Taller 2. Análisis de redes sociales

Student Michel Mendivenson Barragán Zabala
Proffesor Juan Camilo Sosa
Course Statistical Network Analysis
University Universidad Nacional de Colombia

Punto 1:

El grado de un nodo en una red tanto dirigida como no dirigida se puede calcular fácilmente a partir de la matriz de adyacencia $\mathbf{Y} = [y_{i,j}]$. El *out-degree* d_i^{out} y el *in-degree* d_i^{in} del nodo i se pueden calcular respectivamente como:

$$d_i^{\text{out}} = \sum_{j:j\neq i} y_{i,j} \qquad \textbf{y} \qquad d_i^{\text{in}} = \sum_{j:j\neq i} y_{j,i}$$

Muestre que este cálculo funciona tanto para relaciones dirigidas como no dirigidas. Específicamente, muestre que si la red es no dirigida entonces $d_i^{\text{out}} = d_i^{\text{in}}$.

Lo que primordialmente se debe revisar acá es que la suma $d_i^{\text{out}} = \sum_{j:j\neq i} y_{i,j}$ es la suma de los elementos de la matriz de adyacencia correspondientes a la fila i (Al menos para una red binaria) mientras que la sumatoria $d_i^{\text{in}} = \sum_{j:j\neq i} y_{j,i}$ corresponde a la suma de los elementos de la matriz de adyacencia que se encuentran ubicados en la columna i. Con esto, para una red no dirigida (binaria) se vuelve muy fácil demostrar lo que se pide pues para este tipo de red la matriz de adyacencia es simétrica, luego tendremos que si X es esta matriz $X^T = X$ y para cualquier vector fila de la matriz la siguiente relación se cumple $X_{j.} = X_{.j}$ (La fila j es igual a la columna j)y por ende las sumatorias serán iguales.

Punto 2:

Tanto para redes dirigidas como no dirigidas, se define la media global de las interacciones como

$$\bar{y} = \frac{1}{n(n-1)} \sum_{i,j:i\neq j} y_{i,j}$$

donde $\mathbf{Y} = [y_{i,j}]$ es la matriz de adyacencia de la red correspondiente. Tal estadístico corresponde a una descripción muy rudimentaria acerca de la plausibilidad de observar una relación entre dos nodos cualesquiera, dado que no tiene en cuenta la heterogeneidad nodal (algunos nodos son más propensos a enviar/recibir más relaciones).

a) Muestre que para relaciones no dirigidas la media global $\bar{y}$ es igual a la media tanto de la parte triangular superior de $\mathbf{Y}$ como de la parte triangular inferior de $\mathbf{Y}$.

Es posible definir la media global para la triangular inferior de la siguiente forma:

$$\frac{2}{n(n-1)} \sum_{i=1,j=1,i<j}^{n}\sum^{n} x_{i,j}$$

Es decir, la suma de la triangular inferior de la matriz de adyacencia sobre la cantidad de nodos que se presentan en esta sección de y de la triangular superior de la siguiente forma:

$$\frac{2}{n(n-1)} \sum_{i=1,j=1,i>j}^{n}\sum^{n} x_{i,j}$$

Si utilizamos la matriz triangular inferior:

$$\begin{aligned}
\frac{2}{n(n-1)} \sum_{i=1,j=1,i<j}^{n}\sum^{n} x_{i,j} &= \frac{1}{n(n-1)} \sum_{i=1,j=1,i<j}^{n}\sum^{n} x_{i,j} + \frac{1}{n(n-1)} \sum_{i=1,j=1,i<j}^{n}\sum^{n} x_{i,j} \\
&= \frac{1}{n(n-1)} \sum_{i=1,j=1,i<j}^{n}\sum^{n} x_{i,j} + \frac{1}{n(n-1)} \sum_{i=1,j=1,i>j}^{n}\sum^{n} x_{i,j} \quad \text{Por la simetría de la matriz de adyacencia.} \\
&= \frac{1}{n(n-1)} \left(\sum_{i=1,j=1,i<j}^{n}\sum^{n} x_{i,j} + \sum_{i=1,j=1,i>j}^{n}\sum^{n} x_{i,j} \right) \\
&= \frac{1}{n(n-1)} \sum_{i,j:i\neq j} y_{i,j} = \bar{y}
\end{aligned}$$

Note que esta prueba funciona exactamente para la matriz triangular superior.

b) Muestre que tanto para relaciones dirigidas como no dirigidas la media global corresponde a la densidad de la red.

La densidad de una red H no dirigida esta definida de la siguiente forma:

$$den(H) = \frac{|E_H|}{\frac{|V_H|(|V_H|-1)}{2}}$$

donde $|E_H|$ corresponde a la cantidad de enlaces de la red y $|V_H|$ a la cantidad de vértices de la red. Para una red no dirigida de n nodos:

$$\frac{|E_H|}{\frac{|V_H|(|V_H|-1)}{2}} = \frac{2}{n(n-1)} * |E_H| \qquad \text{Debido a } |V_H| = n$$

$$= \frac{2}{n(n-1)} * \sum_{i=1}^{n}\sum_{j=1}^{n} I_E(i,j)$$

$$\text{Con } I_E = \begin{cases} 1; \ (i,j) \in E(H) \\ 0; \ (i,j) \notin E(H) \end{cases}$$

$$= \frac{2}{n(n-1)} \sum_{i=1}^{n}\sum_{j<1}^{n} A_{ij} \qquad \text{Si A es la matriz de adyacencia simétrica asociada a } H$$

$$= \frac{1}{n(n-1)} \sum_{i,j:i\neq j} y_{i,j} = \bar{y} \qquad \text{Por lo visto en el item (a)}$$

Para el caso de una red no dirigida se tendrá que $den(H) = \frac{|E_H|}{|V_H|(|V_H|-1)} = \frac{2}{n(n-1)} * \sum_{i=1}^{n}\sum_{j=1}^{n} I_E(i,j) = \frac{1}{n(n-1)} \sum_{i,j:i\neq j} y_{i,j} = \bar{y}$.

c) Muestre que tanto para relaciones dirigidas como no dirigidas se tiene que $\bar{y} = \bar{d}^{\text{out}} = \bar{d}^{\text{in}}$. Es decir, el grado promedio tanto de entrada como de salida son iguales y a su vez equivalentes a la densidad.

El grado de entrada se define como:

$$\bar{d}^{\text{in}} = \frac{1}{n}\sum_{j=1}^{n} d_j^{\text{in}}, \quad \text{donde } d_j^{\text{in}} = \sum_{i=1}^{n} y_{i,j}.$$

El grado de salida se define como:

$$\bar{d}^{\text{out}} = \frac{1}{n}\sum_{i=1}^{n} d_i^{\text{out}}, \quad \text{donde } d_i^{\text{out}} = \sum_{j=1}^{n} y_{i,j}.$$

Para redes dirigidas, la suma total de las entradas y salidas es igual al número total de enlaces presentes en la red:

$$\sum_{i,j} y_{i,j} = \sum_{j=1}^{n}\sum_{i=1}^{n} y_{i,j} = \sum_{i=1}^{n}\sum_{j=1}^{n} y_{i,j}.$$

Por lo tanto, la suma total de los grados de entrada es igual a la suma total de los grados de salida:

$$\sum_{j=1}^{n} d_j^{\text{in}} = \sum_{i=1}^{n} d_i^{\text{out}} = \sum_{i,j} y_{i,j}.$$

