

Analyse de réseaux: Dolphins, Preferential Attachment et Détection de communautés.

Kenson FAVEUR

2025-02-22

Introduction

L'analyse des réseaux complexes est un domaine clé en science des données, utilisé dans des applications variées telles que les réseaux sociaux, les écosystèmes et l'analyse des interactions biologiques. Ce projet explore plusieurs aspects des réseaux :

- La visualisation du réseau des **Dolphins** et l'analyse de leur connectivité.
- La génération d'un **graphe de preferential attachment** et l'étude de ses propriétés.
- L'analyse des **mesures de centralité** dans ce graphe.
- La **détection de communautés** sur un graphe issu d'un dataset.

Objectifs

Les objectifs de ce projet sont : 1. Visualiser et analyser la structure d'un réseau réel (dolphins). 2. Générer un réseau artificiel basé sur l'attachement préférentiel. 3. Étudier différentes **mesures de centralité** et leur distribution. 4. Appliquer des méthodes de **détection de communautés** et comparer les résultats.

Méthodologie

Visualisation du réseau "Dolphins"

Nous chargeons le réseau dolphins à partir du package {onadata}, puis nous colorons les nœuds en fonction de leur degré.

```
# Chargement Les packages nécessaires.  
library(igraph)  
library(onadata)  
library(ggplot2)  
  
# Chargement du réseau  
data("dolphins", package = "onadata")  
g <- graph_from_data_frame(dolphins)
```

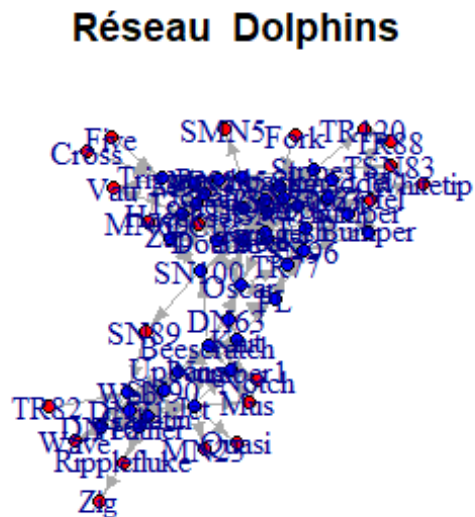
```

# Calcul des degrés des noeuds.
deg <- degree(g)

# Définition des couleurs
node_colors <- ifelse(deg < 4, "red", "blue")

# Plot du réseau
plot(g, vertex.color = node_colors, vertex.size = 8, edge.arrow.size = 0.5,
     main = "Réseau Dolphins")

```



Analyse et interprétation

Le graphe “dolphins” est un réseau réel représentant les interactions entre dauphins. Les nœuds sont colorés selon leur degré : Rouge : nœuds avec un degré < 4 (peu connectés). Bleu : nœuds avec un degré ≥ 5 (bien connectés).

Interprétation: Certains dauphins jouent un rôle central dans le groupe. La distribution des degrés n’est pas uniforme, indiquant une organisation sociale non aléatoire. Les nœuds rouges (faible degré) pourraient représenter des individus moins actifs ou marginaux dans le réseau.

Graphe de Preferential Attachment et analyse de la centralité.

Nous générons un graphe Barabási-Albert et modifions la taille des nœuds en fonction de leur betweenness.

```

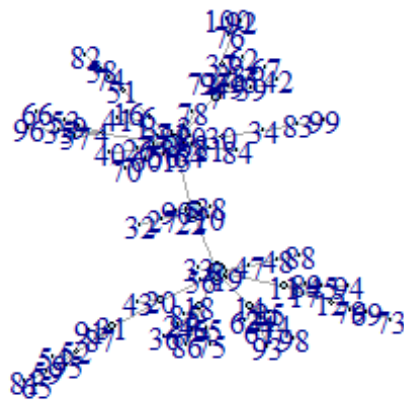
# Génération du graphe
g_pa <- barabasi.game(100, directed=FALSE)

# Calcul de la centralité de betweenness
betweenness_values <- betweenness(g_pa)

# Plot du réseau
plot(g_pa, vertex.size = betweenness_values / max(betweenness_values) * 10 +
2,
     vertex.color = "lightblue", edge.arrow.size = 0.5,
     main = "Réseau de Preferential Attachment (100 nœuds)")

```

Réseau de Preferential Attachment (100 nœuds)



Analyse et interprétation.

