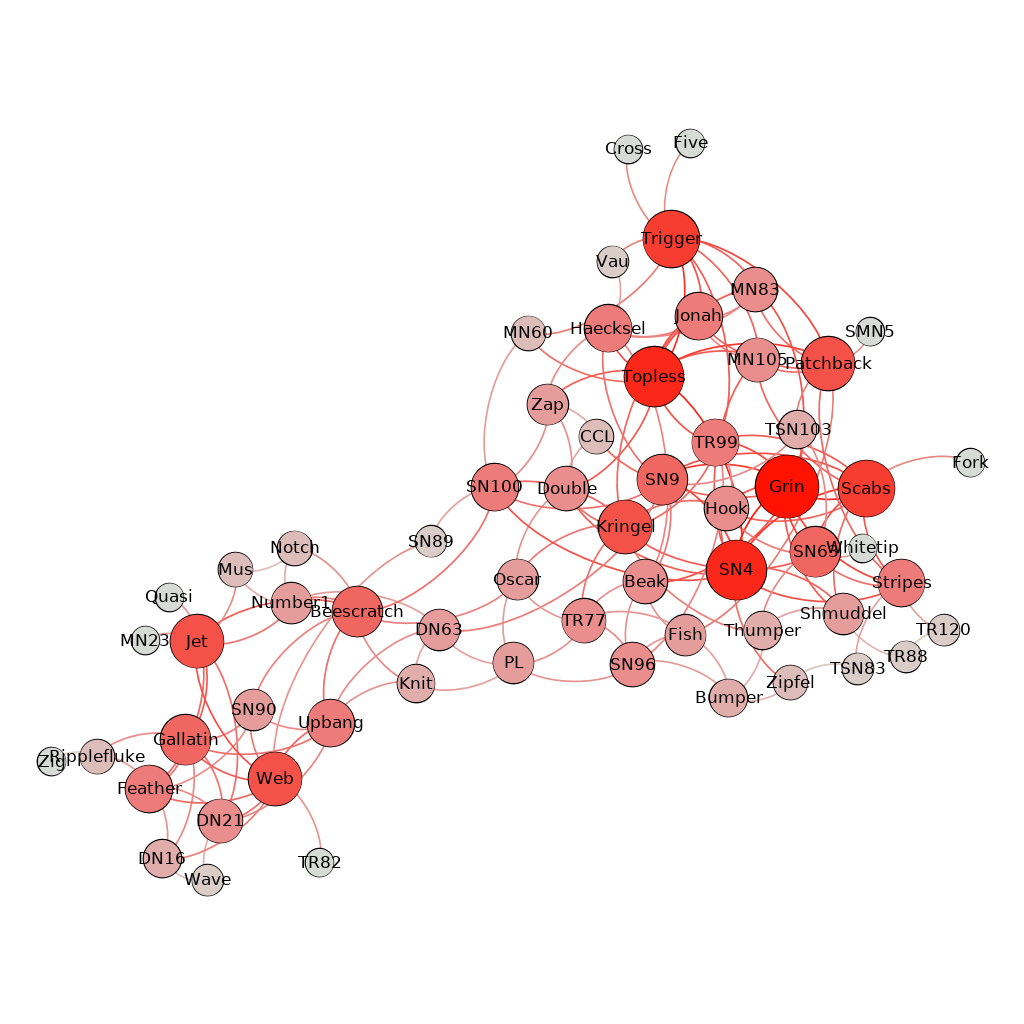
Análisis y Visualización Básica de una Red Social con Gephi

Curso académico 2018-2019

Dolphin Social Network



Trabajo realizado por: Jesús Pérez Martínez 49217794W

j.perezmar@posgrado.uimp.es

Índice de contenidos

[1. Introducción 1](#_Toc530081367)

[2. Análisis y visualización básica de la red 2](#_Toc530081368)

[3. Estudio de la centralidad de los actores 6](#_Toc530081369)

[4. Gráficas adicionales 11](#_Toc530081370)

[5. Conclusión 13](#_Toc530081371)

Índice de ilustraciones

[Ilustración 1 - representación gráfica de la red 2](file:///C:\Users\Xachap\Documents\Universidad\MUIIA\REIGR\Bloque%20I%20-%20Redes%20sociales\Práctica%20Bloque%20I%20Redes%20Sociales\Análisis%20y%20Visualización%20Básica%20de%20una%20Red%20Social%20con%20Gephi.docx#_Toc530081449)

[Ilustración 2 - distribución de grados en la red social 4](#_Toc530081450)

[Ilustración 3 - histograma de distancias medias en la red 4](#_Toc530081451)

[Ilustración 4 - distribución de los coeficientes de clustering 5](#_Toc530081452)

[Ilustración 5 - distribución de grados sobre la red 6](#_Toc530081453)

[Ilustración 6 - distribución del nivel de intermediación sobre la red 7](#_Toc530081454)

[Ilustración 7 - distribución del nivel de cercanía sobre la red 8](#_Toc530081455)

[Ilustración 8 - distribución del nivel del valor del vector propio sobre la red 9](#_Toc530081456)

[Ilustración 9 - representación de la centralidad de grado y vector propio 11](#_Toc530081457)

[Ilustración 10 - representación de las medidas de intermediación y cercanía 12](#_Toc530081458)

[Ilustración 11 - representación de la medida de intermediación y del vector propio 13](#_Toc530081459)

Índice de tablas

[Tabla 1 - diferentes medidas de análisis de la red social 3](#_Toc530081478)

[Tabla 2 - actores con mayor grado de la red 6](#_Toc530081479)

[Tabla 3 - actores con mayor nivel de intermediación de la red 7](#_Toc530081480)

[Tabla 4 - actores con mayor nivel de cercanía de la red 8](#_Toc530081481)

[Tabla 5 - actores con mayor nivel de vector propio de la red 9](#_Toc530081482)

[Tabla 7 - resultados de la clasificación por aparición ponderada en las medidas 10](#_Toc530081483)

# Introducción

La presente red social que analizaremos a lo largo de este documento tiene su origen en el artículo *Evidence for Social Role in a Dolphin Social Network* de David Lusseau. En él, se estudian las relaciones entre varios grupos de delfines que habitan en Doubtful Sound, un fiordo localizado en Nueva Zelanda. Estas relaciones fueron estudiadas entre los años 2000 y 2002, usando identificación por imágenes y análisis de códigos de conducta, y como consecuencia se creó una red social no dirigida de 62 nodos y 159 enlaces, que es la que vamos a estudiar en el presente documento usando la herramienta *Gephi* (versión 0.92 de Linux).

