

Detección de comunidades en una red social de delfines

Emanuel Chironi, Federico Sevlever, Lucas Alonso, and Francisco Correa
Universidad de Buenos Aires
Redes complejas
Trabajo Computacional N 3

I. ENCUENTRE LA PARTICIÓN EN CLUSTERS DE ESTA RED UTILIZANDO LA METODOLOGÍA LOUVAIN, INFOMAP, FAST GREEDY Y EDGE BETWEENNESS. VISUALICE LOS RESULTADOS GRÁFICAMENTE.

Consideramos la red social de 62 delfines de Nueva Zelanda. Particionamos la red en clusters según los criterios mencionados en el enunciado. Coloreamos los nodos en función de las comunidades obtenidas y las representamos gráficamente en las figuras 1, 2, 3 y 4. En cada figura se muestran 3 formas de graficar distintas.

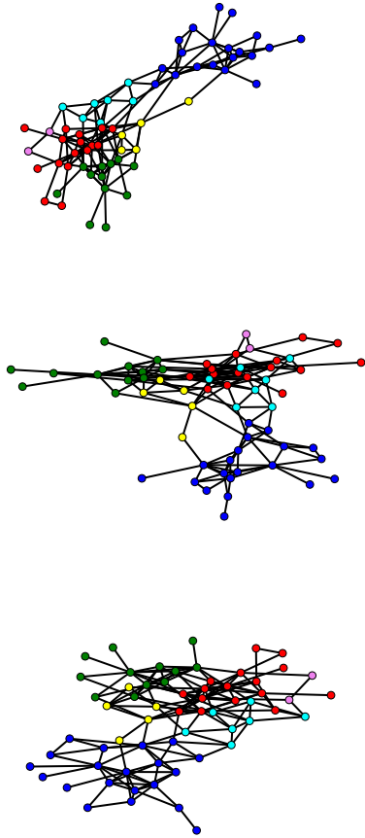


FIG. 1. Distintas representaciones de la red usando el algoritmo de clustering Infomap. Cada color representa un cluster.

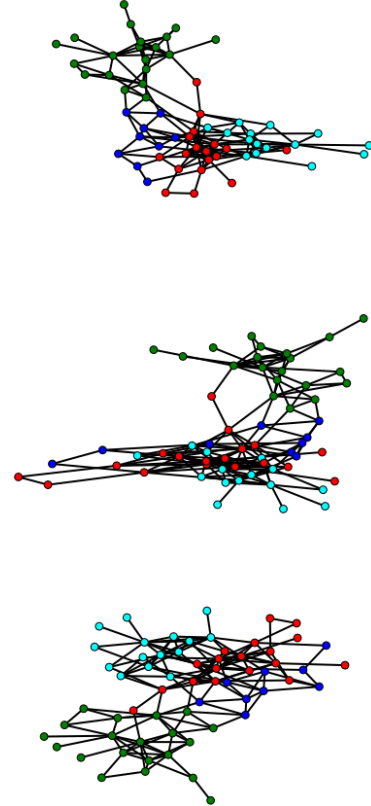


FIG. 2. Distintas representaciones de la red usando el algoritmo de clustering Louvain. Cada color representa un cluster.

II. CARACTERICE LAS PARTICIONES OBTENIDAS EN TRMINOS DE MODULARIDAD Y SILOUHETTE DE CADA PARTICIÓN. COMPARE CON VALORES ESPERADOS EN REDES RECABLEADAS Y ESTABLEZCA SI TIENE DERECHO A LLAMAR MODULAR A ESTA RED.

Mostramos en la Tabla I los resultados de Modularidad y valor medio de Silouhette para cada una de las particiones.

En la figura 5 mostramos los valores de silouhette para cada nodo, ordenados por cluster y de mayor a

