



UNIVERSIDAD DE LA REPÚBLICA
Facultad de Ciencias Económicas y de Administración
Instituto de Estadística

Aplicación del análisis de redes para el estudio del gasto en turistas de cruceros en Uruguay para la temporada 2010-2011

Ramón Álvarez-Vaz; Silvia Altmark

Abril, 2019

Serie Documentos de Trabajo

DT (19/1) - ISSN : 1688-6453

Forma de citación sugerida para este documento:

Álvarez-Vaz, Ramón y Altmark, Silvia (2019). Aplicación del análisis de redes para el estudio del gasto en turistas de cruceros en Uruguay para la temporada 2010-2011[en línea]. Serie Documentos de Trabajo, DT (19/1). Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay.

Aplicación del análisis de redes para el estudio del gasto en turistas de cruceros en Uruguay para la temporada 2010-2011.

Ramón Álvarez-Vaz ¹; Silvia Altmark ²

Departamento de Métodos Cuantitativos, Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, Universidad de la República

RESUMEN³

Desde hace casi 10 años el turismo de cruceros en Uruguay viene creciendo, determinando un importante aporte de divisas cada temporada (abril a octubre de cada año) y para eso es necesario poder caracterizar las variables económicas involucradas en esta actividad, en particular, el gasto. Para este trabajo se comparan la tipología surgida de aplicar métodos de clusters jerárquicos y no jerárquicos, con la que surge al aplicar el análisis de redes (SNA) a los datos del gasto en cruceristas. A partir de los datos de las temporadas 2010-2011, del Ministerio de Turismo y Deporte, obtenidos de una muestra de pasajeros a través de una encuesta cara a cara con diseño muestral complejo, los autores ya habían construido una tipología de cruceristas creadas al aplicar el algoritmo de Ward sobre distancias para variables binarias (gasta o no gasta) en 5 rubros.

Para evaluar como funciona el análisis de redes se seleccionan 4 cruceros, (según el tamaño de cada uno) sobre los que, a partir de los gastos binarios, se construyen grafos y sobre las que se aplican las diferentes métricas para la descripción de los mismos. Se estudia la asociación de las características socio-demográficas de los cruceristas con la tipología previa de gastos y las comunidades identificadas con el SNA para identificar eventuales patrones de comportamiento al cambiar de tipo de crucero.

Palabras clave: Análisis de Redes, Clustering, Gastos de Cruceristas, Métricas.

CÓDIGOS JEL:

C10, C30, C50

Clasificación MSC2010:90B10,62D05, 62H30

¹ *email:* ramon@iesta.edu.uy, ORCID: 0000-0002-2505-4238

² *email:* salt@iesta.edu.uy, ORCID: 0000-0003-3123-2165

³ Este documento esta basado en el trabajo final del curso "DE LAS REDES COMPLEJAS A LAS REDES SOCIALES: INTRODUCCIÓN AL USO DEL BIG DATA", dictado en Espacio Interdisciplinario, 2017

ABSTRACT⁴

For almost 10 years cruise tourism in Uruguay has been growing, determining an important contribution of foreign currency each season (April to October of each year) for that it is necessary to be able to characterize the economic variables involved in this activity, in particular, the expenditure in its different areas. In this working paper the author compare the typology that arose from applying methods of hierarchical and non-hierarchical clusters, with which it arises when applying network analysis (SNA) to data on cruise spending. Based on data from the 2010-2011 seasons, from the Ministerio de Turismo y Deportes (Ministry of Tourism and Sports), to data coming from a sample of passengers through a face-to-face survey with a complex sampling design. The authors had already constructed a typology of cruise passengers applying Ward's algorithm over distances for binary variables (spend or not spend) in 5 items. To evaluate performance of social network analysis, 4 cruises are selected with probability proportional to size on which, based on binary expenditures, graphs are constructed and on which the different metrics are applied to describe them. The association of the socio-demographic characteristics of the cruise passengers with the previous typology of expenditures and the communities identified with the SNA is studied to identify patterns with the different type of cruises.

Key words: Network Analysis, Clustering, Cruise Expenses, Metrics

JEL CODES: C10, C30, C50

Mathematics Subject Classification MSC2010: 90B10,62D05, 62H30 .

⁴This document is based on the final work of the course "FROM COMPLEX NETWORKS TO SOCIAL NETWORKS: INTRODUCTION TO THE USE OF BIG DATA", delivered at Espacio Interdisciplinario, 2017

1. Introducción

El turismo es una actividad muy relevante para la economía del Uruguay, en términos de divisas, valor agregado (PIB) y empleo. De acuerdo a las estimaciones del Ministerio de Turismo, surgidas del Anuario, en 2015, el aporte del sector a la economía uruguaya es del orden del 14 % de las exportaciones totales y el 58 % de las exportaciones de servicios del 7,7 % del PIB, en lo que de alrededor de 4 % en términos de generación del PIB y 8 % con respecto a la contribución a los puestos de trabajo del país.

De acuerdo a los datos relevados por el Ministerio de Turismo a partir de sus encuestas, el turismo de cruceros presenta una evolución creciente en Uruguay, partiendo de 75 arribos, 56.167 pasajeros desembarcados y *U\$S*18,899,464 de gasto en la temporada 2004-2005, desde cuando se dispone de datos. Entre esa temporada y la temporada 2011-2012, la cantidad de personas desembarcadas aumentó más del 500 % y el gasto en *U\$S*, el 11 %. Es importante señalar que la temporada de cruceros se extiende generalmente entre octubre de un año y abril del siguiente.

Teniendo en cuenta estos antecedentes y partiendo de trabajo anteriores, sobre los que se aplicó técnicas de clustering ((Alvarez Vaz et al., 2015)) para crear tipologías de cruceristas, se propone como ejercicio complementar los resultados a través del análisis de redes.

El objetivo principal de este trabajo es el poder caracterizar los grupos de cruceristas, que viajan en los diferentes cruceros seleccionados, con los siguientes objetivos específicos:

- Generar una red por cada tipo de crucero seleccionado y caracterizarla mediante diferentes métricas que se presentan en 2
- Describir las características de cada red mediante el comportamiento de algunos nodos
- Caracterizar los nodos mas importantes de cada red analizada
- Analizar si existe un patrón o estructura que se mantiene al cambiar de tipo de crucero (tamaño del crucero)

El trabajo está organizado de la siguiente forma: en la sección 2 y 3 se presentan brevemente en que consiste el problema en estudio y se desarrollan las técnicas aplicar y los datos que se utilizan, en la sección 4 se presentan los resultados, que se discuten en la 5, para terminar en la última sección 6, donde se presentan las conclusiones y futuros pasos.

2. Marco teórico y Métricas para Caracterización de los nodos y los vértices

En esta sección se presentan las diferentes métricas que se usan para la caracterización de las redes. Para la presentación de las mismas se seguirá la notación de del libro 'Statistical Analysis of Network Data with R' (Kolaczyk and Csárdi, 2014),(Luke, 2015) aunque textos seminales como (Wasserman and Faust, 1994),(Borgatti et al., 2013) son una guía también a seguir.

2.1. Grados de los vértices

Los grados de d_v de un vértice v de un grafo $G(V, E)$ es el número de aristas en E incidentes sobre V . A partir de esta medida se puede definir f_d como la fracción de vértices de $v \in V$ con grado $d_v = d$. El conjunto $\{f_d\}d \leq 0$ es lo que se llama *distribución de grados* de G . Para las redes ponderadas, una generalización útil del grado es la noción de *Fuerza de vértice* que se obtiene simplemente sumando los pesos de los bordes de un vértice dado

2.2. Centralidad de los vértices

Las medidas de centralidad de intermediación tienen por objeto resumir en qué medida un vértice se encuentra 'entre' otros pares de vértices (Freeman, 1979) (**Betweenness centrality**)

$$c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s, t|v)}{\sigma(s, t)} \quad (1)$$

Donde $\sigma(s, t|v)$ es el número total de caminos más cortos entre s y t que pasan a través de v , y $\sigma(s, t)$ es el número total de caminos más cortos entre s y t (independientemente de si pasan o no por v). Esta medida de centralidad puede rescalarse al intervalo $[0, 1]$ mediante un factor de $(N_v - 1)(N_v - 2) / 2$, siendo N_v el número de vértices del grafo $G(V, E)$

Las medidas de centralidad de proximidad intentan capturar la noción de que un vértice es 'Central' si está 'cerca' de muchos otros vértices.(Freeman, 1979),(Brandes, 2001). El enfoque estándar, introducido por (Sabidussi, 1966), es dejar que la centralidad varíe inversamente con una medida de la distancia total de un vértice de todos los demás (**Closeness centrality**)

$$c_{CL}(v) = \frac{1}{\sum_{u \in V} dist(v, u)} \quad (2)$$

Donde $dist(v, u)$ es la distancia geodésica entre los vértices $u, v \in V$. También para comparar entre otras medidas de centralidad, esta medida se puede rescalar al intervalo $[0, 1]$, a través de la multiplicación por un factor $N_v - 1$.

Finalmente, otras medidas de centralidad se basan en nociones de 'prestigio' O 'rango'. Es decir, buscan capturar la idea de que cuanto más centrales sean los vecinos de un vértice, más central es el vértice en sí mismo. Estas medidas pueden expresarse en términos de vectores propios de Soluciones de sistemas lineales de ecuaciones y hay muchas Medidas de centralidad de vectores propios.

De acuerdo a (Bonacich, 1987),(Bonacich and Lloyd, 2001)

$$C_{E_i}(v) = \alpha \sum_{\{u,v\} \in E} C_{E_i}(u) \quad (3)$$

El vector $C_{E_i} = (C_{E_i}(1), \dots, C_{E_i}(N_v))T$ es la solución al autovalor para $AC_{E_i} = \alpha^{-1}C_{E_i}$, donde A es la matriz de adyacencia para el grafo $G(V, E)$. Bonacich sostiene que una elección óptima de α^{-1} es el autovalor más grande de A , y por lo tanto C_{E_i} es el autovector correspondiente. Cuando G es no dirigido el valor propio más alto de A será simple y su autovector tendrá valores distintos de cero y del mismo signo.

2.3. Descripción de los enlaces

Se puede extender la idea de intermediación para los enlaces, aspecto que se denomina (Edge betweenness centrality) y que es una extensión de la intermediación de nodos asignando a cada enlace un valor que refleja el número de caminos más cortos *shortest paths*, que atraviesan ese enlace. Para otras medidas de centralidad que caractericen los enlaces pueden consultarse en (Brandes and Erlebach, 2005).

2.4. Cohesión de la red

Existen varias maneras de evaluar la cohesión de una red, dependiendo del problema, donde puede usarse tr

ias o componentes gigantes así como también lo que se denomina cliques, que no son más que subconjuntos de nodos totalmente cohesivos, en el sentido de que todos los vértices dentro del subconjunto están conectados por enlaces. Se pueden definir cliques de tamaño 1 (que en este caso son los nodos v) mientras que cliques de tamaño 2 representan los enlaces (e); los cliques o subgrafos de tamaño 3 son lo que también (Kolaczyk and Csárdi, 2014) denomina *triangles*, de manera que al ir aumentando el tamaño de los cliques puede ver cual es la estructura del grafo bajo análisis. Por el proceso de construcción antes descrito al aumentar el tamaño del clique, los últimos contienen los niveles más bajo, por lo cual (Kolaczyk and Csárdi, 2014) definen un concepto de *clique máximo*, que surge de

considerar un subgrafo que no es subconjunto de un clique mayor y que da una métrica extra que los autores denominan *clique number* que corresponde al tamaño del *clique máximo*.

Además de los cliques y sus variantes, hay otras clases de subgrafos que pueden resultar interesantes para definir la cohesión de la red. Hay 2 cantidades fundamentales, que surgen en los primeros trabajos en análisis de redes sociales, y que se conocen como díadas y tríadas. Las díadas son pares de nodos que en grafos dirigidos, pueden tomar tres estados posibles:

1. nulo (no Bordes dirigidos)
2. asimétrico (un borde dirigido)
3. mutuo (dos bordes dirigidos).

De manera similar, las tríadas son trietas de nodos que pueden tomar 2^4 estados posibles, partiendo del subgrafo nulo hasta el subgrafo en el cual las tres díadas formadas por los vértices en la tríada tienen bordes dirigidos mutuamente.

También la noción de clusters o comunidades sirve para la caracterización de la red

2.5. Densidad y otras medidas de frecuencia

Las caracterizaciones de la cohesión de red que surgen de los subgrafos y cliques consisten en ver con qué frecuencia y donde se encuentran a través de métricas como las que se presentan a continuación.

La densidad de un gráfico es la frecuencia de los enlaces realizados en relación con el potencial número efectivo de bordes. Por ejemplo, en un grafo (no dirigido) G sin autobucles y sin múltiples enlaces, la densidad de un subgrafo $H = (VH, EH)$ es

$$den(H) = \frac{|E(H)|}{|V(H)|(|V(H)| - 1)/2} \quad (4)$$

donde el valor de $den(H)$ estará entre $[0, 1]$ y proporciona una medida de que tan cerca está el grafo H de ser un clique; En el caso de que G sea un grafo dirigido, el denominador en la ecuación (4) se sustituye por $|V(H)|(|V(H)| - 1)$.

Otro concepto que se maneja para evaluar la densidad de la red es lo que (Kolaczyk and Csárdi, 2014) denomina *clustering coefficient*, donde surge un coeficiente de agrupamiento que puede medirse como

$$clT(G) = \frac{3\tau_{\Delta}(G)}{\tau_3(G)} \quad (5)$$

Donde $\tau_{\Delta}(G)$ es el número de triángulos en el grafo G , y $\tau_3(G)$, el número de tripletas conectadas un subgrafo de tres vértices conectados por dos bordes, también A veces llamado un 2 estrellas). El valor $cl|T(G)$ se denomina alternativamente *transitividad* y es una métrica estándar en el estudio de las redes sociales, también conocido como 'fracción de tripletas transitivas. Por otra parte $cl|T(G)$ es una medida de la agrupación global o de *clustering*, resumiendo la frecuencia relativa con la que las tripletas conectadas forman triángulos cerrados.

2.6. Conectividad

'Una noción de conectividad es la que tiene que ver con el hecho de que si dado un subconjunto de k vértices (o enlaces) se quitan del grafo, el subgrafo restante aún permanece conectado. En particular un grafo $G(V, E)$ se llama *k-vértice-conectado* si el número de vértices $N_v > k$, y al eliminar cualquier subconjunto de vértices $X \in V$ de cardinalidad $|X| < k$ deja un subgrafo conectado. A su vez si $G(V, E)$ se denomina *k-borde-conectado* si $N_v \leq 2$, y al eliminar cualquier subconjunto de aristas $Y \in E$ de cardinalidad $|Y| < K$ deja un subgrafo que está conectado'.

De esa manera se define como *conectividad* de vértice (enlace) de $G(V, E)$ al entero más grande tal que G es *k-vértice-* (*k-borde-*) conectado. (Kolaczyk and Csárdi, 2014) manifiestan que se puede demostrar que la conectividad del vértice está acotada por la conectividad de enlace, la que a su vez está acotada por el grado mínimo d entre los vértices en G .

2.7. Clustering de la red y Particionado Espectral

Cuando se habla de partición de la red $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$, de un conjunto \mathcal{S} se refiere a la división de la misma en clase naturales tales que estas son disjuntas entre sí y a su vez la unión de ellas reproducen el conjunto de partida ($\bigcup_{k=1}^K C_k = \mathcal{S}$). Pero a su vez es importante también evaluar si un subconjunto de nodos (algunas de esas clases) es 'cohesivo' si para lo cual se entiende que es así si los nodos están bien conectados entre sí, y al mismo tiempo están relativamente bien separados de los nodos restantes. Así los algoritmos de particionado buscan una partición $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$ de un grafo $G = (V, E)$, de manera que los conjuntos $E(C_k, C_{k'})$ de enlaces conectando nodos de C_k en $C_{k'}$ sea relativamente pequeña en comparación al conjunto $E(C_k) = E(C_k, C_{k'})$ de enlaces que conectan nodos al interior de C_k .

Una primer forma de evaluar el particionado de la red es a través de clustering jerárquico, de tipo aglomerativo, donde se incorpora una función de costo, que refleja la cohesión, con lo cual surge el concepto de *modularidad* de \mathcal{C} , donde se define $f_{kk}(\mathcal{C})$ como la fracción de enlaces de la red original que conectan nodos de C_i con nodos de C_j

$$mod(C) = \sum_{k=1}^K [f_{kk}(C) - f_{kk}^*]^2, \quad (6)$$

donde f_{kk}^* es el valor esperado de f bajo el supuesto de un modelo aleatorio de asignación de enlaces. Valores grandes de la *modularidad* sugieren que C captura una estructura *no trivial de grupos* (es decir que existen grupos), a la inversa si los enlaces se asignasen al azar.

Otro enfoque común para el particionado de grafos es a través la descomposición en autovalores del *Laplaciano* de un grafo.

El *Laplaciano* de un grafo $G(V, E)$, se construye a partir de la matriz de adyacencia A , quedando una matriz $L = D - A$, donde $D = diag[(d_v)]$ es una matriz diagonal con elementos $D_{vv} = d_v$ que son la secuencia de grados de G .

Se puede afirmar que G va a tener K componentes conectados su y solamente si $\lambda_1(L) = \lambda_2(L) = \dots = \lambda_k(L) = 0$ y $\lambda_{k+1} > 0$, donde $\lambda_1 \leq \lambda_2 \leq \lambda_{N_v}$ son los autovalores de L ordenados de menor a mayor.

Por lo tanto, el número de componentes en un grafo está directamente relacionado con el número de autovalores del Laplaciano distintos de cero.