Comenzaremos haciendo un estudio básico de la red, en el que se expondrán diferentes medidas globales, las cuales corresponden número de nodos y enlaces, grado medio, número máximo de enlaces, diámetro, distancia media, coeficiente medio de clustering y número de componentes conexas. Posteriormente, se realizará una comparativa de medidas locales, acompañada de gráficos adicionales para mostrar los resultados de una manera más representativa, y poder detectar los nodos más importantes de la red.

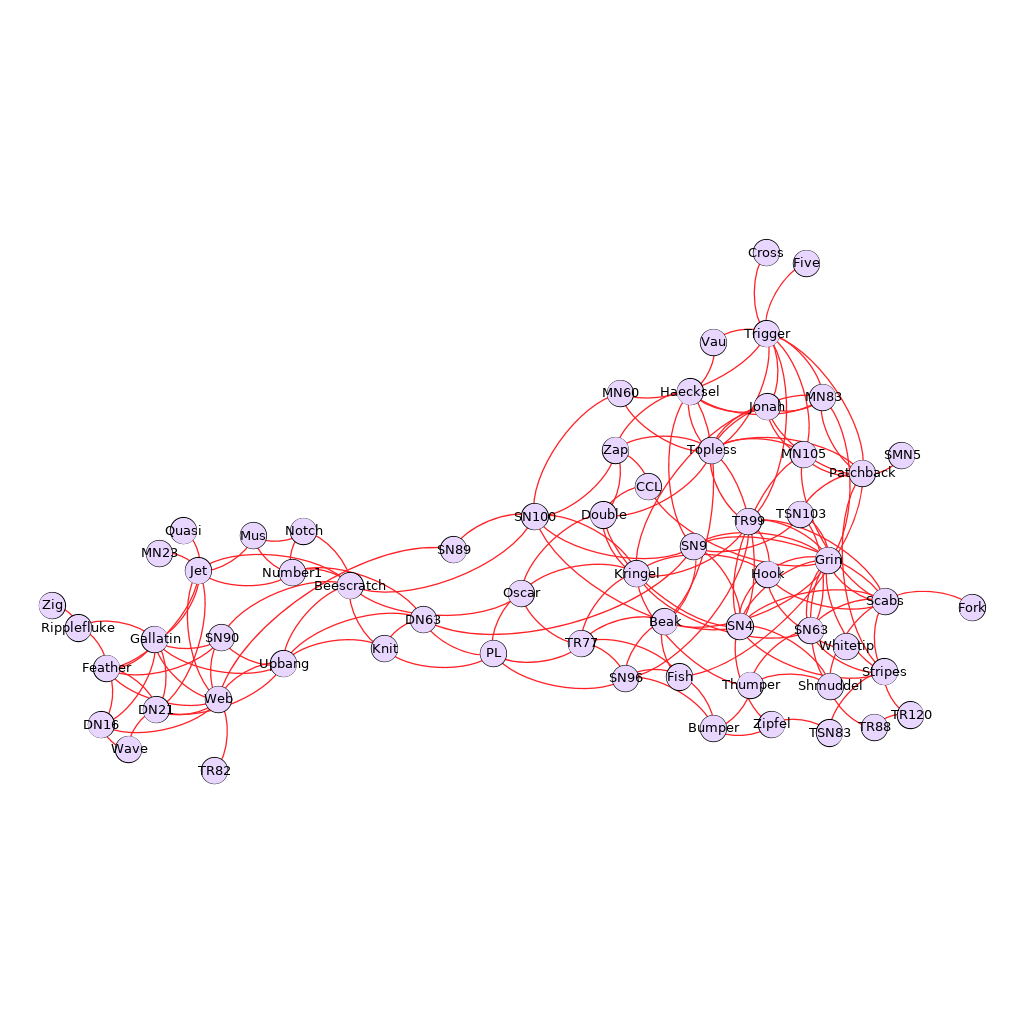
# Análisis y visualización básica de la red

Tras cargar la red en *Gephi*, observamos que es necesario aplicar un algoritmo de *layout*, para favorecer una correcta visualización de la misma. En este caso, el algoritmo usado ha sido *Force Atlas 2*, con los siguientes parámetros:

* **Gravedad de puesta a punto:** 10.0
* **Disuadir Hubs:** true
* **Evitar solapamiento:** true

En la *Ilustración 1* podemos ver una representación genérica con los nodos coloreados y un tamaño fijo, para favorecer la visualización de la red.

Ilustración 1 - representación gráfica de la red



Tras representar gráficamente la red de una manera clara, procederemos a realizar un análisis general usando las herramientas que aparecen en la ventana *Contexto* del programa. Los datos recogidos aparecen en la siguiente tabla:

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| Número de nodos *N* | 62 |
| Número de enlaces *L* | 159 |
| Número máximo de enlaces *Lmax* | 1891 |
| Densidad del grafo *L/Lmax* | 0.084 |
| Grado medio *<k>* | 5.129 |
| Diámetro *dmax* | 8 |
| Distancia media *d* | 3.357 |
| Coeficiente medio de clustering *<C>* | 0.303 |
| Número de componentes conexas | 1 |
| Número de nodos componente gigante | 62 |
| % de aristas componente gigante | 100% |

Tabla 1 - diferentes medidas de análisis de la red social

Los datos más inmediatos que encontramos son el número de nodos y de enlaces (ya comentados previamente). Con el número de nodos *N*, podemos calcular de forma manual el número máximo de enlaces , a partir de la siguiente fórmula:

Con este valor podemos estimar la *Densidad del grafo*, tanto con la fórmula que aparece en la tabla como de forma automática utilizando *Gephi*. En ambos casos obtenemos un valor de 0.084, un número bajo que es representativo de las redes complejas reales.