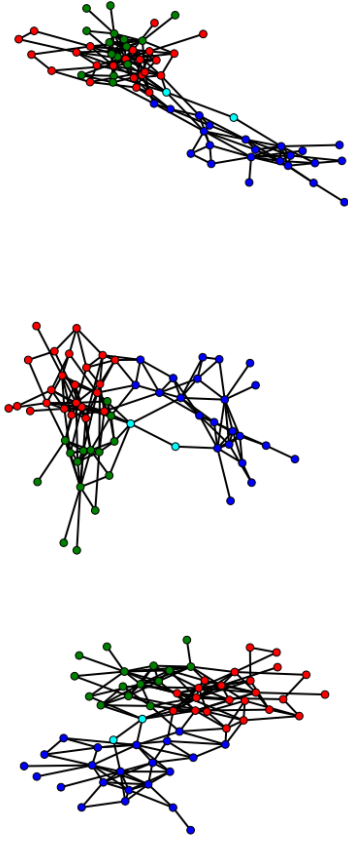


FIG. 3. Distintas representaciones de la red usando el algoritmo de clustering Fast Greedy. Cada color representa un cluster.

menor silhouette. Observamos que todos los valores son positivos, y cada cluster tiene puntos por encima del promedio, lo que indica que las particiones son buenas.

En las figuras 6 y 7 mostramos la distribución de valores de modularidad y del valor medio de silhouette para cada red recableada al azar manteniendo el grado de cada nodo, junto al valor obtenido de la red original. Como se puede observar en las figuras correspondientes, los valores calculados están por fuera de la distribución, esto nos indica que las particiones no son fruto del azar sino que efectivamente existe modularidad en la red.

Método	Modularidad	Silhouette
Infomap	0.52	0.43
Fast Greedy	0.49	0.38
Louvain	0.52	0.40
Edge Betweenness	0.52	0.42

TABLA I. Modularidad y valor medio de Silhouette para cada método de clustering.

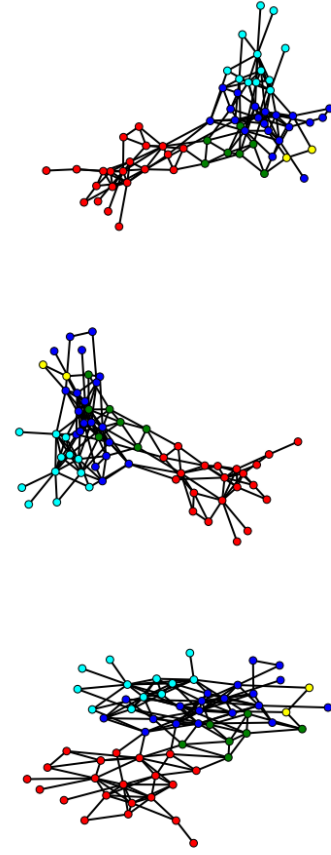


FIG. 4. Distintas representaciones de la red usando el algoritmo de clustering Edge Betweenness. Cada color representa un cluster.

III. CARACTERICE CUANTITATIVAMENTE EL ACUERDO ENTRE LAS PARTICIONES OBTENIDAS UTILIZANDO UNO O MÁS DE LOS OBSERVABLES VISTOS EN CLASE.

Para cuantificar la similitud de las particiones calculamos la información mutua entre cada par de particiones. Mostramos los resultados en la tabla II. Observamos que las particiones en general son parecidas.

	Infomap	Fast G.	Louvain	Edge B.
Infomap	1	0.63	0.71	0.79
Fast Greedy		1	0.51	0.48
Louvain			1	0.52
Edge B.				1

TABLA II. Información mutua entre particiones. Podemos ver que las particiones no son muy diferentes, siendo Infomap y Edge Betweenes las más cercanas mientras que Fast Greedy y Louvain son las más lejanas.

