

# Trabajo Práctico 3

Facundo Emina

Juan Pablo Fiorenza

Redes complejas con aplicaciones a sistemas biológicos  
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

## Resumen

En el análisis de redes resulta de interés el estudio de la conformación de comunidades, es decir, grupos donde sus elementos comparten una mayor conectividad en comparación con los elementos de otros grupos. En nuestro trabajo, se analizó la conformación de comunidades en una red de delfines a partir de los métodos de partición de Louvain, Fast Greedy, Edge Betweenness e Infomap. Asimismo, se estudió si las comunidades formadas por estas particiones compartían alguna relación con la distribución de género, es decir, si delfines de un mismo género tendían a formar comunidades o lo hacían entre géneros distintos. De este análisis se pudo observar la presencia de homofilia en la red.

## 1. Introducción

Al momento de analizar un sistema representado por una red resulta de interés el estudio de la presencia de comunidades en este, así como también las características de los elementos que la conforman para así observar la existencia de alguna correlación entre estas y su unión dentro de las comunidades.

En nuestro trabajo se elaboró un grafo con una red de delfines que poseían distintos géneros, y se analizó la presencia de comunidades realizando la partición de dicho grafo de cuatro maneras diferentes (Louvain, Infomap, Fast Greedy y Edge Betweenness).

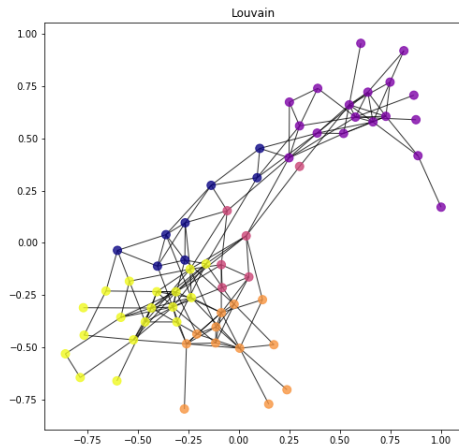
En la sección "Análisis de modularidad y silhouett" se desarrollará el estudio de la modularidad que presentaba nuestra red, teniendo en cuenta las cuatro particiones realizadas y los parámetros calculados de modularidad y silhouette.

En la sección "Comparación entre las particiones" se dará cuenta del grado de similitud entre las particiones teniendo en cuenta el método de precisión. Y por último, en "Correlación entre géneros de los delfines y comunidades" se analizó para las particiones Infomap y Louvain si la conformación de comunidades presentaba alguna relación con los géneros de los delfines.

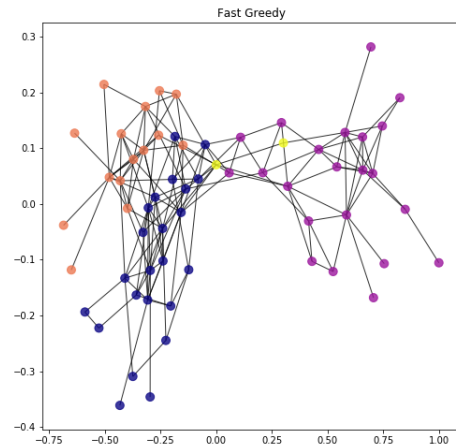
## 2. Análisis de modularidad y silhouette

En el análisis de las particiones de una red en distintas comunidades, resulta interesante observar cuan acertada es la asignación de cada nodo, a la comunidad en la cual se encuentran asignados.

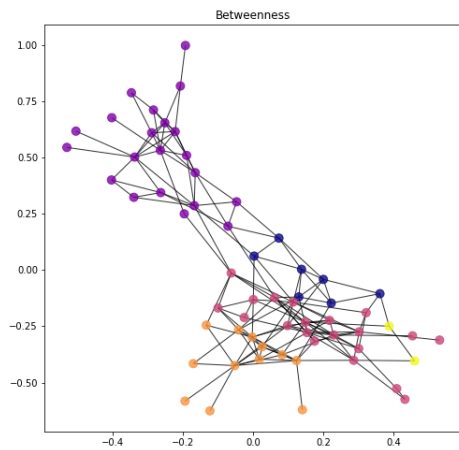
Para ello en primer lugar se graficaron cuatro grafos de la red de delfines, uno para cada partición realizada, esto nos permitió tener una visualización de la densidad de enlaces presentes entre nodos de la misma comunidad (pintados con el mismo color) y nodos de comunidades diferentes. Para ello observar la fig. (1).