Por otra parte, el grado medio tiene un valor de 5.129, lo que nos indica que, de media, cada delfín de la red está relacionado con otros 5. La distribución de grados aparece en la *Ilustración 2*, y en ella podemos ver que se aprecia sutilmente (debido al pequeño número de individuos de la red) la forma de larga estela. Encontramos un número elevado de individuos que tienen pocos contactos, y por el contrario los individuos con muchos enlaces son muy escasos. Para este último caso podemos ver que sólo 13 de ellos tienen 8 o más enlaces con otros enlaces, lo que supone aproximadamente el 21% del total. Por lo tanto, podemos afirmar que la red tiene la propiedad **libre de escala (scale-free)**.

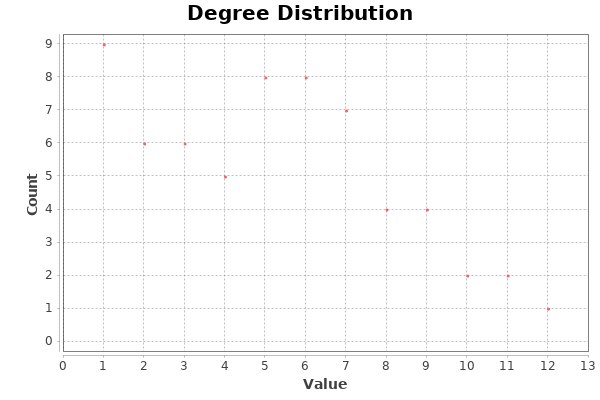


Ilustración 2 - distribución de grados en la red social

El siguiente valor para analizar es el diámetro y la distancia media, con valores 8 y 3.357 respectivamente. En la *Ilustración 3* aparece representada la distribución en forma de histograma[[1]](#footnote-1) de distancias medias de cada uno de los nodos. Como vemos, son muy comunes las distancias pequeñas en la red. Esto nos indica que se ve reflejada la propiedad de mundos pequeños, muy común en redes complejas reales, ya que las distancias entre nodos son logarítmicas.

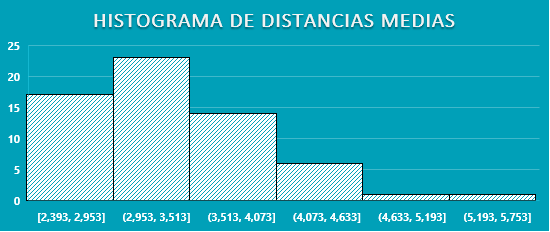


Ilustración 3 - histograma de distancias medias en la red

A continuación, analizaremos la conectividad de la red. Primeramente, el Coeficiente de Clustering, toma el valor de 0.303, un valor no muy alto para ser una red social, pero bastante significativo. Éste nos indica que hay aproximadamente un 30% de probabilidad de que, dado un nodo, dos de sus contactos estén conectados. Podemos ver la distribución del Coeficiente de Clustering en la *Ilustración 4*. Por otra parte, sólo encontramos una componente conexa en esta red.

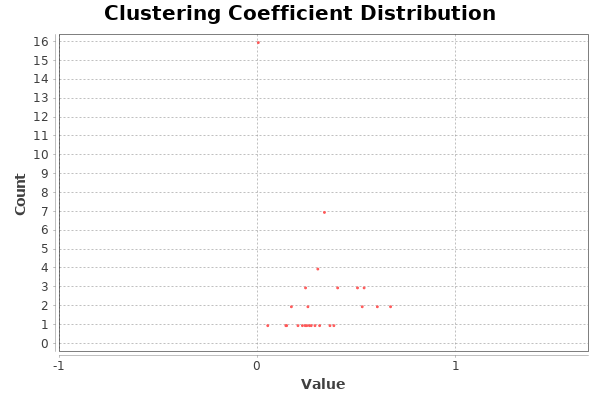


Ilustración 4 - distribución de los coeficientes de clustering

# Estudio de la centralidad de los actores

En este apartado realizaremos un estudio de la centralidad para ver que actores dentro de la red son realmente influyentes, para ello utilizaremos las medidas de grado, intermediación, cercanía y vector propio.

La medida de grado es la más básica de todas, y ya mostrábamos la distribución de grados en la *Ilustración 2* del apartado anterior. Aparece su representación gráfica en la *Ilustración 4*, donde tanto el tamaño como la intensidad del color de cada uno de los nodos es proporcional al grado de éstos. A continuación, en la *Tabla 2* aparecen los 5 actores con mayor grado de la red.

|  |  |  |
| --- | --- | --- |
| **Centralidad de grado** | | |
| **Ranking** | **Nodo** | **Grado** |
| 1º | Grin | 12 |
| 2º | SN4 | 11 |
| 3º | Topless | 11 |
| 4º | Scabs | 10 |
| 5º | Trigger | 10 |