El primero en asociar λ_2 con la conectividad de un grafo fue (Fiedler, 1973), sugirió dividir los nodos separándolos según el signo de sus coordenadas en el autovalor, lo que produce como resultado dos subconjuntos de nodos (creando lo que llama un *cut*)

$$S = \{v \in V : \mathbf{x}_2(v) \leq 0\} \text{ y } \hat{S} = \{v \in V : \mathbf{x}_2(v) < 0\}. \quad (7)$$

Por lo tanto, el vector \mathbf{x}_2 se denomina el vector de Fiedler, y el vector correspondiente al valor propio λ_2 , el valor de Fiedler

2.8. Enlace selectivo(Asortatividad)

Otra aspecto importante para evaluar la topología de una red es la evaluación de lo que se denomina *enlace selectivo entre nodos* de acuerdo a algunas características y que se miden con lo que se conoce como coeficiente de asortatividad (Assortativity coefficients) y que tiene una lógica muy similar a la de los coeficientes de correlación. Este concepto a veces también se conoce como *homofilia*, y expresa la tendencia de las personas a relacionarse con personas que se le parecen.

Cuando la característica que se estudia es de tipo categórico (nomina u ordinal) la medida de es

$$r_a = \frac{\sum_i f_{ii} - \sum_i f_{i+}f_{+i}}{1 - \sum_i f_{i+}f_{+i}} \quad (8)$$

Donde f_{ij} es la fracción de enlaces en $G(V, E)$ que unen un nodo en la i -ésima categoría con un nodo en la j -ésima categoría y f_{i+}, f_{+i} expresan la suma de la i -ésima fila y columna respectivamente, de la matriz resultante \mathbf{f} de frecuencias. (Newman, 2002), (Newman, 2003)

El coeficiente descrito en la ecuación (8) está acotado en el intervalo $[-1, 1]$, expresando que si es cercano a 0, la mezcla de nodos en el grafo no difiere de la que se obtendría al asignar los enlaces al azar, preservando la distribución de grados marginal; cuando el coeficiente se acerca a 1 o -1 existe una mezcla selectiva perfecta.

Cuando los nodos tienen una característica de interés que es continua, para evaluar la *homofilia*, se consideran como (x_e, y_e) los valores que toman los nodos enlazados por el enlace e , para lo cual se usa el coeficiente de correlación de Pearson de los pares (x_e, y_e)

$$r = \frac{\sum_{x,y} xy - (f_{xy} - f_{x+}f_{+y})}{\sigma_x \sigma_y} \quad (9)$$

3. Descripción del problema en estudio

Se trabaja solamente con 2225 cruceristas que son los que hicieron algún gasto.

Dado que las unidades de análisis son grupos de cruceristas (individuales o múltiples) se propone armar redes para cada crucero. Como los cruceros no vienen una sola vez en la temporada es necesario seleccionar los cruceros de la tabla de datos diferenciando para cual de las ocasiones que vino en la temporada corresponde la información.

Hay cruceros de diferentes tamaños en términos de pasajeros, por lo cual se elegirán cruceros de diferentes capacidades y sobre esos se construirán las redes usando como criterios de enlace la cantidad de gastos que comparten.

4. Resultados

La variable corte sirve para saber que cantidad de los 2225 corresponde a un crucero en algunas de las veces que vino en la temporada. Para eso se crea una variable compuesta entre código de crucero y corte para poder tener en la tabla de datos los cruceros cada vez que vienen y es sobre esta variable que se van a seleccionar de 3 a 5 unidades (son cruceros

en cada oportunidad) y sobre eso se armarán las redes de cruceristas. De esta manera se tiene una variable que dimensiona a los diferentes cruceros en cada llegada y teniendo en cuenta que solo se consideran los cruceristas que tuvieron algún gasto. Se trabaja con el software Gephi (Bastian et al., 2009) y el R (R Core Team, 2016), para el que se usan varias bibliotecas (Butts, 2016),(Csardi and Nepusz, 2006),(Kolaczyk and Csárdi, 2017)

```
Base$Crucero.vez<-Base$crucero1*100+Base$corcte
```

```
tabla4<-table(Base$Crucero.vez)
> tabla4
```

```
914  916  918  937  959  962  963 1408 1421 1432 1447 1448 1464 1469
27   28   30   34   31   32   21   23   19   24   23   19   26   29
1477 1479 1644 1940 2119 2136 2170 2409 2417 2425 2431 2455 2523 2533
32   24   38   28   30   34   36   24   24   22   16   31   30   36
2557 2558 2828 2873 2874 2878 3301 3302 3303 3304 3306 3307 3313 3354
21   25   31   29   35   34   27   25   29   30   24   17   21   20
3360 3368 3371 3372 3426 3434 3456 3510 3542 3549 3580 3581 3582 3943
23   26   31   33   27   37   26   25   26   34   26   25   25   33
4453 4912 4915 4938 4946 4966 4967 5150 5205 5220 5222 5241 5261 5627
23   24   25   21   36   32   33   32   24   22   31   29   26   28
5629 5630 5835 6011 6024 6039 6045 6051 6052 6065 6075 6076
22   22   26   24   22   23   25   24   31   30   23   31
```

Pueden existir varias formas de seleccionar cruceros para analizarlos bajo la lógica del SNA y que a su vez sean diferentes entre sí para poder compararlos. Se opta por clasificar a los cruceros de acuerdo a su capacidad de pasajeros, lo cual estaría dando lugar a redes de cruceristas de cruceros pequeños, medianos y grandes, lo que puede ser una característica importante para la estructura interna de cada red.

```
#para ver el tamaño de los cruceros
```

```
tabla5<-aggregate(Base$capacidad,list(Base$Crucero.vez),mean)
tabla5;tabla6<-table(tabla5$x)
```

```
a<-order(tabla5$x)
```

```
tabla7<-tabla5[a,]
```

```
tabla7
```

```
library(car)
```

```
tabla7$estrato<-recode(tabla7$x,'lo:1000=1;1000:2000=2;2000:3000=3;3000:hi=4')
```

```
tabla8<-table(tabla7$estrat)
```

```
tabla8
```

Capacidad	Frecuencia
769	1
777	4
800	1
837	1
1200	1
1339	3
1350	8
1700	4
1874	1
1896	5
2074	6
2099	1
2100	16
2101	1
2394	9
2449	7
3100	1
3200	3
3470	9

Tabla 1: Descripción de los tamaño de los cruceros a ser analizados

Observando los resúmenes de la tabla 1 se ve que hay 4 tamaños de barcos; menos de 1000, entre 1000 y 2000, 2000 a 3000 y 3000 o mas. De esta manera se propone elegir 4 cruceros , 1 de cada tamaño

estrato	Nh	crucero seleccionado
menos de 1000 (1)	7	4453
de 1000 a 2000 (2)	22	1644
de 2000 a 3000(3)	40	4912
3000 o mas (4)	13	5150

Tabla 2: Descripción de los estratos de cruceros

4.1. Selección de los cruceros

Las nuevas unidades 'cruceros veces' se estratifican en 4 de acuerdo a la capacidad de cada barco y de esos 4 estratos se selecciona mediante Muestro aleatorio simple (MAS) cruceros con los que se construyen redes, una por crucero.

```
set.seed(1234)
tabla7$sorteo<-c(sort(runif(7)),sort(runif(22)),sort(runif(40)),sort(runif(13)))
```

El número de cruceristas que se consideran en cada crucero se presenta en la Tabla

Crucero	n_j	Red	Etiqueta
$j = 1$	23	$g1$	4453
$j = 2$	38	$g2$	1644
$j = 3$	34	$g3$	4912
$j = 4$	32	$g4$	5150

Tabla 3: Descripción de los cruceros analizados

En la temporada 2010-2011 se muestrearon un total de 82 cruceros. De la Tabla 3 se ve que aparece el crucero con el que se construirá la red $g1$ y la etiqueta 4453, significa que es el crucero 44 que fue muestreado a mitad de temporada, mientras que el crucero para la red $g3$ es el 49, en el que se aplicó la encuesta casi a comienzos de la misma.

Para el resto del análisis de los resultados, para cada crucero seleccionado se crea una red que se denominará $g1, g2, g3, g4$ y los resultados para las correspondientes métricas de cada red tendrán un sufijo 1, 2, 3, 4 que se identifica con cada crucero seleccionado

4.2. Formas de visualización de las redes

Para visualizar las redes y poner de manifiesto con claridad la estructura que existe en las mismas es importante manejar un concepto que en la literatura del SNA se denomina *layout*. Por ejemplo, si se considera una red de tamaño moderado con 50 nodos, podría usarse una grilla cuadrada de 10 por 10, donde ir posicionando cada nodo. El primer nodo en la red podría ir en cualquiera de las 100 posiciones, el segundo nodo en 99 posiciones, y así sucesivamente, lo que da una cantidad muy grande $3,1 \times 10^{93}$ diferentes diseños de red posibles. Sin embargo muchas de estas posibles configuraciones pueden dar lugar a representaciones gráficas muy confusas, donde van a haber muchos enlaces que se cruzan, con lo cual la idea es elegir un diseño que tenga una probabilidad más alta de ser visualmente aceptable, por lo cual hay necesidad de recurrir a diferentes algoritmos que busquen maximizar ese efecto. (Kolaczyk and Csárdi, 2017),(Csardi and Nepusz, 2006)

- Minimizar los cruces de enlaces.
- Maximizar la simetría de la disposición de los nodos.
- Minimizar la variabilidad de las longitudes de los enlaces.
- Maximizar el ángulo entre los enlaces cuando cruzan o unen nodos.
- Minimizar el espacio total utilizado para la visualización de la red.

En la Figura 1 puede verse cual es el resultado de disponer un *layout* de tipo aleatorio, donde se ven varios enlaces que se cortan (esto debido a que la red para el caso del crucero 1) es chica; además el uso de esta forma de visualizar hace que la disposición sea distinta cada vez que se grafica, al ser aleatoria la disposición espacial de cada nodo (por ende de cada enlace) en la red

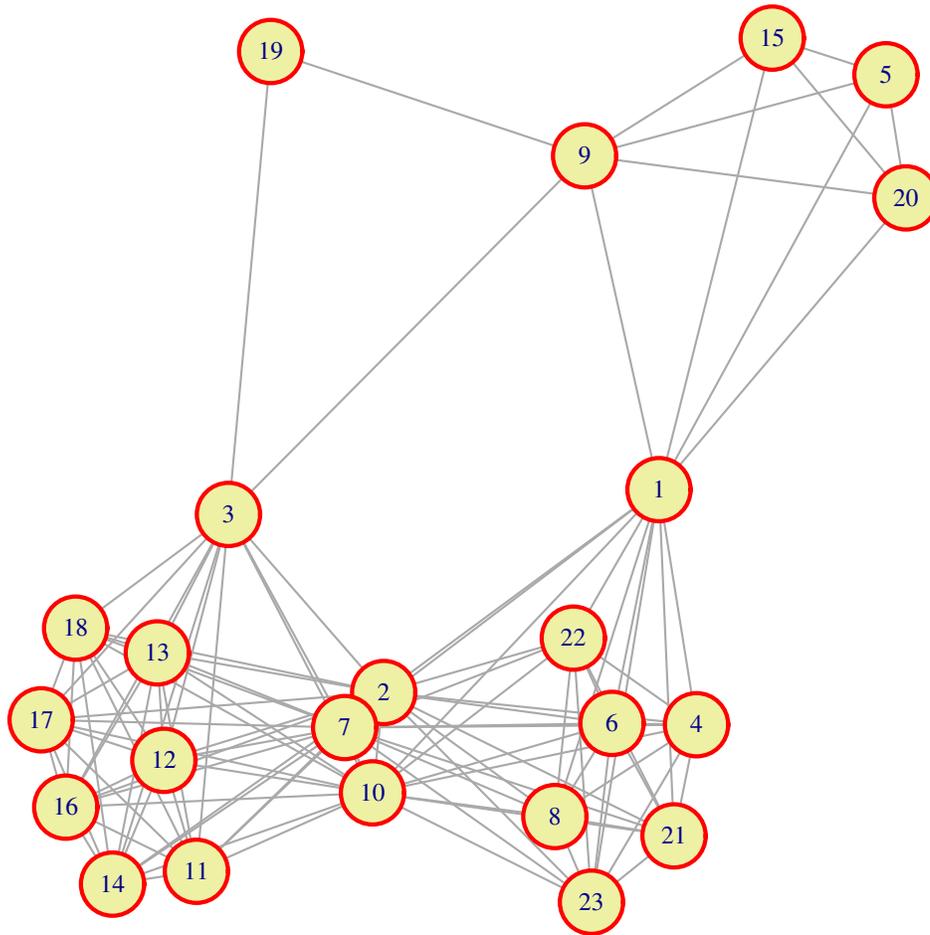


Figura 1: Red del crucero estrato 1 (layout aleatorio)

En este caso como aparece en la Figura 2 el diseño usado para visualizar el grafo, parte de ubicar el nodo 1 y disponer los restantes nodos en sentido trigonométrico negativo y donde de cada nodo parten los enlaces correspondientes. Este diseño tiene el inconveniente de generar grafos con muchas densidad de enlaces que se cruzan, sobre todo si el número de enlaces es muy grande y con lo cual no es claro el patrón de conexiones en la red.

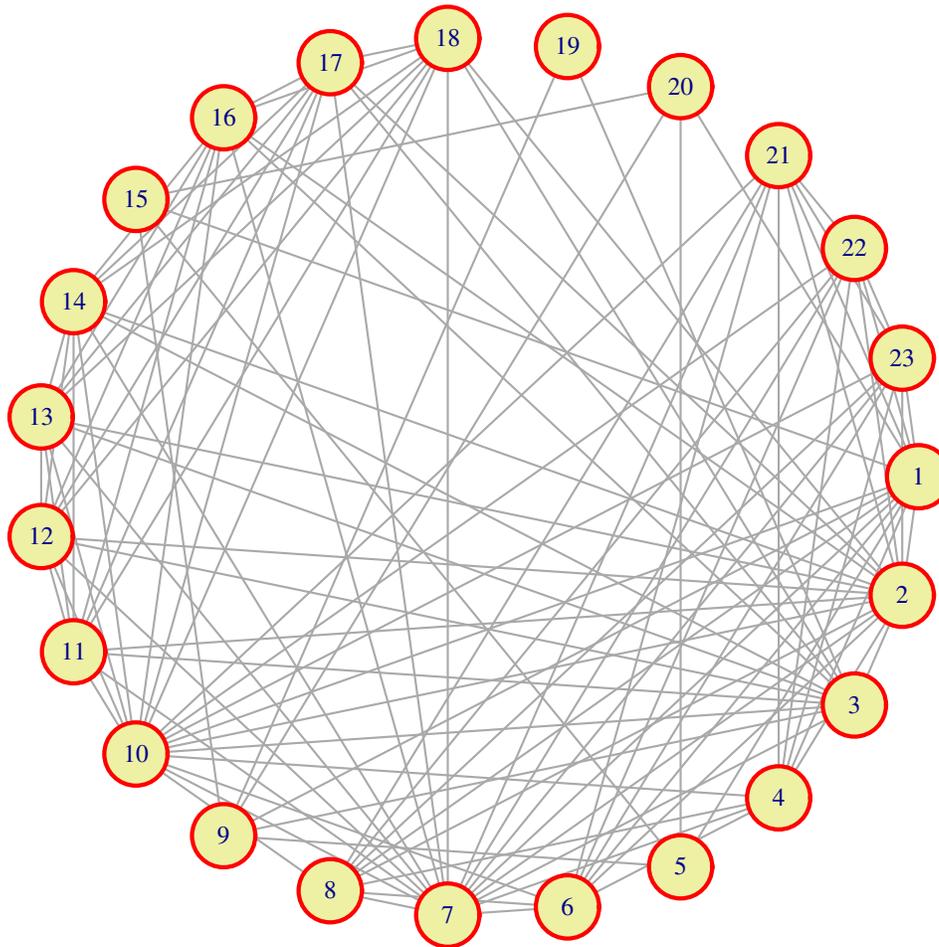


Figura 2: Red g_1 del estrato 1 (layout Circle)

Se ha desarrollado un gran número de enfoques para el diseño automático de gráficos de red. Una clase general de algoritmos, llamados de *fuerza-dirigida*, trabajan a partir de un enfoque en el que los nodos conectados tengan una fuerza de atracción similar a un resorte, al mismo tiempo que asignan fuerzas repulsivas a todos los pares de nodos. Los resortes en este algoritmo actúan para tirar de los nodos conectados más cerca uno del otro, mientras que las fuerzas repulsivas empujan los nodos no conectados lejos uno de otros. El sistema de red resultante se moverá y oscilará durante un tiempo antes de llegar al estado estacionario que tiende a minimizar la energía en la red. Este algoritmo, tiende a producir representaciones de las redes estéticamente agradables (Fruchterman and Reingold, 1991)

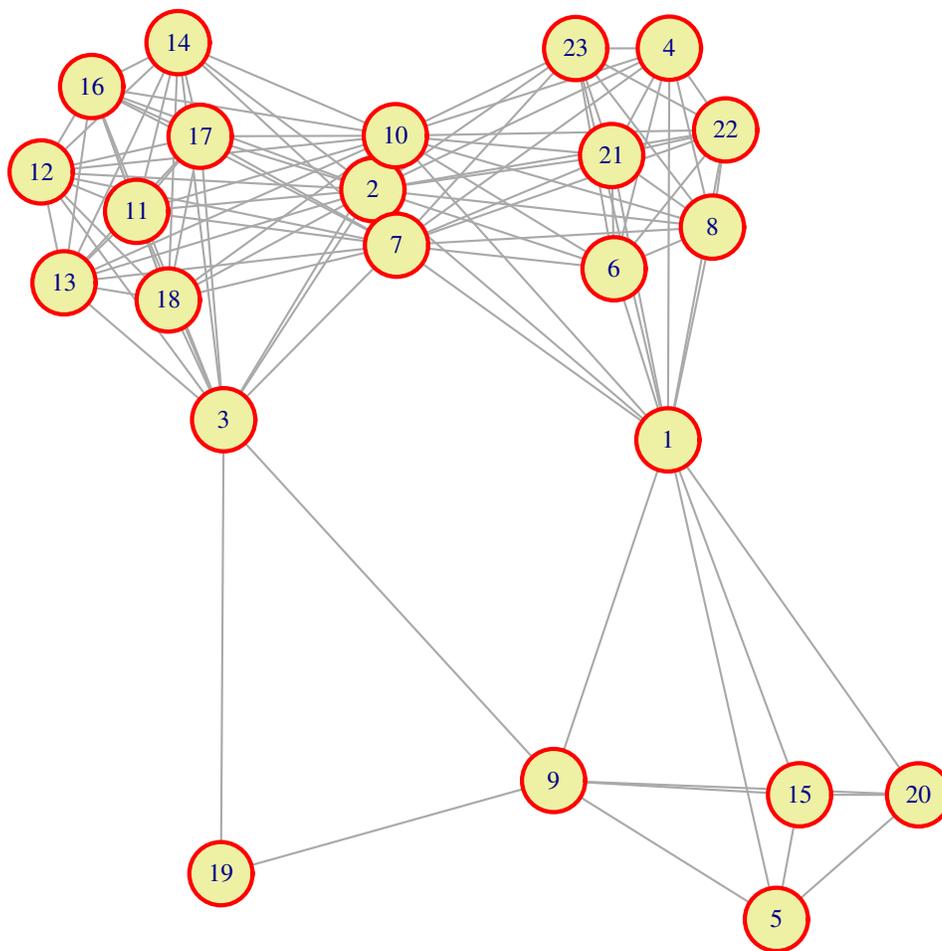


Figura 3: Red g_1 del estrato 1 (layout Fruchterman-Reingold)