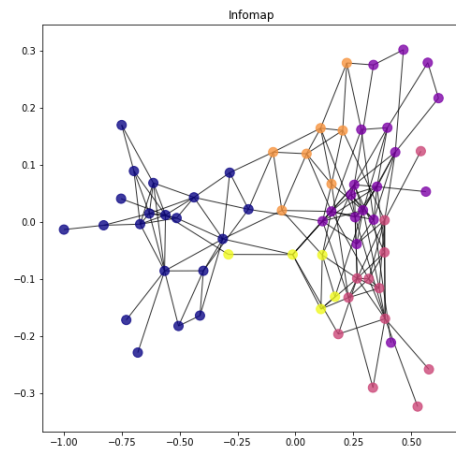
(a) Louvian



(b) Fast Greedy



(c) Betweenness



(d) Infomap

Figura 1: Particiones de la red. Con color los delfines de la misma comunidad.

Para todas las particiones salvo Fast Greedy, se obtuvo 5 comunidades, y 4 comunidades para esta última. Se calculó el valor de modularidad correspondiente a cada partición junto con el valor

promedio de Silhouette, que pueden verse en la tabla 1. Se observa que el mayor valor de modularidad se dio para la partición de Infomap, seguida por Edge Betweenness y luego Louvain. Los valores más bajos de modularidad y silhouette corresponden con la partición que dividió a la red en 4 comunidades.

Partición	Número de comunidades	Modularidad	Silhouette
Louvain	5	0,5188	0,2632
Fast Greedy	4	0,4954	0,1380
Edge Betweenness	5	0,5194	0,2876
Infomap	5	0,5277	0,2664

Tabla 1: Modularidad y silhouette promedio calculado para cada partición

A su vez, se realizó para cada partición un gráfico de los valores de Silhouette para cada nodo. En la figura 2 se puede apreciar la misma. El valor de Silhouette compara cuan similar es un elemento de una dada comunidad a los elementos de esa misma comunidad y cuan disímil es este a los elementos de las otras comunidades, teniendo en cuenta la distancia entre el nodo analizado y los nodos de la misma comunidad y de las otras. Este valor puede oscilar entre -1 y 1, donde un valor alto indica una buena correspondencia con los elementos de la misma comunidad y un valor bajo una mala.

