

Trabajo Computacional No. 3  
Redes Complejas con Aplicaciones a Biología de Sistemas

Carlos Ríos<sup>1</sup>, Rodrigo Sieira<sup>1</sup>, Andrés Troiano<sup>1</sup> y Marco Villagran<sup>1</sup>

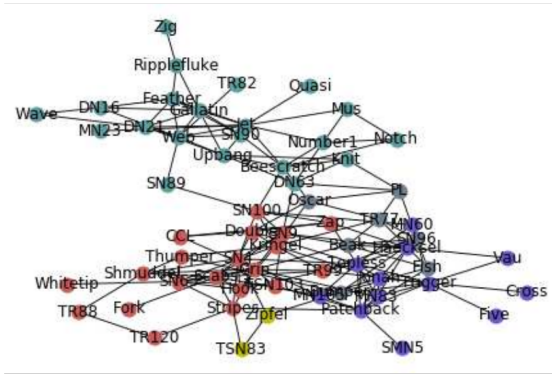
<sup>1</sup>UBA, Buenos Aires, Argentina

7 de noviembre de 2018

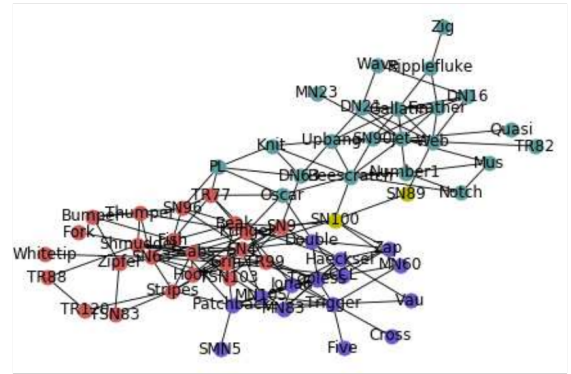
## 1. Red social de 62 delfines de Nueva Zelanda

1.1. Encuentre la partición en clusters de esta red utilizando la metodología *Louvain*, *Infomap*, *Fast Greedy* y *Edge Betweenness*. Visualice los resultados gráficamente.

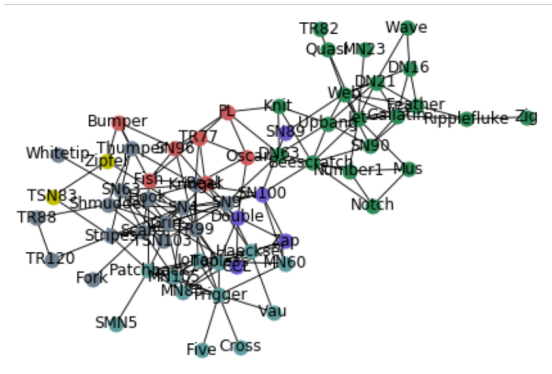
a. Edge-betweenness



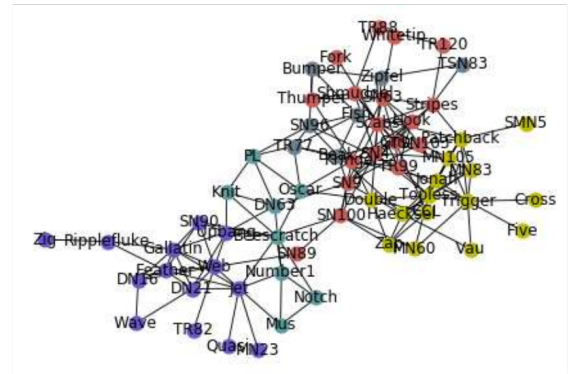
b. Fast greedy



c. Infomap



d. Louvain



**Figura 1:** Visualización de comunidades de la red de delfines de Nueva Zelanda mediante distintos métodos de clustering. La red social se particionó en comunidades utilizando a) *Edge Betweenness*, b) *Fast Greedy*, c) *Infomap* y d) *Louvain*. En los cuatro grafos de la figura se puede apreciar visualmente la separación entre comunidades.