Tabla 2 - actores con mayor grado de la red

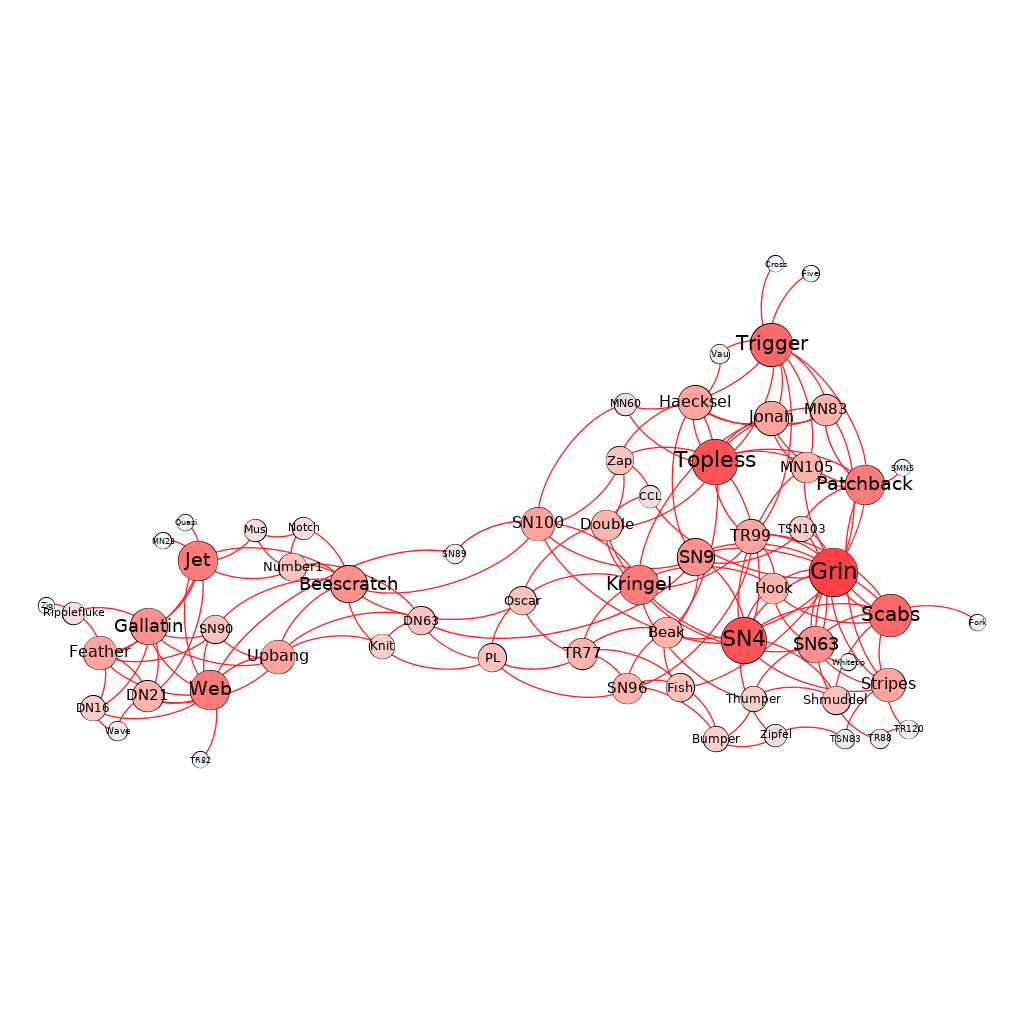


Ilustración 5 - distribución de grados sobre la red

Ahora comenzaremos a analizar la red con dos medidas basadas en distancias, la intermediación y la cercanía. Tal y como podemos ver en la *Tabla 3*, los resultados cambian notablemente si medimos cuales son los nodos con mayor intermediación, siendo el delfín SN4 el único común con el anterior criterio de análisis. Esta medida representa la cuantía de caminos mínimos que pasan por un nodo, por lo que como es lógico y tal como podemos ver en el gráfico de la red en la *Ilustración 5*, los nodos con mayor nivel de intermediación se encuentran en la parte central de la red, haciendo de puente entre los dos extremos.

|  |  |  |
| --- | --- | --- |
| **Centralidad de intermediación** | | |
| **Ranking** | **Nodo** | **Intermediación** |
| 1º | SN100 | 0.248 |
| 2º | Beescratch | 0.213 |
| 3º | SN9 | 0.143 |
| 4º | SN4 | 0.139 |
| 5º | DN63 | 0.118 |