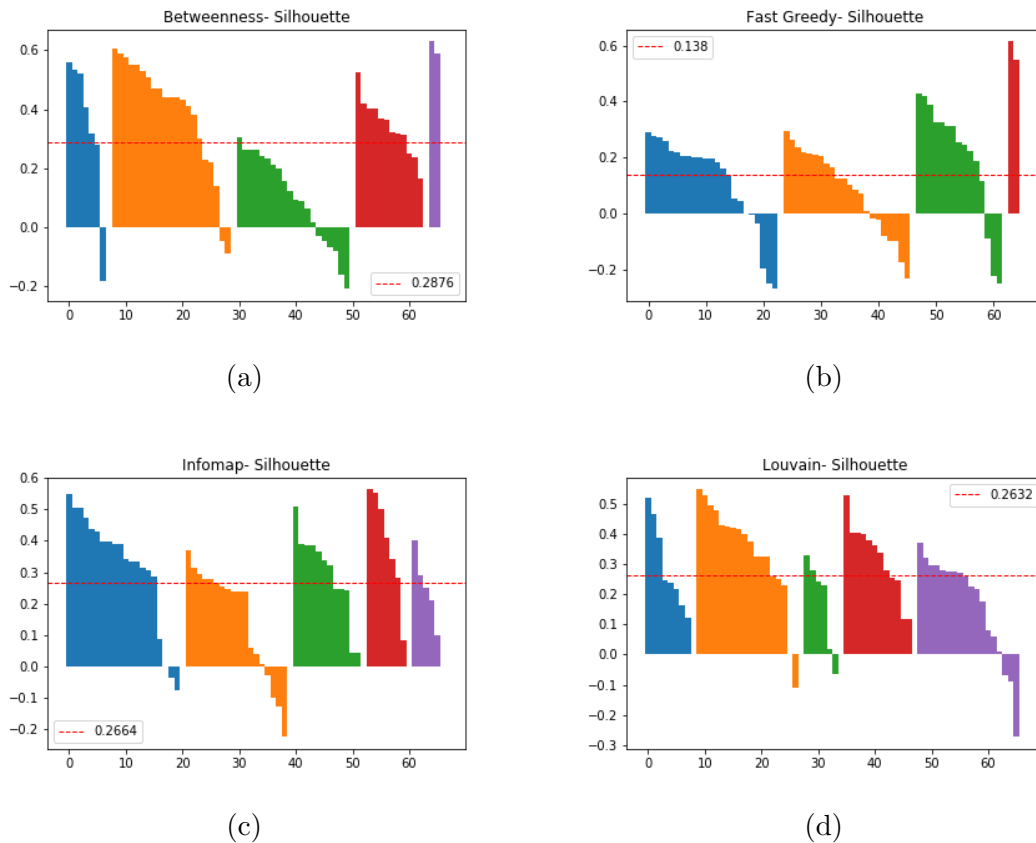


Figura 2: Gráfico de silhouette para cada comunidad para cada partición. Eje X nombre del nodo, eje Y valor de silhouette.

Los gráficos de la figura (2) nos permiten ver cuantas comunidades, expresadas cada una con colores distintos, superaron en cada partición el valor medio de silhouette (línea roja punteada). Se observa que para todas las particiones, mas de la mitad de las comunidades superaron su valor medio, estando este mas cerca de 1 que de -1.

Asimismo, se comparó los valores de modularidad y silhouette de las particiones de Louvain y Edge Betweenness con la distribución de valores obtenidos para un recableado aleatorio de la red, manteniendo su distribución de grado. Lo que se hizo en este proceso fue calcular las particiones (Louvain para un caso y Edge Betweenness para otro) a cada red generada por el recableado aleatorio y luego, plasmar la distribución de modularidad y silhouette en un histograma. Para ambos casos se observó que ambos valores se encontraban fuertemente alejados de la media del histograma (con un p-valor nulo)