## 1.2. Caracterice las particiones obtenidas en términos de modularidad y silhouette de cada partición. Compare con valores esperados en redes recableadas y establezca si tiene derecho a llamar modular a esta red.

Utilizando la libreta de Python `1_ab.ipynb` (ver Apéndice), se evaluaron las particiones obtenidas mediante los cuatro métodos de clustering utilizando medidas internas de las redes tales como modularidad ( $q$ ) y silhouette ( $S$ ). Por otro lado se calcularon dichos valores en **redes aleatorias generadas mediante el recableado de enlaces en 800 iteraciones**.

Manteniendo la distribu

En la figura 2 se muestran los silhouette plot de cada algoritmo, mientras que en las figuras 3 y 4 se muestran histogramas con valores de  $S$  y  $q$  respectivamente, correspondientes a recableados aleatorios. Como se puede observar en la figura 4, se obtuvieron valores de  $q$  cercanos a 0.52 en *Edge Betweenness*, *Infomap* y *Louvain*, mientras que con *Fast Greedy* se obtuvo  $q = 0,495$ . Dichos valores de modularidad fueron mayores que los obtenidos en redes aleatorias, quedando fuera de la distribución obtenida en 800 iteraciones. Mediante el método *Edge Betweenness*, *Infomap* y *Louvain*, se obtuvieron valores promedio de silhouette significativamente mayores que la media de las distribuciones observadas en 800 iteraciones, mientras que con el método *Fast Greedy* se obtuvo un valor  $S(i)$  menor al de la media.

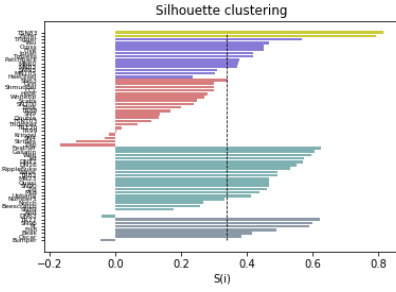
En base a lo observado con el índice  $q$  se puede afirmar que los cuatro métodos de clustering fueron efectivos en detectar comunidades dado que generaron particiones con valores de modularidad positivos, en todos los casos significativamente mayor a la distribución obtenida con redes aleatorias. El índice silhouette promedio, en cambio, sugiere que el método *Fast Greedy* no generó una red modular, dado que el índice fue significativamente menor al de la media. Este resultado es concordante con lo que se observa en el gráfico de silhouette de la figura 2, donde se pueden identificar una cantidad de nodos con valores  $S(i)$  negativos sustancialmente menor a la obtenida con los otros métodos de clustering. Con el resto de los métodos, el valor de  $S(i)$  promedio aportó resultados equivalentes a lo obtenido con el índice de modularidad  $q$ .

Cuál es la hipótesis nu

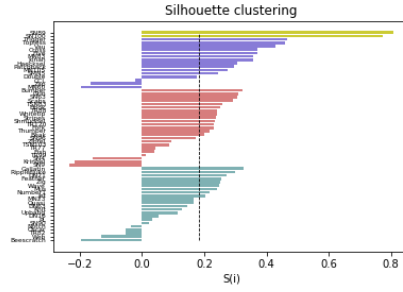
Se calculó el p-valor correspondiente a cada distribución de silhouette. En los casos de *Fast Greedy* y *Louvain* se obtuvieron **p-valores inferiores a 0,05** por lo cual se considera que los valores experimentales de silhouette tienen significancia estadística, **rechazando la hipótesis nula**. En los métodos *Edge Betweenness* e *Infomap* en cambio, se obtuvieron p-valores de 0.179 y 0.140 respectivamente con lo cual no rechazan la hipótesis nula. Para la modularidad no se calculó el p-valor debido a que los valores experimentales quedaron fuera de la distribución aleatoria, como se puede observar en la figura 4.