Tabla 3 - actores con mayor nivel de intermediación de la red

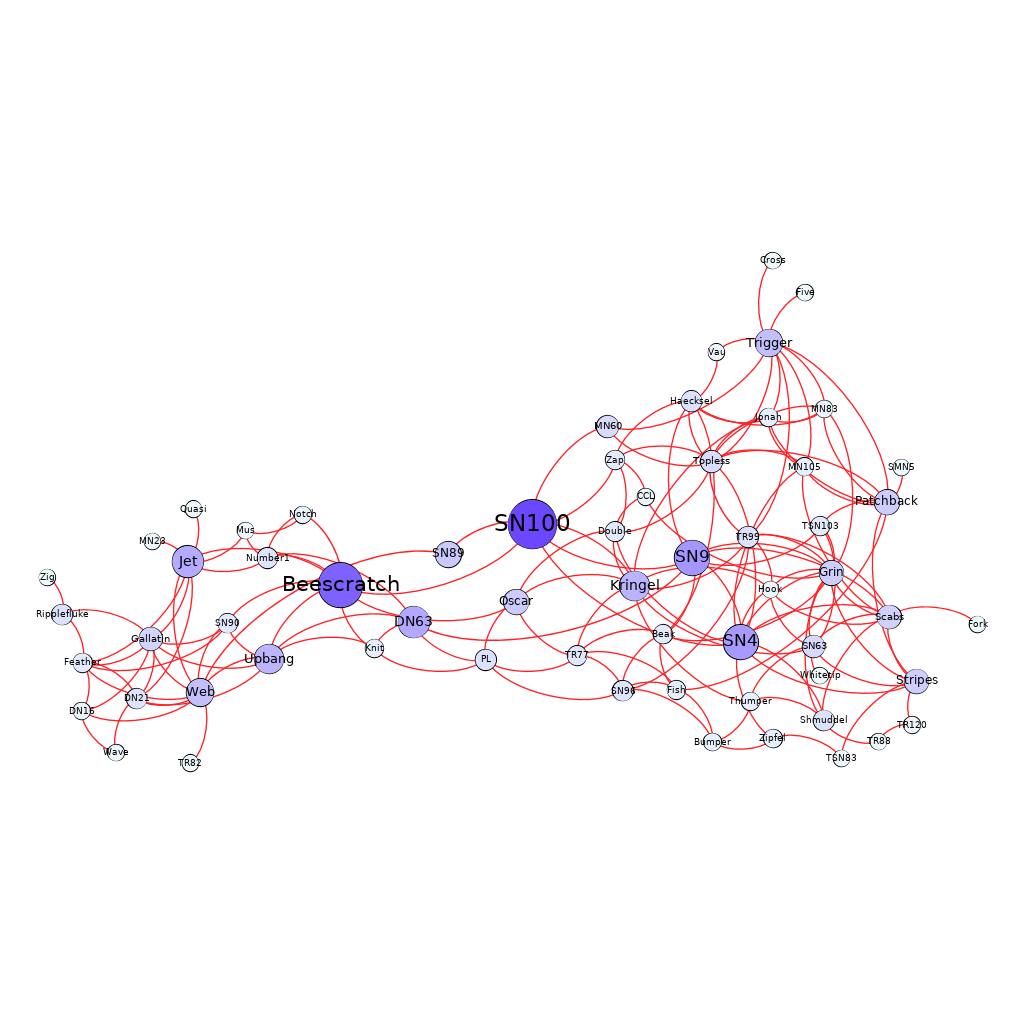


Ilustración 6 - distribución del nivel de intermediación sobre la red

Si nos enfocamos en realizar un estudio de la cercanía, encontramos que, de nuevo, hay algunos nodos que ya han aparecido en los dos casos anteriores, los cuales son SN100, SN9, SN4 y Grin. Estos nodos tienen por término medio las distancias más cortas a todos los demás que encontramos en la red. De nuevo, observamos en la *Ilustración 6* que los nodos con mayor nivel de cercanía se encuentran en la parte central de la red. Aunque en este caso, a diferencia del nivel de intermediación, encontramos un mayor grado de cercanía en la parte derecha de la red. Esto se debe a la mayor densidad de nodos de dicha zona. En la *Tabla 4* vemos, de nuevo, los actores 5 actores con mayor nivel de cercanía.

|  |  |  |
| --- | --- | --- |
| **Centralidad de cercanía** | | |
| **Ranking** | **Nodo** | **Cercanía** |
| 1º | SN100 | 0.418 |
| 2º | SN9 | 0.404 |
| 3º | SN4 | 0.399 |
| 4º | Kringel | 0.391 |
| 5º | Grin | 0.376 |

Tabla 4 - actores con mayor nivel de cercanía de la red

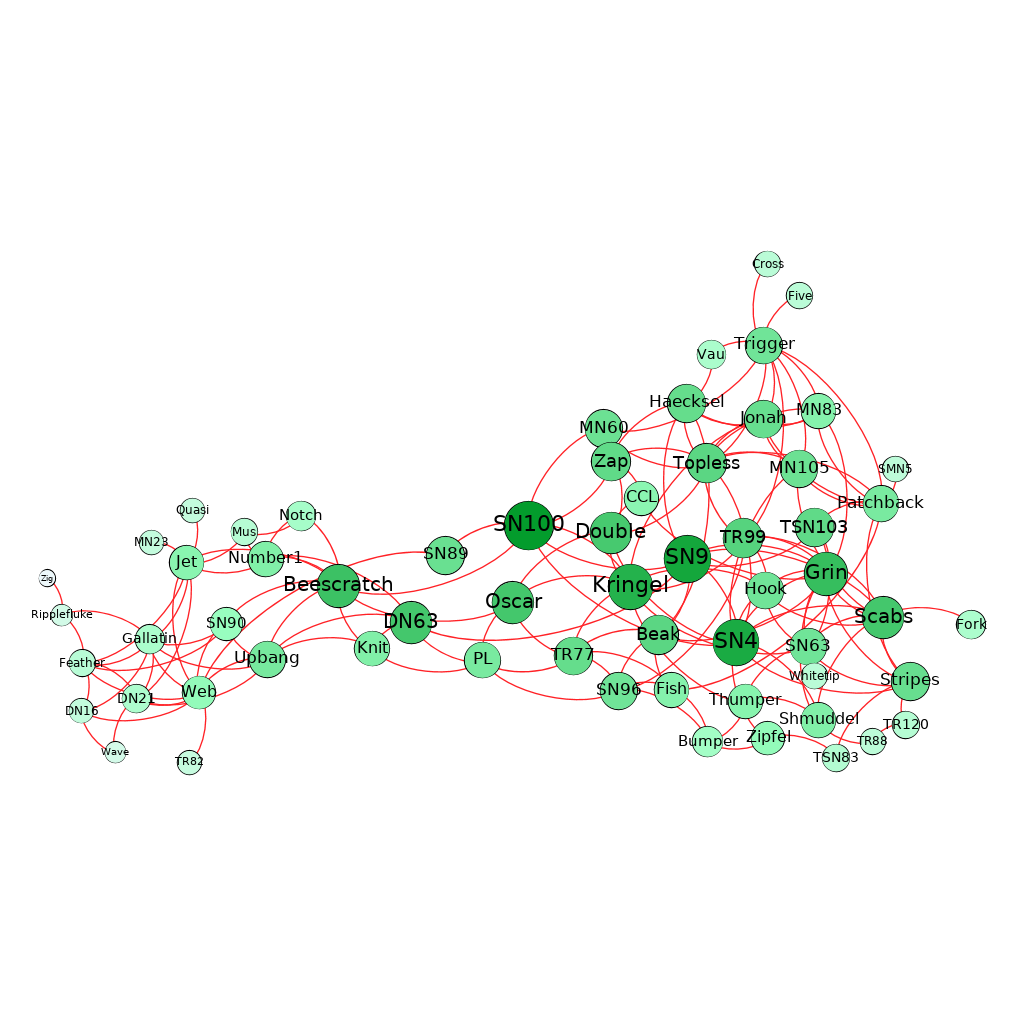


Ilustración 7 - distribución del nivel de cercanía sobre la red

Para finalizar nuestra comparación de resultados haciendo uso de diferentes medidas, analizaremos los resultados obtenidos al estudiar la centralidad del vector propio, considerada una de las medidas locales de centralidad más importantes. En esta medida de grados ponderados, vemos que se repiten algunos de los nodos encontrados en el estudio de la centralidad de grado, como son los delfines Grin, SN4, Topless y Scabs. A continuación, observamos en la *Tabla 5* el ranking de nodos para esta medida, y su representación gráfica en la *Ilustración 7.*

|  |  |  |
| --- | --- | --- |
| **Centralidad de vector propio** | | |
| **Ranking** | **Nodo** | **Vector propio** |
| 1º | Grin | 1.0 |
| 2º | SN4 | 0.959 |
| 3º | Topless | 0.908 |
| 4º | Scabs | 0.882 |
| 5º | TR99 | 0.698 |