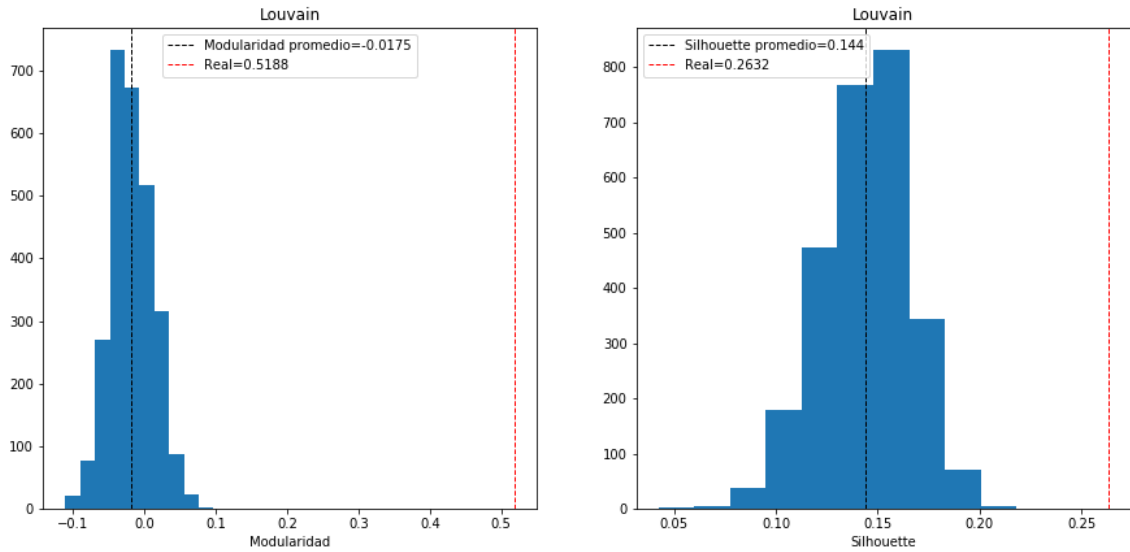


Figura 3: Histogramas para Louvain

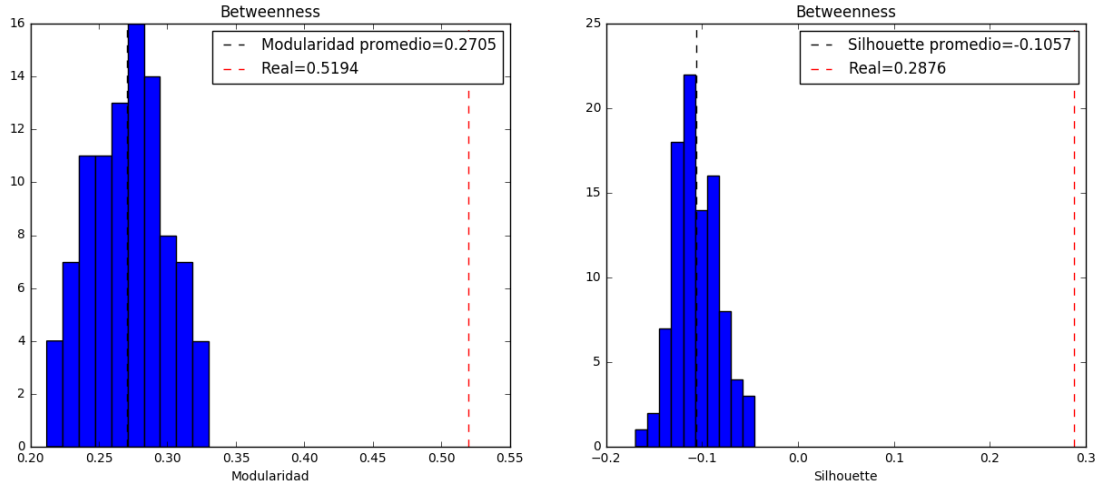


Figura 4: Histogramas para Louvain

Dado que el valor de modularidad y silhouette obtenido para las particiones de Louvian y Edge Betweenness se aleja del valor medio obtenido para la distribución de redes recableadas al azar, podemos descartar la hipótesis de que las comunidades dentro de nuestra red surgen de forma azarosa. Sumado a que el valor medio de modularidad para ambas particiones es positivo y cercano a 1. Además, el valor medio de silhouette se encuentra más cercano al 1 que al -1, lo que indica una buena asignación de comunidades a cada nodo.

### 3. Comparación entre las particiones

Como se demostró en la sección anterior se puede afirmar que los delfines de nuestra red, más allá de la partición tenida en cuenta, presentan una organización basada en comunidades. A pesar de ello, los valores de modularidad y silhouette no fueron iguales para todas las particiones, partiendo de este hecho se prosiguió a realizar un análisis del grado de acuerdo que existe entre ellas. Para ello el método fue el de *precisión*, este básicamente consiste en elaborar una matriz de 2 x 2 conocida como matriz de confusión donde los elementos de la diagonal representan la cantidad de pares de nodos que se encuentran en la misma comunidad para ambas particiones y la cantidad de pares de nodos que se encuentran en diferentes comunidades para ambas particiones; y en las otras posiciones se tabulan la cantidad de pares de nodos que estando en la misma comunidad en una partición están separados en la otra y viceversa. Del análisis de dicha matriz<sup>1</sup> se logró confeccionar la siguiente tabla.

<b>Louvian</b>	0.7176	0.8678	0.8995
<b>Infomap</b>		0.6288	0.6584
<b>Fast Greedy</b>			0.8466
<b>Edge Betweenness</b>			

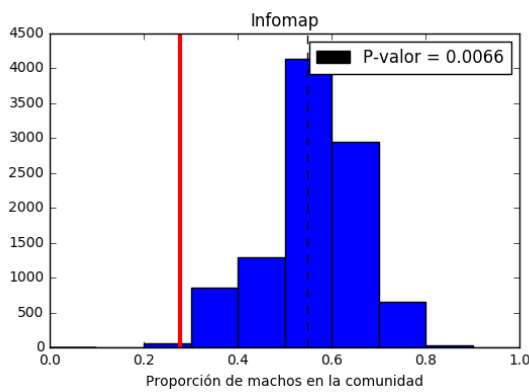
Tabla 2: Comparación entre las particiones, donde Louvian e Infomap son en un 71.76 % parecidas.