Un modèle de Barabási-Albert a été utilisé pour créer un réseau “scale-free”. Les nœuds sont dimensionnés en fonction de leur centralité de betweenness.

Interprétation Les réseaux “scale-free” ont quelques nœuds très connectés (hubs), ce qui reflète des phénomènes du monde réel (réseaux sociaux, Internet, etc.). Les nœuds avec une betweenness élevée contrôlent les flux d’information, suggérant leur importance stratégique. Contrairement au réseau des dauphins, ici les connexions suivent une loi de puissance, où les hubs dominent.

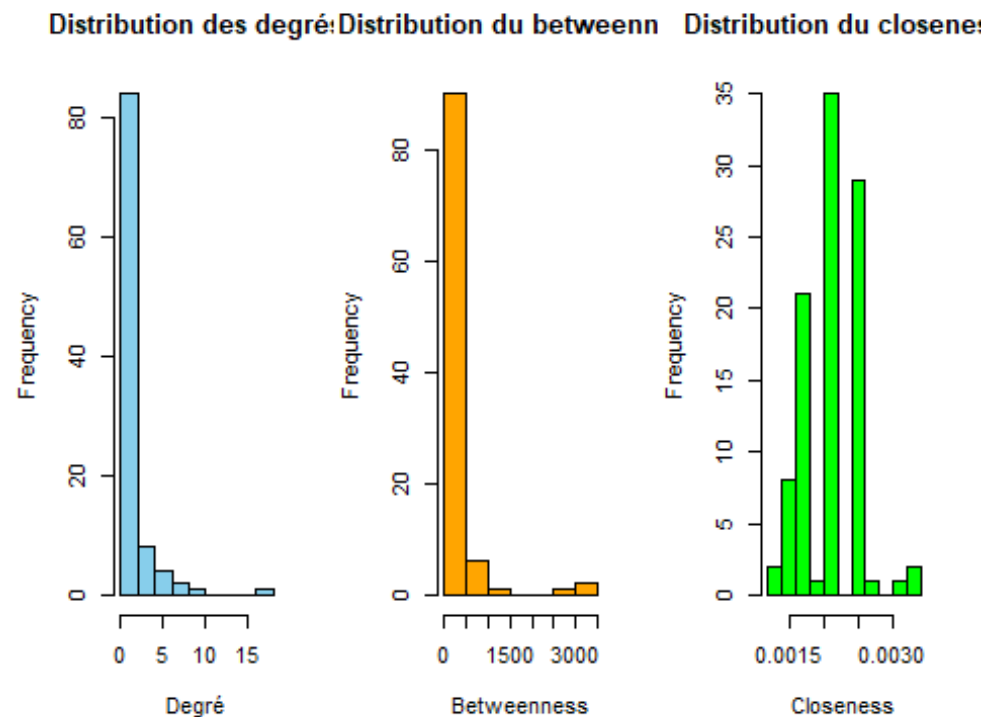
Histogramme des 3 mesures de centralités

Nous analysons la distribution des degrés, betweenness et closeness.

```
# Calcul des centralités
degree_values <- degree(g_pa)
closeness_values <- closeness(g_pa)

# Affichage des histogrammes
par(mfrow = c(1, 3)) # 3 graphiques en ligne

hist(degree_values, col = "skyblue", main = "Distribution des degrés", xlab = "Degré")
hist(betweenness_values, col = "orange", main = "Distribution du betweenness", xlab = "Betweenness")
hist(closeness_values, col = "green", main = "Distribution du closeness", xlab = "Closeness")
```



Analyse et interprétation

Trois mesures de centralité sont analysées : Degré (nombre de connexions). Betweenness (contrôle des flux d'information). Closeness (proximité moyenne aux autres nœuds).