El grado promedio de entrada y de salida se calcula dividiendo estas sumas por el número de nodos n:

$$\bar{d}^{\text{in}} = \frac{1}{n}\sum_{j=1}^{n} d_j^{\text{in}}, \quad \bar{d}^{\text{out}} = \frac{1}{n}\sum_{i=1}^{n} d_i^{\text{out}}.$$

Dado que las sumas totales son iguales, se cumple que:

$$\bar{d}^{\text{in}} = \bar{d}^{\text{out}} = \frac{1}{n}\sum_{i,j} y_{i,j}.$$

Por definición de la media global $\bar{y}$, tenemos:

$$\bar{y} = \frac{1}{n(n-1)} \sum_{i,j:i\neq j} y_{i,j}.$$

Para redes dirigidas, el total de posibles relaciones es $n(n-1)$, y para redes no dirigidas es $\frac{n(n-1)}{2}$. En ambos casos, multiplicando por el denominador correcto:

$$\bar{y} = \frac{1}{n}\frac{1}{n-1}\sum_{i,j:i\neq j} y_{i,j}.$$

Esto coincide con el grado promedio (de entrada y salida), ya que las relaciones simétricas o dirigidas no afectan la equivalencia de la fórmula. Por lo tanto, en cualquier caso:

$$\bar{y} = \bar{d}^{\text{in}} = \bar{d}^{\text{out}}.$$

Además, como se demostró en el ítem **b)**, $\bar{y}$ corresponde también a la densidad de la red. Por lo tanto:

$$\bar{y} = \bar{d}^{\text{in}} = \bar{d}^{\text{out}} = \text{densidad}.$$

Punto 3:

Considere un grafo estrella de orden n y un grafo círculo de orden n. A continuación se representan ambos grafos para $n = 9$. Estos grafos tienen aproximadamente la misma densidad, pero su estructura es muy diferente. Recuerde que la densidad de un grafo se puede calcular como el grado promedio dividido por $n - 1$.

a) Muestre que para el grafo círculo de orden n el grado promedio es $\bar{d} = 2$.

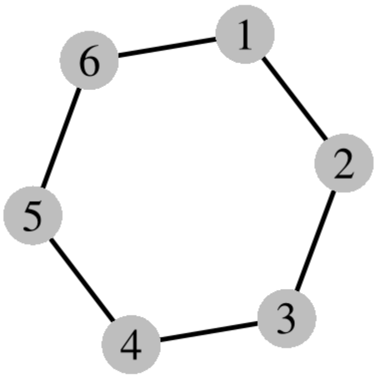


Figura 1: Grafo circular de 6 nodos.

Un grafo circular es un grafo particular del tipo de grafos regulares cuyo conteo de conexiones por nodo es 2. Un grafo de este tipo se verá como el de la figura 1 con más o menos nodos: Por lo cual podremos calcular el promedio de grado de los nodos así:

$$
\begin{aligned}
\bar{d} &= \frac{1}{n} \times \sum_{i=1}^{n} d_i \\
&= \frac{1}{n} \times \sum_{i=1}^{n} 2 \\
&= \frac{1}{n} \times n \times 2 \\
&= 2
\end{aligned}
$$

b) Muestre que para el grafo estrella de orden n el grado promedio es $\bar{d} = 2\frac{n-1}{n} \to 2$ cuando $n \to \infty$.

Respecto al grafo de tipo estrella, es posible definirlo como un grafo en el que un nodo está conectado a todos los demás $(n-1)$ mientras que los nodos restantes se conectan solamente al primero. De esta forma, el grado promedio

$$
\begin{aligned}
\bar{d} &= \frac{1}{n} \times \sum_{i=1}^{n} d_i \\
&= \frac{1}{n} \times \left[(n-1) + \sum_{i=2}^{n} 1 \right] \qquad d_1 = n-1, d_{i>1} = 1 \\
&= \frac{1}{n} \times [(n-1) + (n-1)] \\
&= \frac{2(n-1)}{n}
\end{aligned}
$$

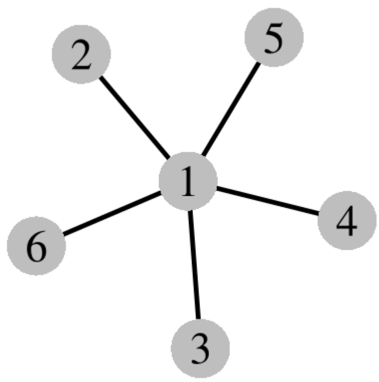


Figura 2: Grafo de estrella de 6 nodos.

Y finalmente, se cumple que $\underset{n\to\infty}{Lim} \left(2\frac{n-1}{n}\right) = 2 \underset{n\to\infty}{Lim} \left(\frac{n-1}{n}\right) = 2 \underset{n\to\infty}{Lim} \ 1 - \frac{1}{n} = 2$

Punto 4:

¿Cuáles de las siguientes secuencias son caminatas en el grafo que se presenta a continuación? ¿Cuáles senderos? ¿Cuáles circuitos? ¿Cuáles ciclos?

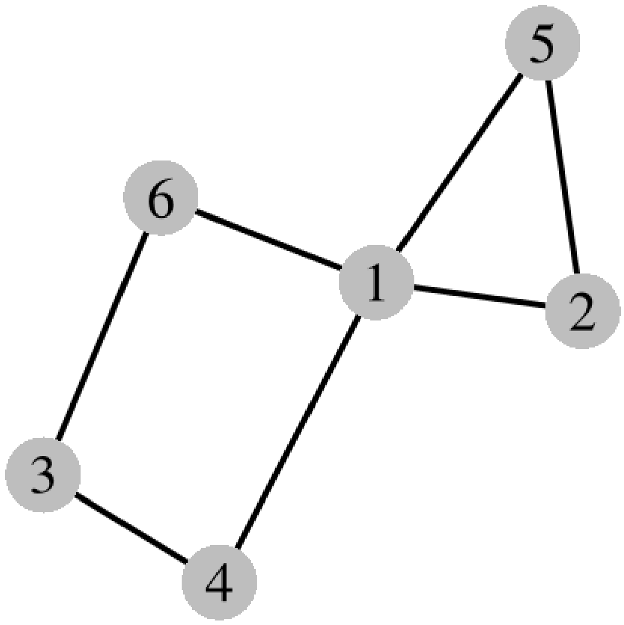


Figura 3: Grafo punto 4.

Teniendo en cuenta que los tres tipos de grafos se definen así:

- ⋆ **Sendero:** Una caminata cuyo primer y último vértice pueden ser distintos y en cuyo recorrido no se repiten ni vértices ni aristas.
- ⋆ **Circuito:** Una caminata a través del grafo que no repite aristas, pero en la que los vértices puedes ser repetidos.
- ⋆ **Ciclo:** Una caminata en que no se repite ni un vértice ni una arista y cuyo vértice inicial y final son los mismos.

Podemos decir de los siguientes subgrafos:

a) $\mathbf{2-1-6-3-4}$: Es un sendero, y un circuito pues no se repiten aristas ni vértices, pero no es ciclo pues no termina en el mismo vértice de salida.

b) $\mathbf{2-1-6-3-4-1-5}$: Solamente se clasifica como circuito pues no se repiten aristas, pero sí vértices (El uno).

c) $\mathbf{2-1-2-5-1-4}$: No es ninguno, debido a que se repiten aristas (1−−2) y vértices (1,2).