Donde puede observarse claramente que las particiones de Louvian y Edge Betweenness fueron las que mayor acuerdo tuvieron, es decir, presentan la mayor cantidad de pares de nodos que se encuentran compartiendo la misma comunidad en ambas o se encuentran en comunidades diferentes para ambas también.

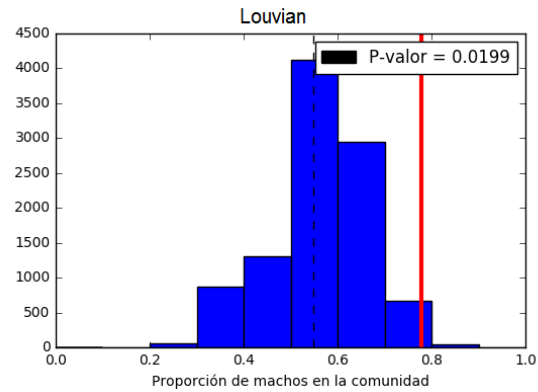
## 4. Correlación entre géneros de los delfines y comunidades

Otro de los análisis que se realizó fue observar si la formación de comunidades guardaba alguna relación con el género de los delfines (macho, hembra, NA). Para ello se estudio la proporción de delfines del mismo género presentes en cada comunidad y se la comparó con la proporción que habría si la unión entre los delfines se realizara de manera aleatoria, pero conservando la distribución de grado de la red.

El análisis previamente comentado se realizó para las particiones Infomap y Louvian, se utilizó como criterio para la selección de esta última la similitud entre las particiones Louvian, Fast Greedy y Edge Betweenness expresada en la tabla (2), pudiéndose observar un acuerdo de mas del 80 %. Por otra parte, debido a que Infomap poseía un nivel de similitud menor al 80 % con todas las otras tres, se decidió también incluirla en el análisis.



(a) Análisis de una comunidad obtenida con la partición Infomap



(b) Análisis de una comunidad obtenida con la partición Louvian

Figura 5: Distribución de la proporción de delfines del mismo género para la red aleatoria y para la red real

Donde la línea vertical roja representa la proporción de delfines machos en la red real, esta resulto mucho mas baja que la media en la distribución aleatoria para la comunidad analizada de Infomap y mayor a la media para la de Louvian. De esta manera se interpreta que la proporción de delfines hembras era alta en el primer caso y lo mismo para la proporción de machos en el segundo caso; concluyéndose que las comunidades estaban formadas en su mayoría por delfines del mismo género. A su vez, en otras comunidades de las mismas particiones la proporción de machos o hembras también dió mayor a la media, lo cual permitió seguir sosteniendo la misma hipótesis de la tendencia a la homofilia dentro de las comunidades.

## 5. Conclusiones

Como conclusión se puede afirmar que los delfines de la red analizada tienden a formar comunidades, dentro de las cuales sus miembros presentan una alta conectividad. A su vez, se observó que distintas particiones de la red guardaban un alto acuerdo entre las comunidades que se formaban, en particular esto sucedió para las particiones realizadas con Louvian, Fast Greedy y Edge Betweenness, reforzando la hipótesis de que las comunidades que contenían a cada delfín eran las acertadas para estos. Por último, apoyándonos en el análisis de las comunidades formadas, se observó que estas guardan una relación con los géneros de los delfines, detectándose en particular dentro de las comunidades la presencia de homofilia, es decir, la preferencia de los delfines de formar comunidades con aquellos que tienen el mismo género.

## 6. Referencias

[1] Chernomoretz, Ariel. (2018). Redes Complejas con aplicaciones a Biología de Sistemas”. [http://materias.df.uba.ar/redesa2018c2/files/2018/10/15\\_Clusters\\_2.pdf](http://materias.df.uba.ar/redesa2018c2/files/2018/10/15_Clusters_2.pdf)