Hicieron p-valor a dos p

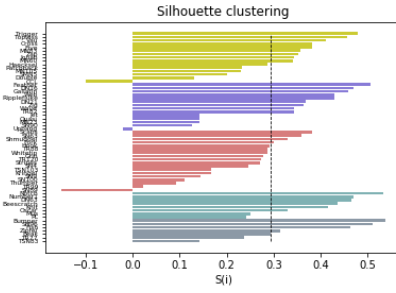
a. Edge betweenness



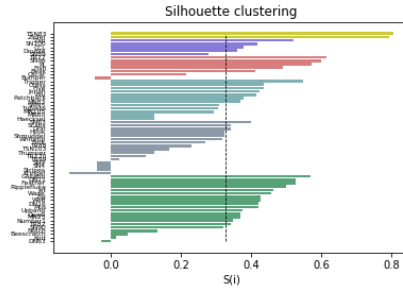
b. Fast greedy



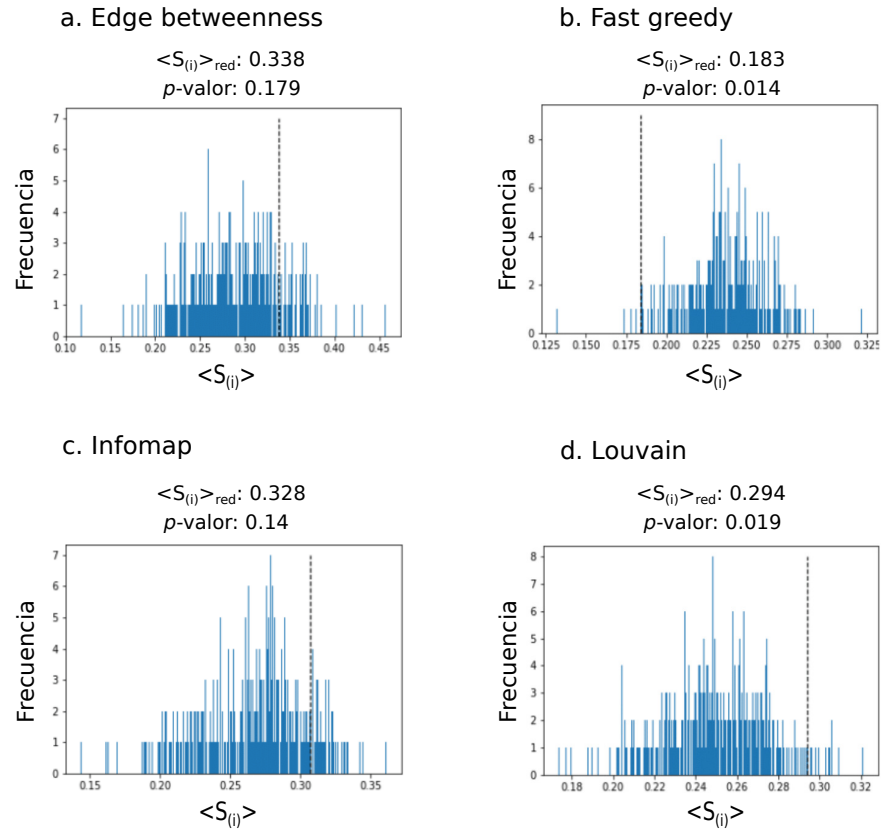
c. Infomap



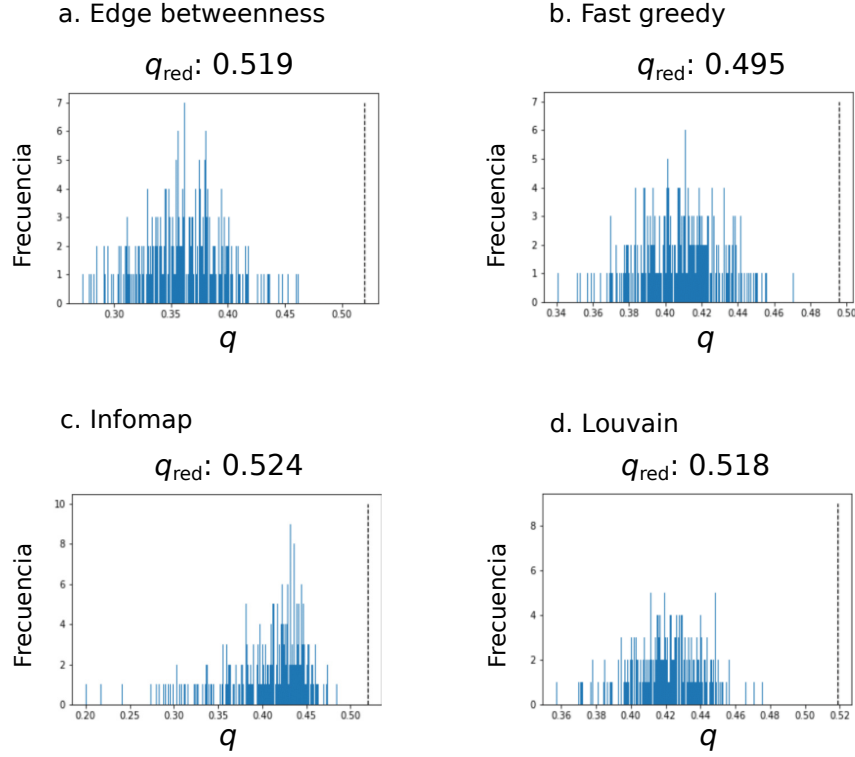
d. Louvain



**Figura 2:** Evaluación de distintas particiones de la red social de delfines de Nueva Zelanda. Silhouette plot de las particiones generadas por a) *Edge Betweenness*, b) *Fast Greedy*, c) *Infomap* y d) *Louvain*.



**Figura 3:** Análisis de Silhouette en redes aleatorias. Se determinó el valor promedio de silhouette para la red social de delfines de Nueva Zelanda (línea punteada) y para redes aleatorias obtenidas en 800 iteraciones (barras azules).



**Figura 4:** Análisis de modularidad en redes aleatorias. Se determinó el valor de modularidad  $q$  para la red social de delfines de Nueva Zelanda (línea punteada) y para redes aleatorias obtenidas en 800 iteraciones (barras azules).

### 1.3. Caracterice cuantitativamente el acuerdo entre las particiones obtenidas utilizando uno o más de los observables vistos en clase.

Se realizó una evaluación de las particiones obtenidas utilizando información externa, tomando como partición de referencia las comunidades generadas con cada uno de los distintos métodos. Utilizando el código `1_c.py` (ver Apéndice) se evaluaron las particiones utilizando el enfoque de la teoría de la información mutua, donde se analiza la cantidad de información en bits contenida en las distintas particiones con respecto