Alumna: Paula Brück

Padron: 107533

TEORIA DE ALGORITMOS II (75.30)

1.er Parcial Domiciliario

PUNTO 1:

- Teniendo en cuenta la red a analizar , la cual representa las conexiones de diferentes países a través de los vuelos directos que se realizan entre ellos , podemos calcular :
    - ***El diametro de la red:***
        - El diámetro es el largo máximo de todos los caminos mínimos.
        - Las redes tienden a tener un diámetro muy pequeño.
        - En redes grandes, se puede determinar con el algoritmo de BFS (Breadth-first search).
        - Las redes más dispersas suelen tener un diámetro mayor que aquellas que son más densas ya que existen menos caminos entre cada par de nodos.
        - Equivale al valor máximo de excentricidad para todos los nodos de la red.

$$E(i) = \max_{j \in V(G)/i} d(i, j) \qquad d_{\max} = \max \{E(i) : i \in V(G)\}$$

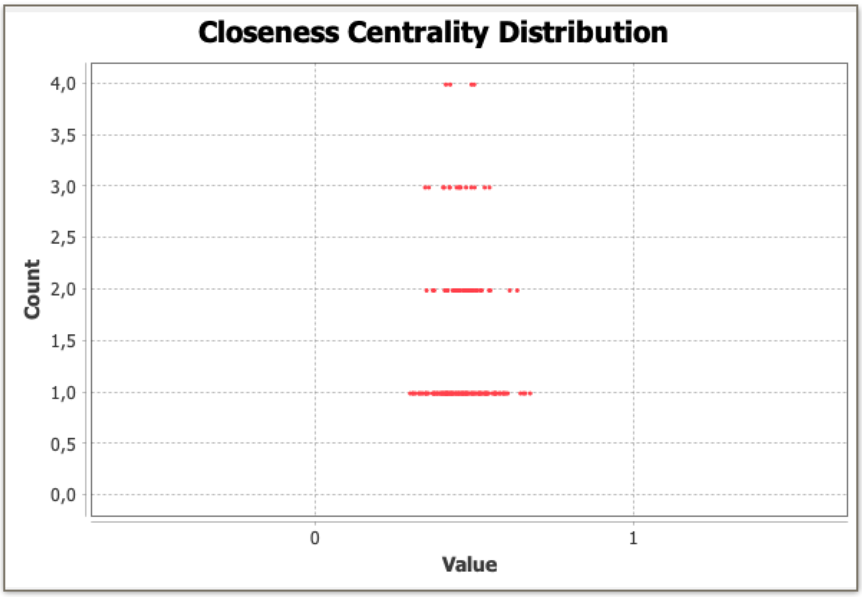
- En el contexto del análisis de Redes Sociales (SNA) esta métrica da una idea de la proximidad entre pares de actores en la red, indicando cómo de lejos están en el peor de los casos.

- Considerando todo esto y analizando nuestra Red de vuelos podemos decir que:

Alumna: Paula Brück

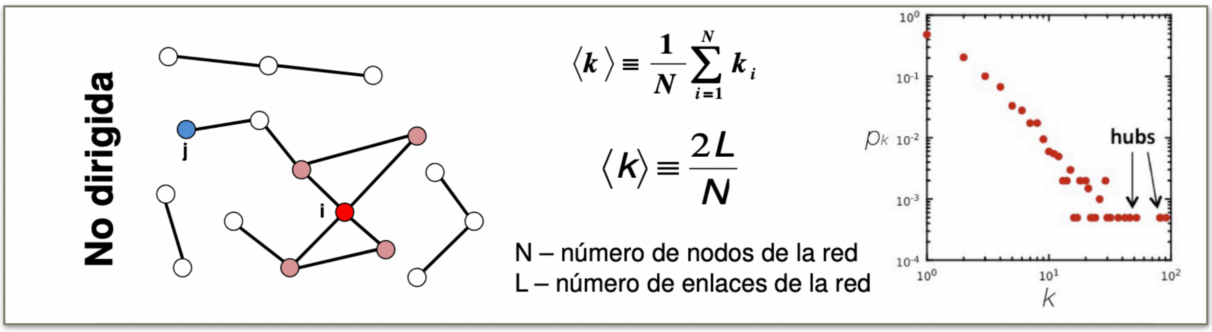
Padron: 107533

- El diámetro de la red es de 5. Este valor representa la máxima distancia existente entre dos nodos en toda la red siendo la distancia media 2,5.
  - El histograma de distancias es el siguiente:



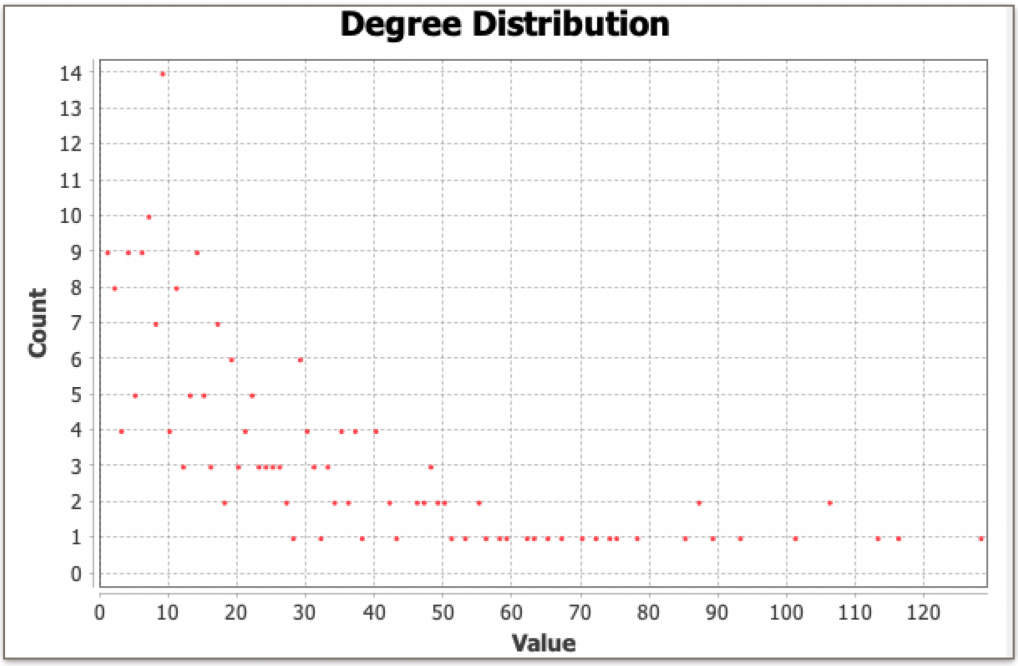
- El grado promedio de la red:

- También conocido como el Grado de un vértice, hace referencia a el numero de aristas de un nodo.



- En esta Red el grado Promedio es 24,908 lo cual quiere decir que cada nodo de la red está conectado con otros 24 en media y las gráficas de densidades de los grados son las siguientes:

- ***Grado promedio:***



- ***El coeficiente de clustering promedio de la red:***

- El coeficiente de Clustering (agrupamiento) de un vértice en un grafo cuantifica que tanto está de agrupado con sus vecinos. Si el vértice está agrupado como un clique, entonces su valor es máximo, mientras que un valor pequeño indica un vértice poco agrupado en la red.

$$C_i = \frac{2e_i}{k_i(k_i-1)} \qquad C = \frac{1}{N}\sum C_i$$

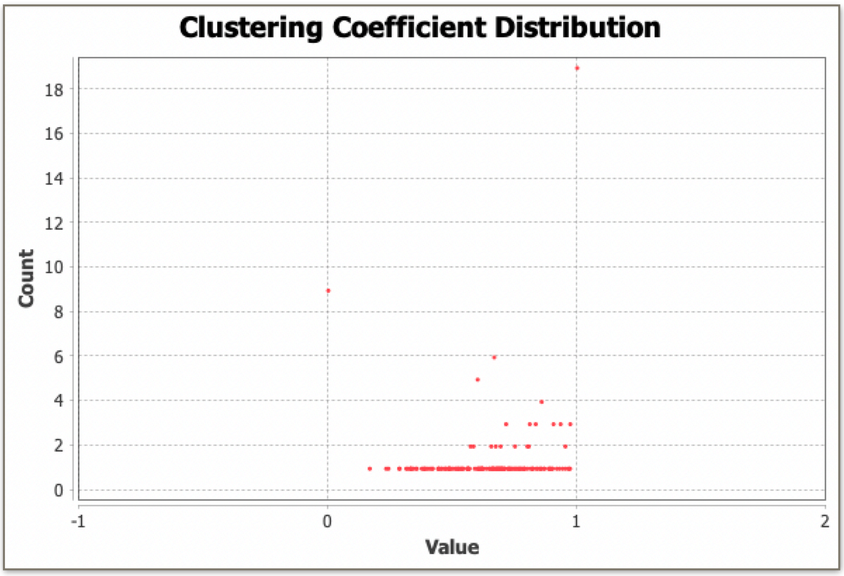
- e_i vendría a ser la cantidad de aristas que unen vecinos de i.

- En el caso de nuestra Red al representar los vuelos directos entre países y por lo tanto al ser no dirigida el promedio del coeficiente de

Alumna: Paula Brück

Padron: 107533

Clustering , es decir el promedio de los coeficientes individuales para cada vértice, es de 0,66.



PUNTO 2:

- Homofilia:

- Es conocida como la tendencia de las personas por la atracción a sus homónimos.