Punto 5:

Considere el conjunto de datos dado en `comtrade.RData` (este archivo contiene una arreglo de cuatro dimensiones denominado `comtrade`), asociado con el crecimiento anual del comercio (diferencia en dólares en escala logarítmica respecto al año 2000). Este conjunto de datos involucra 30 países, 10 años desde 1996 hasta 2005, y 6 clases de productos diferentes, como se muestra a continuación:

```
load("comtrade.RData")
dimnames(comtrade)[c(1,3,4)]

## [[1]]
##  [1] "Australia"          "Austria"           "Brazil"
##  [4] "Canada"             "China"             "China, Hong Kong SAR"
##  [7] "Czech Rep."         "Denmark"           "Finland"
## [10] "France"             "Germany"           "Greece"
## [13] "Indonesia"          "Ireland"           "Italy"
## [16] "Japan"              "Malaysia"          "Mexico"
## [19] "Netherlands"        "New Zealand"       "Norway"
## [22] "Rep. of Korea"      "Singapore"         "Spain"
## [25] "Sweden"             "Switzerland"       "Thailand"
## [28] "Turkey"             "United Kingdom"    "USA"
##
## [[2]]
## [1] "Chemicals"
## [2] "Crude materials, inedible, except fuels"
## [3] "Food and live animals"
## [4] "Machinery and transport equipment"
## [5] "Manufact goods classified chiefly by material"
## [6] "Miscellaneous manufactured articles"
##
## [[3]]
##  [1] "1996" "1997" "1998" "1999" "2000" "2001" "2002" "2003" "2004" "2005"
```

a) **Calcule el aumento medio global $\bar{y}$ a lo largo de los 10 años en bienes manufacturados. Para ello considere la matriz de adyacencia `Y` dada por:**

```
Y <- apply(X = comtrade[,,c(5,6),], MARGIN = c(1,2), FUN = mean)
```

Teniendo en cuenta que ya se proporciona la matriz de adyacencia, la media global es simplemente calculada con usar la fórmula y además teniendo en cuenta que la diagonal principal de la matriz contiene valores `NA` el cálculo se reduce a correr las siguientes dos líneas de código:

```
n = nrow(Y)
media = (1/(n * (n-1)) * sum(Y, na.rm = T))
```

Finalmente, obtenemos que la media global del aumento a lo largo de 10 años es de 0,03778362.

b) **Calcule la media de todas las observaciones de cada fila de `Y`, es decir, calcule la media fila $\bar{y}_{i\bullet} = \frac{1}{n-1}\sum_{j:j\neq i} y_{i,j}$ para cada país. Realice una histograma de los promedios fila $\bar{y}_{i\bullet}$. Los promedios fila caracterizan diferentes niveles de actividad de los nodos en términos de la sociabilidad. ¿Cómo se pueden interpretar los promedios fila $\bar{y}_{i\bullet}$?**

Para calcular la media global por fila, ya teniendo la matriz `Y` basta con correr las siguientes dos líneas de código:

```
mFila = apply(MARGIN=1, X = Y, FUN = function(x) sum(x, na.rm = T)/29)
```

y con estos resultados generamos el siguiente histograma:

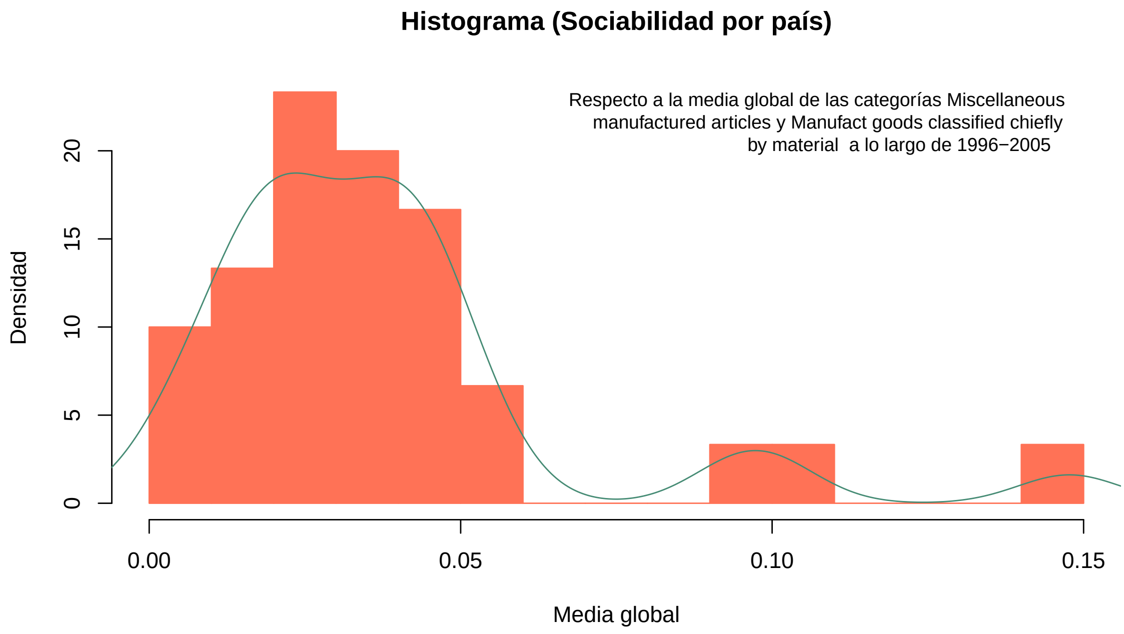


Figura 4: Histograma de sociabilidad.

Teniendo en cuenta que la sociabilidad de cada país en este contexto se refiere al crecimiento del comercio que los países emiten hacia el exterior, podemos observar que la media global de los últimos 10 años no refleja un cambio tan drástico aunque todos los países presentan crecimiento positivo en este rubro. Además, hay algunos países cuyo crecimiento en términos de comercio hacia el exterior podría considerarse como anormal respecto a la gran mayoría de países (Más adelante los identificaremos).

c) Calcule la media de todas las observaciones de cada columna de `Y`**, es decir, calcule la media columna** $\bar{y}_{\bullet j} = \frac{1}{n-1}\sum_{i:i\neq j} y_{i,j}$ **para cada país. Realice una histograma de los promedios columna** $\bar{y}_{\bullet j}$**. Los promedios columna caracterizan diferentes niveles de actividad de los nodos en términos de la popularidad. ¿Cómo se pueden interpretar los promedio columna** $\bar{y}_{\bullet j}$**?**

Utilizando una metodología similar a la del punto anterior tendremos que:

```
mCol = apply(MARGIN=2, X = Y, FUN = function(x) sum(x, na.rm = T)/29)
```

y con estos resultados generamos el siguiente histograma:

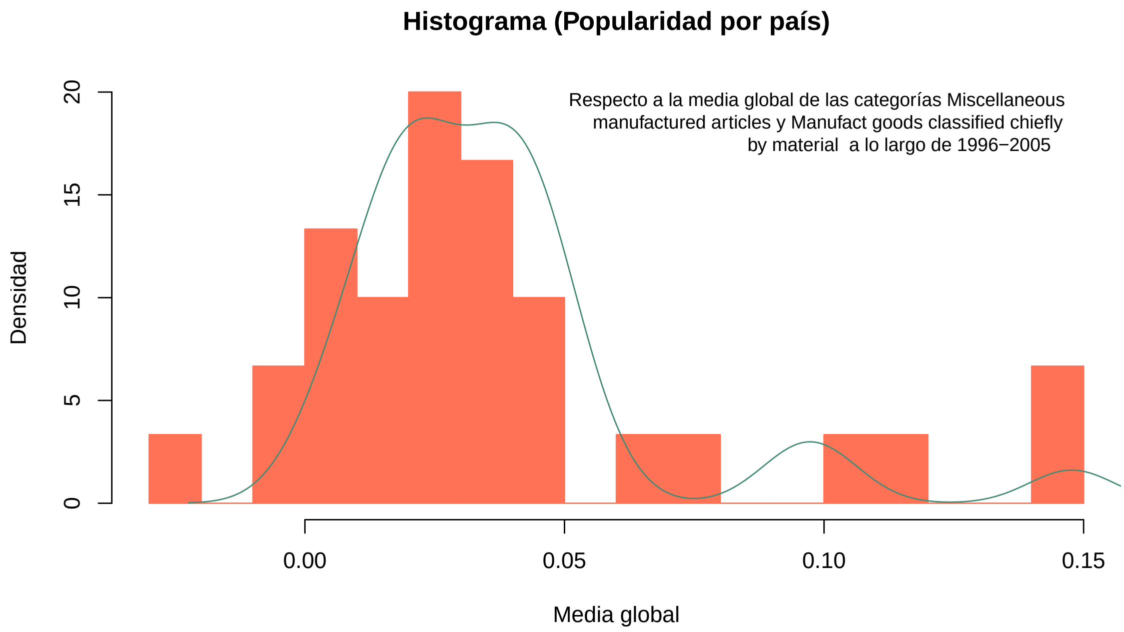


Figura 5: Histograma de popularidad.

El crecimiento del comercio desde el exterior presenta un comportamiento similar al del crecimiento del comercio hacia el exterior con la diferencia de que ahora se presentan incluso decrecimientos en algunos países, es decir, en promedio para algunos países el comercio desde el exterior a disminuido. Aún así, también se presentan los niveles anormales (También en niveles positivos) respecto a la tendencia de la mayoría de países.

d) **Calcule tanto la media de los promedios fila $\bar{y}_{i\bullet}$ como la media de los promedios columna $\bar{y}_{\bullet j}$. ¿Qué se puede concluir acerca de la tendencia local en este caso?**

e) **Calcule tanto la DE de los promedios fila $\bar{y}_{i\bullet}$ como la DE de los promedios columna $\bar{y}_{\bullet j}$. ¿Qué se puede concluir acerca de la heterogeneidad local en este caso?**

f) **Calcule el coeficiente de correlación entre los promedios fila $\bar{y}_{i\bullet}$ y los promedios columna $\bar{y}_{\bullet j}$. Realice un dispersograma de los promedios columna $\bar{y}_{\bullet j}$ (eje y) frente a los promedios fila $\bar{y}_{i\bullet}$ (eje x), junto con la recta $y = x$ como punto de referencia. ¿Qué se puede concluir?**

Sólo por mejor presentación se resuelven los 3 items anteriores simultáneamente corriendo las siguientes líneas de código de R:

```
mediaFila = mean(mFila)       # Medias
mediaCol = mean(mCol)
sdFila = sd(mFila)            # Desviaciones
sdCol = sd(mCol)
corrMedia = cov(mFila, mCol)/(sdFila * sdCol)
```

Obtenemos los siguientes resultados:

	Sociabilidad	Popularidad
Media	0.0378	0.0378
Desviación estándar	0.0302	0.0410
Correlación	0.7003	

Cuadro 1: Medidas de tendencia, dispersión y relación lineal.

En términos de aumento medio durante los últimos 10 años de todos los países en cuanto a su nivel de exportaciones e importaciones se ha presentado una mejoría que ha sido más dispersa en términos de exportaciones que en términos de importaciones. Sin embargo, ambos rubros presentan dispersiones altas (80 % y 110 % aproximadamente) y respecto a esto, los países que presentan un menor crecimiento en comercio hacia el exterior también parecen presentarlo en cuanto a comercio desde el exterior pues la correlación entre estas dos variables es alta.

Esta correlación es identificable gráficamente, pues en el gráfico a continuación se identifica una relación lineal importante entre los niveles demostrados por los países en las dos categorías.

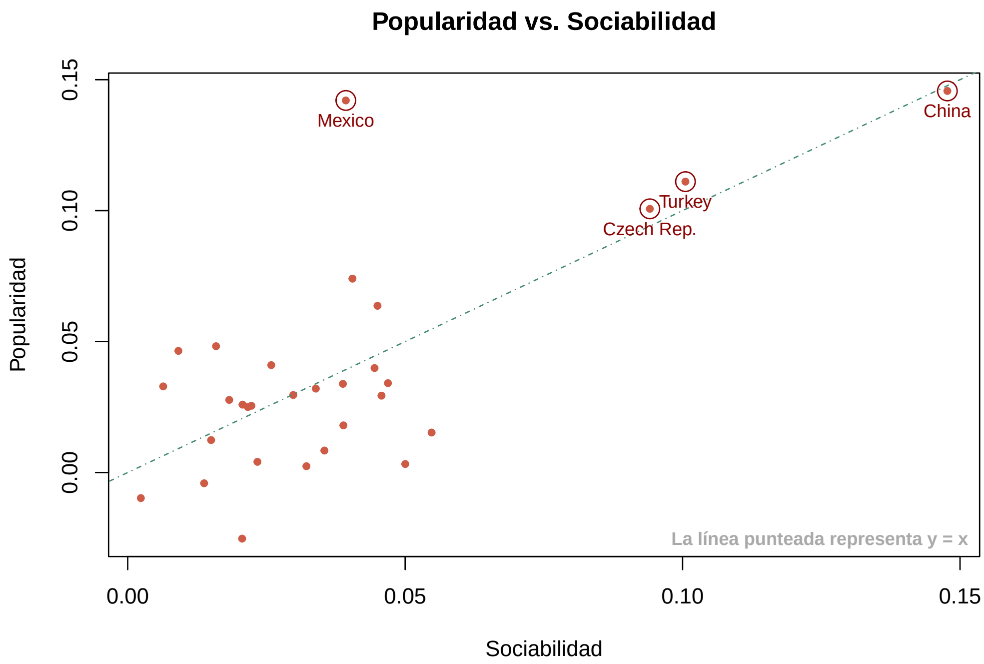


Figura 6: Dispersograma popularidad y sociabilidad.

En cuanto a los países con comportamiento anormal, se identifican a China, Turquía y la República Checa como países con crecimientos anormales tanto en el rubro de importaciones como de exportaciones. Sin embargo, incluso un poco más extraño es el caso de México pues aunque su aumento en nivel de comercio hacia el exterior está cercano al de la mayoría de los países, su aumento en comercio desde el exterior ha aumentado al nivel de los países mencionados antes.

Punto 6:

Considere el conjunto de datos dado en `conflict.RData` recopilado por Mike Ward y Xun Cao del departamento de Ciencias Políticas de la Universidad de Washington, asociado con datos de conflictos entre países en los años 90. El archivo `conflict.RData` contiene una lista con tres arreglos, `X`, `Y`, y `D`. `X` tiene tres campos: `population` (población en millones), `gdp` (PIB en millones de dolares), `polity` (puntuación política, un índice de democracia). `Y` hace referencia a una matriz $\mathbf{Y} = [y_{i,j}]$ en la que $y_{i,j}$ representa el número de conflictos iniciados por el país i hacia el país j. Finalmente, `D` es un arreglo de tres dimensiones dimensiones cuya tercera dimensión contiene indices entre cada par de países asociados con: comercio (dimensión 1), importaciones (dimensión 2), organizaciones intergubernamentales (dimensión 3), y distancia geográfica (dimensión 4).

a) **Hacer una visualización decorada de la red de conflictos teniendo en cuenta diferentes diseños.**

El grafo de forma general se ve de la siguiente forma:

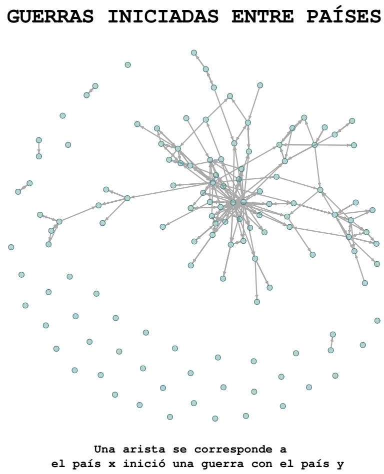


Figura 7: Red de conflictos entre países.