4.3. Red de crucero 1 $g(1)$

```

> E(g1)
+ 110/110 edges:
[1] 1-- 2 1-- 4 1-- 5 1-- 6 1-- 7 1-- 8 1-- 9 1--10
[9] 1--15 1--20 1--21 1--22 1--23 2-- 3 2-- 4 2-- 6
[17] 2-- 7 2-- 8 2--10 2--11 2--12 2--13 2--14 2--16
[25] 2--17 2--18 2--21 2--22 2--23 3-- 7 3-- 9 3--10
[33] 3--11 3--12 3--13 3--14 3--16 3--17 3--18 3--19
[41] 4-- 6 4-- 7 4-- 8 4--10 4--21 4--22 4--23 5-- 9
[49] 5--15 5--20 6-- 7 6-- 8 6--10 6--21 6--22 6--23
[57] 7-- 8 7--10 7--11 7--12 7--13 7--14 7--16 7--17
[65] 7--18 7--21 7--22 7--23 8--10 8--21 8--22 8--23
[73] 9--15 9--19 9--20 10--11 10--12 10--13 10--14 10--16
+ ... omitted several edges
> V(g1)
+ 23/23 vertices:
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
[21] 21 22 23
    
```

grados	Frecuencia
4	1
6	3
8	1
11	6
12	7
14	1
15	1
19	3
$\bar{g}_1 \simeq 12$	Total=23

Tabla 4: Distribución de los grados de g_1

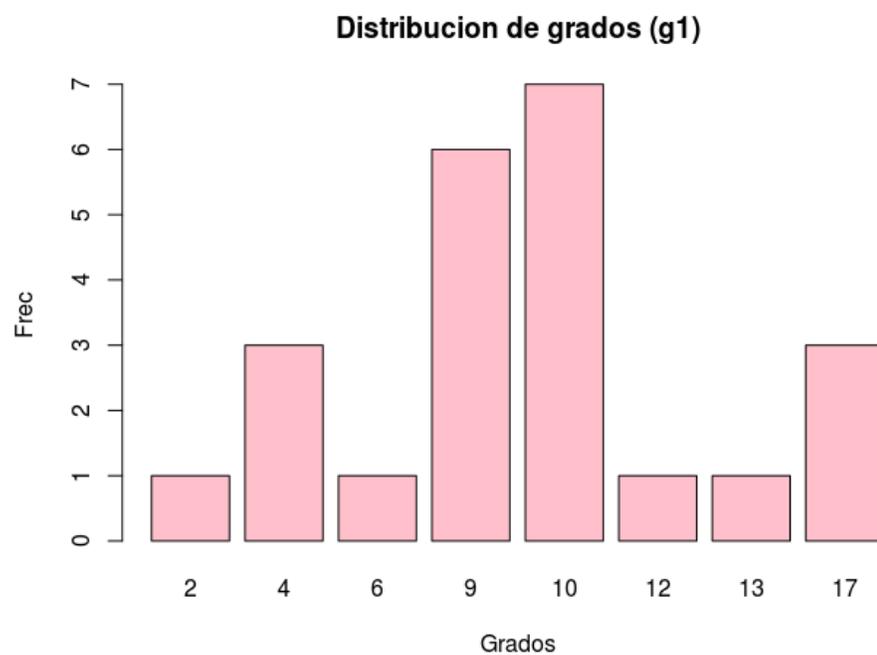


Figura 4: Distribución de los grados de g_1

Los algoritmos usado en en este caso para detectar Comunidades estructurales tratan de encontrar subgrafos densos, usando heurísticas, usando paseos aleatorios (*random walk*) y donde la idea es que, los paseos aleatorios cortos tienden a estar en la misma comunidad. (Pons and Latapy, 2015)

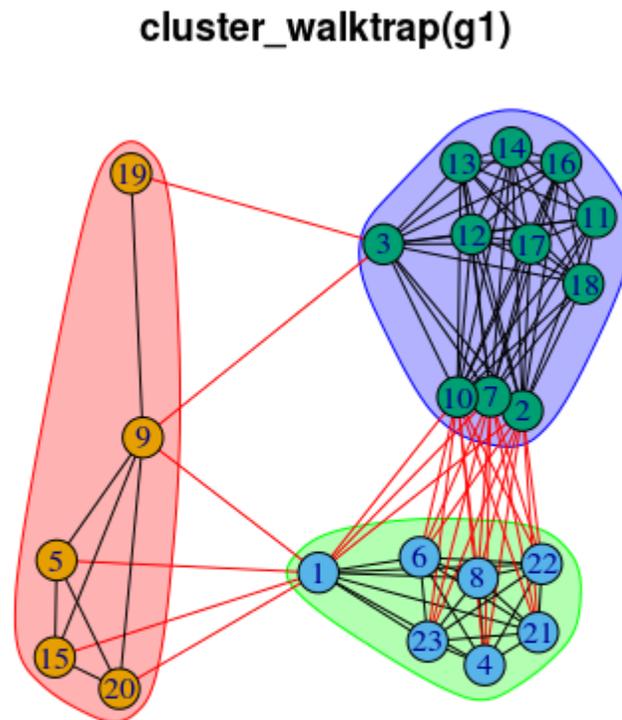


Figura 5: Clustering para Red g_1 del estrato 1 con algoritmo walktrap

Para poder determinar los clusters en la red se usa una función que tratar de evidenciar grafos densos, que también se llaman *comunidades*, optimizando el escore de modularidad, a través del algoritmo *fast greedy* que aparece en Finding community structure in very large networks, <http://www.arxiv.org/abs/cond-mat/0408187>.(Clauset et al., 2004)

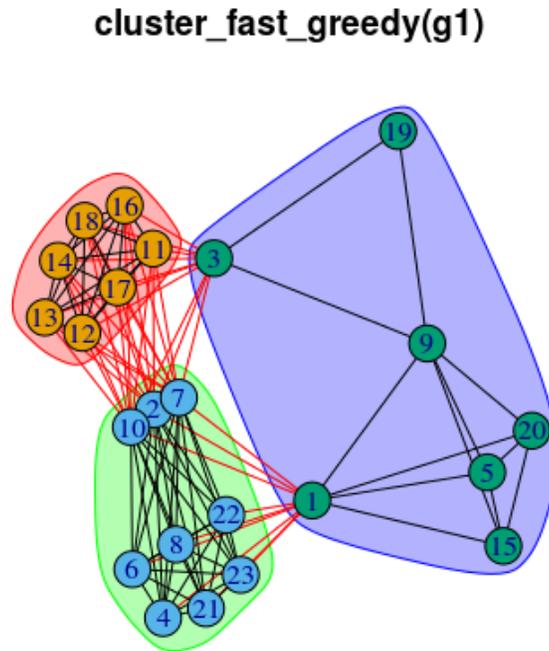


Figura 6: Clustering para Red g_1 del estrato 1 con algoritmo fast_greedy

En este caso el criterio para determinar los clusters o comunidades basado en *edge betweenness*, es que probablemente los enlaces que conectan módulos separados tiene alto *edge betweenness* ya que los caminos más cortos de una comunidad a otra deben atravesarlos. Por lo tanto si gradualmente se van quitando enlaces con el más alto *edge betweenness*, se obtendría un mapa jerárquico, que en este caso es un árbol raíz o dendrograma, donde los tallos representan los nodos individuales y la raíz el grafo enteramente.

El algoritmo *cluster-edge-betweenness* en forma iterativa calcula la intermediación de enlaces del grafo, quitado el enlace con mas alta intermediación y volviendo a recalcular, hasta .(Newman and Girvan, 2004)

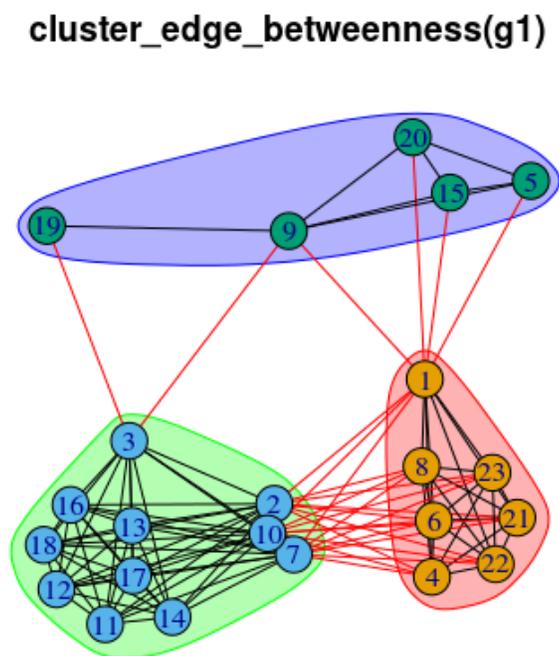


Figura 7: Clustering para Red g_1 del estrato 1 con algoritmo *edge_betweenness*

El algoritmo *label_prop* es el más rápido, con performance de tiempo lineal, que se usa para detectar comunidades en redes. Comienza etiquetando los nodos con etiquetas únicas y luego adaptando las misma por voto mayoritario en la vecindad (*neighborhood*) del nodo.(Raghavan et al., 2007)

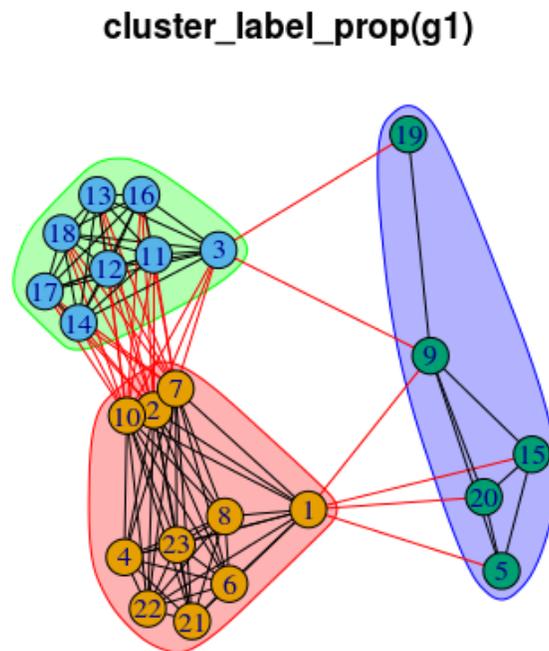


Figura 8: Clustering para Red g_1 del estrato 1 con algoritmo *label_prop*

algoritmo	modularidad (\bar{m})	Nro de clusters
cluster_walktrap (cw.1)	0.2880	3
cluster_fast_greedy (cfg.1)	0.2654	3
cluster_edge_betweenness (eb.1)	0.2880	3
cluster_label_prop (clp.1)	0.2923	3

Tabla 5: Modularidad para g_1 con diferentes algoritmos

4.3.1. Caracterización de la red

En la tabla 6 se ven las características de los 23 nodos, donde se consignan las métricas de centralidad (cercanía e intermediación), grados de cada nodo, así como el número de triángulos que conforman cada uno de ellos.

En la Figura 9 se puede ver como es la relación entre el número de grados de cada nodo y la betweenness, donde resalta el nodo 1 que tiene 13 grados siendo el nodo más intermediario; en el gráfico los puntos que aparecen con forma de estrella en rojo significa que en esa coordenada (valor de número de grado y betweenness) cae mas de un nodo, por ejemplo para el valor 10 grados hay tiene nodos.

Para el caso de la cercanía se ve que la relación entre esta y el número de grados es creciente, siendo los nodos 2 y 7 los mas centrales y con mayor grado.

Nodos	grados	betweenness intermediación	closeness cercanía	triángulos
1	13	0.23	0.71	42
2	17	0.12	0.76	80
3	12	0.12	0.69	46
4	9	0.00	0.61	36
5	4	0.00	0.47	6
6	9	0.00	0.61	36
7	17	0.12	0.76	80
8	9	0.00	0.59	36
9	6	0.06	0.58	7
10	17	0.10	0.73	80
11	10	0.00	0.59	45
12	10	0.00	0.59	45
13	10	0.00	0.59	45
14	10	0.00	0.59	45
15	4	0.00	0.47	6
16	10	0.00	0.59	45
17	10	0.00	0.59	45
18	10	0.00	0.59	45
19	2	0.00	0.46	1
20	4	0.00	0.47	6
21	9	0.00	0.61	36
22	9	0.00	0.61	36
23	9	0.00	0.61	36