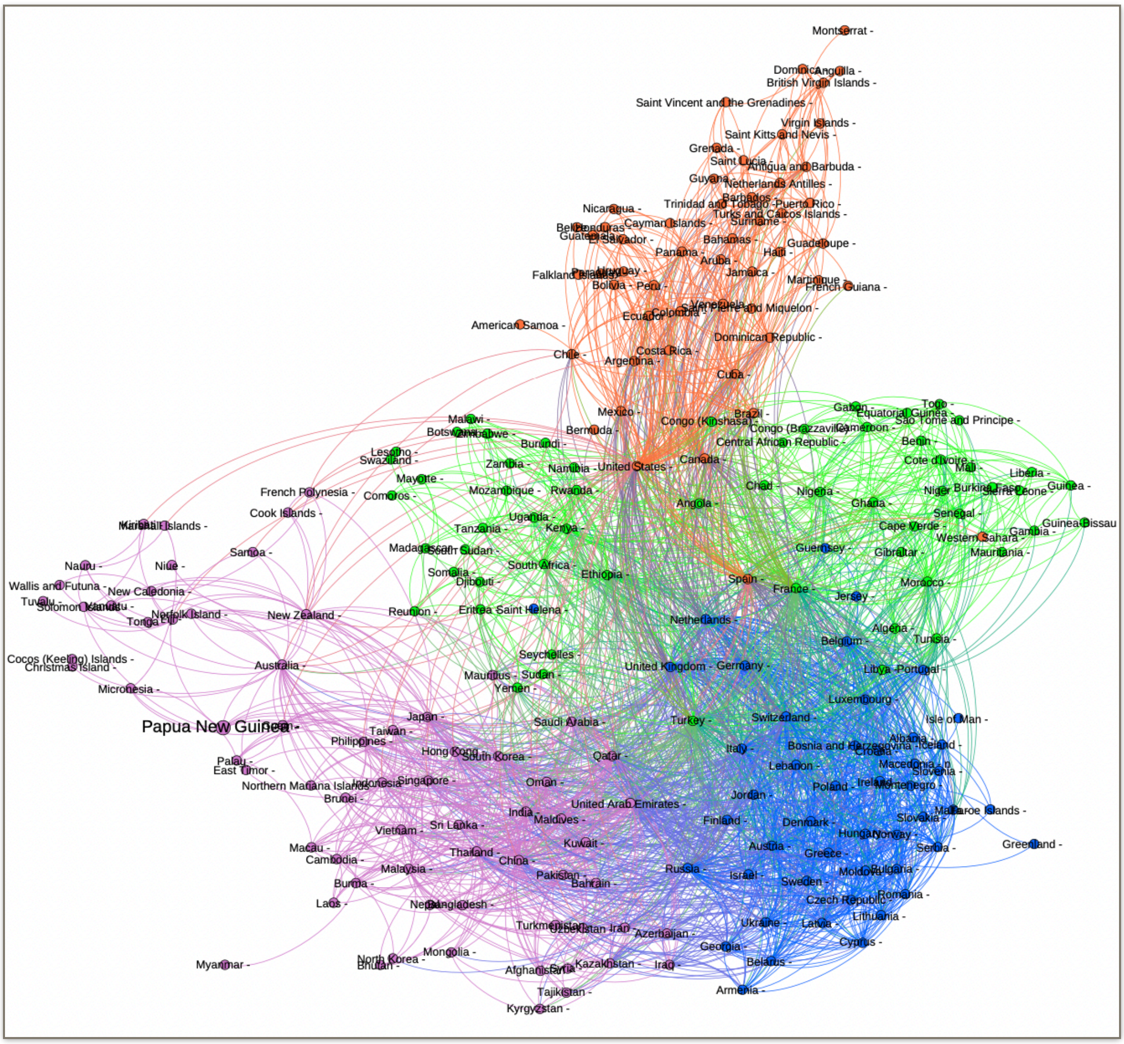
- La similitud puede ser respecto a diferentes atributos como edad, sexo, creencias, educación, estrato social, profesión , lugar donde viven , opiniones.

- Nos explica otra forma en la que los vínculos se forman.

- En el caso de esta red de vuelos podemos analizar la homofilia por continentes , es decir el espacio geográfico al que pertenecen los países que se conectan mediante los vuelos directos.
- Nuestra red de vuelos cuenta con 229 Nodos (países) y con 2852 aristas (vuelos directos).
- Para analizar esto podemos considerar en esta red 4 comunidades distintas :

Alumna: Paula Brück

Padron: 107533

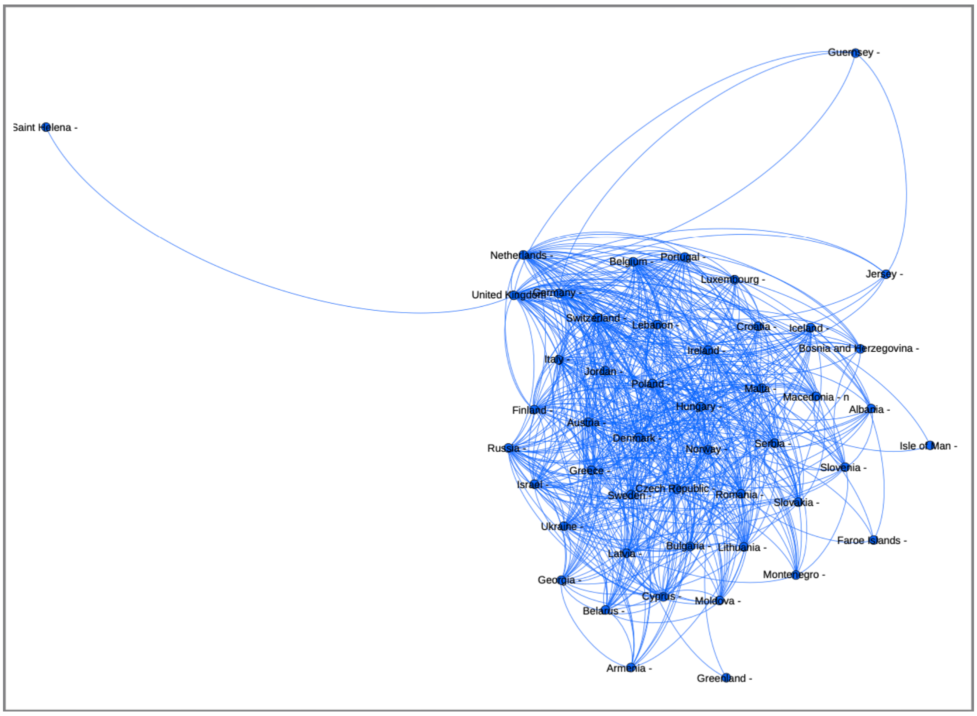


- Paises de Europa (Grupo 0)
- Paises de America (Grupo 1)
- Paises de Asia (Grupo 2)
- Paises de Africa (Grupo 3)

- Analizando los vuelos podemos ver que en el mayor de los casos los vuelos directos de un país se producen a países que pertenecen al mismo continente. Es decir que habría homofilia por ubicación geográfica , los países de un mismo continente (mas cercanos) tendrían más vuelos directos entre ellos. Para demostrar esto tenemos que considerar la probabilidad de que dos nodos conectados por una arista sean de distinto grupo va a ser : p_i . (1- p_i) . Entonces sí queremos ver la sumatoria total tenemos que hacer : sum p_i . (1 - p_i) .