Para la decoración de la red se tuvieron en cuenta la popularidad y la sociabilidad de los nodos y no el grado total de los vértices debido a que el grado total nos da información mezclada: Un país puede tener un grado alto si es atacado frecuentemente o si ataca de forma frecuente. Además también se tuvieron en cuenta las variables nodales representadas en el archivo `dat$X`:

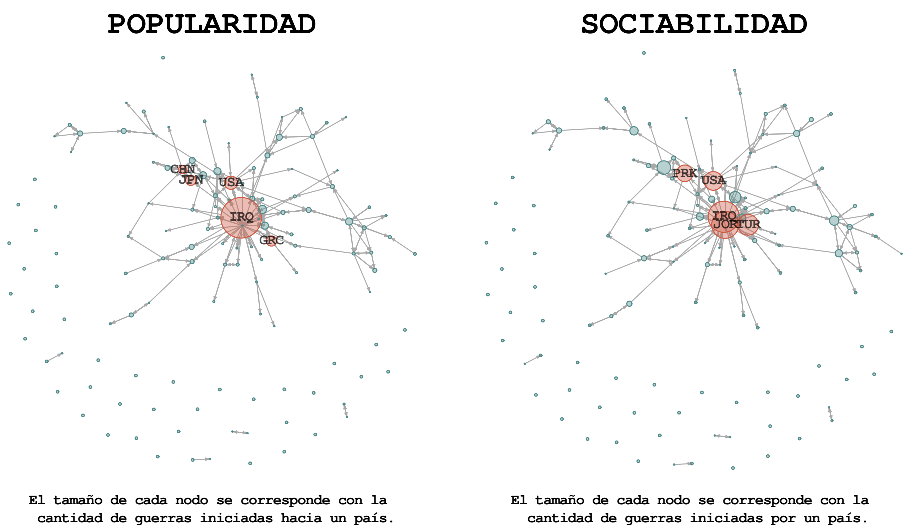


Figura 8: Sociabilidad vs popularidad

Los grados son lo suficientemente bajos para ser utilizados directamente para asignar de forma proporcinal el tamaño de los nodos, en cuanto a los tamaños relacionados con las variables nodales debido a la disparidad en variables como el PIB y la población se .^{es}tandarizaron" las variables dividiendolas por la observación de mayor valor para que las diferencias y en especial su escala sean notables en el gráfico. Finalmente, en cuanto a la variable de puntaje político, la variable también fue estandarizada respecto al mayor valor de la variable, pero como el puntaje puede tomar valores negativos se utilizó el valor absoluto de esta, sin embargo, los nodos de color verde son aquellos 5 países que tienen un mayor puntaje mientras que los nodos de color rojo se corresponden a los 5 países con menor puntaje en este indicador:

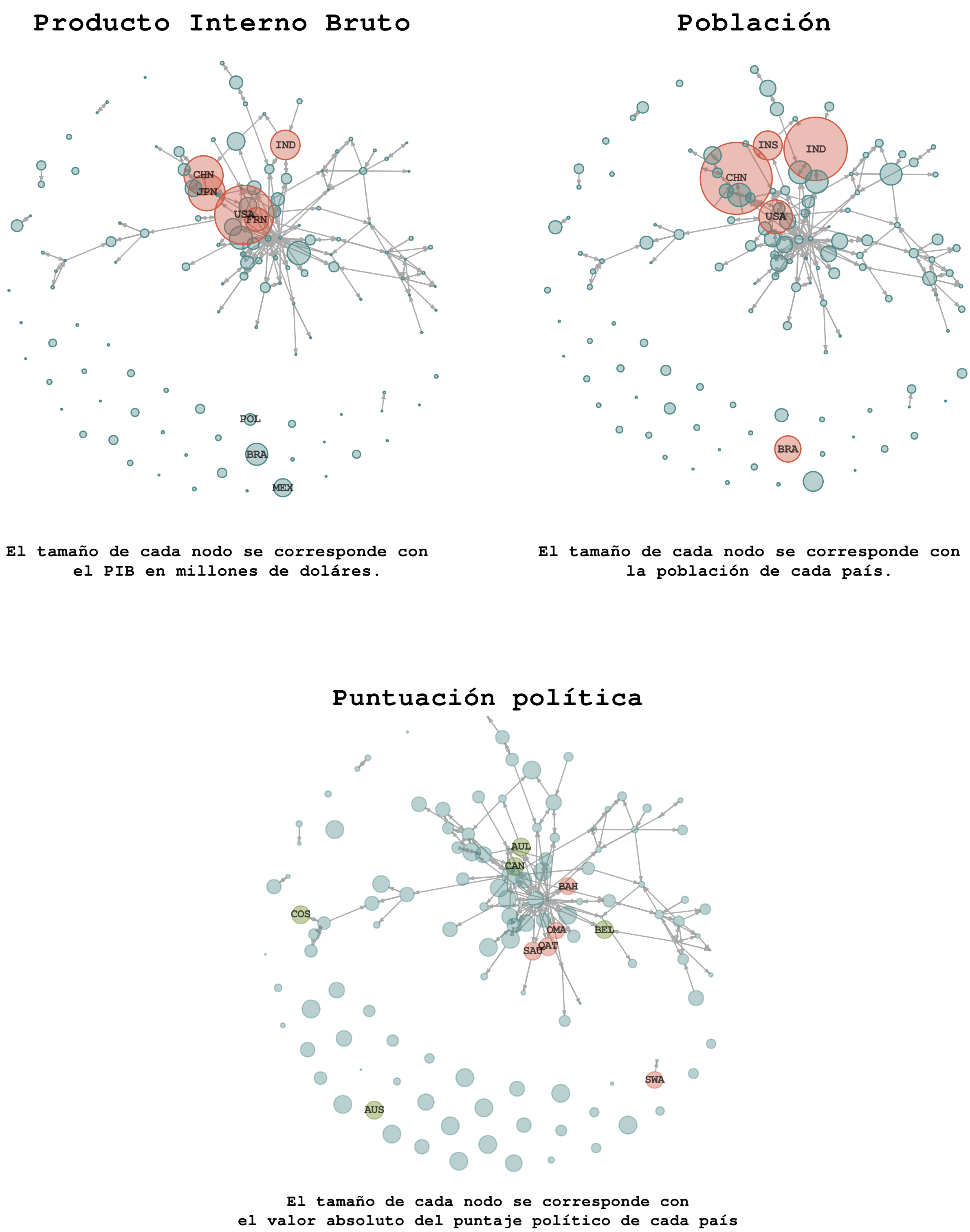


Figura 9: PIB, población y puntaje político.

b) Calcule la media global

Teniendo en cuenta que la instrucción de R que nos devuelve la diagonal de una matriz `diag(dat$Y)` nos devuelve un vector de ceros, no es necesario preocuoparnos por la no inclusión de la diagonal al calcular la media global de la matriz de adyacencia para la red de conflictos:

```
n = nrow(dat$Y)
mediaConflict = sum(dat$Y)/(n * (n - 1))
```

Obtenemos una media global de 0,0182 aproximadamente.

c) Obtenga y grafique la distribución del *out-degree* y del *in-degree*. Calcule e interprete la media y la desviación estándar de esta distribución.

