

# 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. Se codificaron los clusters de cada partición por medio de un color, de manera que los nodos de un mismo color corresponden a un mismo cluster. Las distintas particiones, correspondientes a los distintos métodos propuestos en el enunciado, se representaron gráficamente (ver las figuras 1, 2, 3 y 4). En cada figura se muestran 3 formas distintas de graficar los nodos sobre el plano. **A diferencia de otras situaciones, en las cuales el uso de distintos layouts permite mejorar la visualización de los resultados, en este caso particular no se aprecian diferencias significativas.** Asimismo, es interesante sealar que no se advierten variaciones notorias en la calidad de las particiones obtenidas por distintos procedimientos de clustering. Dado que la simple inspección no permite notar diferencias, a continuación se trabajara con parámetros cuantitativos que permiten medir la calidad del proceso de clustering con mayor precisión.

## II. CARACTERICE LAS PARTICIONES OBTENIDAS EN TÉRMINOS 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 obtenidas en el punto anterior.

En la figura 5 se muestran los valores de silouhette que se encontraron para cada nodo, ordenados por cluster y de mayor a menor silouhette. Observamos que todos los valores son positivos, y cada cluster tiene puntos por encima del promedio, lo que indica que las particiones son buenas.

Por otra parte, la tabla I resume los valores de modularidad y silouhette para cada una de las particiones obtenidas por medio de los distintos métodos de clustering. Se puede observar que los valores de ambos parámetros son muy cercanos entre sí. Esto muestra que

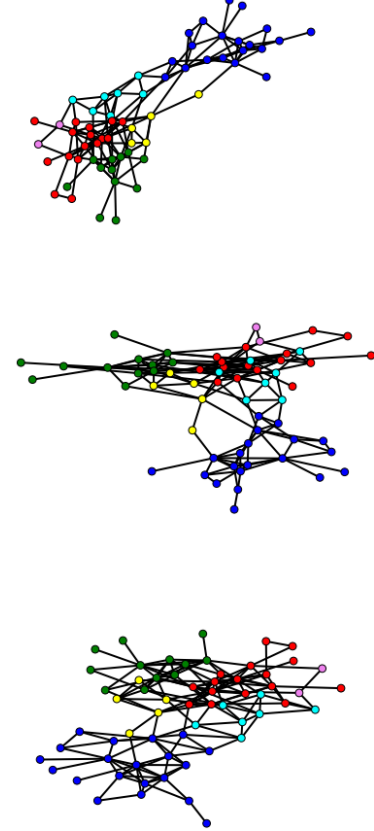


FIG. 1. Distintas representaciones (layouts) de la red, la cual ha sido particionada usando el algoritmo de clustering Infomap. Cada color representa un cluster.

los métodos ofrecen resultados muy semejantes para la red de interés y es consistente con las observaciones realizadas sobre el final del punto anterior. No solamente no se advierten grandes variaciones visualmente, sino que tampoco hay diferencias relevantes en cuanto a la calidad de las particiones, cuando se las compara con criterios cuantitativos.

Finalmente, a efectos de evaluar cuan significativas son las particiones en sí mismas, se procedió a realizar una comparación con los resultados que cabría esperar por azar. En las figuras 6 y 7 mostramos la distribución de valores de modularidad y del valor medio de silouhette para cada red recableada al azar manteniendo el grado

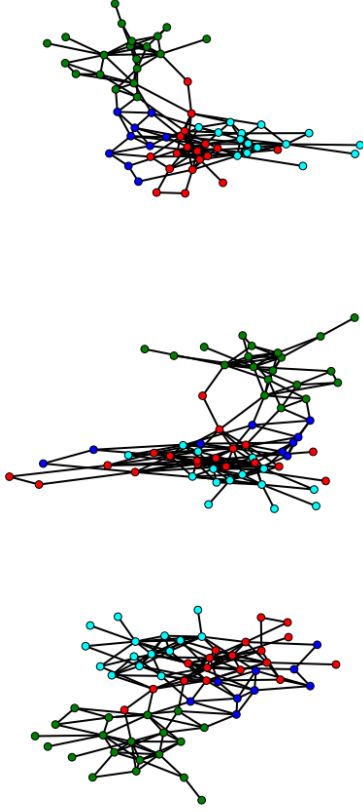


FIG. 2. Distintas representaciones (layouts) de la red, la cual ha sido particionada usando el algoritmo de clustering Louvain. Cada color representa un cluster.

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.

