

# Trabajo computacional 3

Di Filippo Juan, Catoni Josefina, Yalovetzky Romina

Octubre 2018

## 1. Introducción

En el presente trabajo se estudió una red social, compuesta por las relaciones entre 62 delfines de Nueva Zelanda, con el objetivo de analizar en profundidad la partición en clusters de la misma. Para efectuar dicha partición se trabajó con una base de criterios estándar: Infomap, Fastgreedy, Louvain y Edge-Betweenness. Además, se realizó un modelo nulo a partir de hacer estadística de la red recableada, manteniendo la distribución de grado constante, y se contrastó la red original con dicho modelo en base a dos parámetros: la modularidad y el coeficiente de Silhouette.

Por otro lado, se caracterizó cuantitativamente el acuerdo entre las particiones obtenidas utilizando la matriz de .... y ... ?

Finalmente, se estudió la relación entre el género de los delfines y la estructura de comunidades del grupo. Para ello, se blablalaba.

## 2. Resultados

### 2.1. Criterios

#### 2.1.1. Edge betweenness

El algoritmo aplicado de edge betweenness consiste en ir removiendo los enlaces progresivamente. Lo hicimos hasta que la red queda con un único enlace aplicando el algoritmo de Girvan Newman. En cada iteración de quitado de enlaces calculamos la modularidad. Entonces, podemos ver en Fig. 1 para qué cantidad de comunidades (y cómo están conformadas) ésta magnitud se maximiza. Nos quedamos con ésta distribución en comunidades según éste criterio.

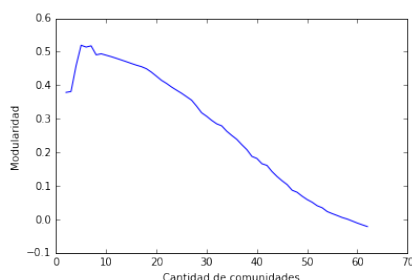


Figura 1

### 2.2. Comparación de los criterios

Una forma de caracterizar a la distribución en comunidades es a partir de la modularidad. Cuán más grande sea éste valor, la distribución en comunidades resulta ser la más óptima. Vemos en la siguiente tabla los valores obtenidos para los criterios estudiados.

Criterio	Modularidad
Edge Betweenness	0.519
Louvain	
Fast Greedy	
Infomap	

Cuadro 1: Valores de modularidad para distintos criterios de clustering

Uno de los observables que nos permite hacer una comparación gráfica de los métodos es calcular la cantidad de nodos que hay en cada comunidad según las distribuciones como vemos en Fig. 2.

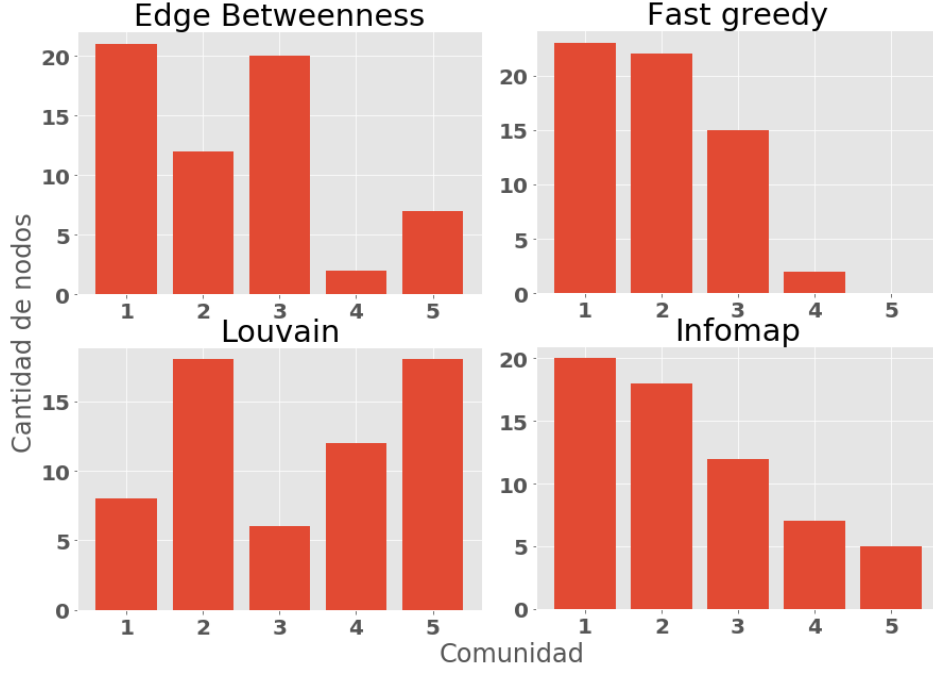


Figura 2

Otro observable es la matriz de la probabilidad conjunta. Esta expresa para cada par de distribución en comunidades la probabilidad conjunta de que un nodo elegido al azar pertenezca a la comunidad C1 de la primera partición y a la C2 de la segunda.

$$p(c_1, c_2) = \frac{N_{c_1, c_2}}{\sum_{c_1, c_2} N_{c_1, c_2}} \quad (1)$$

A continuación vemos la probabilidad conjunta entre partición edge betweenness y las otras estudiadas. En cada una de las filas se muestra a las 5 comunidades de la partición edge betweenness (EB) y en las columnas las respectivas comunidades de las demás particiones.

		Fast Greedy					Louvain					Info map				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
EB	C1	0	32.3	0	1.6	0	3.2	29	1.6	0	0	32.3	0	0	0	1.6
	C2	0	0	19.4	0	0	0	0	0	19.4	0	0	0	19.4	0	0
	C3	25.8	0	4.8	1.6	0	0	0	6.5	0	25	0	25.8	0	0	6.5
	C4	3.2	0	0	0	0	0	0	0	0	3	0	3.2	0	0	0
	C5	8.1	3.2	0	0	0	9.7	0	1.6	0	0	0	0	0	11.3	0

Cuadro 2: Matrices de probabilidad conjunta para la partición Edge Betweenness (sus comunidades en filas) y Fast Greedy, Louvain e Info Map (respectivas comunidades en columnas)

## Referencias

- [1] JEONG, Hawoong, et al. Lethality and centrality in protein networks. Nature, 2001, vol. 411, no 6833, p. 41.
- [2] Why Do Hubs Tend to Be Essential in Protein Networks?; Xionglei He, Jianzhi Zhang; Plos genetics, 2006.
- [3] Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality; Elena Zotenko, Julian Mestre, Dianne P. O'Leary, Teresa M. Przytycka; PLOS Computational Biology, 2008.