- Haciendo este calculo podremos obtener el valor ideal/teórico sin homofilia . Luego tomando en consideración: la cantidad de aristas que se cruzan de grupo / la cantidad de aristas totales de nuestra red de vuelos podremos calcular el porcentaje real de sin homofilia el cual nos ayudara a determinar si por ubicación geográfica existe realmente homofilia.
- Para esto debemos analizar los cruces de aristas que se producen de un grupo a otro grupo distinto.

Grupo 0 :



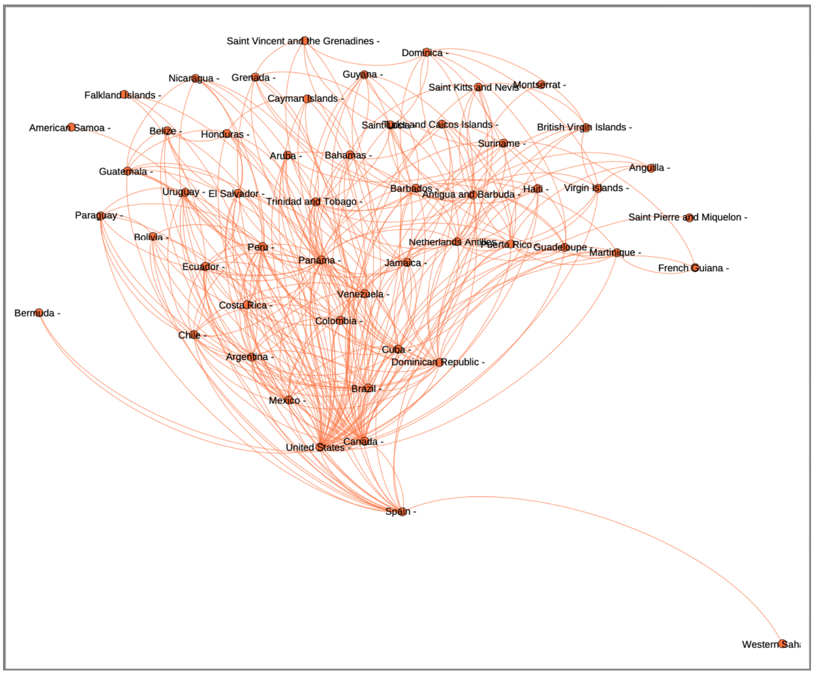
Nodos: 51

Cantidad de aristas que cruzan a otro grupo: 510

Grupo 1:

Alumna: Paula Brück

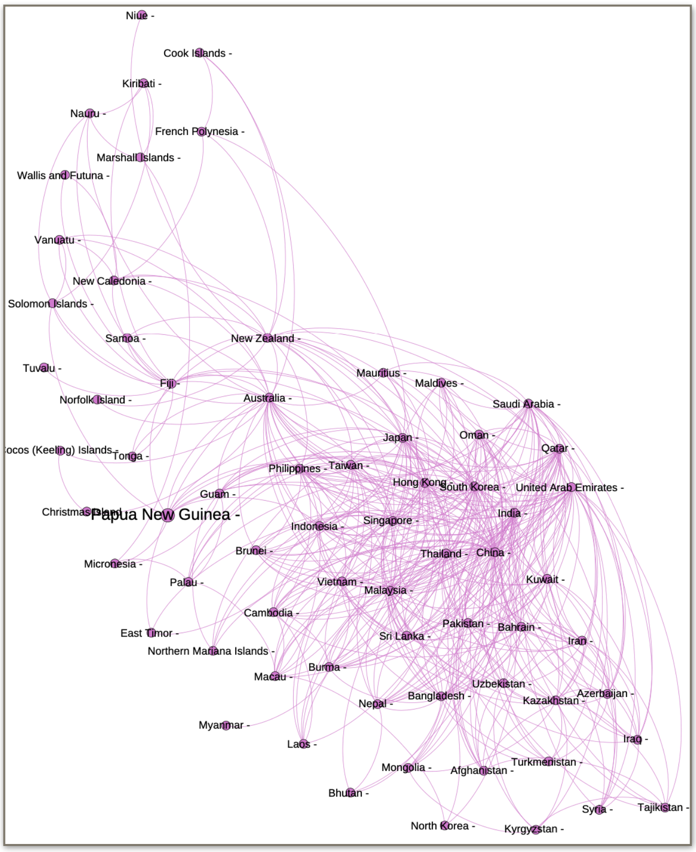
Padron: 107533



Nodos: 52

Cantidad de aristas q cruzan: 95

Grupo 2 :