Tabla 6: Caracterización de los nodos de g_1

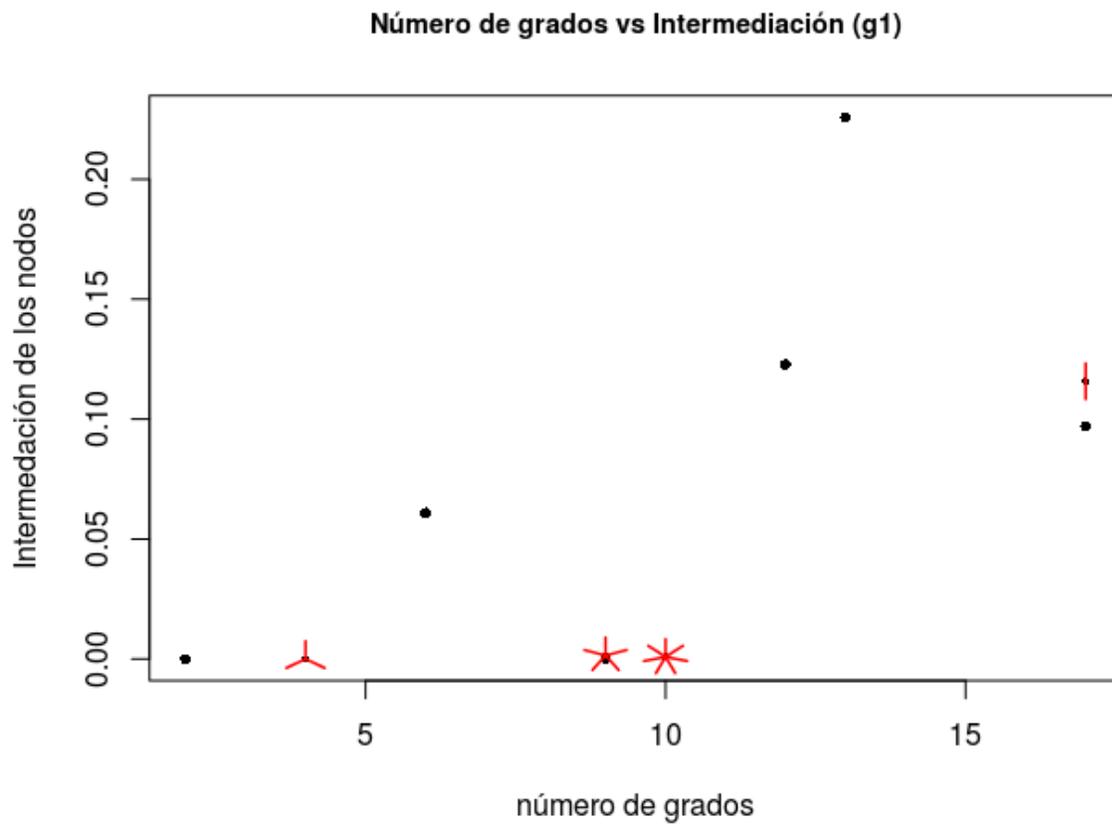


Figura 9: Relación entre grados y betweenness para Red g_1

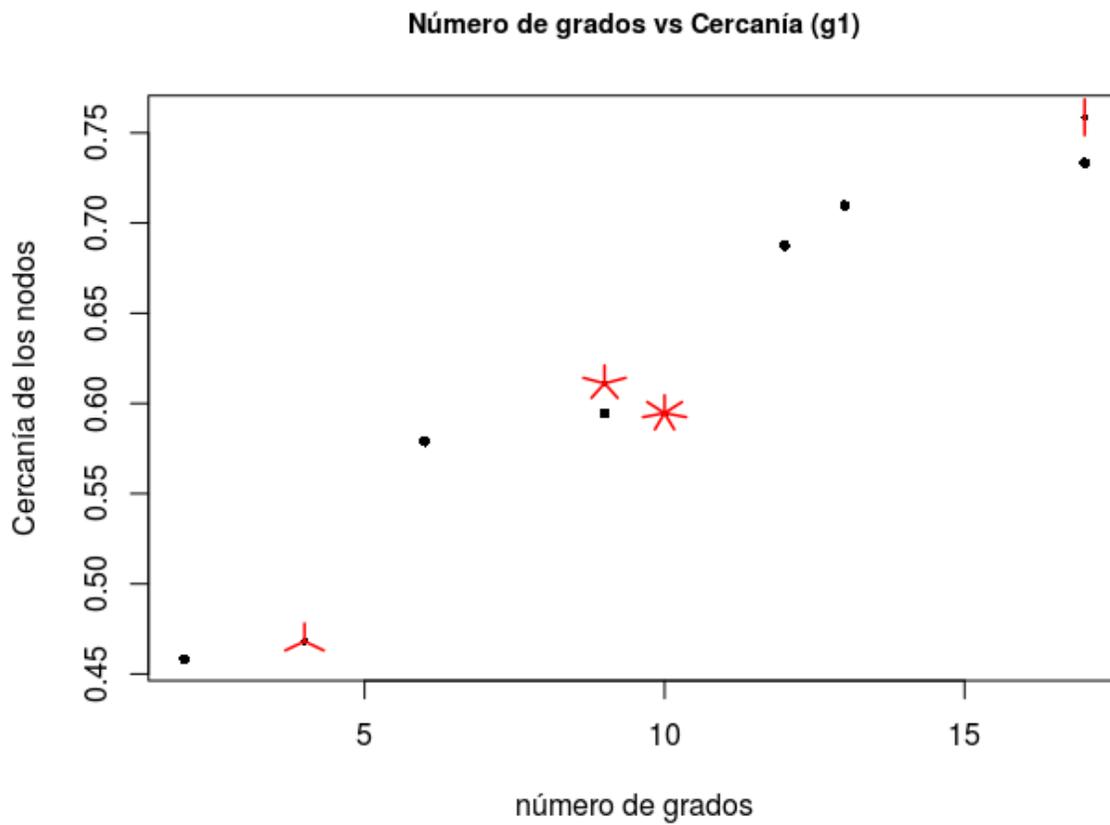


Figura 10: Relación entre grados y betweenness para Red g_1

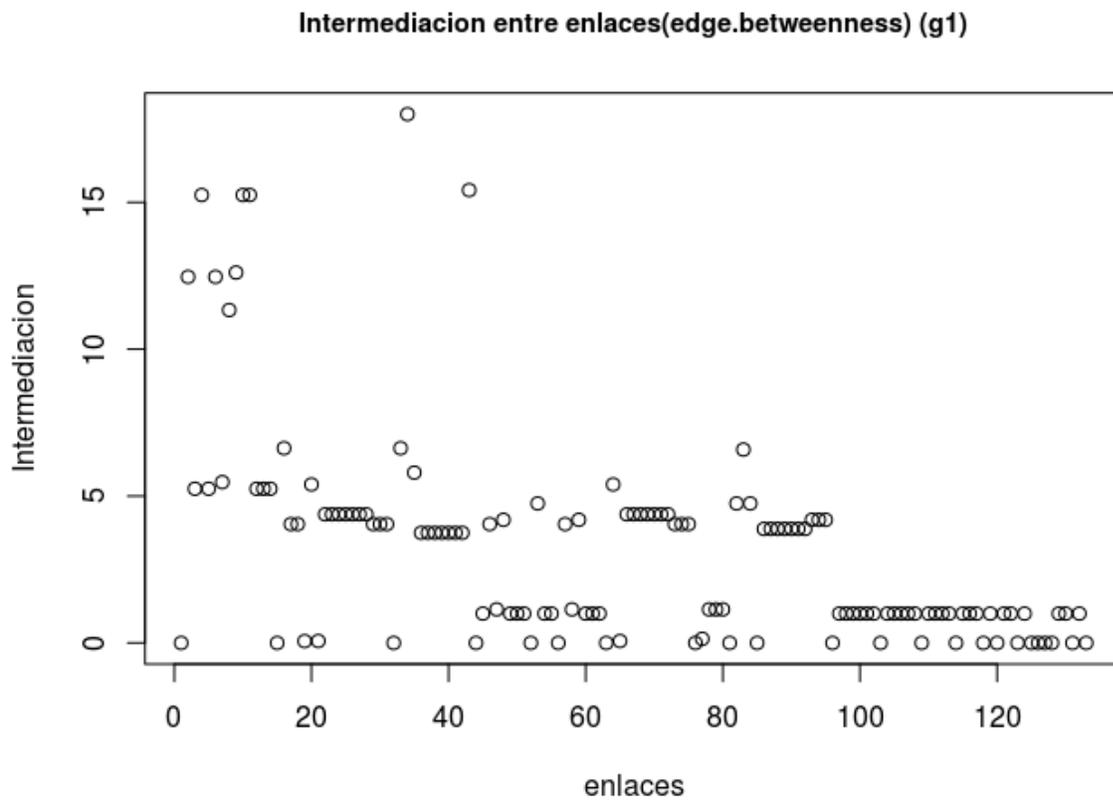


Figura 11: betweenness para los enlaces para Red g_1

```
>eb.g1[30:45]
[1] 6.629167 18.000000 5.800000 3.750000 3.750000 3.750000
[7] 3.750000 3.750000 3.750000 3.750000 15.416667 1.000000
[13] 4.045833 1.142857 4.192857 1.000000
E(g1)[30:45]
+ 16/110 edges:
[1] 3-- 7 3-- 9 3--10 3--11 3--12 3--13 3--14 3--16 3--17 3--18
[11] 3--19 4-- 6 4-- 7 4-- 8 4--10 4--21
> E(g1)[order(eb.g1, decreasing=T)[1:10]]
+ 10/110 edges:
[1] 3-- 9 3--19 1-- 5 1--15 1--20 1--10 1-- 2 1-- 7 1-- 9 2-- 3
```

Para encontrar sub grafos o cliques

```
> E(g1)[order(eb.g1, decreasing=T)[1:10]]
+ 10/110 edges:
[1] 3-- 9 3--19 1-- 5 1--15 1--20 1--10 1-- 2 1-- 7 1-- 9 2-- 3
> table(sapply(cliques(g1,max=6), length))#distribucion de los tamanos de cliques
 1  2  3  4  5  6
23 110 295 545 715 672
> table(sapply(maximal.cliques(g1), length))#cliques ,maximales
 3  5 10 11
 1  1  1  1
> clique.number(g1)#tamano de clicke
[1] 11
> largest_cliques(g1)
[[1]]
+ 11/23 vertices:
[1] 2 7 10 3 11 12 13 14 16 17 18
```

Para estudiar la densidad de los diferentes nodos

```
> graph.density(g1,loops = FALSE)
[1] 0.4347826
> graph.density(ego.c1_19,loops = FALSE)
[1] 1
> graph.density(ego.c1_15,loops = FALSE)
[1] 1
> graph.density(ego.c1_12,loops = FALSE)
[1] 1
```

```
> graph.density(ego.c1_7,loops = FALSE)
[1] 0.6339869
```

4.3.2. Particionado del grafo g_1

Para la evaluación del eventual particionado de g_1 se utilizan la descomposición de valores propios del laplaciano, así como es el comportamiento del vector de Fiedler, donde interesa identificar la presencia de un valor propio nulo.

```
k.lap <- graph.laplacian(g1)
eig.anal <- eigen(k.lap)
plot(eig.anal$values, col="blue", ylab="Valores propios",xlab="Indice",
cex.main=0.99,main="Valores propios del Laplaciano (g1)")
abline(h=0,col="grey")
```

En este caso la idea es poder evaluar una eventual bipartición de la red, en función de los valores que asumen los nodos para el vector de Fiedler y algún atributo de los nodos, en este caso el número de gastos

```
f.vec <- eig.anal$vectors[, 22]
faction <- get.vertex.attribute(g1, "nrogastos")
f.colors <- as.character(length(faction))
f.colors[faction == 1] <- "black"
f.colors[faction == 2] <- "red"
f.colors[faction == 3] <- "cyan"
plot(f.vec, pch=16, xlab="Numero de Nodo",
ylab="Entradas para el Vector de Fiedler", col=f.colors,
cex.main=0.99,main="relación entre VF y número de gastos en cada nodo")
abline(0, 0, lwd=2, col="lightgray")
```

La Figura 13 muestra que no existe un biparticionado, ya que el vector propio de Fiedler no queda separado en 2 subgrafos (el eje 0 no separa los nodos con los mismos colores) de acuerdo al número de gastos, lo que estaría indicando que hay más de 2 comunidades en términos de número de gastos desde la perspectiva de SNA.

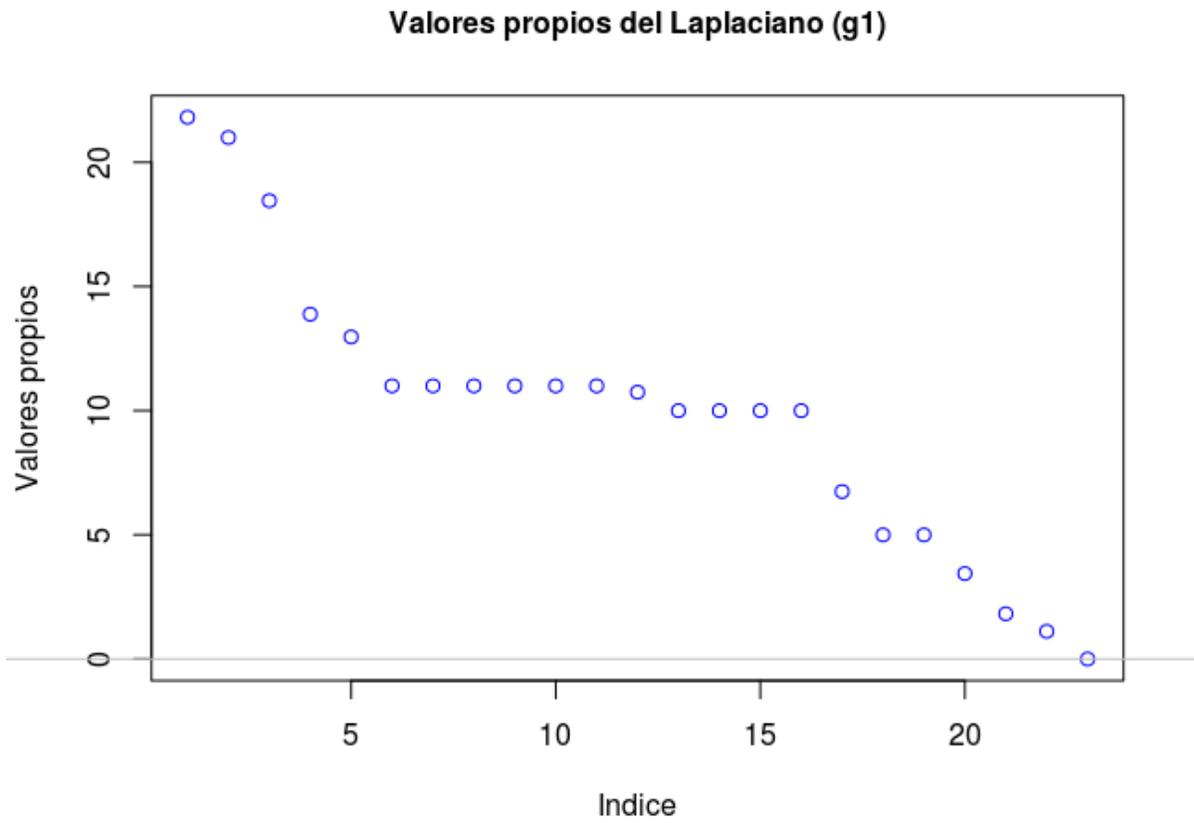


Figura 12: Valores propios del Laplaciano para g_1

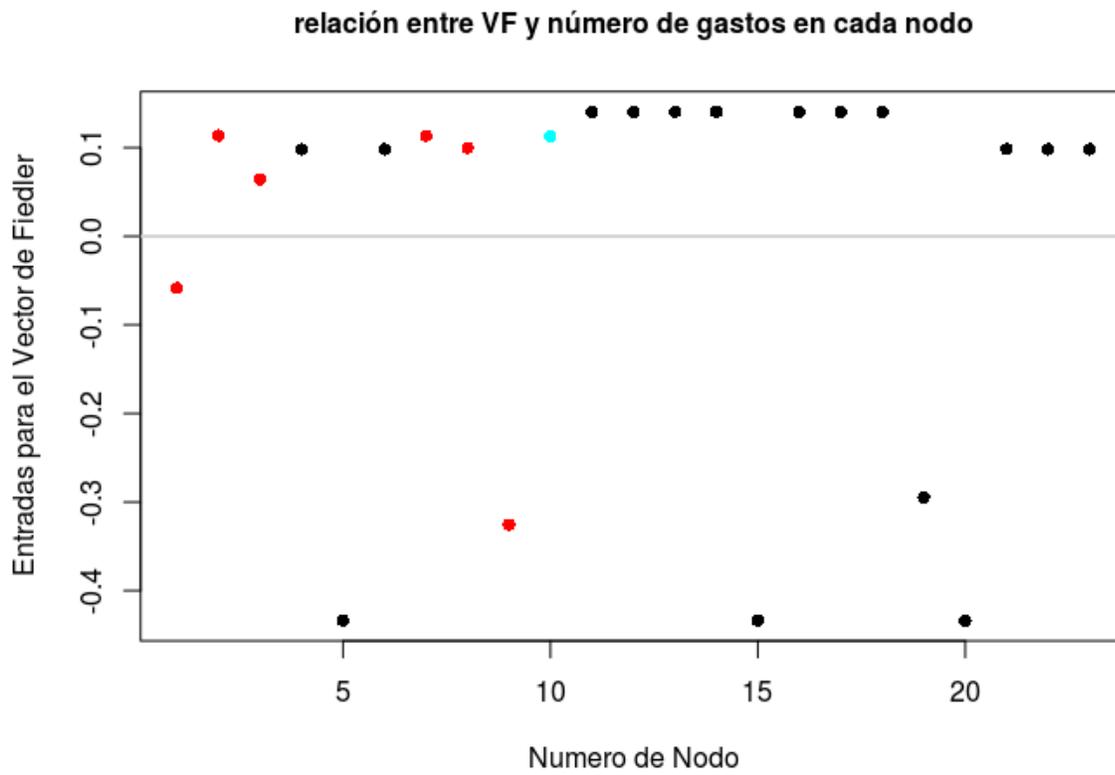


Figura 13: Vector de Fiedler para g_1

	cw.1			cfg.1			eb.1			clp.1		
	1	2	3	1	2	3	1	2	3	1	2	3
Nro gastos=1	4	5	7	7	5	4	5	7	4	5	7	4
Nro gastos=2	1	2	3	0	3	3	2	3	1	4	1	1
Nro gastos=3	0	0	1	0	1	0	0	1	0	1	0	0

Tabla 7: Cluster(Nro gastos) vs membresía para diferentes algoritmos de clustering de redes: g_1

En la tabla 7 puede verse como es la relación entre el número de gastos por cruceristas y la configuración de clusters que queda para g_1 usando los diferentes algoritmos de detección de comunidades presentados antes.

	cw.1			cfg.1			eb.1			clp.1		
	1	2	3	1	2	3	1	2	3	1	2	3
Cluster.bin=2	4	6	10	7	7	6	6	10	4	8	8	4
Cluster.bin=3	0	1	1	0	2	0	1	1	0	2	0	0
Cluster.bin=4	1	0	0	0	0	1	0	0	1	0	0	1

Tabla 8: Cluster(Cluster.Bin) vs membresía para diferentes algoritmos de clustering de redes: g_1

En la Tabla 8 puede verse como es la relación entre el los grupos creados mediante las técnicas de clustering convencional y la configuración de clusters que queda para g_1 usando los diferentes algoritmos de detección de comunidades presentados antes.

Para poder comparar como es la eventual asociación entre el clustering generado por el análisis de redes y la tipología previa que se tenía de los cruceristas mediante técnicas de clustering convencionales, en este caso mediante algoritmos jerárquico de Ward, usando distancias binarias, se presentan las siguientes tablas

	Cluster.bin				Total
	1	2	3	4	
Base Completa	134	1748	212	131	
crucero 1	0	20	2	1	23
crucero 2	2	34	1	1	38
crucero 3	2	21	1	0	24
crucero 4	0	21	7	4	32
Total	3	96	16	6	117

Tabla 9: Distribución de Cluster.Bin para totalidad de cruceros y para los 4 cruceros

El cluster que en la totalidad de cruceros tiene mas frecuencia es el 2, con 78,5% aspecto que se repite para los 4 cruceros seleccionados para la construcción de las redes, donde el grupo 2 tiene una presencia mayoritaria, en general con un 82%.

nodo	Gasto Tour	Gasto Alim	Gasto Shopp	Gasto Resto	Gasto Trans
19	1.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	1.00	0.00
12	0.00	0.00	1.00	0.00	0.00
7	0.00	1.00	1.00	0.00	0.00
Total crucero	3.00	10.00	11.00	5.00	2.00

4.3.3. Comparación de la red g_1 vs random graph y small world

La idea es probar que tan lejos está g_1 de ser una red de tipo '*random graph*' y lo mismo de tipo '*small-world*' (Kolaczyk and Csárdi, 2017). Para eso se simulan 2 tipos de redes con las mismas características de g_1 , es decir de tipo $G(24, 110)$. En un caso la red simulada es con el modelo Erdos-Renyi (Erdos and Renyi, 1959) y luego se simula otro grafo con misma secuencia de grados que g_1

```
library(sand)
set.seed(42)
degs <- degree(g1)
g.er <- erdos.renyi.game(24, 110,type="gnm")
g.grg <- degree.sequence.game(degs, method="vl")
> mean(degree(g1))[1] 9.565217
> mean(degree(g.rg))[1] 9.565217
> mean(degree(g.grg))[1] 9.565217
> mean(betweenness(g1))[1] 7.550492
> mean(betweenness(g.rg))[1] 6.304348
> mean(betweenness(g.grg))[1] 6.304348
> mean(closeness(g1))[1] 0.02747352
> mean(closeness(g.rg))[1] 0.02900967
> mean(closeness(g.grg))[1] 0.02932709
diameter(g1)[1] 3
> diameter(g.er)[1] 3
> diameter(g.grg)[1] 3
> average.path.length(g1)[1] 1.671937
> average.path.length(g.er)[1] 1.608696
> average.path.length(g.grg)[1] 1.573123
```

Grafo aleatorio con modelo Erdos-Renyi $G(23,110)$

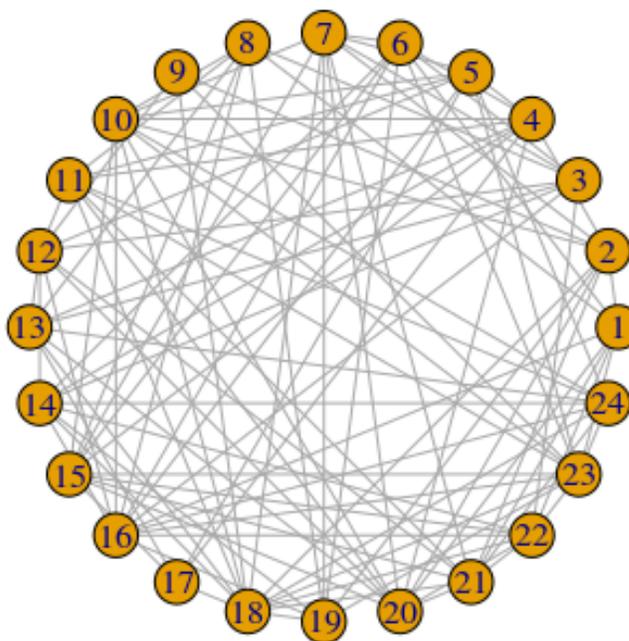


Figura 14: Grafo aleatorio con modelo Erdos-Renyi $G(23, 110)$

En la Figura 14 se puede ver la estructura de la red simulada bajo el modelo Erdos-Renyi, con la topología característica de un grafo aleatorio.