Tabla 5 - actores con mayor nivel de vector propio de la red

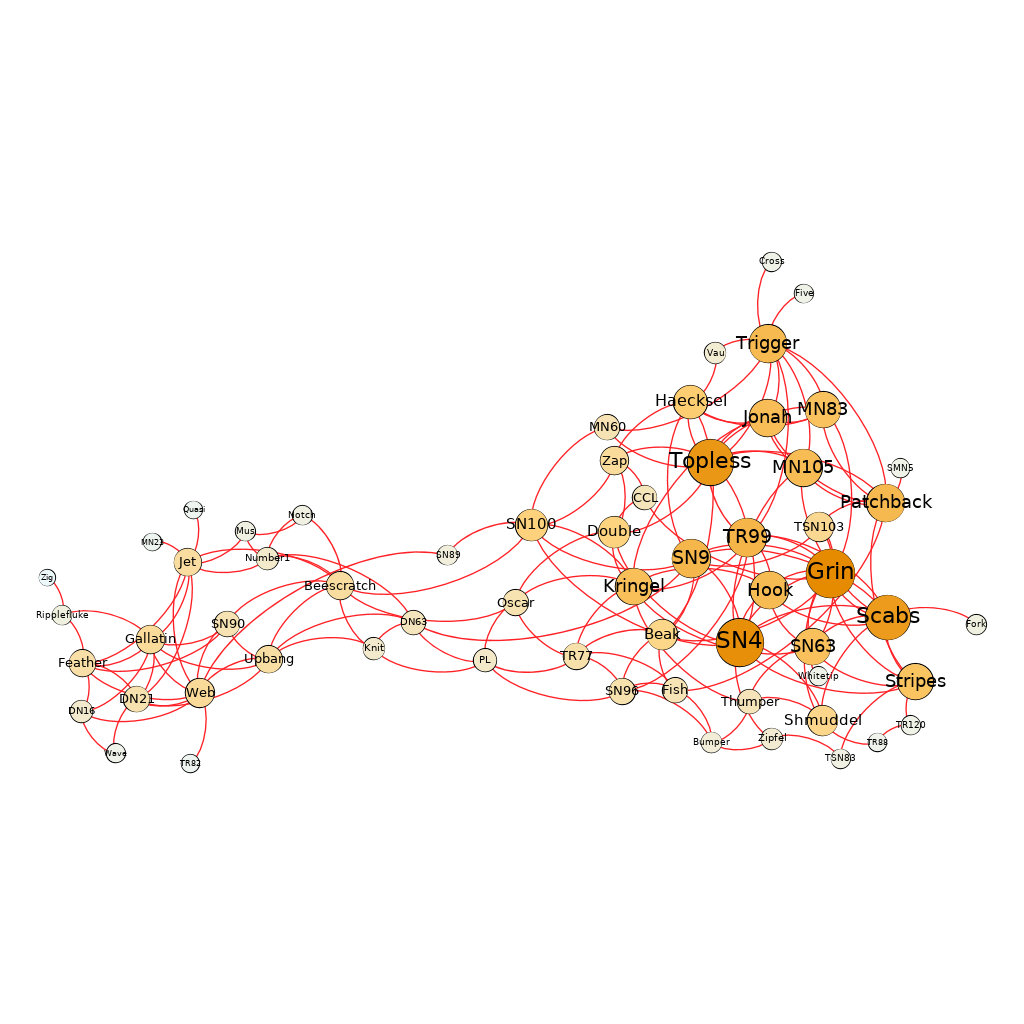


Ilustración 8 - distribución del nivel del valor del vector propio sobre la red

Habiendo realizado las siguientes medidas para encontrar los 5 nodos más influyentes para cada caso, podemos realizar una clasificación de cada delfín atendiendo al número de veces que ha aparecido en cada una de las medidas y su ranking. Así pues, estableceremos la siguiente fórmula:

Dónde el ranking de un nodo *j* viene dado por el sumatorio de las puntuaciones *x* para cada una de las medidas locales denotadas por *i*. Cada puntuación se calcula de acuerdo con las siguientes sentencias condicionales:

* Si un nodo *j* ha quedado en primera posición para una medida local, contará como 0,5.
* Si un nodo *j* ha quedado en segunda posición para una medida local, contará como 0,4.
* Si un nodo *j* ha quedado en tercera posición para una medida local, contará como 0,3.
* Si un nodo *j* ha quedado en cuarta posición para una medida local, contará como 0,2.
* Si un nodo *j* ha quedado en quinta posición para una medida local, contará como 0,1.
* En otro caso, si un nodo *j* no aparece entre los cinco mejores para una medida local dada, contará como 0 y no sumará.

Por lo tanto, los resultados de la clasificación son los siguientes:

|  |  |  |
| --- | --- | --- |
| **Posición** | **Nodo** |  |
| 1º | SN4 | 1.3 |
| 2º | Grin | 1.1 |
| 3º | SN100 | 1.0 |
| 4º | SN9 | 0.7 |
| 5º | Topless | 0.6 |
| 6º | Scabs | 0.4 |
| 7º | Beescratch | 0.4 |
| 8º | Kringel | 0.2 |
| 9º | Trigger | 0.1 |
| 10º | DN63 | 0.1 |
| 11º | TR99 | 0.1 |

Tabla 7 - resultados de la clasificación por aparición ponderada en las medidas

Esta clasificación nos da una visión clara del número de apariciones de los nodos para cada medida local. Los nodos con mayor importancia son Grin, SN4 y SN100, los cuales han aparecido en la mitad o más de las medidas locales realizadas, y en casi todos los casos con una importancia notable.

Los siguen SN9 y Topless, apareciendo dos veces cada uno, el primero en medidas basadas en distancias y el segundo en las dos medidas basadas en grado. Los nodos siguientes sólo aparecen una vez, o bien dos veces en cuarta posición, por lo que tienen una importancia mucho menor.

# 4. Gráficas adicionales

Gracias a las diferentes opciones de la ventana *Apariencia* que proporciona *Gephi*, podemos mostrar la representación de dos medidas en el mismo grafo. En nuestro caso, usaremos el tamaño y la intensidad del color para representar cada pareja de medidas. Cuanto más grande sea el nodo, indicará mayor valor para la medida representada, y lo mismo para la intensidad del color. Comenzaremos comparando las medidas de centralidad de grado y vector propio, representadas en la *Ilustración 9* por color y tamaño respectivamente.