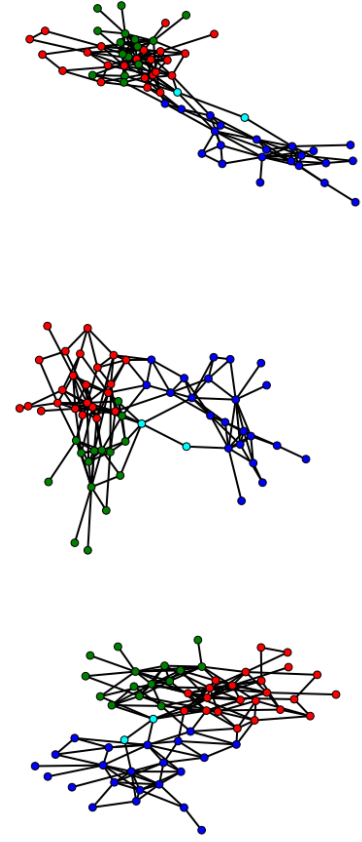


FIG. 3. Distintas representaciones (layouts) de la red, la cual ha sido particionada usando el algoritmo de clustering Fast Greedy. Cada color representa un cluster.

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

En los ítems 1 y 2 se realizaron particiones de la red de delfines utilizando distintos métodos. Estas particiones se compararon tanto visualmente como en términos de las diferencias en los factores de calidad de clustering de cada una de ellas. Se encontró que no había diferencias significativas. Aquí analizamos la relación entre las distintas particiones respecto de cómo se reparten los nodos específicos entre los distintos clusters de dos o más particiones. Estas comparaciones se realizaron entre todos los pares posibles de particiones. Este tipo de análisis mide la correlación entre las particiones, en el sentido de que mide cuánta información provee sobre la segunda partición la información que se tiene sobre la primera. Esto se logra midiendo un parámetro denominado información mutua entre cada par de particiones. Mostramos los resultados en la tabla II. Observamos que las particiones en general son parecidas, dado que sus valores son

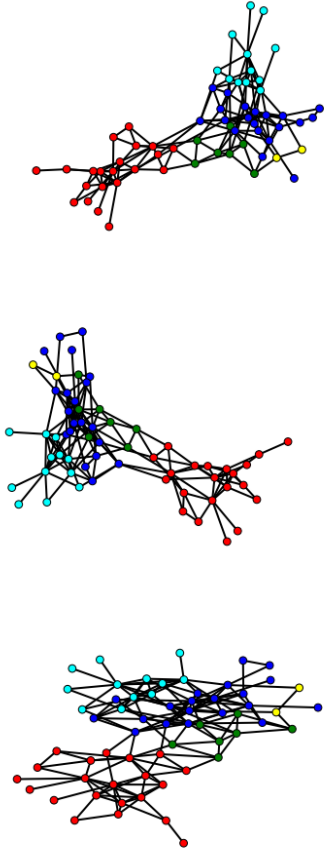


FIG. 4. Distintas representaciones (layouts) de la red, la cual ha sido particionada usando el algoritmo de clustering Edge Betweenness. Cada color representa un cluster.

cercanos y altos. Es particularmente interesante notar la alta información mutua que infomap tiene tanto con Louvain como con Edge Betweenness.

	Infomap	Fast G.	Louvain	Edge B.
<b>Infomap</b>	1	0.63	0.71	0.79
<b>Fast Greedy</b>		1	0.51	0.48
<b>Louvain</b>			1	0.52
<b>Edge B.</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.