En la Figura 15 se puede ver la estructura de la red simulada que tiene la misma secuencia de grados.

Grafo g1 observado $G(23,110)$

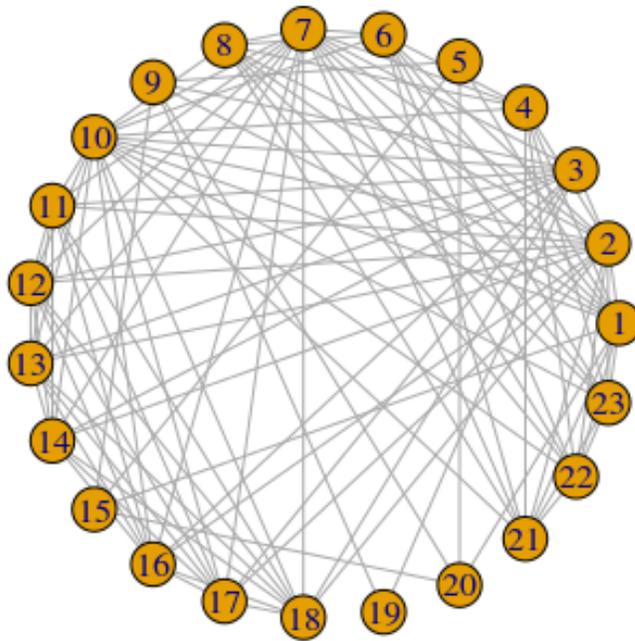


Figura 15: Grafo g1 observado $G(23,110)$

Sin embargo esas 2 características (que sea de tipo $G(24, 110)$) no alcanzan para que las topologías de las 3 redes sean las mismas. Eso puede verse en como es la relación para la betweenness y la closeness de los nodos; la curva en negro representa a $g1$, mientras que la curva roja es la del grafo con modelo Erdos-Renyi ($g.rg$), y la curva verde es la que corresponde a ($g.grg$) con misma secuencia de grados.

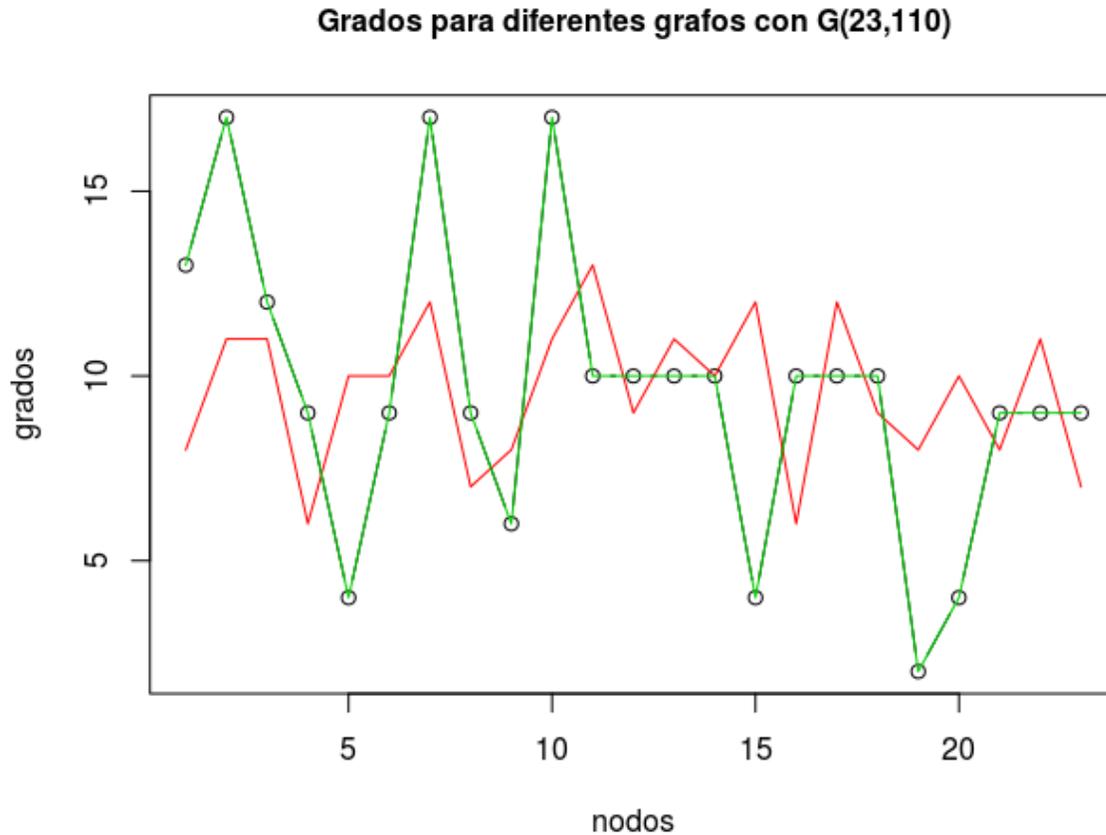


Figura 16: Comparación de grados para grafos de tipo $G(23, 110)$

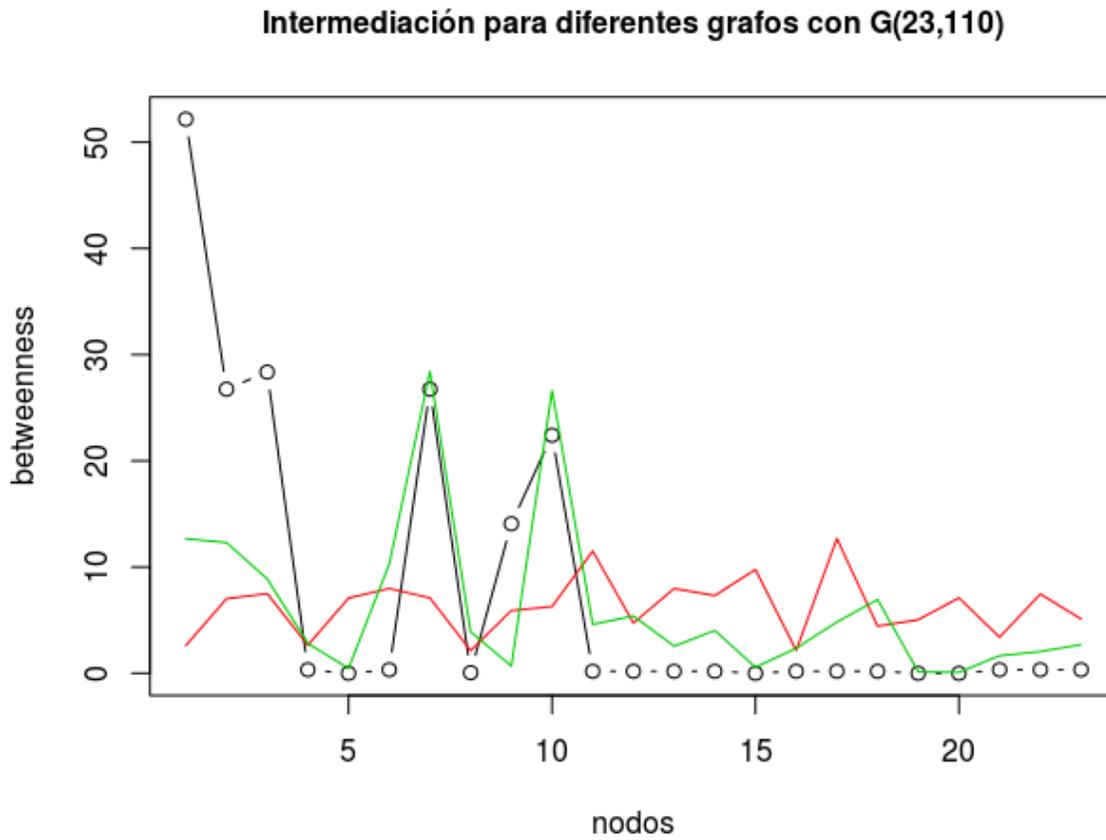


Figura 17: Comparación de Intermediación para grafos de tipo $G(23, 110)$

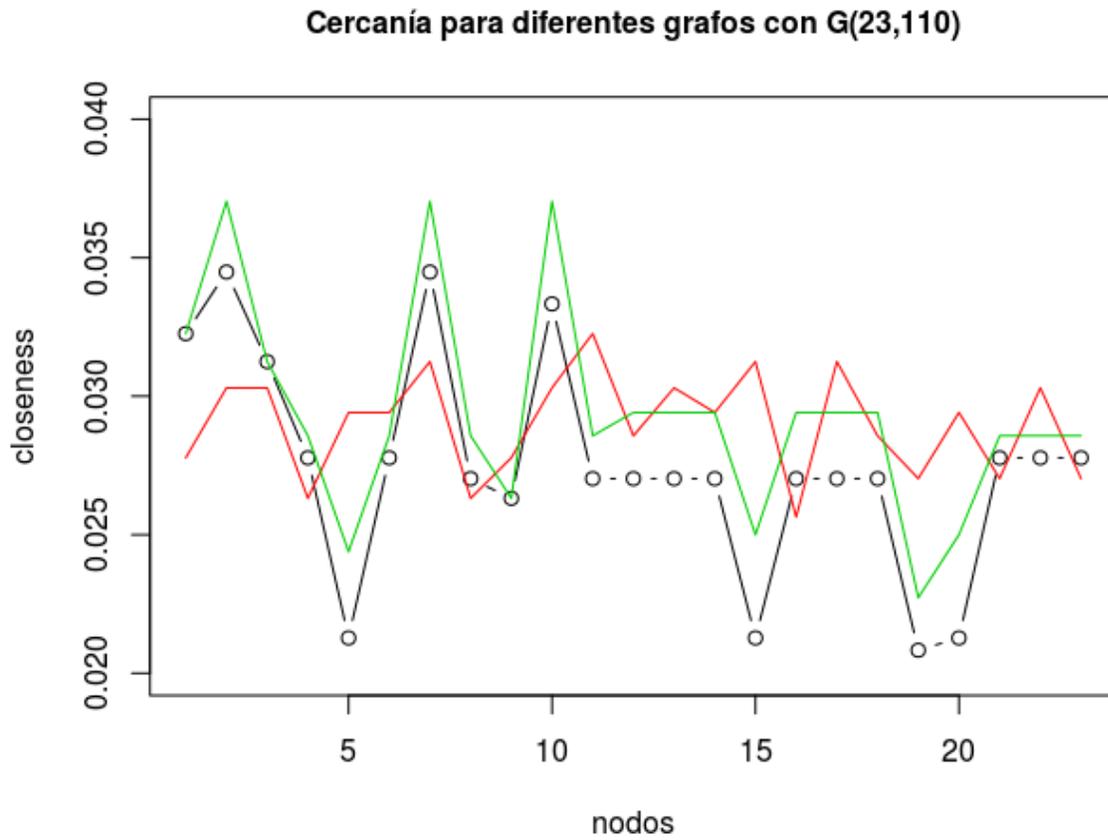


Figura 18: Comparación de closeness para grafos de tipo $G(23,110)$

Para poder evaluar el comportamiento de la red bajo estudio (g_1) vs 2 redes simuladas con estructura randómicas, se simula 1000 redes de cada tipo para las cuales se evalúan el número de comunidades generadas, las que se pueden comparar con la cantidad de comunidades identificadas en la sección anterior.

```
set.seed(1234)
nv <- vcount(g1)
ne <- ecount(g1)
degs <- degree(g1)
ntrials <- 1000
num.comm.rg <- numeric(ntrials)
for(i in (1:ntrials)){
g.rg <- erdos.renyi.game(nv, ne, type="gnm")
c.rg <- fastgreedy.community(g.rg)
num.comm.rg[i] <- length(c.rg)
}
num.comm.grg <- numeric(ntrials)
for(i in (1:ntrials)){
g.grg <- degree.sequence.game(degs, method="vl")
c.grg <- fastgreedy.community(g.grg)
num.comm.grg[i] <- length(c.grg)
}
rslts <- c(num.comm.rg,num.comm.grg)
indx <- c(rep(0, ntrials), rep(1, ntrials))
counts <- table(indx, rslts)/ntrials
barplot(counts, beside=TRUE, col=c("blue", "red"),
xlab="Numero de Comunidades",
ylab="Frecuencia Relativa",
legend=c("Tamaño Fijo ", " Grados Prefijados"))
```

En la Figura 19 puede verse que para el caso de grafo con modelo (ER), la cantidad de comunidades que predomina es 3, mientras que para el caso de los 1000 grafos simulados con misma secuencia de grados, el reparto es diferente otorgando más frecuencia a grafos con 4 comunidades; estos resultados deben compararse contra la cantidad de comunidades identificadas en la sección 4.3.1 que son 3

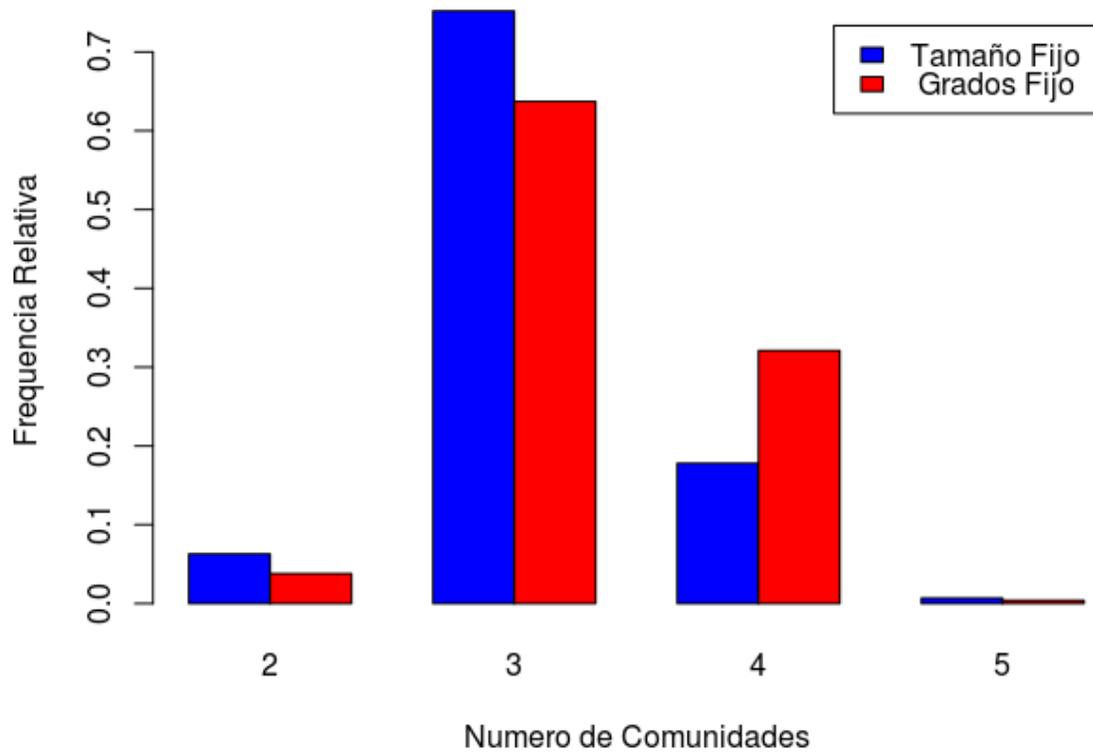


Figura 19: Verificación del Número de Comunidades para $G(24, 110)$ mediante simulación

Para evaluar el comportamiento de $g1$, vs una red de tipo *small world* de dimensiones $G(23, 110)$, se simulan 1000 redes y se comparan los valores de los coeficientes de clustering y la transitividad de las redes simuladas y de $g1$

```
ntrials <- 1000
nv <- vcount(g1)
ne <- ecoun(g1)
cl.rg <- numeric(ntrials)
apl.rg <- numeric(ntrials)
set.seed(1234)
for (i in (1:ntrials)) {
  g.rgs <- erdos.renyi.game(nv, ne, type="gnm",
  directed=TRUE)
  cl.rg[i] <- transitivity(g.rgs)
  apl.rg[i] <- average.path.length(g.rgs)
}
summary(cl.rg)
summary(apl.rg)

transitivity(g1)
average.path.length(g1)
```

Luego de efectuada la simulación puede verse que el nivel de clusterización es mucho menor y el rango no contiene el valor de $g1$ que es 0.79. Lo mismo sucede para la longitud media que en la simulación muestra un rango de [1,93; 2,27], mientras que el valor observado para $g1$ es de 1.67.

De esta manera se puede ver que $g1$ no se comporta como una red *small world*

```
> summary(cl.rg)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.2987 0.3665 0.3830 0.3833 0.4015 0.4695
> summary(apl.rg)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.935  2.053  2.077  2.080  2.105  2.275
> transitivity(g1)
[1] 0.7923008
> average.path.length(g1)
[1] 1.671937
```

4.4. Red de crucero 2 (g2)

Para las redes g_2, g_3, g_4 se presentan la estructura de cada red visualizando cada una con un layout de tipo aleatorio, consignando para cada una de ellas el perfil de gasto de 4 nodos característicos que son los que tiene menor y mayor grado, centralidad e intermediación, pudiendo a su vez comparar cual es la característica de gastos de los 5 rubros para ese crucero (cada red). A su vez para cada red se presentan las comunidades detectadas, a través del algoritmo *fast_greedy*.

