where <u>highly connected communities</u> of a particular species .



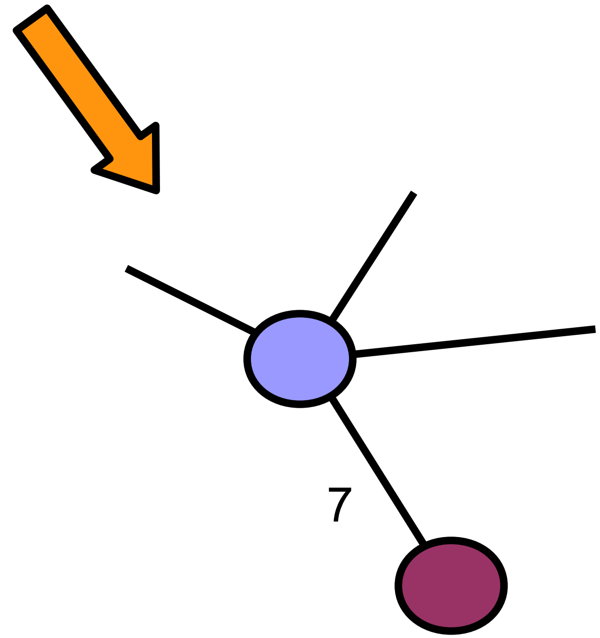
C(1)=3/(4*3/2)=1/2

C(2)=0



F usion

a lgorithm



Stop: - no more increment of the metric is possible

- after a certain number of steps

→ **Hierarchy of clusters!** ←

Blondel et at (2008),

"Fast unfolding of communities in large networks"