Interprétation des histogrammes: La distribution du degré est hétérogène, confirmant la présence de hubs. La distribution du betweenness est encore plus inégale, signifiant

que seuls quelques nœuds contrôlent l'information. Le closeness suit une distribution relativement homogène, indiquant que la plupart des nœuds sont bien reliés au réseau.

En conclusion, les hubs sont essentiels dans un réseau scale-free, tandis que dans le réseau Dolphins, la connectivité est plus homogène.

Détection de communautés et de visualisation.

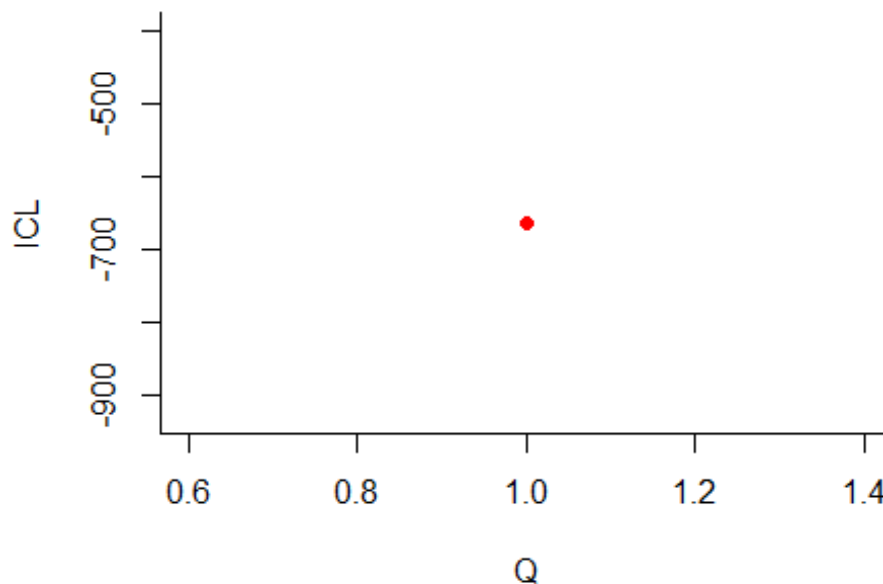
Nous appliquons l'algorithme de Spectral Clustering .

```
# Chargement des packages
library(igraph)
library(onadata)
library(sbm)
library(Matrix)
library(igraphdata)

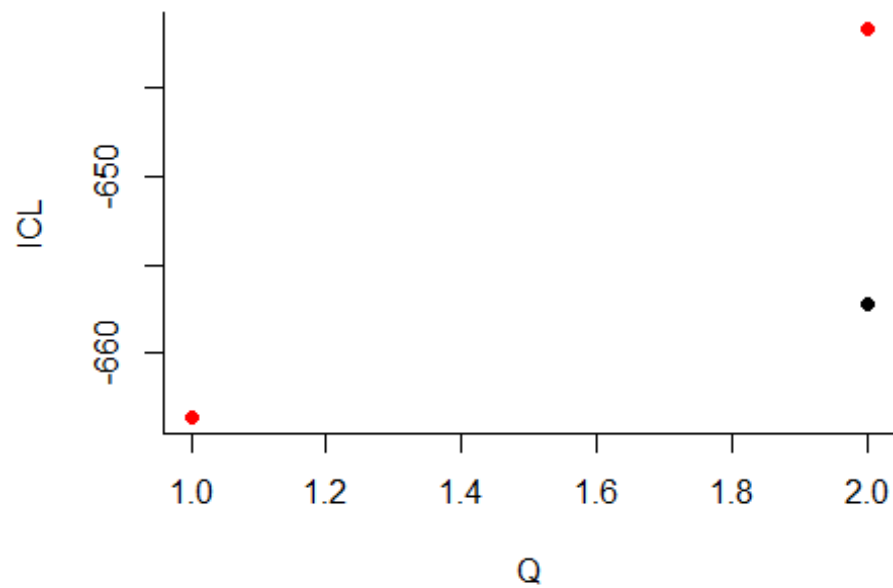
# Conversion en matrice d'adjacence pour SBM
adj_matrix <- as.matrix(as_adjacency_matrix(g))

# Modèle SBM
sbm_model <- estimateSimpleSBM(adj_matrix, model = "bernoulli")

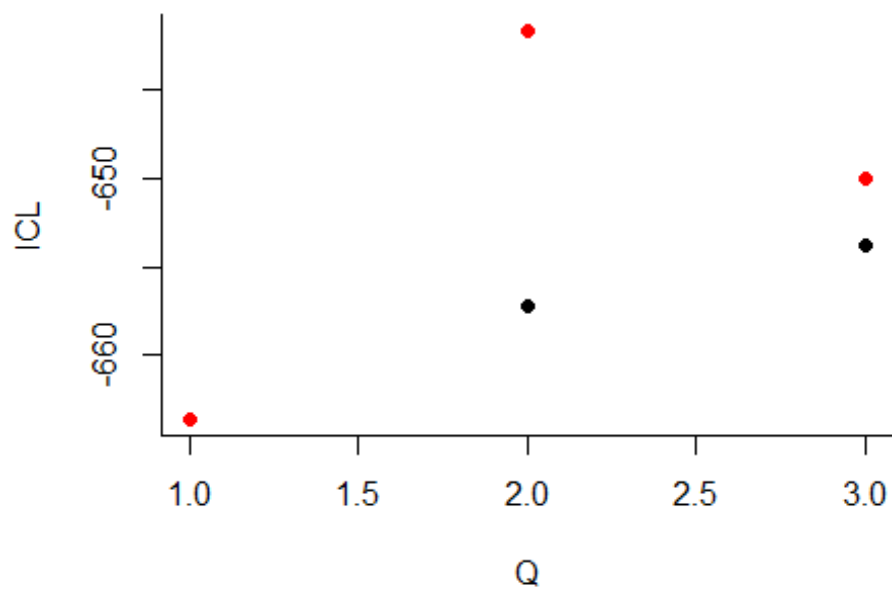
## -> Estimation for 1 groups
##
```