También para cada grafo se presenta como es la densidad de cada uno y como cambia esta en la vecindad de los nodos caracterizados

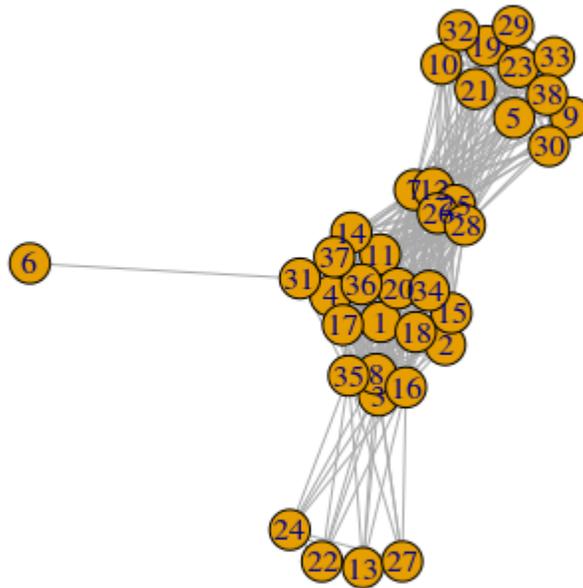


Figura 20: Red g_2 del estrato 2

nodo	Gasto Tour	Gasto Alim	Gasto Shopp	Gasto Resto	Gasto Trans
6	0.00	0.00	0.00	0.00	1.00
13	0.00	0.00	0.00	1.00	0.00
15	0.00	1.00	0.00	0.00	0.00
28	0.00	1.00	1.00	0.00	0.00
Total del crucero	1.00	22.00	16.00	8.00	2.00

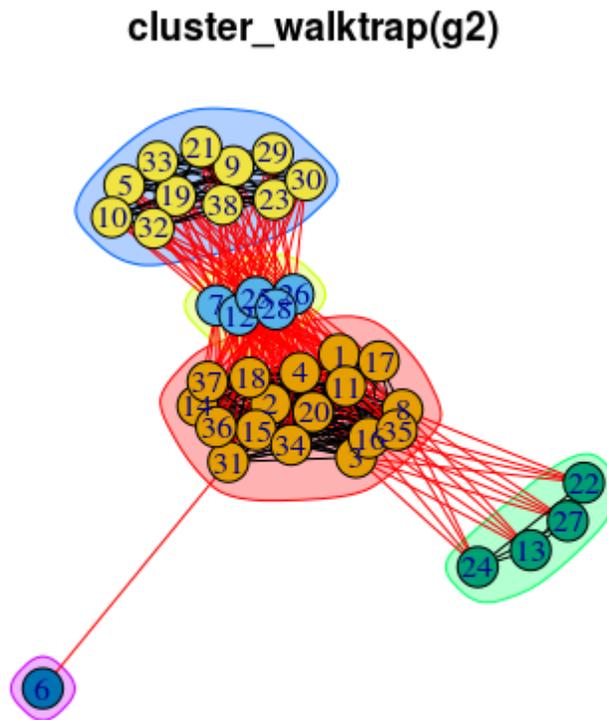


Figura 21: Clustering para Red g_2 del estrato 2 con algoritmo fast_greedy

```
> E(g2)[order(eb.g2,decreasing=T)[1:10]]
+ 10/364 edges:
[1] 6--31 3--13 3--22 3--24 3--27 8--13 8--22 8--24
[9] 8--27 13--16
> table(sapply(cliques(g2,max=5), length))#distribucion de los tamanos de cliques

1      2      3      4      5
38    364   2142   9199  30757
> table(sapply(maximal.cliques(g2), length))#cliques ,maximales

2  8 16 22
1  1  1  1
> clique.number(g2)#tamano de clicke
[1] 22
> largest_cliques(g2)
[[1]]
+ 22/38 vertices:
[1] 7 12 25 26 28 1 2 3 4 8 11 14 15 16 17 18 20 31 34 35
[21] 36 37

> graph.density(g2,loops = FALSE)
[1] 0.5177809
> graph.density(ego.c2_6,loops = FALSE)
[1] 1
> graph.density(ego.c2_13,loops = FALSE)
[1] 1
> graph.density(ego.c2_15,loops = FALSE)
[1] 1
> graph.density(ego.c2_28,loops = FALSE)
[1] 0.6458333
```

4.5. Red de crucero 3 (g3)

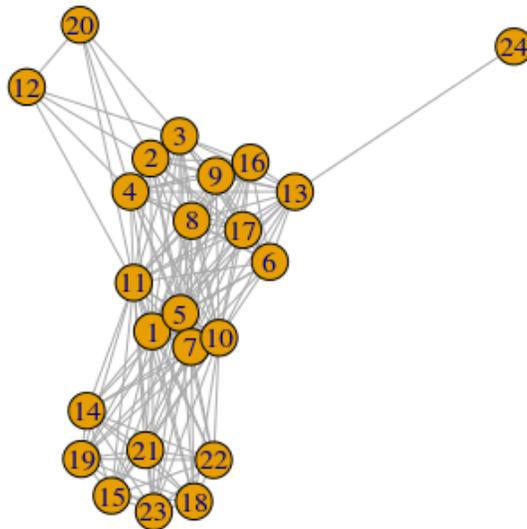


Figura 22: Red g_3 del estrato 3

```
> E(g3)[order(eb.g3,decreasing=T)[1:10]]
+ 10/157 edges:
[1] 13--24 11--12 11--20 11--13 7--13 10--13 1--13 5--13
[9] 11--14 11--15
> table(sapply(cliques(g3,max=5), length))#distribucion de los tamanos de cliques

 1    2    3    4    5
24 157  590 1505 2799
> table(sapply(maximal.cliques(g3), length))#cliques ,maximales

 2    6   12   14
```

nodo	Gasto Tour	Gasto Alim	Gasto Shopp	Gasto Resto	Gasto Trans
24	0.00	0.00	0.00	0.00	1.00
5	0.00	1.00	1.00	0.00	0.00
19	0.00	1.00	0.00	0.00	0.00
11	0.00	1.00	1.00	1.00	0.00
Total de crucero	0.00	12.00	14.00	6.00	2.00

cluster_fast_greedy(g3)

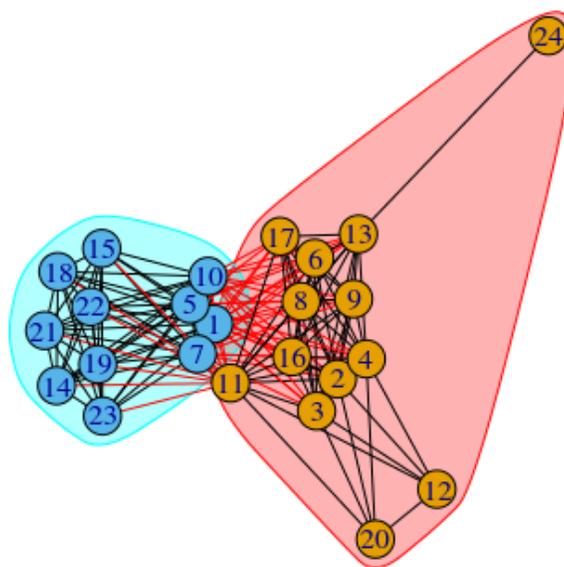


Figura 23: Clustering para Red g_3 del estrato 3 con algoritmo fast_greedy

```
1 1 1 1
> clique.number(g3)#tamano de clicke
[1] 14
> largestCliques(g3)
[[1]]
+ 14/24 vertices:
[1] 11 1 5 7 10 2 3 4 6 8 9 13 16 17

> graph.density(g3,loops = FALSE)
[1] 0.5688406
> graph.density(ego.c3_24,loops = FALSE)
[1] 1
> graph.density(ego.c3_5,loops = FALSE)
[1] 0.7
> graph.density(ego.c3_19,loops = FALSE)
[1] 1
> graph.density(ego.c3_11,loops = FALSE)
[1] 0.6166008
```


nodo	Gasto Tour	Gasto Alim	Gasto Shopp	Gasto Resto	Gasto Trans
6	0.00	0.00	1.00	0.00	0.00
15	1.00	0.00	0.00	1.00	0.00
7	0.00	1.00	0.00	0.00	0.00
24	0.00	1.00	1.00	0.00	0.00
Total de crucero	6.00	16.00	20.00	5.00	5.00

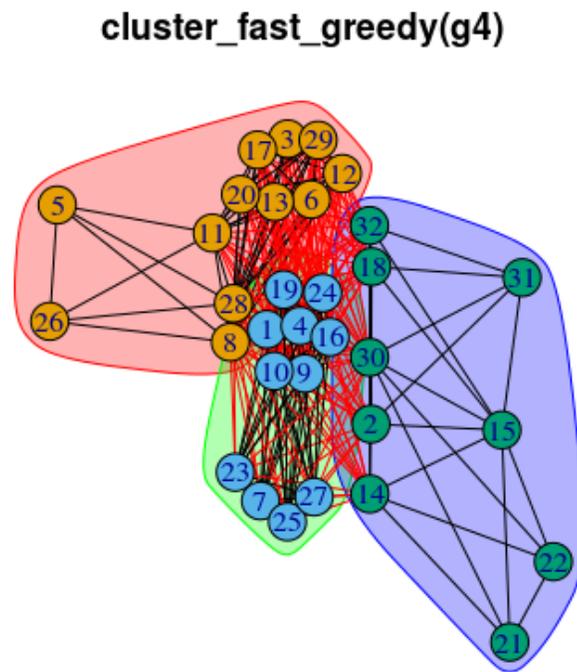


Figura 25: Clustering para Red g_4 del estrato 4 con algoritmo $fast_{greedy}$

```
4 5 6 13 16 20
1 2 1 1 1 1
> clique.number(g4)#tamano de clicke
[1] 20
> largestCliques(g4)
[[1]]
+ 20/32 vertices:
[1] 30 1 4 8 9 10 16 19 24 28 18 32 3 6 11 12 13 17 20 29

> graph.density(g4,loops = FALSE)
[1] 0.5866935
> graph.density(ego.c4_6,loops = FALSE)
[1] 1
> graph.density(ego.c4_15,loops = FALSE)
[1] 0.6944444
> graph.density(ego.c4_17,loops = FALSE)
[1] 1
> graph.density(ego.c4_24,loops = FALSE)
[1] 0.8215385
```

4.7. Comparación de las 4 redes

Una forma de comparar las diferencias entre las redes es evaluar como funciona el nivel de homofilia. Para eso se seleccionan 2 atributos que son la cantidad de gastos que tiene cada nodo y a su vez el tramo de gasto (quintil de gasto) en el que hizo ese nodo algunos de los gastos

```
> assortativity.nominal(g1, (V(g1)$nrogastos==1)+1, directed=FALSE)
[1] -0.1431979
> assortativity.nominal(g2, (V(g2)$nrogastos==1)+1, directed=FALSE)
[1] -0.09108134
> assortativity.nominal(g3, (V(g3)$nrogastos==1)+1, directed=FALSE)
[1] -0.1344944
> assortativity.nominal(g4, (V(g4)$nrogastos==1)+1, directed=FALSE)
[1] -0.152885

> assortativity.nominal(g1, (V(g1)$GtoTot.quintn==1)+1, directed=FALSE)
[1] -0.056
> assortativity.nominal(g2, (V(g2)$GtoTot.quintn==1)+1, directed=FALSE)
[1] -0.04984858
> assortativity.nominal(g3, (V(g3)$GtoTot.quintn==1)+1, directed=FALSE)
[1] 0.1658135
> assortativity.nominal(g4, (V(g4)$GtoTot.quintn==2)+1, directed=FALSE)
[1] -0.0717974
```

En la tabla 10 se puede ver un resumen de las diferentes métricas evaluadas para los 4 grafos y con que software se hizo. Si bien solamente aparece reseñada los valores y con que software se hizo, en general ambos permiten la evaluación de las mismas

métricas	g1	g2	g3	g4	software
nodos	23	38	24	32	igraph
enlaces	110	364	157	291	igraph
R_1 (Razon enlaces/nodos)	4,8	9,6	6,5	9,1	
grado medio	11,6	21,2	15,1	20,2	gephi
grado medio con pesos	12,6	22,6	17,3	24,8	gephi
betweenness	7.55	10.87	5.94	8.19	igraph
closeness	0.027	0.017	0.029	0.021	igraph
eigen-centrality	0.4851	0.5701	0.5364	0.5608	igraph
transitividad de los nodos	0.792	0.837	0.807	0.834	igraph
excentricidad máxima	3	3	3	3	igraph
edge-density	0.434	0.517	0.568	0.586	igraph
conectividad de los enlaces	2	1	1	4	igraph
diámetro	3	3	3	3	gephi
radio	2	2	2	2	gephi
densidad	0,526	0,572	0,656	0,651	gephi
modularidad	0,294	0,272	0,205	0,118	gephi
componentes conexos	1	1	1	1	gephi
coef. Medio de clustering	0,83	0,892	0,854	0,845	gephi
long media del camino	1,672	1,566	1,464	1,429	gephi

Tabla 10: Resumen de métricas para $g1$, $g2$, $g3$, $g4$

5. Discusión

En la sección 4.3, se hace con mucho detalle una identificación de la topología del grafo $g1$, a través de la evaluación de las diferentes métricas que caracterizan a los nodos y los enlaces y a la estructura de la misma con el estudio de la cantidad de comunidades, el particionado a través de valores propios. También se estudia y se compara para el grafo $g1$ contra redes aleatorias o de pequeño mundo con la mismos parámetros $G(24, 110)$. Por un tema de extensión solamente para los restantes grafos se presentan las métricas principales que se consignan en la tabla 10, aunque el procedimiento es equivalente y se presenta en los scripts de programación de R en el apéndice A.

Si se analizan los valores de la Tabla 10, las métricas globales de los 4 grafos muestran grado medio que crece con la cantidad de nodos, siendo $g1$ el que tiene menor ratio enlace/nodo, menor grado medio y menor densidad de enlaces, mientras que es el que tiene mayor modularidad y menor densidad de los 4 grafos. Por otro lado es el grafo $g3$ el que tiene menor intermediación de centralidad. En términos de cercanía las 4 redes muestran un comportamiento similar, así como de transitividad de nodos. También el nivel de clus-terización de las redes es similar y se destaca $g1$ por ser el que tiene mayor longitud media de camino y menor centralidad de valores propios. A su vez todos los grafos presentan, como era de esperar, un único componente conexo.

En resumen, todas las métricas calculadas nos arrojan a $g2$ como el más denso y mejor conectado. Vale destacar la excepción de la medida de modularidad, en la que $g1$ presenta el valor más alto, indicando una estructura interna de mayor complejidad.

Para enriquecer el análisis se describen las características socio demográficas y por gasto de los 4 clusters binarios creados mediante técnicas de clustering convencional, considerando la totalidad de los cruceristas y cruceros (2225) para luego evaluar para los 4 grafos (1 por crucero) cuales son las comunidades.

- El cluster 1 tiene casi un 95 % de cruceristas con un solo gasto, siendo el cluster 4 el que diversifica el gasto, al tener casi 47 % de gastos en 2 rubros y 23 % de gastos en 3 rubros.
- el cluster 1 se caracteriza por ser el que solamente tiene el 100 % del gasto en el quintil 5, mientras que el cluster 2 escalona el gasto en los 4 primeros quintiles, casi sin presencia de cruceristas que gasten en el quintil 5 (menos del 4 %); el cluster 3 tiene fuerte presencia de gastos en los quintiles 4 y 5 en partes casi iguales y por ultimo el cluster 4 se diferencia por no tener cruceristas en los quintiles 1 y 2.
- con respecto al % de hombres que hay en cada grupo de cruceristas , el cluster 1 es el que tiene menos % de viajeros hombres, mientras que es el cluster 4 el que

muestra mayor presencia de cruceristas hombres.

- Para evaluar la edad de los grupos de cruceristas se logra identificar que el cluster 4 es el más joven con casi 20 % de sus integrantes que son menores a 65 años, mientras que el cluster 1 es el que tiene mayoría de viajeros adultos, con casi 75 % de ellos mayores a 65 años.
- En cuanto a la composición de los clusters en términos de número de pasajeros, el cluster 2 es el que más se diferencia de los otros 3 al ser le que tiene casi 16 % de sus integrantes que son viajeros que lo hace solos, mientras que es el cluster 4 es que tiene mayoría de cruceristas que viajan en grupos de a 3 o más.

Teniendo en cuenta esa tipología creada como si los 2225 no viajaran en diferentes cruceros a lo largo de la temporada, es importante cruzar la pertenencia de cada nodo (un grupo de cruceristas de cada red, que coincide con la venida de un crucero en un día de la temporada) con la comunidad identificada. Para esta caracterización se usa solamente el algoritmo de *fast_greedy*, que hay que recordar que buscaba maximizar la modularidad.

		1	2	3
	2	7	7	6
Cluster.Bin	3	0	2	0
	4	0	0	1
	Total	7	9	7

Tabla 11: Relación entre grupos (Cluster.Bin) y Comunidades detectadas para $g1$

Para $g1$ se puede ver que hay 3 comunidades repartidas en forma homogénea y que a la interna del cluster 2 que es el mayoritario, también se reparten en forma pareja

		1	2
	1	0	2
	2	20	14
Cluster.Bin	3	1	0
	4	1	0
	Total	22	16

Tabla 12: Relación entre grupos (Cluster.Bin) y Comunidades detectadas para $g2$

Para el caso del grafo $g2$ a diferencia de como quedaba la distribución por clusters, donde era había un grupo que se llevaba el 90 % en este caso surgen 2 comunidades Para el caso del grafo $g3$ a diferencia de como quedaba la distribución por clusters, donde era había un grupo que se llevaba el 90 % en este caso surgen 2 comunidades

	1	2	
	1	2	0
Cluster.Bin	2	11	10
	3	0	1
Total	13	11	

Tabla 13: Relación entre grupos (Cluster.Bin) y Comunidades detectadas para $g3$

	1	2	3	
	2	10	7	4
Cluster.Bin	3	2	4	1
	4	0	0	4
Total	12	11	9	

Tabla 14: Relación entre grupos (Cluster.Bin) y Comunidades detectadas para $g4$

Para el caso del grafo $g4$ a diferencia de como quedaba la distribución por clusters, donde era había un grupo que se llevaba el 90% en este caso surgen 2 comunidades.