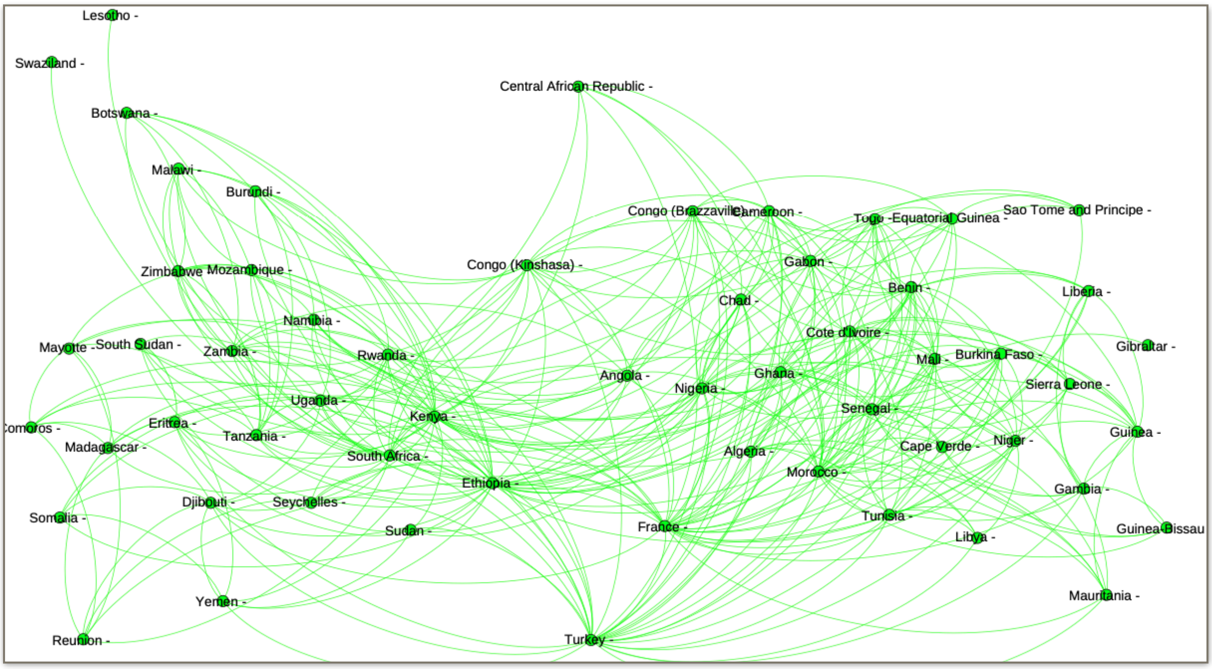
Alumna: Paula Brück

Padron: 107533

Nodos: 66

Cantidad de aristas q cruzan: 126

Grupo 3:



Nodos: 60

Cantidad de aristas q cruzan: 264

- Luego de calcular las aristas que cruzan de un mismo grupo a otro distinto sumamos cada uno de los resultados para así obtener el total de aristas que cruzan en nuestra red.

Aristas Totales que cruzan:

510 ———> Grupo 0

\+ 95 ———> Grupo 1

126 ———> Grupo 2

264 ———> Grupo 3

---

995 ———> Total

Alumna: Paula Brück

Padron: 107533

- Luego calculamos el valor ideal/ teórico de sin homofilia de la siguiente manera:

$P = p0\ (1- p0) + p1\ (1- p1) + p2\ (1- p2) + p3\ (1- p3)$

- El calculo de p_i es el resultado de la cantidad de nodos de el grupo i / la cantidad de nodos totales de la red.

$p0 = n0\ /\ nT$, $p1 = n1\ /\ nT$, $p2 = n2\ /\ nT$, $p3 = n3\ /\ nT$

$p0 = 51/229 = 0{,}22$, $p1 = 52/229 = 0{,}23$, $p2= 66/229 = 0{,}29$,

$p3= 60/229 = 0{,}26$

- Reemplazando p0, p1, p2 y p3 en P:

$P = 51/229(1-51/229) + 52/229(1-52/229) + 66/229(1- 66/229) + 60/229(1-60/229)$

$P = 0{,}17 + 0{,}18 + 0{,}21+ 0{,}19$

P = 0,75 ———> Ideal sin homofilia (threshold teórico)

- Para calcular el valor real de sin homofilia presente en nuestra red tenemos que hacer:

$$\frac{\text{Aristas Totales que cruzan}}{\text{Aristas Totales}} = \frac{995}{2852} = \mathbf{0{,}35}$$ ———> Realidad de sin homofilia

- Entonces teniendo el valor ideal/ teórico y el valor real de sin homofilia calculamos el porcentaje de sin homofilia:

0,75 ———> 100%

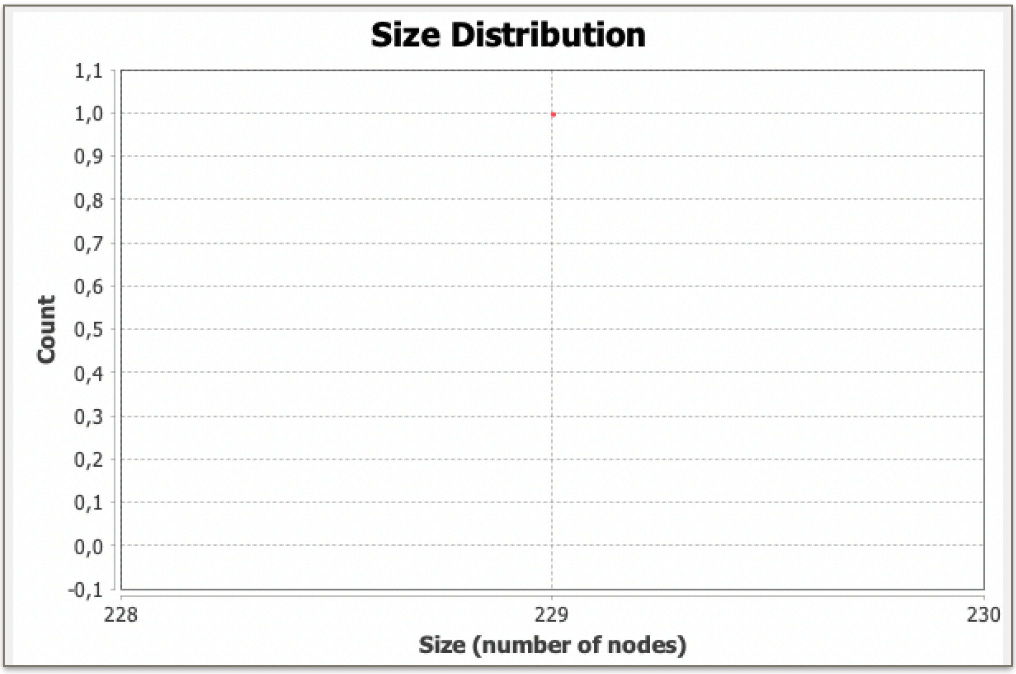
0,35 ———> X = **46,6 %** sin homofilia

- Es decir que si hay un 46,6% sin homofilia el restante **53,4%** es el porcentaje de la homofilia presente según la ubicación geográfica de nuestra red de vuelos.
- Por lo tanto podemos decir como conclusión que sí se cumple que la mayoría de los países tienen sus vuelos directos a países cercanos que se encuentran en el mismo continente y que por esta característica hay homofilia .