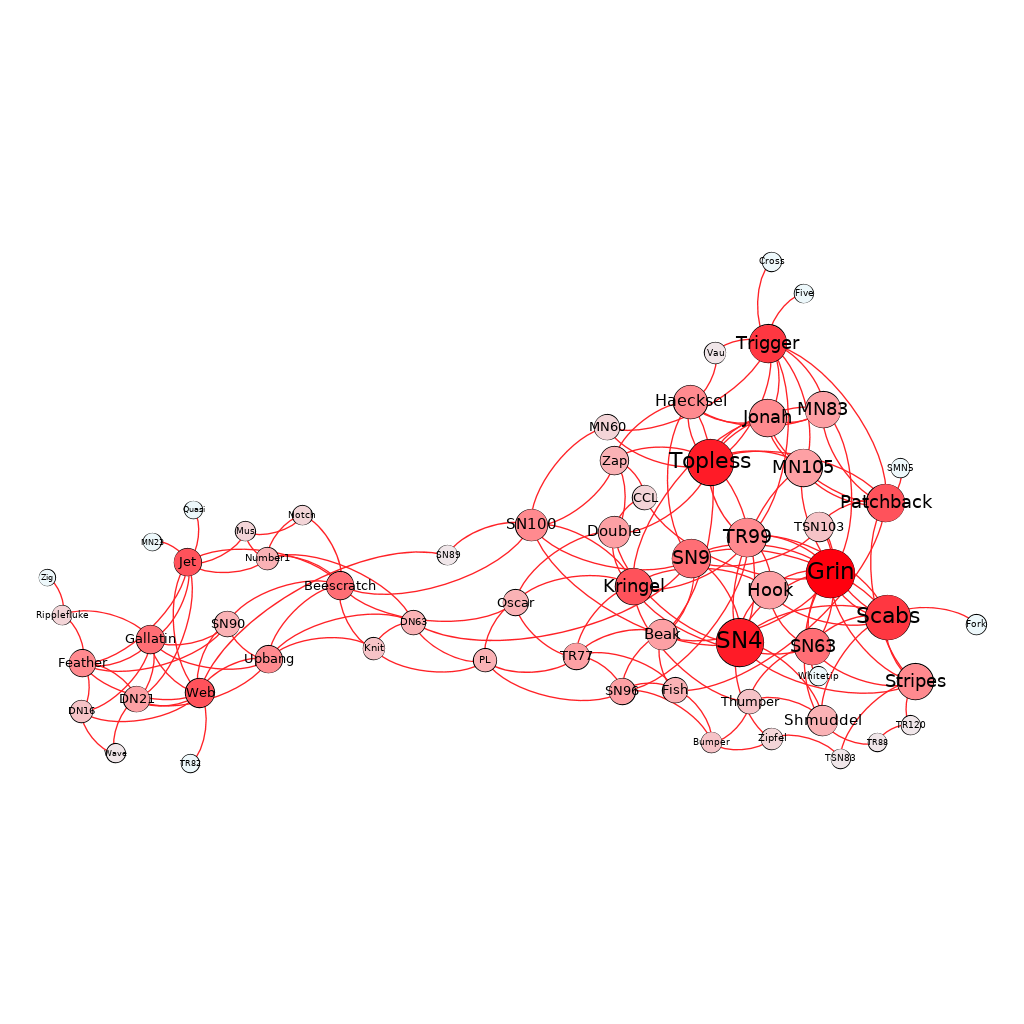


Ilustración 9 - representación de la centralidad de grado y vector propio

Tal y como vimos en el [*Estudio de la centralidad de los actores*](#_Estudio_de_la), los valores obtenidos por la medida de la centralidad de grado y por la de vector propio son muy parecidos, hecho que se refleja claramente en la *Ilustración 9*. Prácticamente es el mismo grafo que el obtenido en la *Ilustración 5* y en la *Ilustración 8*. Ahora compararemos en la *Ilustración 10* las medidas de intermediación (representada en tamaño) y cercanía (representada con intensidad de color).