a las demás. Como se puede observar en la tabla 1, las particiones obtenidas con *Fast Greedy* fueron las que menor cantidad de información compartieron con las del resto de los métodos. En dicho código además se realiza un cálculo de los valores de precisión.

Se analizaron las particiones utilizando el enfoque de precisión en la asignación de grupos, donde se penaliza la ocurrencia de nodos que pertenecen a una

misma comunidad de referencia asignados a comunidades diferentes, y también la ocurrencia de nodos que pertenecen a comunidades de referencia diferentes asignados en una misma comunidad. Como se puede observar en la tabla 2, en este caso también se observó que las particiones obtenidas con *Fast Greedy* resultaron en valores de índice de precisión menores a 0,9 mientras que entre el resto de las particiones los valores fueron más cercanos a 1. En concordancia con la evaluación de particiones con medidas internas, en este caso también se observa que el método *Fast Greedy* brindó resultados menos confiables en comparación al resto de los métodos.

	Información mutua			
	EB	FG	Inf	Louvain
EB	1.00	0.66	0.91	0.78
FG	0.66	1.00	0.77	0.74
Inf	0.91	0.77	1.00	0.88
Louvain	0.78	0.74	0.88	1.00

**Tabla 1:** Análisis de acuerdo entre particiones de la red social de delfines de Nueva Zelanda mediante el enfoque de la teoría de la información mutua. Se calculó el índice de información mutua entre particiones generadas por *Edge Betweenness* (EB), *Fast Greedy* (FG), *Infomap* (Inf) y *Louvain*.