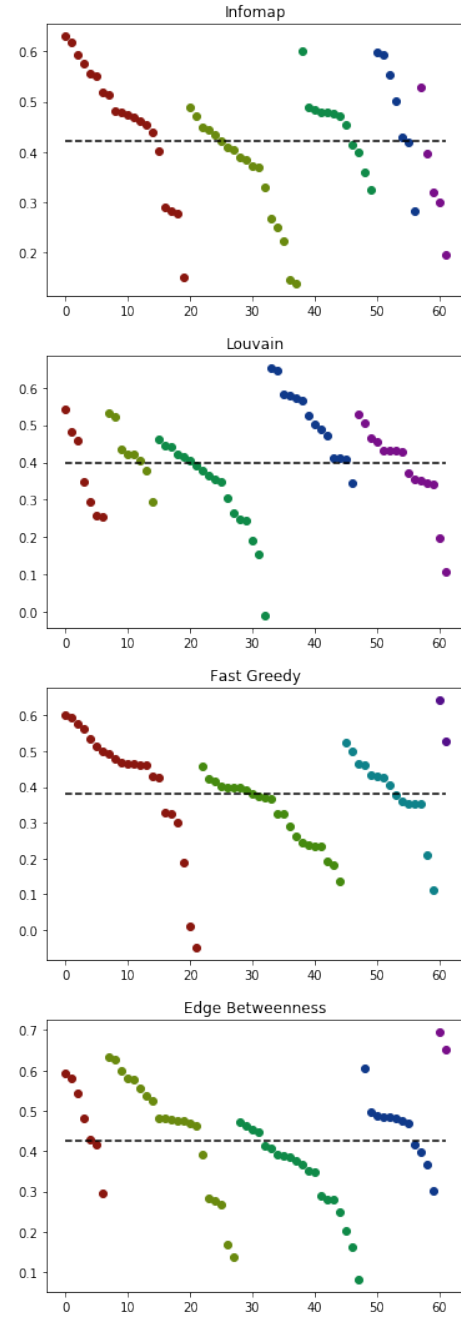


FIG. 5. Silhouette por nodo para el conjunto de nodos de la red de delfines de Nueva Zelanda; cada punto está coloreado de acuerdo al cluster al que pertenece. La línea punteada indica el valor medio de Silhouette calculado sobre todos los nodos. Vemos que para todas las particiones hay nodos por encima y por debajo del valor medio en cada cluster, salvo en los clusters de dos nodos.

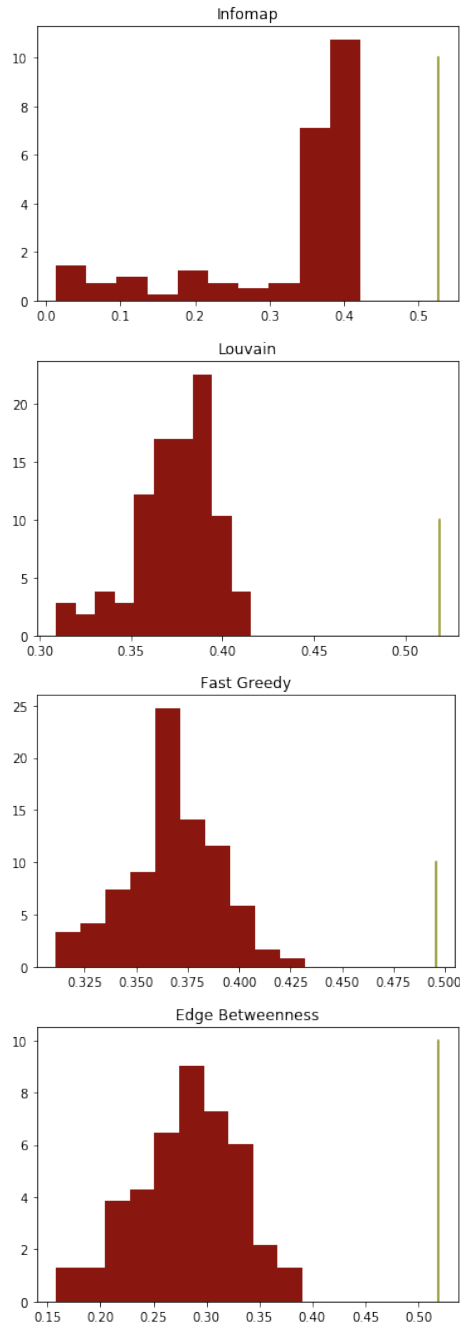


FIG. 6. Modularidad de la red por tipo de partición. En línea verde la modularidad y el histograma bordó representa la distribución de modularidad sobre 100 redes recableadas aleatoriamente, manteniendo el grado de cada nodo.