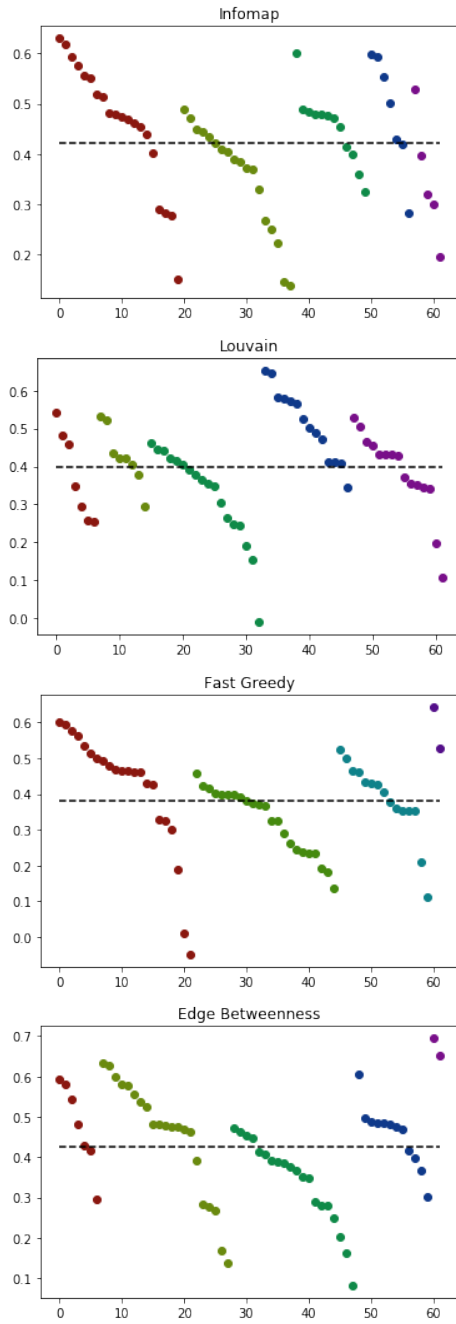


FIG. 5. Silhouette por nodo, cada punto está coloreado de acuerdo al cluster que pertenece. En línea punteada el valor medio del Solhouette sobre todos los nodos. Vemos que para todas las particiones hay nodos por encima y debajo del valor medio en cada cluster, salvo los clusters de dos nodos.

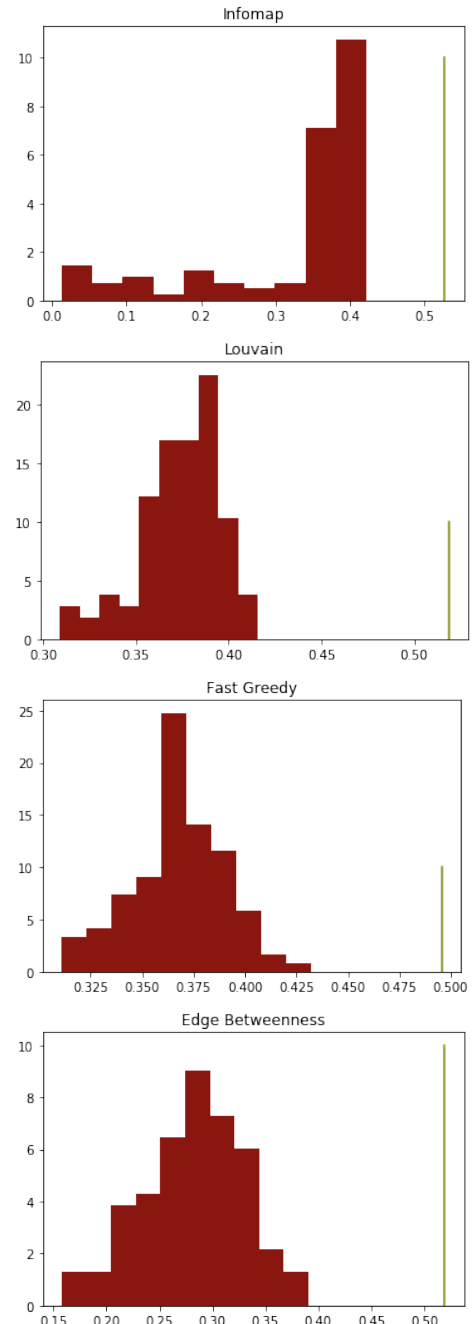


FIG. 6. Modularidad de la red por tipo de partición. En línea verde la modularidad y el histograma bordó representa la distribución de modularidad sobre 100 redes recableadas aleatoriamente, manteniendo el grado de cada nodo.