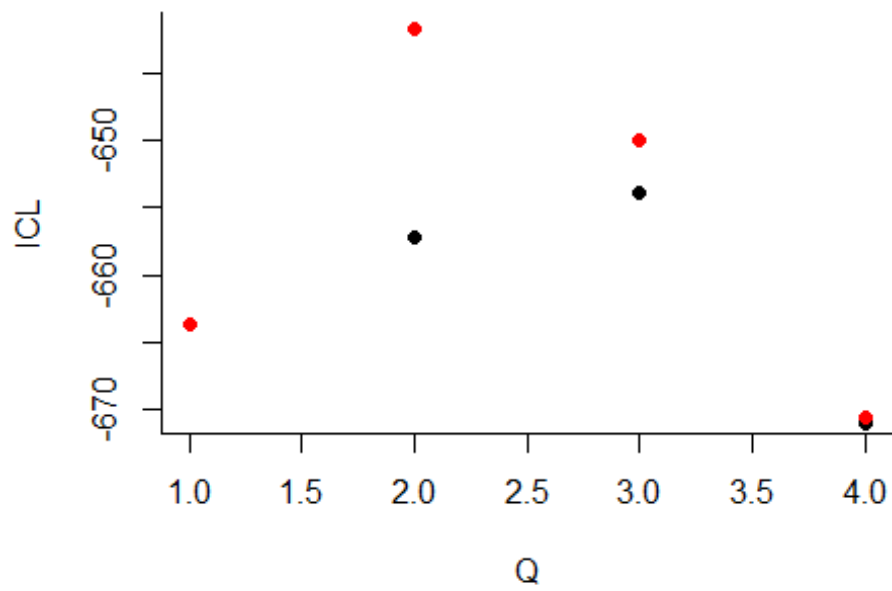
```
## -> Computation of eigen decomposition used for initalizations
##
## -> Pass 1
##     -> With ascending number of groups
##     -> For 2 groups
##
```