PUNTO 3:

- Puentes Locales y Globales:

- **Puentes Globales:**
- Es aquella arista que sin la cual el grafo no es conexo, y suele ser improbable de encontrar en una red social . En esta red social no hay puentes globales ya que nuestra red es una componente conectada.
- **Puentes Locales:**
- Es dado por una arista que une dos vértices sin un adyacente en común, es decir que no forman un triangulo como con Tradic Closure (Si dos personas tienen un amigo en común, hay una importante probabilidad de que eventualmente se vuelvan amigos/conocidos)



Alumna: Paula Brück

Padron: 107533

- Los puentes locales que hay en esta red son 11:

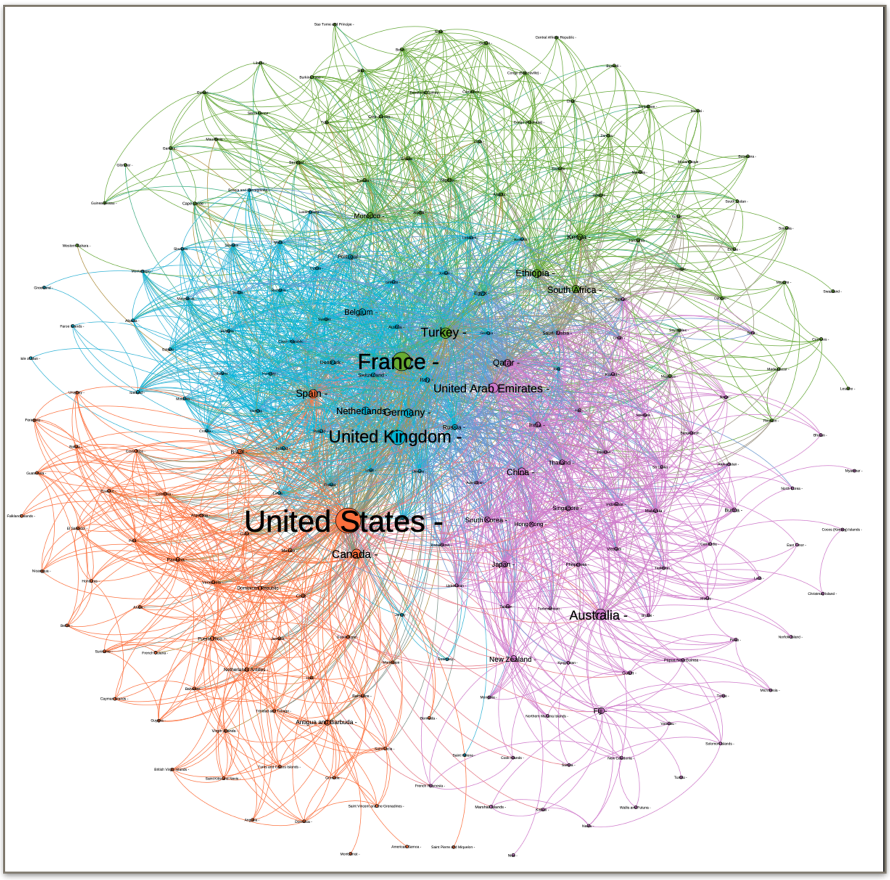
```
Fiji<———————> Tuvalu
Netherlands<———————> Antilles
United States<———————>American Samoa
United Kingdom<———————>Saint Helena
Canada<———————>Saint Pierre and Miquelon
Antigua and Barbuda<———————>Montserrat
New Zealand<———————>Niue
South Africa<———————>Lesotho
South Africa<———————>Swaziland
Virgin Islands<———————>British
Burma<———————>Myanmar
```

PUNTO 4:

- ***Tipo de Centralidad:***
- Existen varias medidas distintas de centralidad como por ejemplo:
  - Grado
  - Intermediacion(betweenness)
  - Cercania(closeness) y armonica (harmonic)
  - Excentricidad
  - Centralidad de vector propio
- Para esta red ya que representa los vuelos directos entre países me parece más útil calcular la centralidad de Intermediación para poder determinar influencia. Si buscamos cual país es mas importante , influyente podremos decir que será cual tenga mayor conexiones con otros ya que esto significara que una mayor cantidad de países se encuentran conectados a través de vuelos directos entre ellos y esto suele estar dado por la demanda de personas que requieren o desean viajar a dicho país o desde dicho país . Cuanto más conocido, importante, mayor nivel económico más razones hay para que diferentes países

requieran contar con vuelos a estos ya que una mayor cantidad de personas se movilizan a dicho país.