Con el objeto de graficar las redes decoradas ya se calcularon los grados de salida y de entrada y se guardaron como atributos de los vértices del grafo g por lo que realizar lo pedido en este punto se reduce a aplicar `hist`, `sd` y `mean` sobre estos atributos. Obtenemos pues los siguientes histogramas:

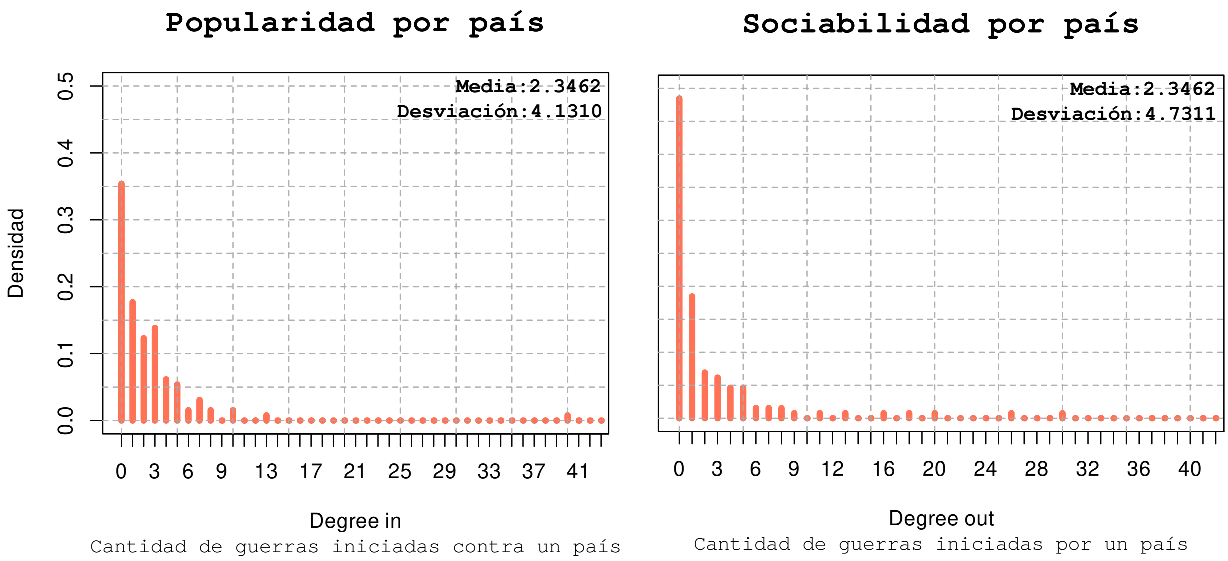


Figura 10: Histogramas grados de salida y de entrada

En términos de cantidad de conflictos entre países, en promedio existen 2 conflictos entre país y país. Con una diferencia mínima entre la cantidad de conflictos recibidos y la cantidad de conflictos iniciados es muy similar lo que no indica que sea baja (Tanto el grado de salida como de entrada tienen un coeficiente de variación de alrededor del 200 %). Esto debido a la presencia de valores atípicos para ciertos países como se verá más adelante.

d) Calcule el coeficiente de correlación entre los valores del *out-degree* y el *in-degree*. Realice un dispersograma de los grados de entrada (eje y) frente a los grados de salida (eje x), junto con la recta $y = x$ como punto de referencia. ¿Qué se puede concluir?

Siguiendo un método similar al utilizado en el punto 5 obtenemos el siguiente dispersograma y valor de correlación:

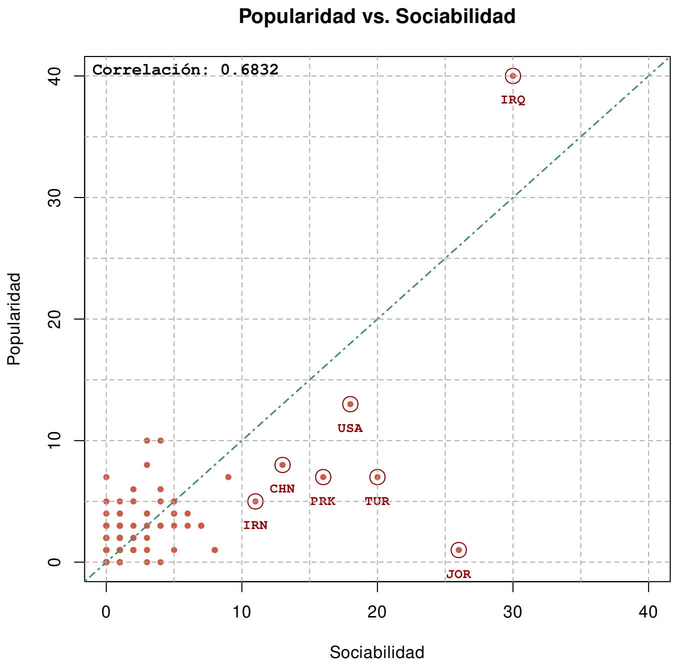


Figura 11: Dispersograma nivel de sociabilidad y popularidad guerra entre países.

El nivel de correlación probablemente es debido a países como Iraq, EEUU y China con valores extremos en comparación con la mayoría de los países respecto a guerras iniciadas y guerras recibidas presentan valores altos pero similares para ambas categorías además de que los países que no entran en guerra o entran en guerra muy pocas veces suelen no provocar guerras ni ser víctimas de otros países.

El caso de Jordania, es un caso atípico incluso para los valores de países como Iraq pues es un país que ha iniciado muchas guerras, pero con el que otros países casi nunca han iniciado guerras.

e) Identifique los países mas activos.

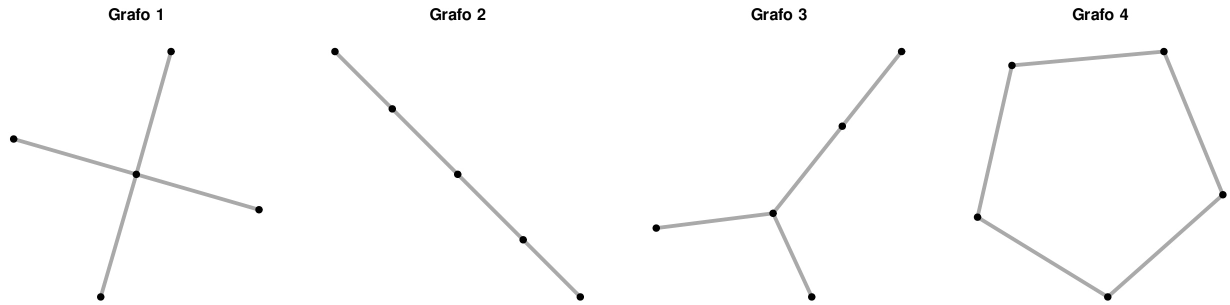
Top Sociabilidad			Top Popularidad		
	Sociabilidad	**Popularidad**		**Sociabilidad**	**Popularidad**
IRQ	30.00	40.00	IRQ	30.00	40.00
JOR	26.00	1.00	USA	18.00	13.00
TUR	20.00	7.00	GRC	4.00	10.00
USA	18.00	13.00	JPN	3.00	10.00
PRK	16.00	7.00	CHN	13.00	8.00
CHN	13.00	8.00	SYR	3.00	8.00
IRN	11.00	5.00	TUR	20.00	7.00
UGA	9.00	7.00	PRK	16.00	7.00
ISR	8.00	1.00	UGA	9.00	7.00
VEN	8.00	1.00	HAI	0.00	7.00

Cuadro 2: Top sociabilidad y popularidad.