	Precision			
	EB	FG	Inf	Louvain
EB	1.00	0.84	0.95	0.91
FG	0.84	1.00	0.89	0.88
Inf	0.95	0.89	1.00	0.95
Louvain	0.91	0.88	0.95	1.00

**Tabla 2:** Análisis de acuerdo entre particiones de la red social de delfines de Nueva Zelanda mediante el método precisión. Se calculó el índice de precisión entre particiones generadas por *Edge Betweenness* (EB), *Fast Greedy* (FG), *Infomap* (Inf) y *Louvain*.

**1.4. Analice cuantitativamente la relación entre el género de los delfines y la estructura de comunidades del grupo. Puede utilizar para ello, por ejemplo, tests de sobre-representación y/o sub-representación. ¿Qué hipótesis puede aventurar sobre propiedades comportamentales de este grupo de delfines a partir de lo encontrado?**

Con el objetivo de analizar la relación entre el género de los delfines y la estructura de las comunidades obtenidas en cada método de partición, utilizando el código `1_d.py` (ver Apéndice), se determinó en cada comunidad la proporción de machos o hembras, y se comparó con la proporción correspondiente en redes aleatorias construidas reasignando género a cada nodo en 1000 iteraciones. Se calculó el p-valor obtenido en cada caso como la fracción del área de la distribución aleatoria que contiene al valor calculado para la comunidad de la red real.

Como se puede observar en la tabla 3, para cada uno de los métodos de clustering utilizados se observan por lo menos 2 comunidades donde tanto la proporción de machos como la de hembras se diferencia del valor esperado por azar con una significancia mayor al 5 %. Se realizaron histogramas de frecuencias de proporción de machos/hembras en redes aleatorias y su comparación con el valor medido en la red real (ver Apéndice), donde se puede observar claramente que cada uno de los métodos de clustering identificó: *a*) una comunidad donde la proporción de machos es significativamente mayor a la esperada por azar, y *b*) otra comunidad donde por el contrario la proporción de hembras es mayor a la esperada por azar. Estos resultados sugieren que en la red social de delfines de Nueva Zelanda existe una comunidad probablemente compuesta en su mayor parte por un grupo de machos jóvenes (caso *a*) y otra comunidad con mayoría de hembras (caso *b*) probablemente liderada por un grupo reducido de machos dominantes. El resto de las comunidades no presentan diferencias significativas con respecto a las proporciones de género esperadas por azar, sugiriendo que los individuos que las componen no presentan un comportamiento de competencia por liderazgo para el apareamiento. A partir de estas observaciones concluimos que los cuatro métodos de clustering utilizados fueron efectivos para identificar estos posibles indicios de patrones comportamentales.

## **2. Apéndice**

Los códigos utilizados se pueden encontrar en <https://github.com/CarlosAndresRiosChavez/ProyectoRedes/tree/master/tp3/scripts>.

Los histogramas de reasignación aleatoria de géneros (preservando las particiones originales) mencionadas en la sección 1.4 se encuentran en <https://github.com/CarlosAndresRiosChavez/ProyectoRedes/tree/master/tp3/plots>.



	P-val							
	EB		FG		Inf		Louvain	
	H	M	H	M	H	M	H	M
0	0,16	0,088	0	0	0	0	0,076	0,03
1	0,109	0,05	0	0,003	0	0	0,001	0,011
2	0,001	0,004	0,301	0,326	0,114	0,034	0,067	0,011
3	0	0	0	0	0,235	0,129	0	0
4	0	0,464	n.a.	n.a.	0,024	0	0,169	0,1
5	n.a.	n.a.	n.a.	n.a.	0	0,456	n.a.	n.a.

**Tabla 3:** Relación entre género y comunidades de la red social de delfines de Nueva Zelanda. Para cada una de las comunidades 0 a 5 (filas) obtenidas por los distintos métodos de clustering *Edge Betweenness* (EB), *Fast Greedy* (FG), *Infomap* (Inf) y *Louvain*, se calculó la proporción de machos (M) y hembras (H) con su correspondiente p-valor para redes aleatorias con género reasignado en 1000 iteraciones. Se resalta en gris a los p-valores menores que 0,05.