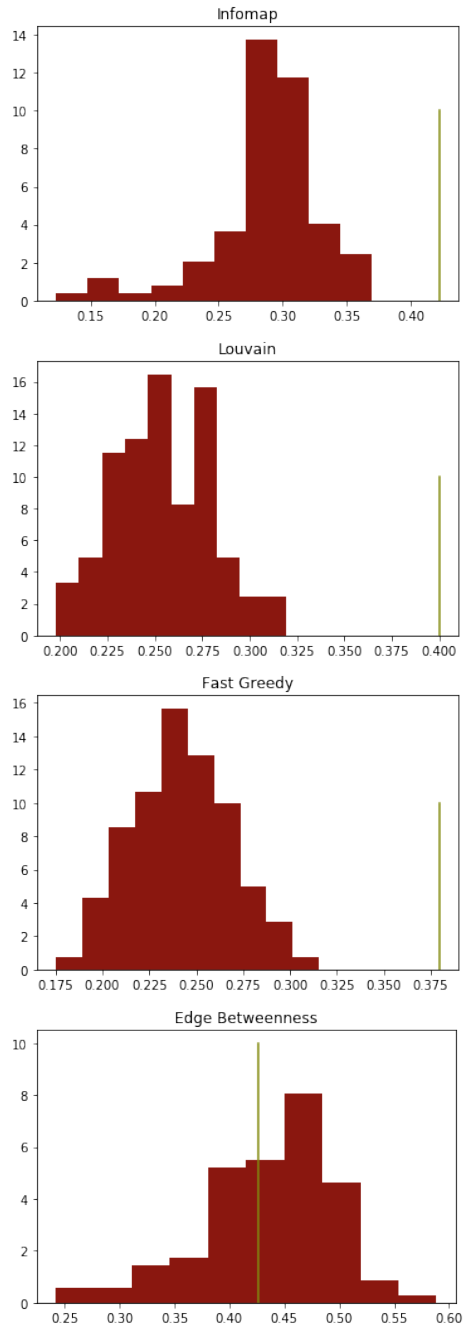


FIG. 7. Silhouette medio de la red por tipo de partición. En línea verde el Silhouette medio de la red y el histograma bordó representa la distribución de Silhouette medio sobre 100 redes recableadas aleatoriamente, manteniendo el grado de cada nodo.

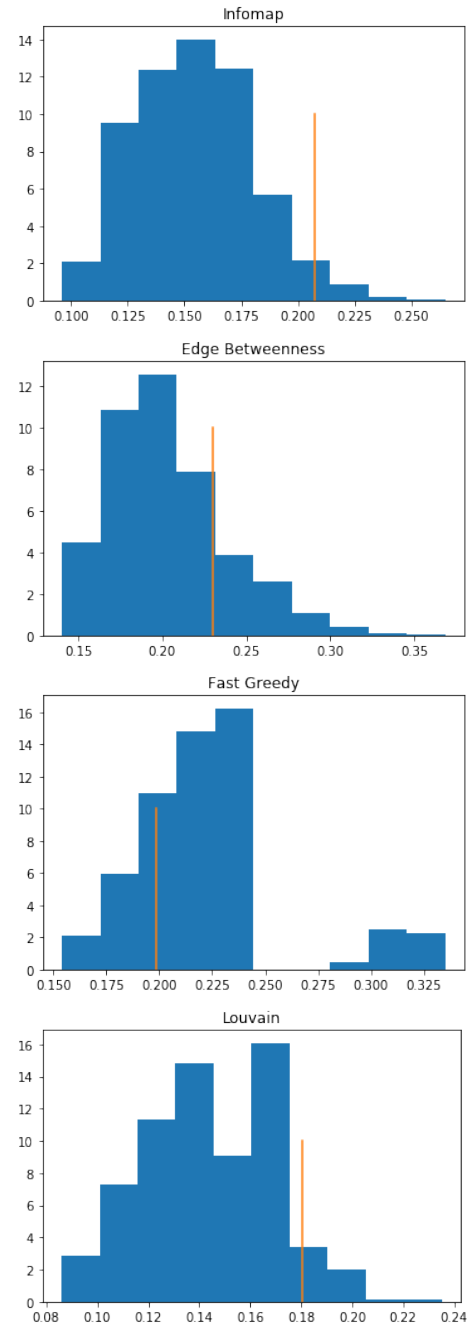


FIG. 8. En azul, información mutua de la red con distribución al azar de generos de nodos de la red. En línea naranja valor de información mutua de comunidad y género provisto por la base de datos.

IV. 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?

Para este ítem hemos etiquetado los géneros de delfines con "1" para male y "2" para female. Una vez hecho esto se compararon estas etiquetas con las etiquetas de color según la separación en comunidades de dicha red para saber si hay una correlación entre comunidad y género.

En la figura 8 podemos observar que la distribución de géneros al azar no difiere significativamente respecto de la distribución de géneros (y comunidades) provista por la base de datos, de manera que no podemos decir que los clusters son homofílicos.