El país más activo como víctima e iniciador de guerras es Iraq. Sin embargo, algunos otros países también exhiben comportamientos interesantes como por ejemplo Jordania que es un país que ha iniciado muchas guerras, pero que juega muy pocas veces de víctima o Estados Unidos que tiene una cantidad de guerras como iniciador y como víctima más o menos similar.

Punto 7:

Para todos los vértices de los cuatro grafos que se presentan a continuación, calcular el grado y las medidas de centralidad. Para cada grafo completar e interpretar la siguiente tabla. Interpretar los resultados.



En `igraph`, todas estas medidas están ya implementadas como `degree`, `closeness`, `betweness` y `eigen_centrality` para el grado y la centralidad por cercanía, por intermediación y propia respectivamente:

	Grado	Cercanía	Intermediación	Propia
Grafo 1	1.60	0.16	1.20	0.60
Grafo 2	1.60	0.13	2.00	0.75
Grafo 3	1.60	0.15	1.60	0.65
Grafo 4	2.00	0.17	1.00	1.00

Cuadro 3: Media del grado y las medidas de centralidad

La media del grado no es una buena representación de la estructura de un grafo pues grafos muy distintos pueden dar en la media de los grados resultados muy parecidos o inclusos iguales. Respecto a las medidas de las medidas por centralidad, todas las medidas presentan diferencias en los grafos respectivamente a lo que representa cada una de las medidas: La centralidad por cercanía muestra como con más cercanía media a los grafos en los que menos pasos sean necesarios para conectar cualquier par de nodos, la media de la centralidad por intermediación nos muestra los grafos más propensos a desconectarse siendo el grafo 2 el grafo con mayor media por obvias razones y finalmente la media de la centralidad propia nos permite ver que grafos tienen nodos que se conectan a su vez con nodos más importantes siendo el de mejor media el grafo 4 por ser completamente conectado.

	Grado	Cercanía	Intermediación	Propia
Grafo 1	1.34	0.05	2.68	0.22
Grafo 2	0.55	0.03	1.87	0.23
Grafo 3	0.89	0.04	2.30	0.23
Grafo 4	0.00	0.00	0.00	0.00

Cuadro 4: Desviación estándar del grado y las medidas de centralidad

La combinación de la media y el grado en cambio sí nos puede dar muy buenas pistas sobre lo que pasa al interior de la estructura de una red, en este caso el grafo 1 en promedio tiene 1.6 conexiones por nodo, pero de las conexiones dentro de esos nodos tenemos una variabilidad de casi el 85 % lo cual nos dice que unos nodos tendrán muchas máx conexiones que otros y lo mismo ocurre con el grafo 3. En cuanto a la dispersión de las medidas de centralidad las medidas de cercanía y propia parecen ser las menos dispersas de las 4 por lo que probablemente puedan ser más confiables para casos en que la dispersión afecte las demás medidas y la intermediación parece ser la medida más dispersa de las 4 debido a que hay nodos con una sola conexión muy seguramente.

Punto 8:

Considere los datos relacionales acerca de los conflictos internacionales del archivo `conflict.RData` despúes de simetrizarla débilmente y remover los nodos aislados:

a) **Hacer una visualización decorada de la red.**

Teniendo en cuenta que en el punto 6 ya se realizaron algunas visualizaciones decoradas básicas utilizando algo de información nodal y que al simplificar la red lo que representa la red cambia ligeramente a una red de si alguna vez algún país estuvo en guerra con algún otro la siguiente gráfica divide a los nodos en tres grupos de acuerdo al índice de democracia y divide las aristas en si el enlace se dio entre un país con calificación positiva, negativa o cero.

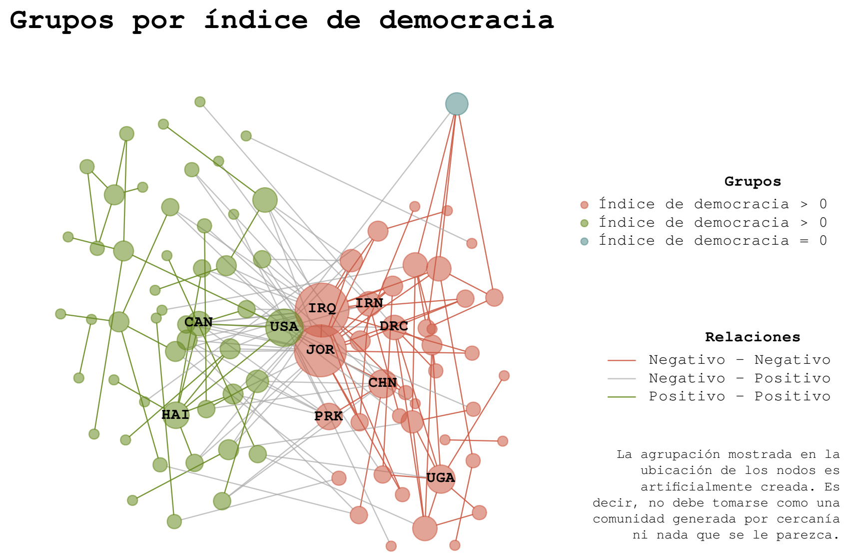


Figura 12: Países que han estado en guerra.

La agrupación vista en la red no corresponde a ninguna comunidad propia de la red sino que en cambio la red fue organizada de esta forma para poder revisar si hay alguna diferencia entre la cantidad de guerras entre países con índices de democracia positivos y negativos. Efectivamente, hay una diferencia importante entre las guerras que se generan en países con índices positivos y negativos. Incluso sin revisar los porcentajes, basta con ver que de los 10 países con más historial de guerras (Los que tienen nombres), 7 tienen calificación negativa en el índice de democracia.

Tipo de relación	Porcentaje
Negativo - Negativo	36.875
Positivo - Negativo	42.5
Positivo - Positivo	20.625

Cuadro 5: Porcentaje de enlaces entre grupos por índice de democracia.

Efectivamente, parece ser que la calificación en la democracia tuviese algo que ver con la posibilidad de entrar en guerra entre países.

b) **Caracterizar local y estructuralmente la red, en términos de la distancia, la centralidad, la cohesión, la conectividad y el agrupamiento. Utilizar todas las métricas disponibles.**

c) **Interpretar los resultados.**

En términos de **cliques**, la red muestra una estructura muy fragmentada. La distribución de la frecuencia de los tamaños de cliques se detalla a continuación:

1	2	3	4
130	160	67	9

Cuadro 6: Frecuencia de tamaño de cliques

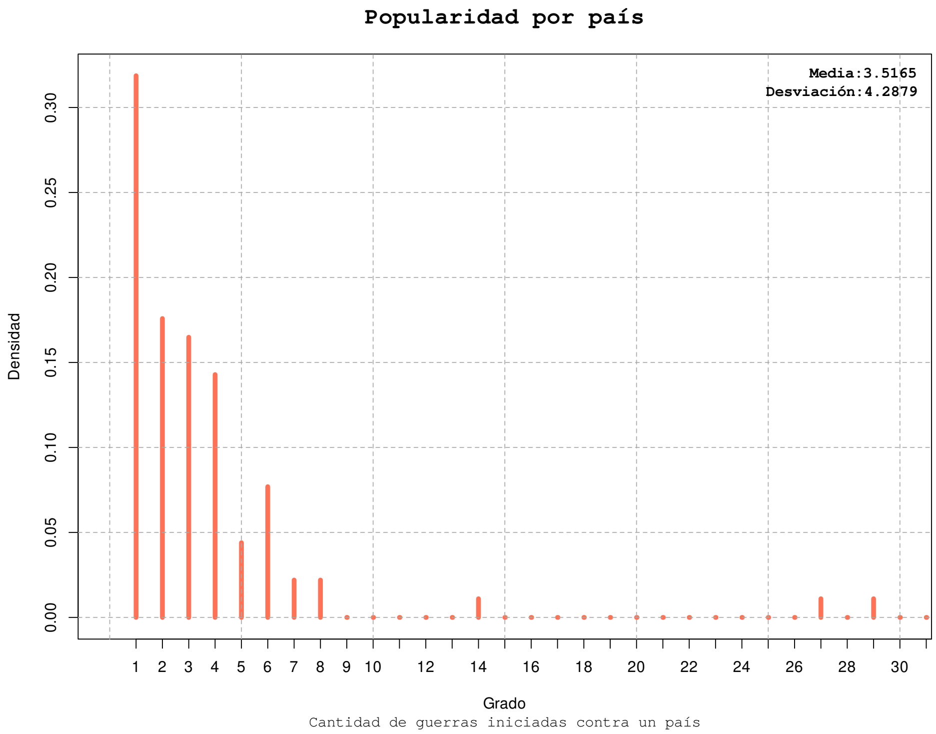


Figura 13: Distribución del grado de los nodos no desconectados.

Esto significa que, a lo sumo, se encuentran grupos de 4 países conectados entre sí en temas relacionados con conflictos bélicos a lo largo de la historia. Esta característica puede deberse a limitaciones geográficas o a intereses económicos, ya que los países de interés estratégico para otros suelen mantenerse constantes a lo largo del tiempo.

En cuanto a los **clanes maximales**, se identificaron alrededor de 98 en una red compuesta por 91 países. Esto refuerza la idea de que la red está altamente desconectada en términos de cliques.

En lo que respecta a otras medidas estructurales, los resultados se resumen en la siguiente tabla:

Medida	Valor
Densidad	0.0391
Transitividad global	0.16341
Compente gigante	83
Puntos de articulación	1

Cuadro 7: Medidas estructurales

En general, la red es poco densa, lo que indica que es la excepción, y no la regla, que los países se involucren en conflictos entre sí. La transitividad global, aunque baja (0.1633), sugiere que existen algunas agrupaciones locales dentro de la red. Finalmente, la componente gigante abarca 83 nodos, lo que muestra que, a pesar de su fragmentación, la mayoría de los países forman parte de una gran subred conectada que es propensa a desconexión si se elimina tan sólo un vértice.

Finalmente, respecto a la caracterización local de la red, tenemos los siguientes gráficos:

El grado es sesgada a la derecha con algunos países alejados de la mayoría de países. Como se vió anteriormente, estos países deberían corresponder a los países de Iraq y demás, pero como en este caso el grado representa haber entrado en conflicto alguna vez con otro país el top 5 se modifica por el caso de Jordania:

	Grado
IRQ	29.00
JOR	27.00
USA	14.00
CHN	8.00
UGA	8.00

Cuadro 8: Top 5 países que han participado en más conflictos internacionales.

En el análisis específico de las guerras producidas entre países no debería ignorarse la división que teníamos antes de simplificar la red pues como se puede observar, perdimos información muy importante con esta simplificación. Por otro lado, como solamente quedaron los nodos conectados, podemos decir que de los países que alguna vez han protagonizado una guerra la mayoría de ellos a lo sumo han participado de 4.

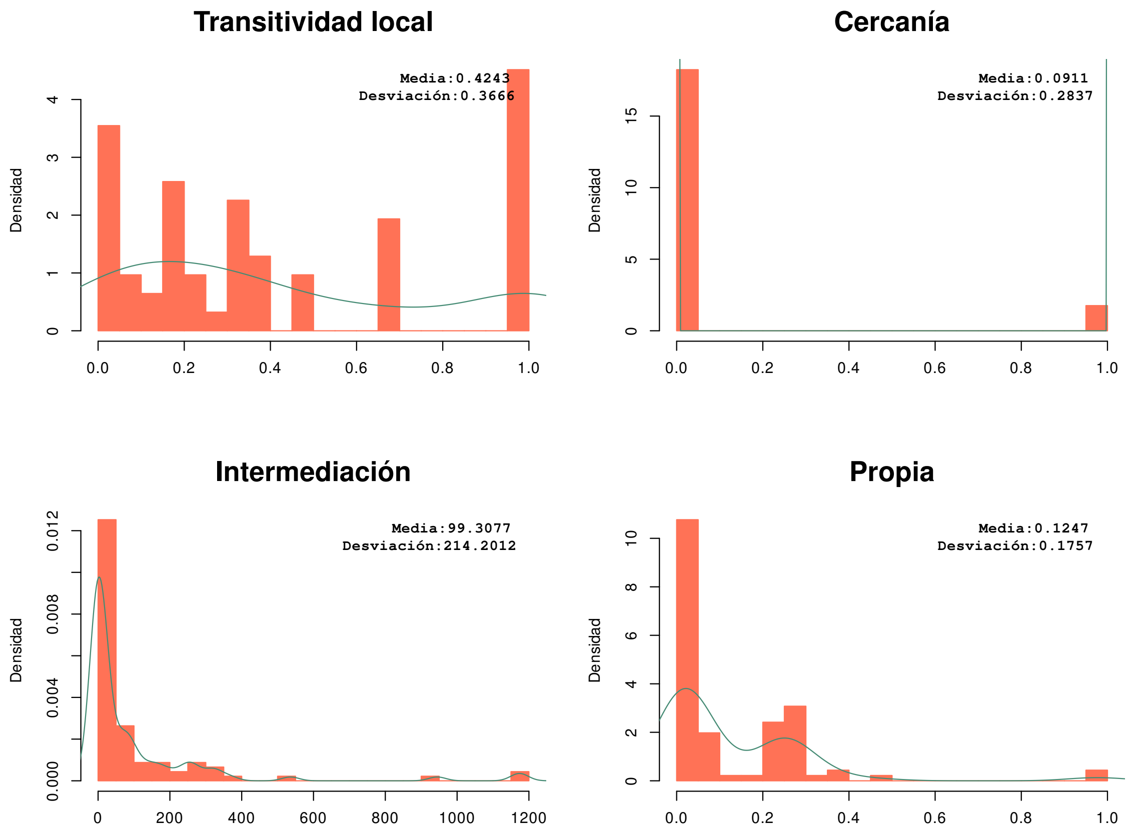


Figura 14: Caracterización local de la red de conflictos internacionales

En términos de todas las medidas tomadas, existe un país que es demasiado importante en el grafo (Que ha tenido conflictos con demasiados países) mientras que en general las medidas se mantienen un poco más bajas para los demás países lo que causa que la dispersión de las medidas se dispare llegando incluso a arruinar completamente la estimación de la densidad para la medida de cercanía (Línea azul). La más homogénea en cuanto a dispersión de todas las medidas parece ser la medida de transitividad local (Muchos países se encuentran en triplas cerradas) mientras que no muchos países parece estar entre dos países importantes por o que la medida de intermediación es predominantemente baja.