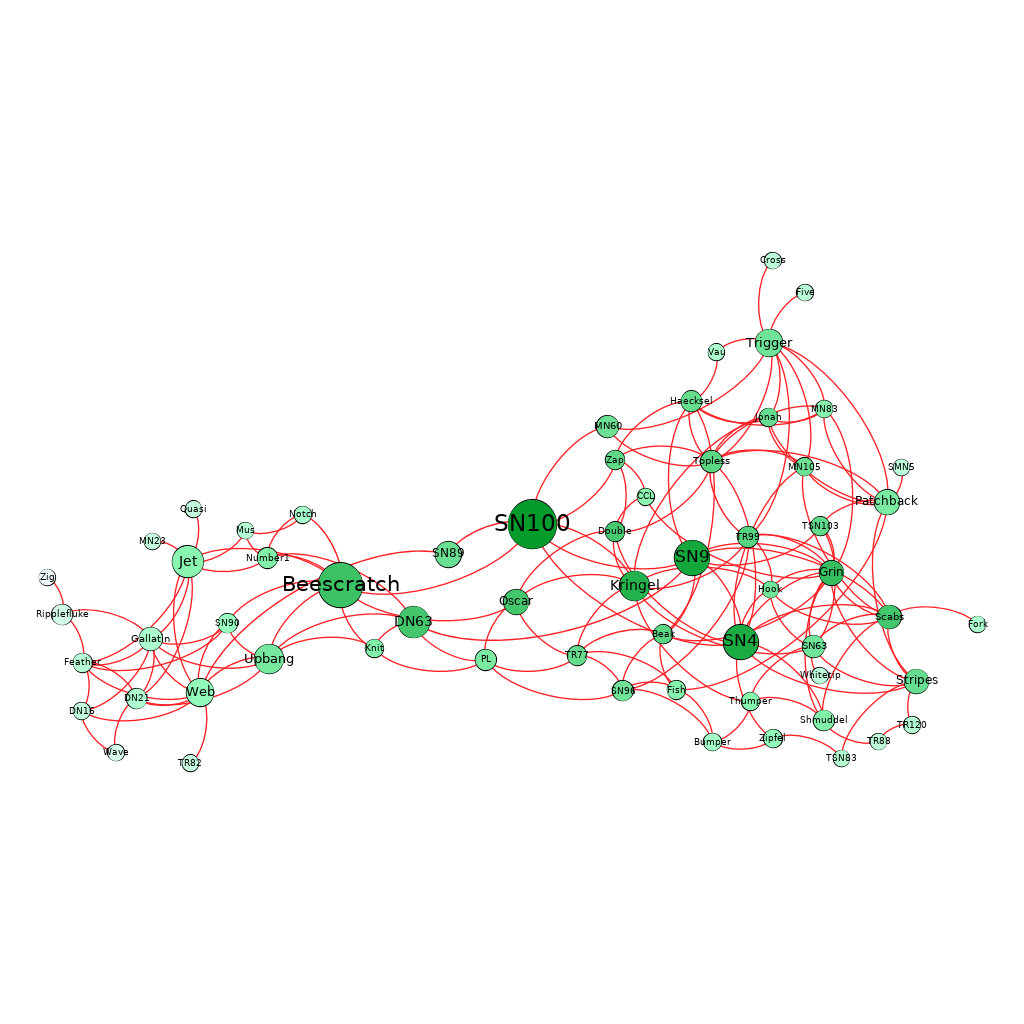


Ilustración 10 - representación de las medidas de intermediación y cercanía

Para esta comparativa obtenemos un resultado más representativo, al mezclar las dos medidas de cercanía en el mismo grafo. Los nodos de cada uno de los extremos son especialmente pálidos y pequeños en comparación con los que se encuentran en el centro.

A continuación, mostraremos la intermediación junto al vector propio, al ser ésta última más representativa que la medida de centralidad de grado. En la *Ilustración 11* aparece la centralidad de intermediación representada en la intensidad del color, y la centralidad de vector propio en tamaño.

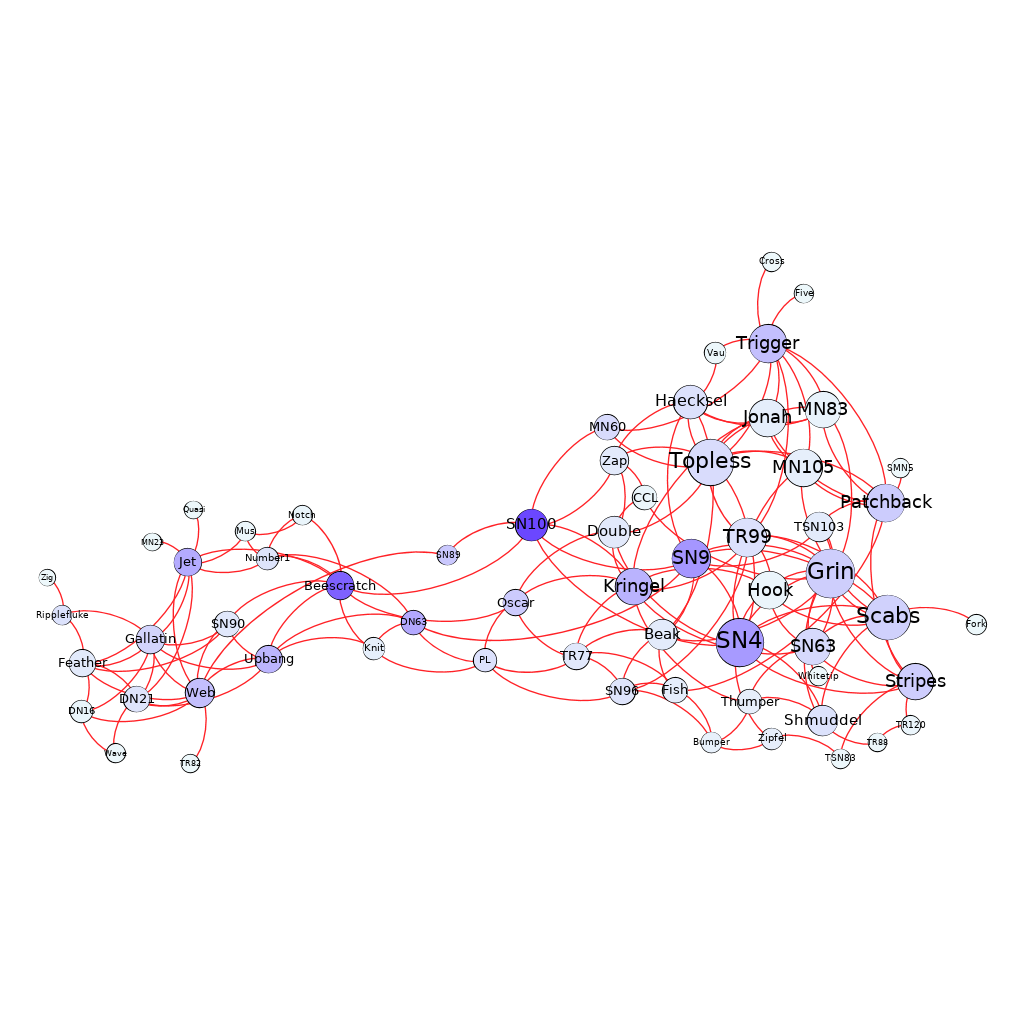


Ilustración 11 - representación de la medida de intermediación y del vector propio

Sólo se realiza la comparativa con la medida de intermediación, ya que el resultado de la cercanía es el mismo al compararlo con la centralidad del vector propio.

# 5. Conclusión

**Bibliografía:**

[1] [Óscar Cordón García. Unidad 2. Bloque I: Redes Sociales (consultado el 15/11/18, se necesita autenticación) (2018).](https://poliformat.upv.es/portal/site/ESP_0_2620/tool/0fe39039-ffa7-405a-8ab1-244ee88ffc53)

[2] [D. Lusseau. Evidence for social role in a dolphin social network (consultado el 15/11/18) (2003).](https://arxiv.org/abs/q-bio/0607048)

1. Este histograma ha sido generado con *Excel*, a partir del archivo *.csv* generado con *Gephi*. Los datos han sido calculados a partir de la medida de centralidad de cada nodo [↑](#footnote-ref-1)