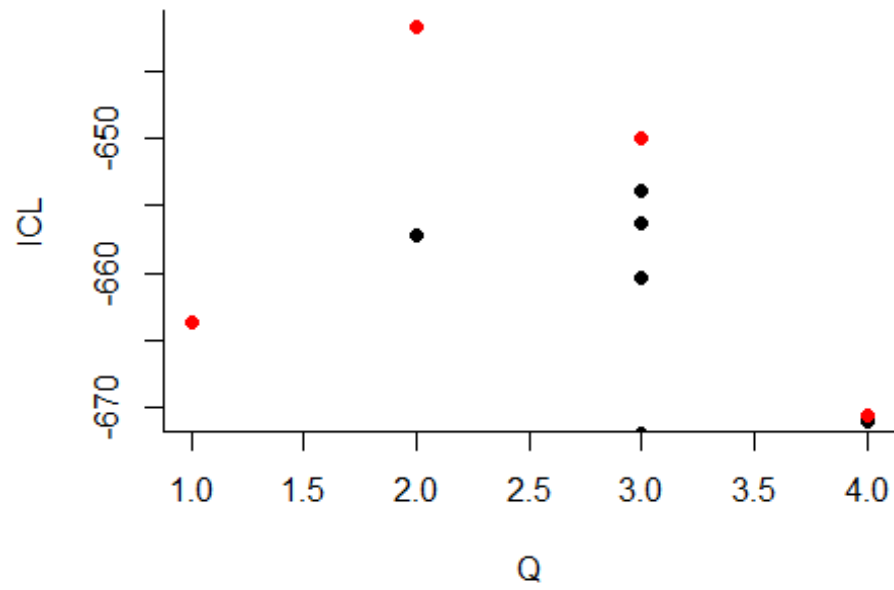
```
##     -> For 3 groups
##
```



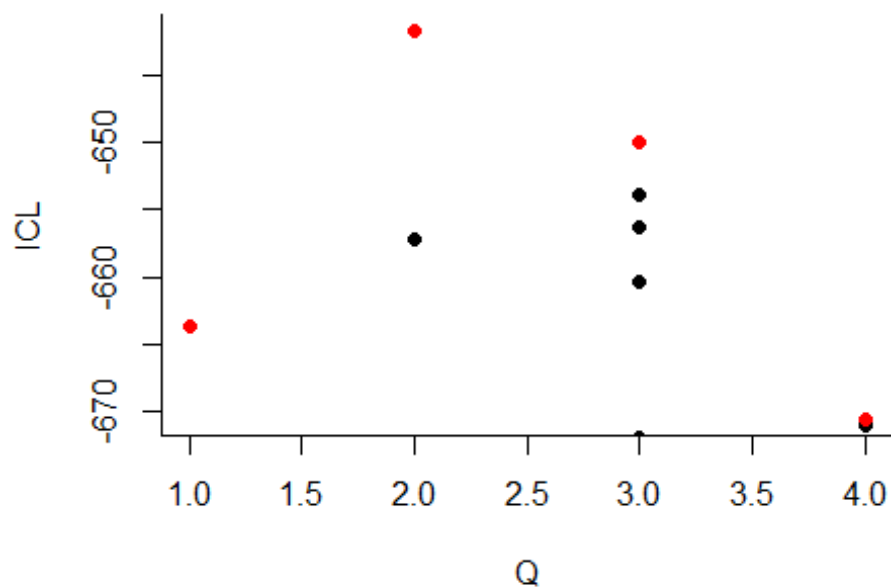
-> For 4 groups
##



```
## -> With descending number of groups
## -> For 3 groups
##
```



```
## -> For 2 groups
##
```