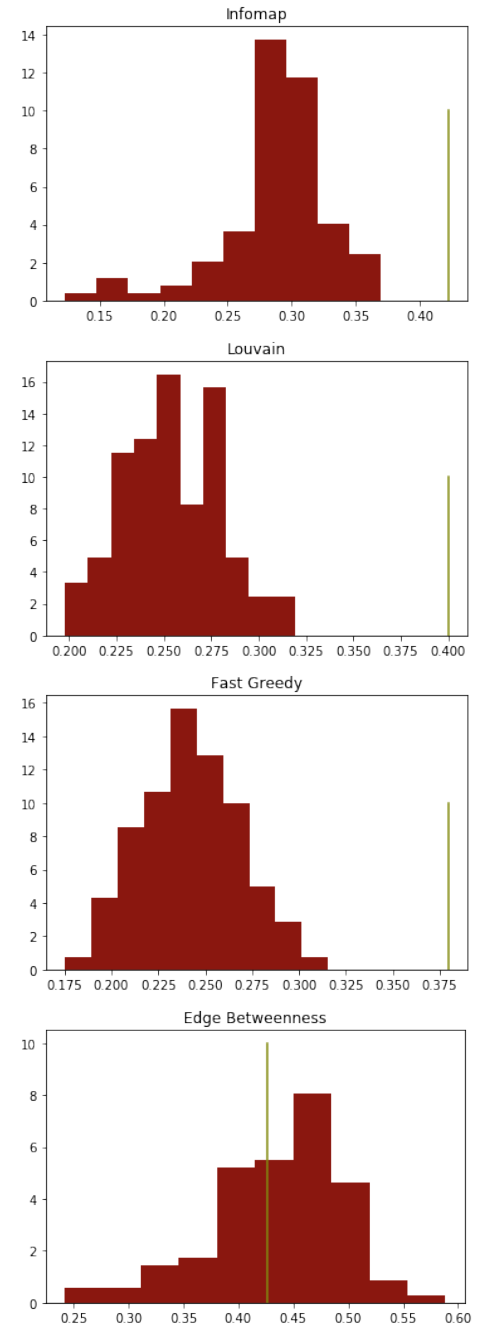


FIG. 7. Silhouette medio de la red por tipo de partición. La línea verde representa el Silhouette medio de la red, mientras que el histograma bordó representa la distribución de Silhouette medio sobre 100 redes recableadas aleatoriamente, manteniendo el grado de cada nodo.

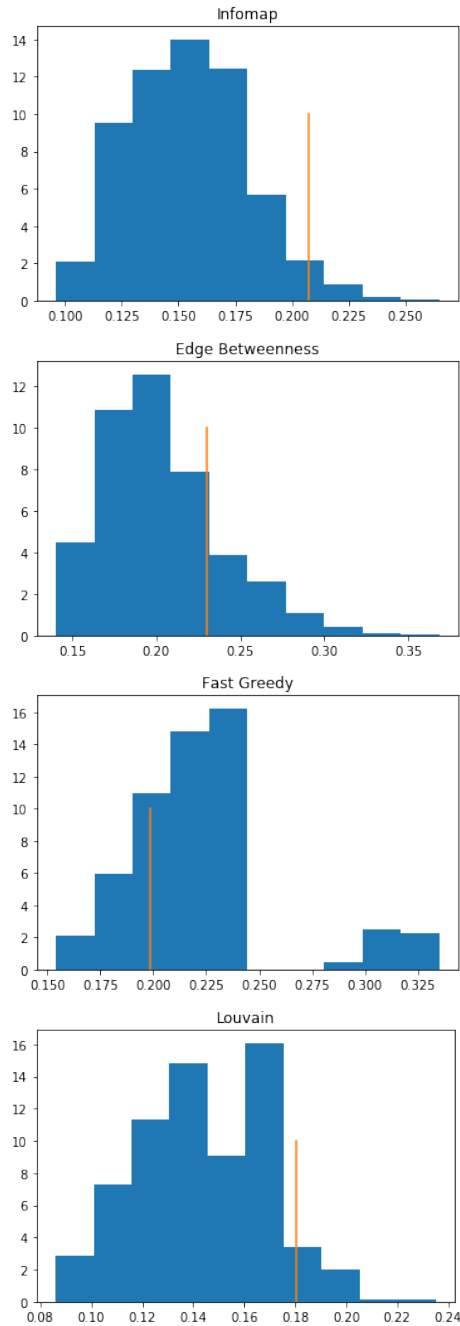


FIG. 8. En azul, información mutua de la red con distribución al azar de generos de nodos de la red. En línea naranja valor de información mutua de comunidad y género provisto por la base de datos.

#### 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?

La idea de este ítem era estudiar la existencia de correlaciones entre el género de los delfines y el modo como se distribuyen en comunidades. Se consideró la distribución de géneros como si fuera una partición en sí misma. Entonces, los géneros de los delfines se etiquetaron del mismo modo que las particiones: "1" para "male", "2" para "female" y "3" para los casos de género "N/A". Una vez que el género se hubo codificado como partición la correspondiente distribución de género se comparó por medio de la información mutua. Como resultado se obtuvieron valores de información mutua para cada una de las cuatro comparaciones posibles.

Sin embargo, a efectos de establecer la significancia de los valores obtenidos, comparamos los valores de información mutua con los que cabría esperar por azar. Para eso reasignamos el género de los nodos al azar, y recalculamos la información mutua. Realizamos 10 000 iteraciones para obtener la distribución de información mutua entre las particiones reales y la asignación de géneros al azar.

En la figura 8 podemos observar que la distribución de géneros al azar no difiere significativamente respecto de la distribución de generos (y comunidades) provista por la base de datos, de manera que no podemos decir que los clusters son homofílicos. Por otro lado, es interesante notar que los valores medios son pequeños. Esto no tiene una interpretación sencilla, dado que no hay un valor medio de referencia para esta clase de comparación puntual.