En la Tabla 15 se pueden ver los perfiles en términos de gastos de los nodos más relevantes de cada grafo, donde los que se seleccionaron son aquellos que tienen número de enlaces, closeness y betweenness, en valores mínimos, medios y máximos.

Por lo tanto para el caso del nodo 19 de grafo $g1$ se está ante el caso de un nodo casi aislado, de los pocos que gasta en Tour, mientras que el nodo 7, con mayor número de enlaces y closeness, donde comparten el gasto en alimentación y shopping.

Para el caso del grafo $g2$ los nodos que sobresalen son el 6 que se caracteriza por tener un alto nivel de centralidad, solo conectado con otro nodo, que gasta en transporte, mientras que un nodo como el 28, que gasta en alimentación y shopping tiene 32 grado, una centralidad mediana a pesar de no tener casi intermediación.

Para el grafo $g3$ los nodos que se se identifican son el 24 que solo gasta en Transporte, también solo conectado con otro nodo, muy bajo valor de centralidad, mientras que el nodo 11 que tiene el mayor grado, con casi 0.75 de centralidad, y que se caracteriza por tener gasto en alimentación y shopping.

En cuanto al grafo $g4$ los nodos que se distinguen son el 6 que tiene un solo gasto (en shopping), con un número medio de grados, alta centralidad, baja intermediación en contraposición con el nodo 24 que es el de un grupo de cruceristas que gastan en alimentación y shopping, un número alta de grados y de centralidad y también de triángulos.

nodo	G.Tour	G. Alim	G. Shopp	G. Resto	G. Trans	Comunidad
19	1.00	0.00	0.00	0.00	0.00	3
15	0.00	0.00	0.00	1.00	0.00	3
12	0.00	0.00	1.00	0.00	0.00	1
7	0.00	1.00	1.00	0.00	0.00	2
T.crucero 1	3.00	10.00	11.00	5.00	2.00	3 comunidades
6	0.00	0.00	0.00	0.00	1.00	1
13	0.00	0.00	0.00	1.00	0.00	1
15	0.00	1.00	0.00	0.00	0.00	1
28	0.00	1.00	1.00	0.00	0.00	2
T. crucero 2	1.00	22.00	16.00	8.00	2.00	2 comunidades
24	0.00	0.00	0.00	0.00	1.00	1
5	0.00	1.00	1.00	0.00	0.00	2
19	0.00	1.00	0.00	0.00	0.00	2
11	0.00	1.00	1.00	1.00	0.00	1
T. crucero 3	0.00	12.00	14.00	6.00	2.00	2 comunidades
6	0.00	0.00	1.00	0.00	0.00	1
15	1.00	0.00	0.00	1.00	0.00	3
7	0.00	1.00	0.00	0.00	0.00	2
24	0.00	1.00	1.00	0.00	0.00	2
T. crucero 4	6.00	16.00	20.00	5.00	5.00	3 comunidades
TOTAL	267	1037	1411	210	154	3079

Tabla 15: Identificación de nodos usado algoritmo *fast_greedy*

El rubro mayoritario de gastos en general es tal como se ve en la última línea de la Tabla 15, el que corresponde a shopping que corresponde a casi la mitad de los gastos de todos los rubros, seguido por alimentación que absorbe una tercera parte de los mismos, siendo que los restantes rubros son francamente minoritarios. Esta relación sin embargo no se mantiene a la interna de los cruceros seleccionados para hacer el análisis de SNA, lo que estaría indicando que hay una cierta estructura en el perfil de gasto que varía al cambiar de crucero. Cuando se cambia de crucero (de tamaño de crucero) si bien la presencia del cluster 2, que corresponde a la tipología previa es mayoritaria, no siempre los nodos se alinean con esta partición, lo que estaría indicando que la variabilidad intra crucero condiciona la partición en comunidades que no se corresponden con los grupos creados usando todos los cruceristas de la temporada.

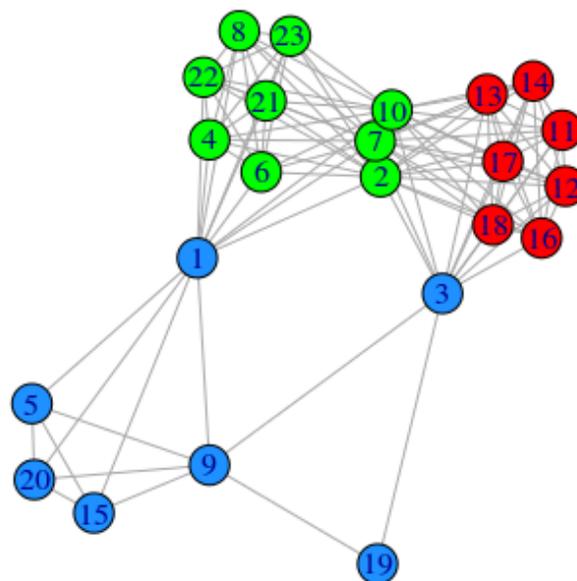


Figura 26: Clustering para Red g_1 contrastado de acuerdo a las comunidades formadas mediante algoritmo fast_greedy

En la Figura 26 queda claramente establecidas las 3 comunidades mediante el algoritmo fast_greedy, dando lugar a una partición que luego al contrastar por otros atributos como son el nivel de gastos, el número de gastos y el cluster binario no se mantiene.

Para entender mejor el contraste de colores, en la Figura 27 el nodo en rojo es un nodo que gasta en el primer quintil, el verde en el segundo, el azul en el tercero, en amarillo el cuarto quintil y en magenta el quinto quintil. Por ejemplo la comunidad 3 que aparece en azul en la Figura 26 en términos de niveles de gastos tiene los 5 quintiles, al igual que la comunidad 2, que aparece en verde, mientras que se diferencia la comunidad 1 en rojo que muestra mayoritariamente gasto pertenecientes al primer tramo.(quintil 1)

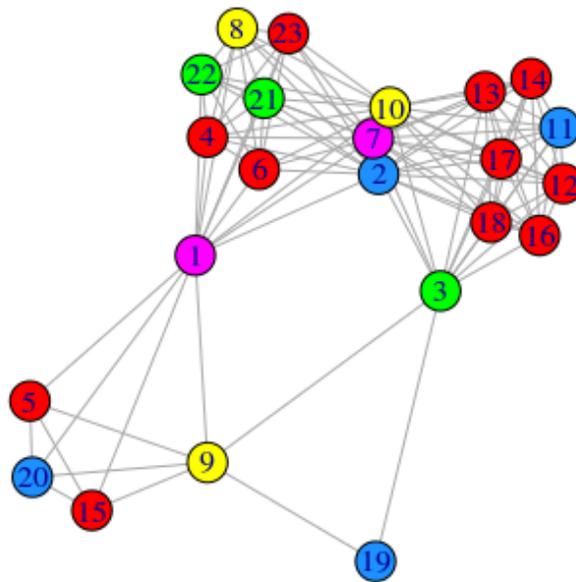


Figura 27: Clustering para Red g_1 contrastado de acuerdo al quintil de gasto

También la comunidad 1 se diferencia de las otras 2 al mostrar en la Figura 28 que solo gasta en solo rubro (en este caso el color rojo es 1 que gasta en 1, el verde que gasta en 2 y en azul en 3), donde además el nodo 10 de la comunidad 2 se diferencia de los 22 restantes nodos al ser el único que gasta en 3 rubros.

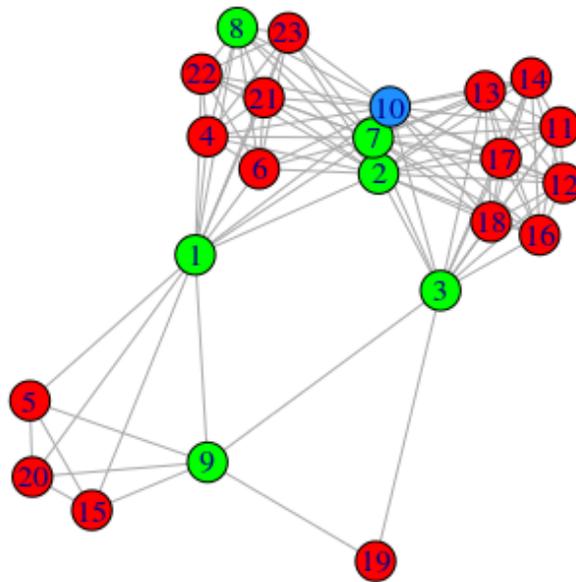


Figura 28: Clustering para Red g_1 contrastado de acuerdo al número de gastos

Cuando se analiza el comportamiento de los nodos de acuerdo a la membresía en las comunidades y al clustering binario original, el nodo 19 de diferencia de los restantes al ser el único que pertenece al cluster 4 (en este caso el color rojo corresponde a 2, verde al 3 y azul al 4), siendo además una característica que que la comunidad 1 se corresponde íntegramente con el cluster 2. (Hay que recordar que el cluster binario 2 era el más frecuente en la totalidad de los cruceristas)

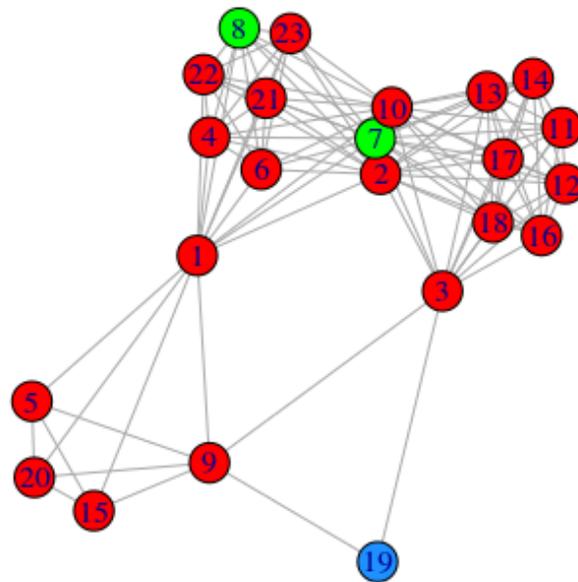


Figura 29: Clustering para Red g_1 contrastado de acuerdo al clustering binario

Finalmente se puede decir que para cada red, los nodos más característicos (los que en cada caso se presentaron en términos de intermediación, grados de conexión) muestran densidades locales diferentes y mayores con respecto a la densidad global de cada red.

6. Conclusiones

A modo de conclusiones preliminares luego de desarrollado el análisis puede decirse que las 4 redes luego de ser caracterizadas en términos de las métricas de nodos y enlaces, muestran comportamientos diferentes entre sí. También las comunidades detectadas muestran una segmentación que difiere si ésta se hace en forma global (clustering binario). A su vez en la segmentación creada con el SNA aparecen comportamientos diferenciados en términos de los niveles o números de gastos y que no se asocian directamente con la lógica usada para crear las redes en cada crucero y que hay que recordar que era en base a si compartían algunos de los 5 rubros de gastos.

Por otra parte en términos de la aplicación se trabajó con el software Gephi, que tiene la gran ventaja de poder visualizar grafos con muchos nodos y enlaces. Es una herramienta donde los graficos son bastante parametrizables y que calcula una serie de métricas clásicas en el SNA. Tiene la desventaja de ser una herramienta básicamente interactiva, con lo cual es necesario que el usuario haga los cambios personalmente. En cambio el R tiene la ventaja de poder generar incluso más cantidad de métricas y sobre todo de ser un entorno de programación, donde se pueden incluso programar mediante funciones otras métricas, simular diferentes tipos de redes y sobre todo trabajar en SNA desde una perspectiva del modelado y la inferencia, aprovechando la potencia que tiene el R como lenguaje de programación orientado a objetos, que permite el tratamiento de estos con diferentes bibliotecas desarrolladas en R. Sin embargo la performance del R para la visualización de grafos, cuando estos son de tamaño mediano, es lenta.

Teniendo en cuenta los comentarios previos es que se propone seguir el trabajo con los siguientes direcciones

- Tomar algún estrato de tamaño de crucero y ampliar la muestra creando de esa manera más redes y evaluando si el patrón detectado con una sola se mantiene
- Crear la tipología con la lógica de clustering convencional (clasificación no supervisada) solamente sobre los nodos de la red y compararla con las comunidades de la red
- Cambiar la forma de muestreo (en vez de MAS hacerlo proporcional al tamaño de cada crucero)
- Evaluar si un crucero que repite en la temporada, genera un grafo diferente y que puede depender del puerto de desembarque.
- Proponer un análisis de redes validando modelos estadísticos (recordar que esto es solo descripción), donde algunos de los atributos evaluados en la caracterización se

pueden usar como variables explicativas, usando la teoría de los modelos exponenciales aleatorios en grafos (ERGM),(Kolaczyk and Csárdi, 2014)

- Ver como se puede usar los expansores en la lógica del SNA ya que hay que recordar que los datos provienen de una muestra con diseño muestral complejo, donde cada nodo (cruceista encuestado) tiene un factor de expansión que incide en 2 aspectos claves: para hacer inferencia y también en los análisis que puedan considerar el análisis factorial, ya que la inercia a descomponer depende de las distancias de los nodos al baricentro y de los pesos de los mismos. (Kolaczyk, 2009),(Kolaczyk and Csárdi, 2014)

Bibliografía

- Alvarez Vaz, R., Altmark, S., and Santiñaque, F. (2015). Caracterización de los componentes del gasto de los turistas de cruceros en Uruguay mediante técnicas de datamining. In *II Jornadas de Econometría, Facultad de Ciencias Económicas*. UBA.
- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.
- Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, (5):1170.
- Bonacich, P. and Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23:191 – 201.
- Borgatti, S. P., Everett, M. G., and Johnson, J. (2013). *Analyzing Social Networks*. SAGE Publications Ltd.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25(2):163–177.
- Brandes, U. and Erlebach, T. (2005). *Network analysis: methodological foundations*. Number 3418 in LCNS, Tutorial. Springer, Berlin ; New York. OCLC: ocm58474176.
- Butts, C. T. (2016). *sna: Tools for Social Network Analysis*. R package version 2.4.
- Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E*, 70:066111.
- Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695.
- Erdos, P. and Renyi, A. (1959). On random graphs i. *Publ. Math. Debrecen*, 6:290–297.
- Fiedler, M. (1973). Algebraic connectivity of graphs. *Czech. Math. J.*, 23(98):298–305.
- Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215.
- Fruchterman, T. M. J. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164.
- Kolaczyk, E. (2009). *Statistical analysis of network data : methods and models*. Springer, New York London.

- Kolaczyk, E. and Csárdi, G. (2014). *Statistical analysis of network data with R*. Springer, New York.
- Kolaczyk, E. D. and Csárdi, G. (2017). *sand: Statistical Analysis of Network Data with R*. R package version 1.0.3.
- Luke, D. (2015). *A user's guide to network analysis in R*. Springer, Cham.
- Newman, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701.
- Newman, M. E. J. (2003). Mixing patterns in networks. *Phys. Rev. E*, 67:026126.
- Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113.
- Pons, P. and Latapy, M. (2015). Computing communities in large networks using random walks.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raghavan, U. N., Albert, R., and Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E*, 76:036106.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603.
- Wasserman, S. and Faust, K. (1994). *Social network analysis: methods and applications*. Number 8 in Structural analysis in the social sciences. Cambridge University Press, Cambridge ; New York.