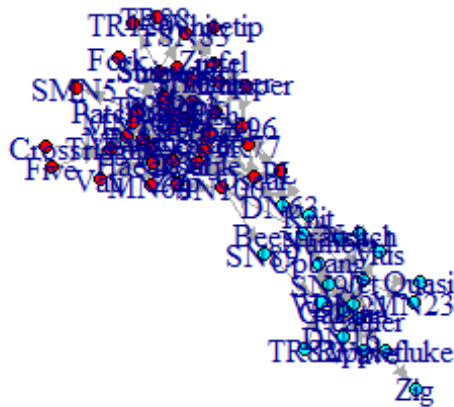
```
## -> Pass 2
##   -> With ascending number of groups
##     -> For 2 groups
##     -> For 3 groups
##     -> For 4 groups
##   -> With descending number of groups
##     -> For 3 groups
##     -> For 2 groups

# Attribution des communautés
sbm_clusters <- sbm_model$memberships

# Définition des couleurs des communautés
node_colors_sbm <- rainbow(length(unique(sbm_clusters)))[sbm_clusters]

# Plot SBM
plot(g, vertex.color = node_colors_sbm, vertex.size = 8, edge.arrow.size =
0.5,
     main = "Communautés détectées par SBM")
```

Communautés détectées par SBM



Détection des communautés avec Spectral Clustering.

```
# Calcul de la matrice de Laplacien normalisé
L <- laplacian_matrix(g, normalized = TRUE)

# Spectral clustering (k = 3 communautés)
eig <- eigen(L)
k <- 3 # Nombre de communautés
spectral_clusters <- kmeans(eig$vectors[, 1:k], centers = k)$cluster

# Attribution des couleurs
node_colors_spectral <-
  rainbow(length(unique(spectral_clusters)))[spectral_clusters]

# Plot Spectral Clustering
plot(g, vertex.color = node_colors_spectral, vertex.size = 8, edge.arrow.size
= 0.5,
     main = "Communautés détectées par Spectral Clustering")
```

Communautés détectées par Spectral Clustering



Détection des communautés avec Louvain pour comparaison.

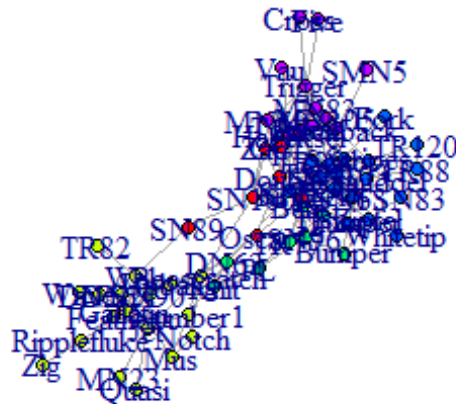
```
# Transformer Le graphe en non orienté
g_undirected <- as.undirected(g, mode = "collapse")

# Détection des communautés avec Louvain
louvain_com <- cluster_louvain(g_undirected)

# Attribution des couleurs
node_colors_louvain <-
rainbow(length(unique(membership(louvain_com))))[membership(louvain_com)]

# Plot Louvain
plot(g_undirected, vertex.color = node_colors_louvain, vertex.size = 8,
edge.arrow.size = 0.5,
main = "Communautés détectées par Louvain (graphe non dirigé)")
```

Communautés détectées par Louvain (graphe non di



Interprétation sur les trois méthodes utilisées: SBM, Spectral Clustering et Louvain.

Trois méthodes ont été testées : SBM (Stochastic Block Model) Spectral Clustering Louvain (comparatif) Comparaison des méthodes SBM modélise les interactions comme des probabilités entre groupes. Spectral Clustering utilise les valeurs propres du Laplacien. Louvain fait la maximisation de la modularité.

Interprétation: Louvain produit les communautés les plus cohérentes, bien adaptées aux réseaux sociaux. SBM est utile si on veut modéliser les relations comme des probabilités d'interaction. Spectral Clustering est intéressant pour des réseaux où les communautés sont bien séparées.

Conclusion

Le réseau Dolphins est homogène avec une structure sociale répartie. Le graphe de preferential attachment montre une organisation dominée par des hubs. L'analyse des centralités met en évidence l'importance de certains nœuds dans les flux d'information. Louvain semble le plus adapté à la détection de communautés dans des réseaux sociaux.

Perspectives : Tester d'autres mesures de centralité. Comparer les résultats sur un réseau plus grand. Analyser l'évolution temporelle du réseau Dolphins.