Increase modularity

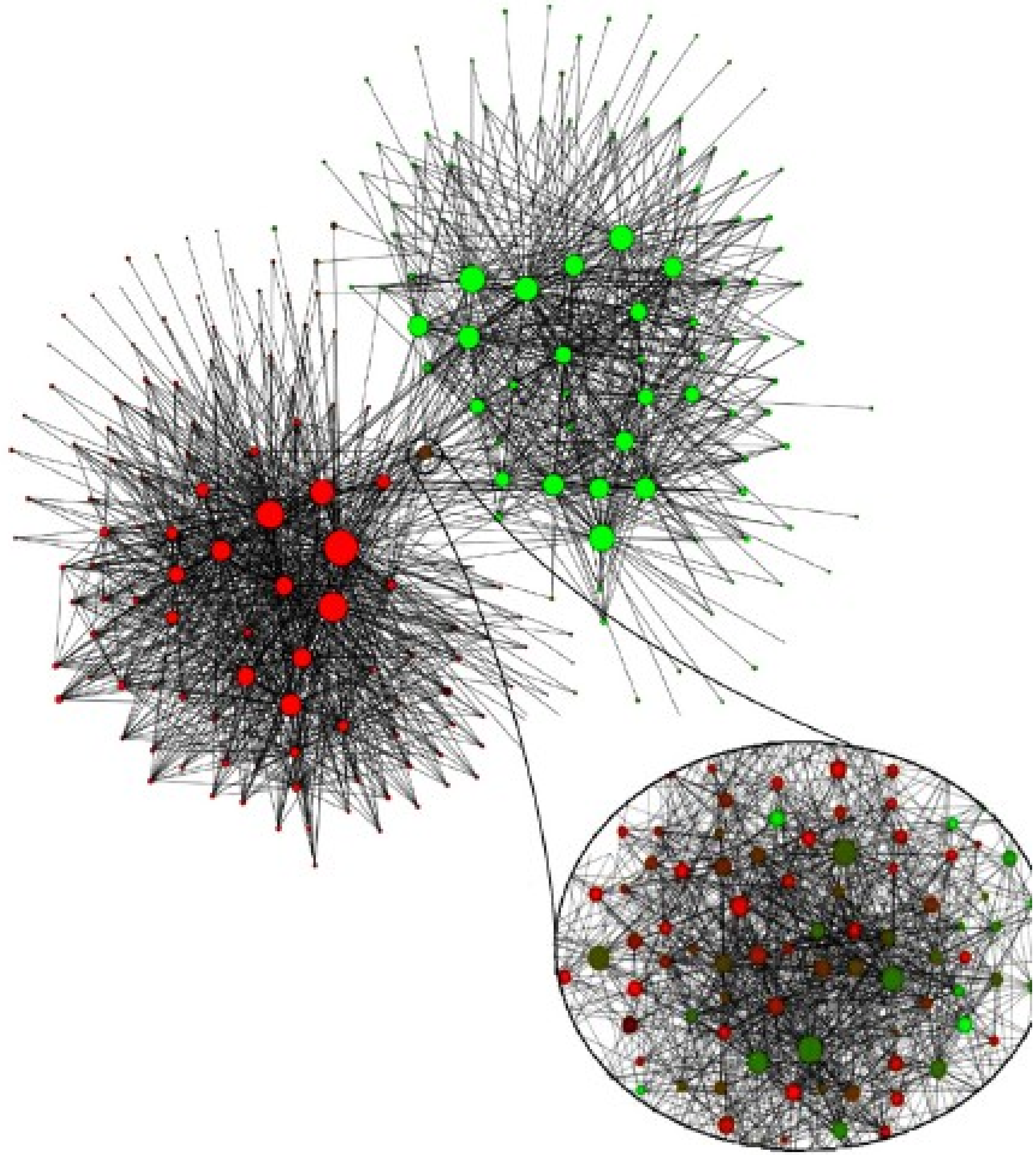


Figure 2. Graphical representation of the network of communities extracted fro
a Belgian mobile phone network. About 2 million customers are represented
this network. The size of a node is proportional to the number of individuals in t
corresponding community and its colour on a red–green scale represents the ma
language spoken in the community (red for French and green for Dutch). Only t
communities composed of more than 100 customers have been plotted. Noti
the intermediate community of mixed colours between the two main langua
clusters. A zoom at higher resolution reveals that it is made of several su
communities with less apparent language separation.

"Network analysis identifies weak and strong links in a metapopulation system"

A Rozenfeld et al 2008:

→ *Posidonia genetic analysis and robustness analysis of the network*

Collaboration with A.Chenuil and A.Weber at IMBE

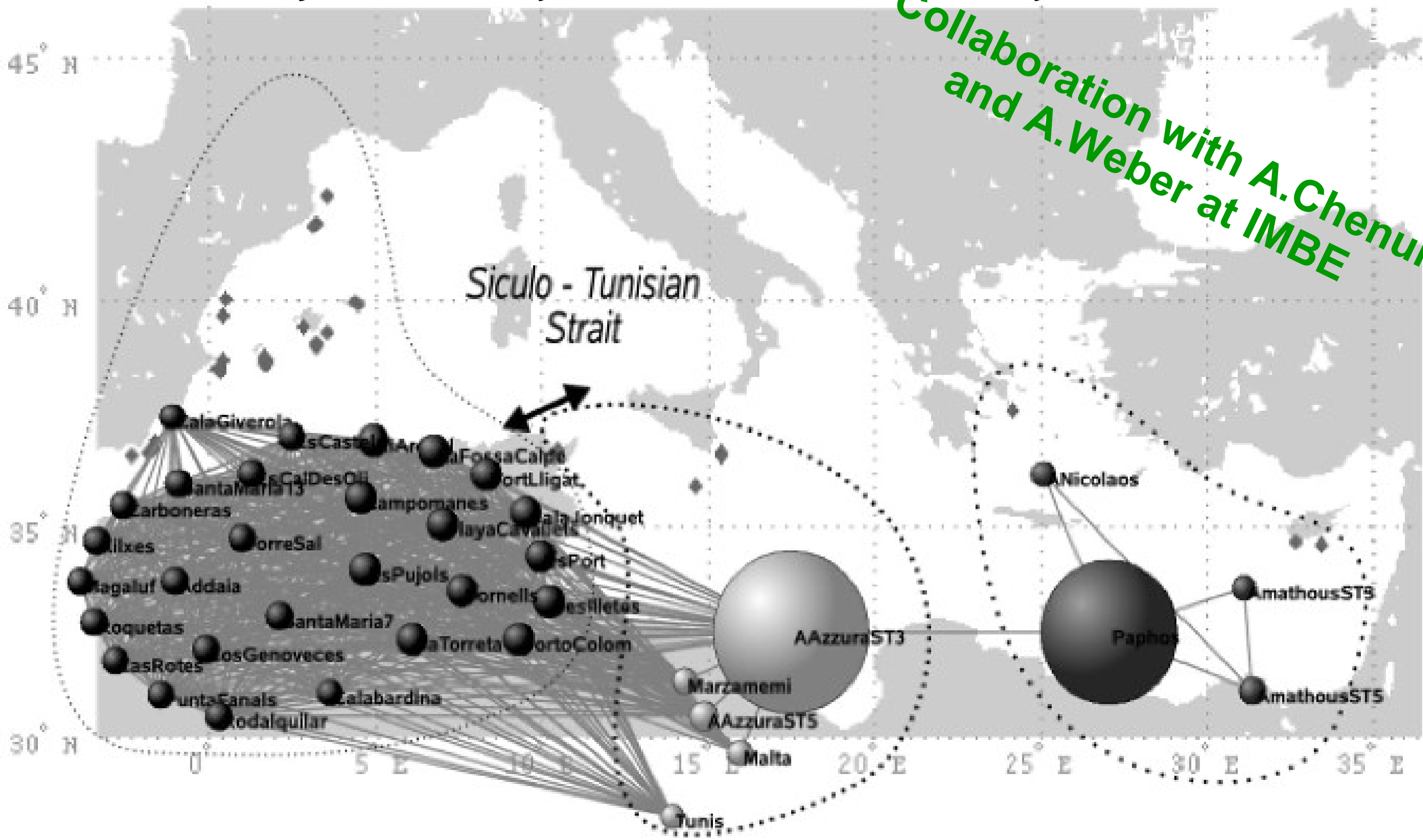


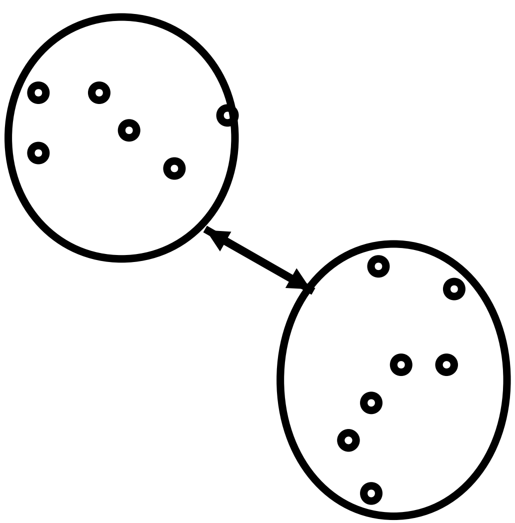
Fig. 1. The network of Mediterranean meadows in which only links with Goldstein distances smaller than the percolation distance *Dp* = 91 (see Fig. 5) are present. Nodes representing populations are roughly arranged according to their geographic origin. The precise geographic locations are indicated as diamonds in the background map. One can identify 2 clusters of meadows, corresponding to the Mediterranean basins (east and west), separated by the Siculo–Tunisian Strait. The size of each node indicates its betweenness centrality (i.e., the proportion of all shortest paths getting through the node).

The genetic distance is symmetric, the current-based one is not.

Which definition of distance we shall use?

I thought to 3 possibilities:

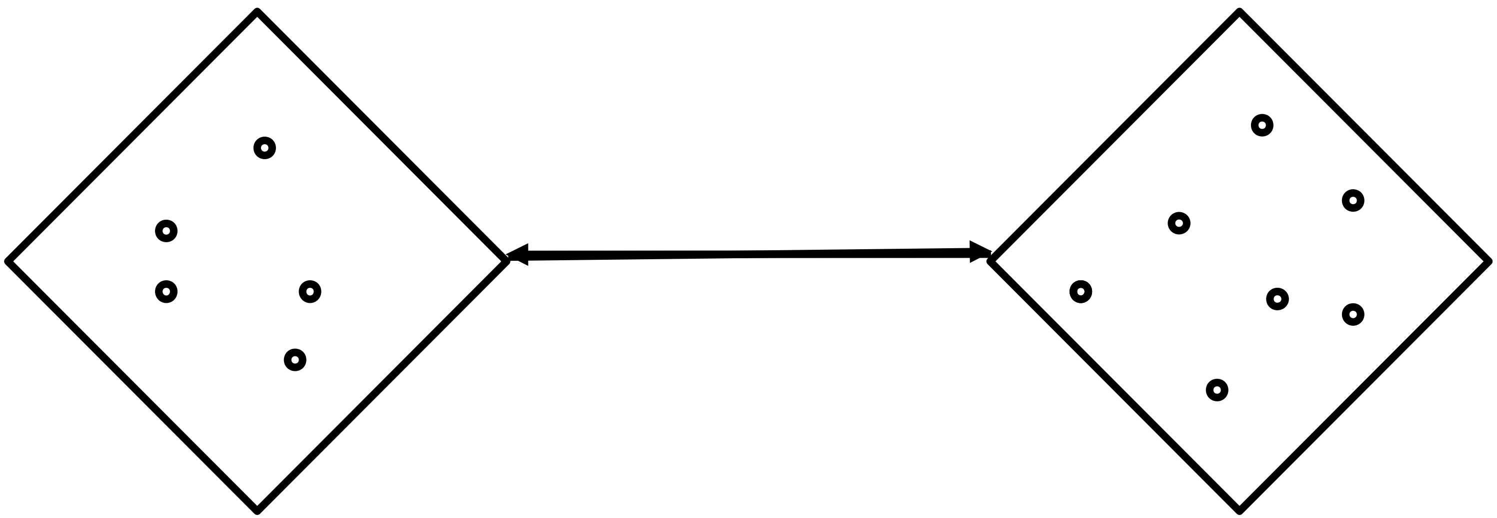
1) a mean genetic variation of the population respect to a model and then the difference between the populations:



$$D_1(1,2) = \left| \frac{\sum_{i \in 1} (g_i - \bar{g})}{I} - \frac{\sum_{j \in 2} (g_i - \bar{g})}{J} \right| = D_1(2,1)$$

2) all vs all:

$$D_2(1,2)=\frac{\sum_{i\in 1}\sum_{i\in 2}(g_i-g_j)}{I\cdot J}=D_2(2,1)$$



(?)*g* is a vector in the 4-dimensional genetic space

Relatively to case 1: if we have Div(G)>>0 we have a genetic differentiation(?)

3) an <u>asymmetric</u> combination of the precedent methods:

$$D_3(1,2) = \frac{D_2(1,2) - \sum_{k,i \in 1 \wedge k \neq i} (g_i - g_k)}{I \cdot J} \neq D_3(2,1)$$

The smaller value points towards the more variable site: it is a sourse of genetic variability.