A. Apéndices Estadísticos

En las secciones que siguen se presenta para cada red los valores de las métricas de intermediación y centralidad para cada nodo, así como la identificación de las comunidades, usando los algoritmos de *walktrap*, *edge_betweenness* y *label_prop*.

cluster_fast_greedy(g2)

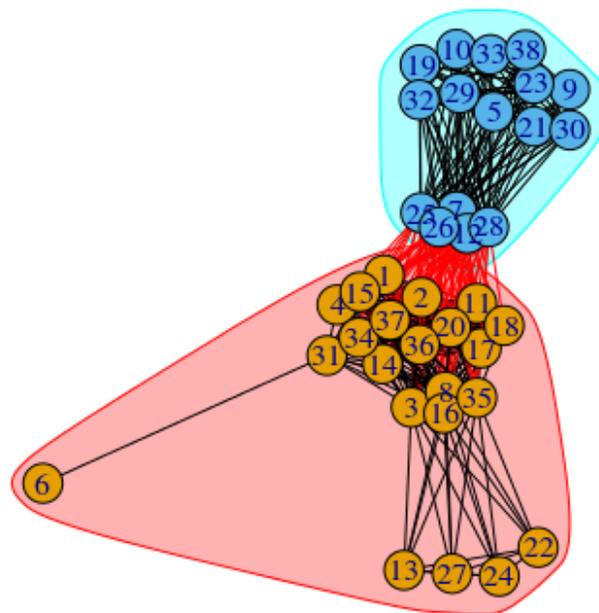


Figura 30: Clustering para Red g_1 del estrato 1 con algoritmo walktrap

	nodo	grados	betweenness	closeness	triángulos
1	1.00	21.00	0.00	0.70	210.00
2	2.00	21.00	0.00	0.70	210.00
3	3.00	25.00	0.05	0.71	228.00
4	4.00	21.00	0.00	0.70	210.00
5	5.00	15.00	0.00	0.58	105.00
6	6.00	1.00	0.00	0.42	0.00
7	7.00	32.00	0.07	0.80	309.00
8	8.00	25.00	0.05	0.71	228.00
9	9.00	15.00	0.00	0.58	105.00
10	10.00	15.00	0.00	0.58	105.00
11	11.00	21.00	0.00	0.70	210.00
12	12.00	32.00	0.07	0.80	309.00
13	13.00	7.00	0.00	0.47	21.00
14	14.00	21.00	0.00	0.70	210.00
15	15.00	21.00	0.00	0.70	210.00
16	16.00	25.00	0.05	0.71	228.00
17	17.00	21.00	0.00	0.70	210.00
18	18.00	21.00	0.00	0.70	210.00
19	19.00	15.00	0.00	0.58	105.00
20	20.00	21.00	0.00	0.70	210.00
21	21.00	15.00	0.00	0.58	105.00
22	22.00	7.00	0.00	0.47	21.00
23	23.00	15.00	0.00	0.58	105.00
24	24.00	7.00	0.00	0.47	21.00
25	25.00	32.00	0.07	0.80	309.00
26	26.00	32.00	0.07	0.80	309.00
27	27.00	7.00	0.00	0.47	21.00
28	28.00	32.00	0.07	0.80	309.00
29	29.00	15.00	0.00	0.58	105.00
30	30.00	15.00	0.00	0.58	105.00
31	31.00	22.00	0.05	0.71	210.00
32	32.00	15.00	0.00	0.58	105.00
33	33.00	15.00	0.00	0.58	105.00
34	34.00	21.00	0.00	0.70	210.00
35	35.00	25.00	0.05	0.71	228.00
36	36.00	21.00	0.00	0.70	210.00
37	37.00	21.00	0.00	0.70	210.00
38	38.00	15.00	0.00	0.58	105.00

Tabla 16: Caracterización de los nodos de g_2

cluster_edge_betweenness(g2)

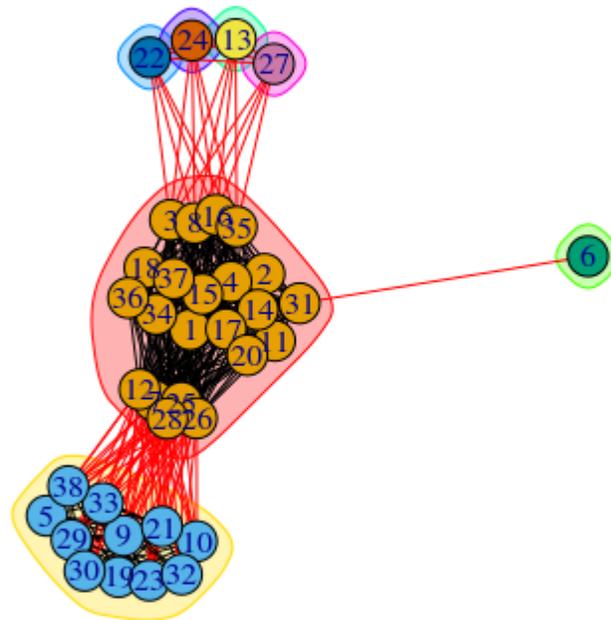


Figura 31: Clustering para Red g_2 del estrato 1 con algoritmo edge_betweenness

cluster_label_prop(g2)

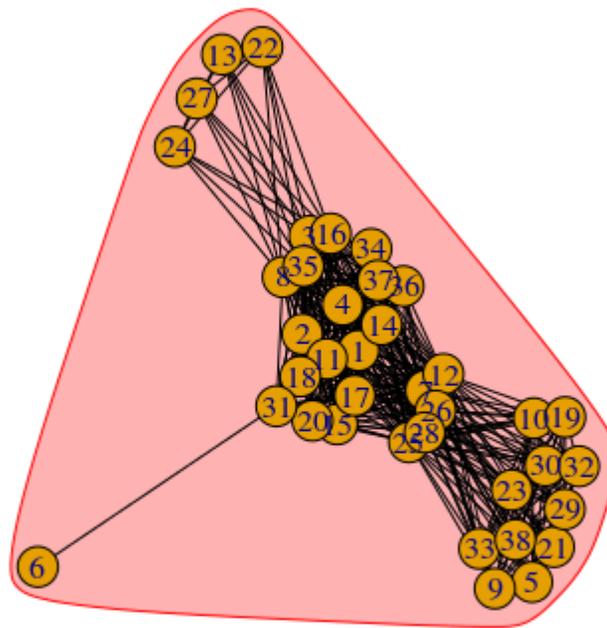


Figura 32: Clustering para Red g_1 del estrato 1 con algoritmo label_prop

	nodo	grados	betweenness	closeness	triángulos
1	1.00	20.00	0.06	0.77	127.00
2	2.00	15.00	0.03	0.68	85.00
3	3.00	15.00	0.03	0.68	85.00
4	4.00	15.00	0.03	0.68	85.00
5	5.00	20.00	0.06	0.77	127.0
6	6.00	13.00	0.00	0.70	78.00
7	7.00	20.00	0.06	0.77	127.0
8	8.00	13.00	0.00	0.70	78.00
9	9.00	13.00	0.00	0.70	78.00
10	10.00	20.00	0.06	0.77	127
11	11.00	22.00	0.11	0.74	134
12	12.00	5.00	0.00	0.55	10.00
13	13.00	14.00	0.09	0.72	78.00
14	14.00	11.00	0.00	0.64	55.00
15	15.00	11.00	0.00	0.64	55.00
16	16.00	13.00	0.00	0.70	78.00
17	17.00	13.00	0.00	0.70	78.00
18	18.00	11.00	0.00	0.64	55.00
19	19.00	11.00	0.00	0.64	55.00
20	20.00	5.00	0.00	0.55	10.00
21	21.00	11.00	0.00	0.64	55.00
22	22.00	11.00	0.00	0.64	55.00
23	23.00	11.00	0.00	0.64	55.00
24	24.00	1.00	0.00	0.43	0.00

Tabla 17: Caracterización de los nodos de g_3

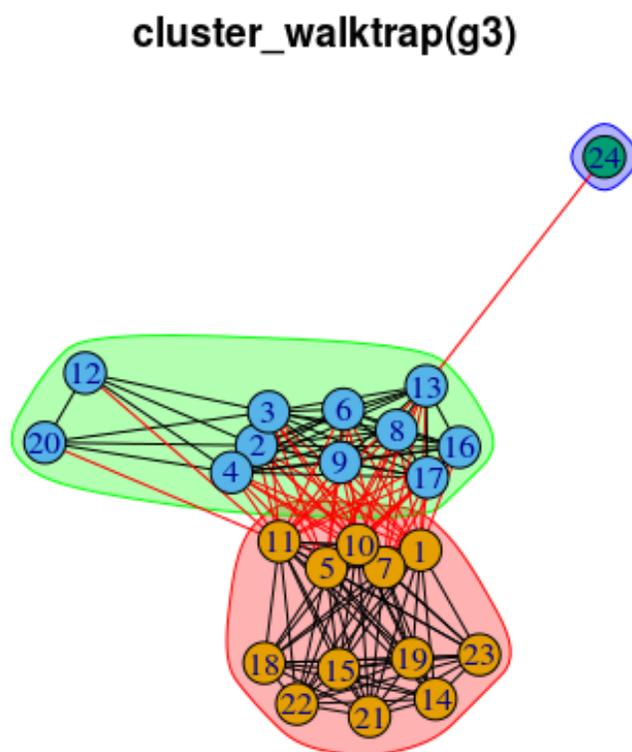


Figura 33: Clustering para Red g_3 del estrato 3 con algoritmo walktrap

cluster_edge_betweenness(g3)

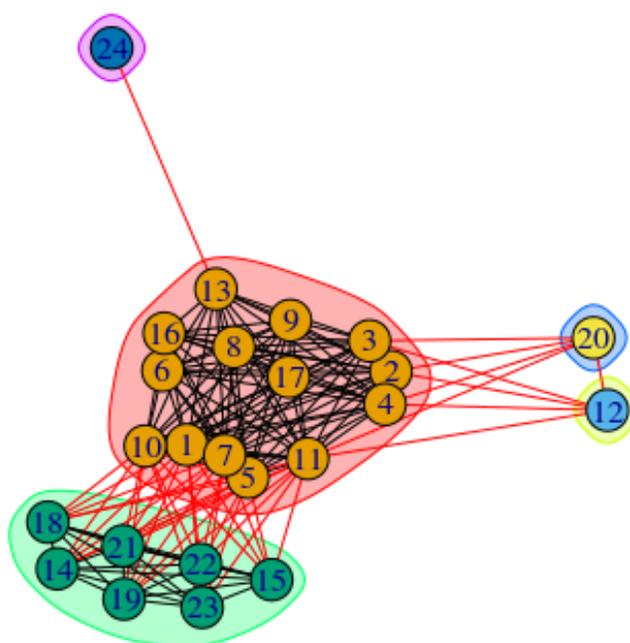


Figura 34: Clustering para Red g_3 del estrato 3 con algoritmo edge_betweenness

cluster_label_prop(g3)

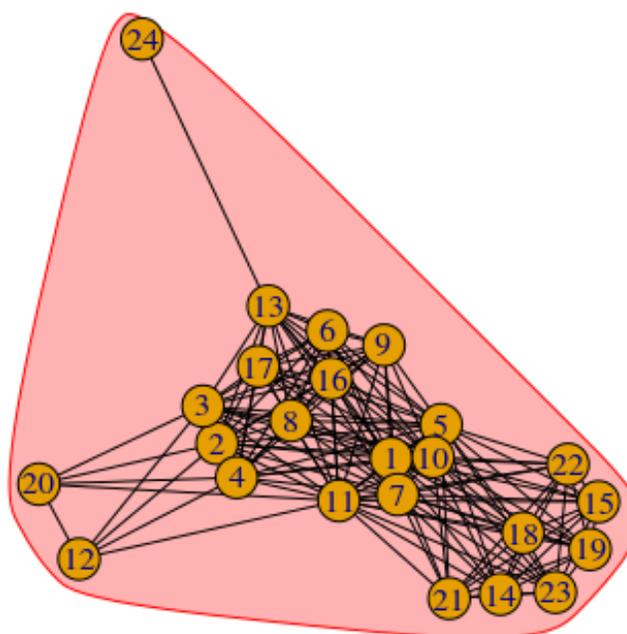


Figura 35: Clustering para Red g_3 del estrato 3 con algoritmo label_prop

	nodo	grados	betweenness	closeness	triángulos
1	1.00	25.00	0.01	0.67	242.00
2	2.00	19.00	0.03	0.70	134.00
3	3.00	19.00	0.01	0.72	171.00
4	4.00	25.00	0.01	0.67	242.00
5	5.00	4.00	0.00	0.50	6.00
6	6.00	19.00	0.01	0.72	171.00
7	7.00	15.00	0.01	0.66	105.00
8	8.00	27.00	0.04	0.69	247.00
9	9.00	25.00	0.01	0.67	242.00
10	10.00	25.00	0.01	0.67	242.00
11	11.00	21.00	0.06	0.72	176.00
12	12.00	19.00	0.01	0.72	171.00
13	13.00	19.00	0.01	0.72	171.00
14	14.00	18.00	0.07	0.69	112.00
15	15.00	8.00	0.01	0.54	17.00
16	16.00	25.00	0.01	0.67	242.00
17	17.00	19.00	0.01	0.72	171.00
18	18.00	22.00	0.03	0.74	189.00
19	19.00	25.00	0.01	0.67	242.00
20	20.00	19.00	0.01	0.72	171.00
21	21.00	4.00	0.00	0.52	6.00
22	22.00	4.00	0.00	0.52	6.00
23	23.00	15.00	0.01	0.66	105.00
24	24.00	25.00	0.01	0.67	242.00
25	25.00	15.00	0.01	0.66	105.00
26	26.00	4.00	0.00	0.50	6.00
27	27.00	15.00	0.01	0.66	105.00
28	28.00	27.00	0.04	0.69	247.00
29	29.00	19.00	0.01	0.72	171.00
30	30.00	29.00	0.07	0.66	255.00
31	31.00	5.00	0.00	0.53	10.00
32	32.00	22.00	0.03	0.74	189.00

Tabla 18: Caracterización de los nodos de g_4

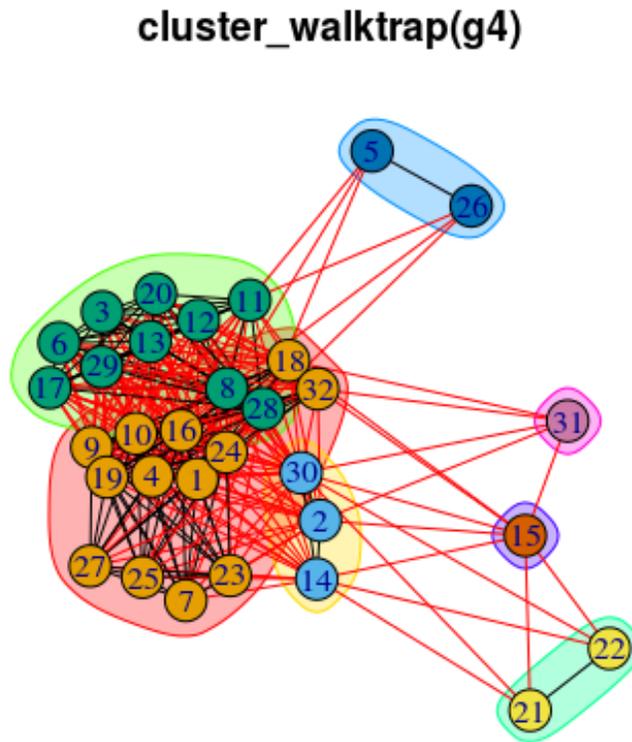


Figura 36: Clustering para Red g_4 del estrato 4 con algoritmo walktrap

cluster_edge_betweenness(g4)

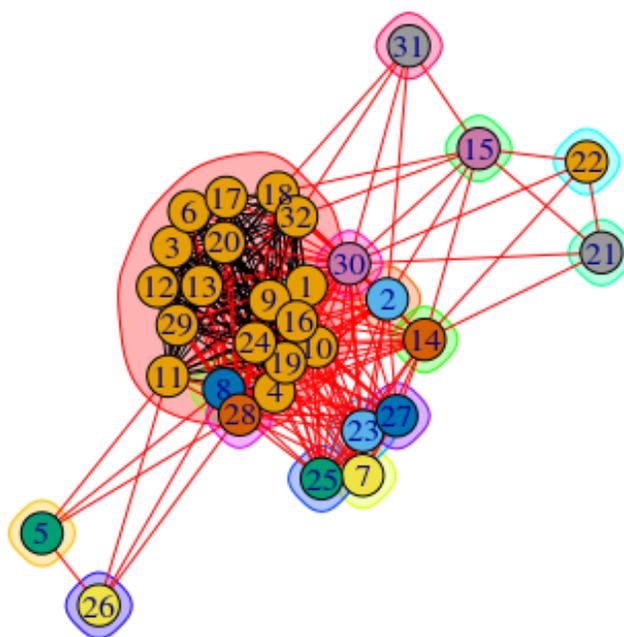


Figura 37: Clustering para Red g_4 del estrato 4 con algoritmo edge_betweenness

cluster_label_prop(g4)

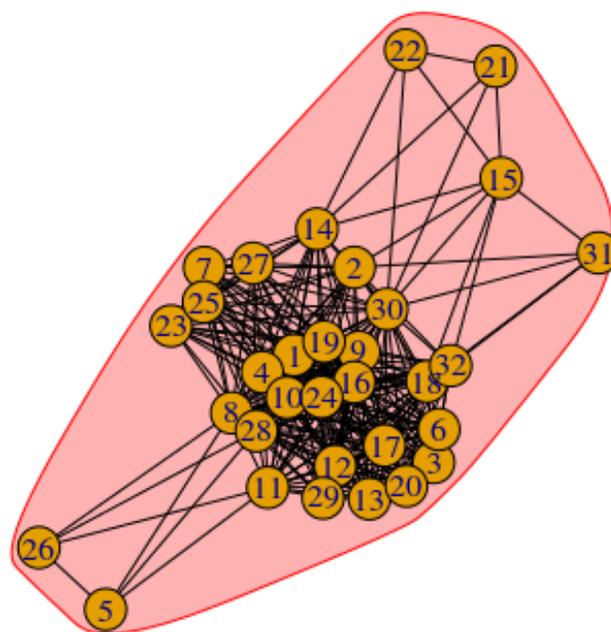


Figura 38: Clustering para Red g_4 del estrato 4 con algoritmo label_prop

Referencias Bibliográficas

- Alvarez Vaz, R., Altmark, S., and Santiñaque, F. (2015). Caracterización de los componentes del gasto de los turistas de cruceros en Uruguay mediante técnicas de datamining. In *II Jornadas de Econometría, Facultad de Ciencias Económicas*. UBA.
- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.
- Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, (5):1170.
- Bonacich, P. and Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23:191 – 201.
- Borgatti, S. P., Everett, M. G., and Johnson, J. (2013). *Analyzing Social Networks*. SAGE Publications Ltd.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25(2):163–177.
- Brandes, U. and Erlebach, T. (2005). *Network analysis: methodological foundations*. Number 3418 in LCNS, Tutorial. Springer, Berlin ; New York. OCLC: ocm58474176.
- Butts, C. T. (2016). *sna: Tools for Social Network Analysis*. R package version 2.4.
- Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E*, 70:066111.
- Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695.
- Erdos, P. and Renyi, A. (1959). On random graphs i. *Publ. Math. Debrecen*, 6:290–297.
- Fiedler, M. (1973). Algebraic connectivity of graphs. *Czech. Math. J.*, 23(98):298–305.
- Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215.
- Fruchterman, T. M. J. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164.
- Kolaczyk, E. (2009). *Statistical analysis of network data : methods and models*. Springer, New York London.

- Kolaczyk, E. and Csárdi, G. (2014). *Statistical analysis of network data with R*. Springer, New York.
- Kolaczyk, E. D. and Csárdi, G. (2017). *sand: Statistical Analysis of Network Data with R*. R package version 1.0.3.
- Luke, D. (2015). *A user's guide to network analysis in R*. Springer, Cham.
- Newman, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701.
- Newman, M. E. J. (2003). Mixing patterns in networks. *Phys. Rev. E*, 67:026126.
- Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113.
- Pons, P. and Latapy, M. (2015). Computing communities in large networks using random walks.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raghavan, U. N., Albert, R., and Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E*, 76:036106.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603.
- Wasserman, S. and Faust, K. (1994). *Social network analysis: methods and applications*. Number 8 in Structural analysis in the social sciences. Cambridge University Press, Cambridge ; New York.

Instituto de Estadística

Documentos de Trabajo



Eduardo Acevedo 1139. CP 11200 Montevideo, Uruguay

Teléfonos y fax: (598) 2410 2564 - 2418 7381

Correo: ddt@iesta.edu.uy

www.iesta.edu.uy

Área Publicaciones

Abril, 2019

DT (19/1)