- La centralidad de Intermediacion es una forma de detectar la cantidad de influencia que tiene un nodo sobre el flujo de información en un gráfico.
- El algoritmo calcula las rutas más cortas no ponderadas entre todos los pares de nodos en un gráfico. Cada nodo recibe una puntuación, según el número de caminos más cortos que pasan por el nodo. Los nodos que se encuentran con mayor frecuencia en las rutas más cortas entre otros nodos tendrán puntuaciones de centralidad de intermediación más altas.
- Es por esto que me parece una buena opción para esta red de vuelos.
- ***Grafico de Centralidad de Intermediacion:***



- Analizando el gráfico podemos ver claramente cómo los países como Estados Unidos ,Francia , Reino Unido, Turquía son algunos de los mas influyentes . Al ser el gráfico de la centralidad de intermediación se esta teniendo en cuenta todas las veces que un país aparece en la ruta de caminos mas cortos de otro país. Es decir que en este gráfico el ranking de influencia va a estar dado por aquellos países que tienen mayor conexión con otros .

- ***PUNTO 5:***
- ***Simulación Modelado Erdös-Rényi:***
  - El modelo aleatorio de Erdös-Rényi es un método usado para la generación de grafos aleatorios en el cual se tiene un grafo de n vértices y cada arista (u,v) aparece aleatoriamente con una probabilidad p.
  - Utilizando los parámetros de esta red obtenemos una simulación con Erdös-Rényi:
  - Nuestro numero de nodos en este caso es de n= 229.
  - Para calcular la distribución de grados de que un nodo tenga k conexiones en la red aleatoria generada con el modelo Erdös–Rényi, primero tenemos que calcular la probabilidad p de que una pareja elegida al azar esté enlazada entre sí.
  - Para esto sabiendo que el grado promedio es igual a la probabilidad por la cantidad de nodos menos 1 ($\tilde{k} = p(n-1)$) y conociendo el valor del grado promedio y de n , podemos despejar y calcular el valor de p.

$\tilde{p}$= k/n-1 ————> p = 24,908/(229-1)

p= 0.11 ————> probabilidad de que una pareja elegida al azar esté enlazada entre sí.

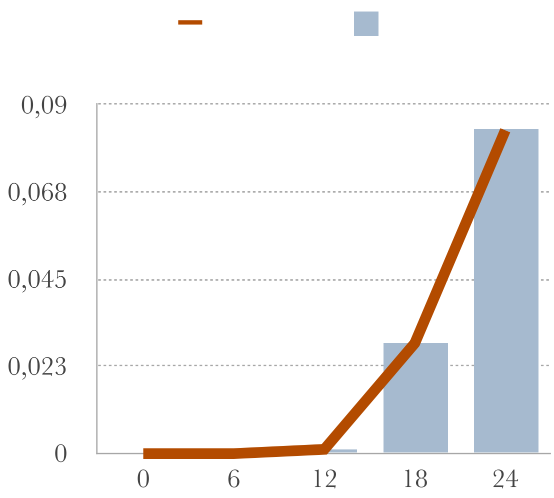
- Una vez obtenido el valor de p podemos calcular la probabilidad P de que un nodo tenga k conexiones:

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Alumna: Paula Brück

Padron: 107533

0 >= k <= 24 (grado promedio de mi red)	P(k)= (n-1/k)p^k.(1-p)^n-1-k
P(0)	0
P(1)	0
P(2)	0
P(3)	0
P(4)	0
P(5)	0
P(6)	0
P(7)	0
P(8)	0
P(9)	0,0001
P(10)	0,0002
P(11)	0,0005
P(12)	0,0011
P(13)	0,0023
P(14)	0,0044
P(15)	0,0077
P(16)	0,0127
P(17)	0,0196
P(18)	0,0285
P(19)	0,0389
P(20)	0,0502
P(21)	0,0615
P(22)	0,0715
P(23)	0,0791
P(24)	0,0835



P(k)= sum P(k_i) = 0,46

- Simulación para la distribución de grados para la red con grado promedio 24,908 nodos y p = 0,11
  - Coeficiente de Clustering:

Alumna: Paula Brück

Padron: 107533

- El de un nodo en particular se calcula haciendo dos veces la cantidad de aristas que generan los triángulos / la cantidad de triángulos que se podrían generar en total de dicho vértice.

$$C_i = \frac{2e_i}{k_i(k_i-1)}$$

- Para poder calcular la e_i (esperanza: cada pareja se conecta con probabilidad p, y las distintas parejas del nodo i que tiene grado k_i)

$$E[e_i] = p\frac{k_i(k_i-1)}{2}$$

- Ahora reemplazando E[e_i] en C_i puedo calcular la esperanza del coeficiente de Clustering promedio:

$$E[C] = \frac{p \cdot k_i(k_i-1)}{k_i(k_i-1)} = p = \frac{\tilde{k}}{n-1}$$

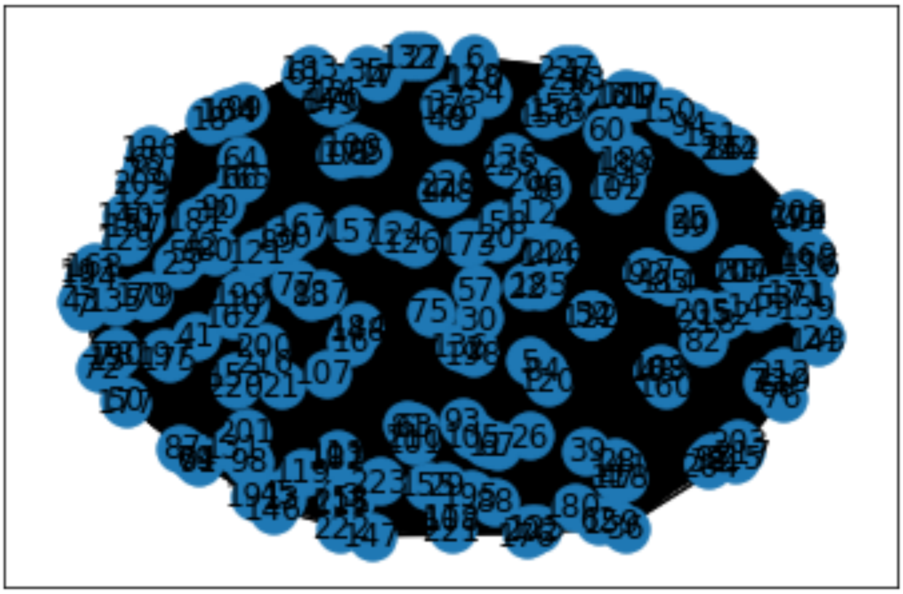
- Y así llegamos a la conclusión que teóricamente C = k/n
- En nuestra red seria:
  - C(teorico)= 24,908/229 = 0,11
  - C(simulado) = sumC_i/sumi = 3,82
- Si analizamos las distancias podemos decir que :
  - Las redes sociales tiene una expansión media ya que no están muy desconectados por lo cual no es baja y a su vez se generan comunidades por lo cual no es alta ya que si estuvieran todos los nodos muy conectados no se podrían diferenciar en grupos .
  - El diámetro crece de manera logarítmica siendo :

Diam(Gnp) = O (log n /log (np)) = O(log 229 / log 24,908)= 1,69

- Si el grado promedio de nuestra red es log(n)/n ya hay una gran componente conexa

Alumna: Paula Brück

Padron: 107533



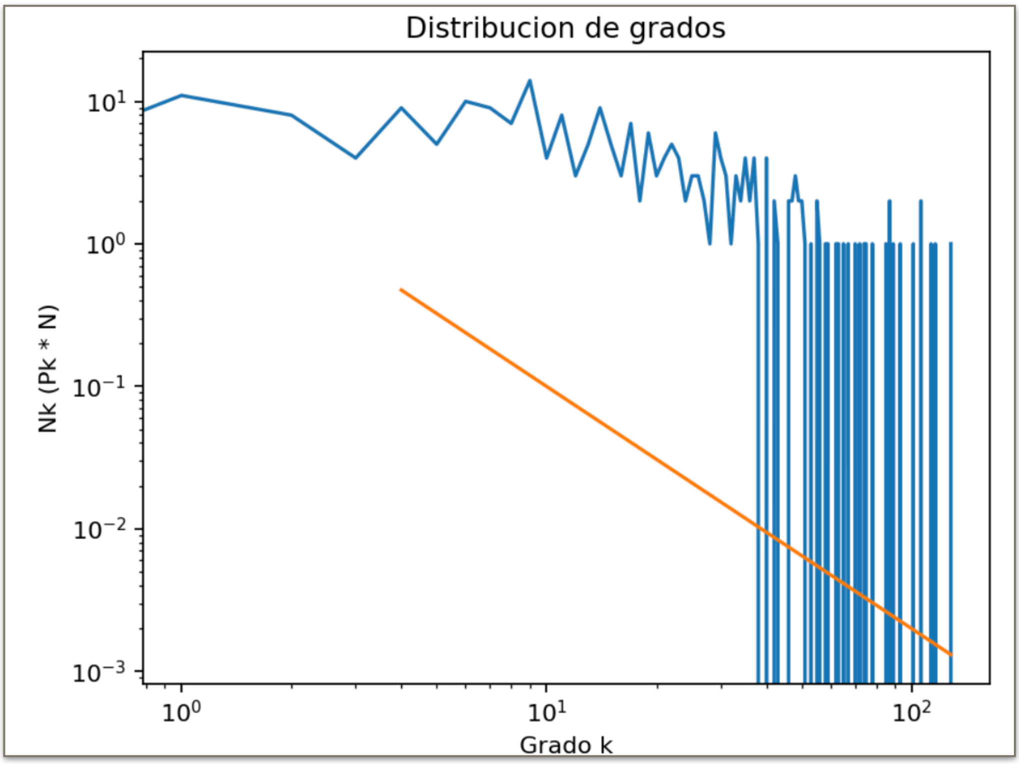
- ***Simulación Modelado Preferential-Attachment:***
  - El modelo de preferential attachment se basa en que aquello que tenga mas de una cosa seguirá teniendo más que aquello que tenia menos.
  - En el caso de redes vamos a basarnos en que en una red aquellos nodos que tienen mayor grado irán aumentando cada vez más su grado que aquellos nodos que tienen menor grado.
  - Utilizando los parámetros de esta red obtenemos una simulación con Preferential-Attachment:
    - Este modelo lo que hace es teniendo en nuestra red 229 nodos va agregándolos en orden
    - Luego al llegar al nodo j se conocen todos los grados (d_i) de aquellos nodos que se encontraban antes ($i < j$) entonces el nuevo j lo que hace es crear m aristas
    - Entonces por cada arista , con una probabilidad p , se elige un nodo al azar para unirlo.
    - La probabilidad de que j se conecte al nodo i es proporcional a su grado (d_i) siendo la probabilidad = 1-p
    - Este modelo genera una ley de potencias :

$$\alpha = 1 + \frac{1}{1-p}$$

- Si medimos el alfa por máxima verosimilitud podemos reemplazar y conocer el valor de p :
  - Para calcular a por máxima verosimilitud aplicamos:

$$\overline{\alpha} = 1 + n \left[\sum_{i}^{n} \ln \left(\frac{d_i}{x_m} \right) \right]^{-1}$$

  - Donde n es la cantidad de nodos de la red , d_i es el grado de cada uno de los vertices y x_m el grado a partir del cual comenzamos a medir.
  - De esa forma podemos decir que alfa = 2,45.



  - Luego reemplazando a en p :

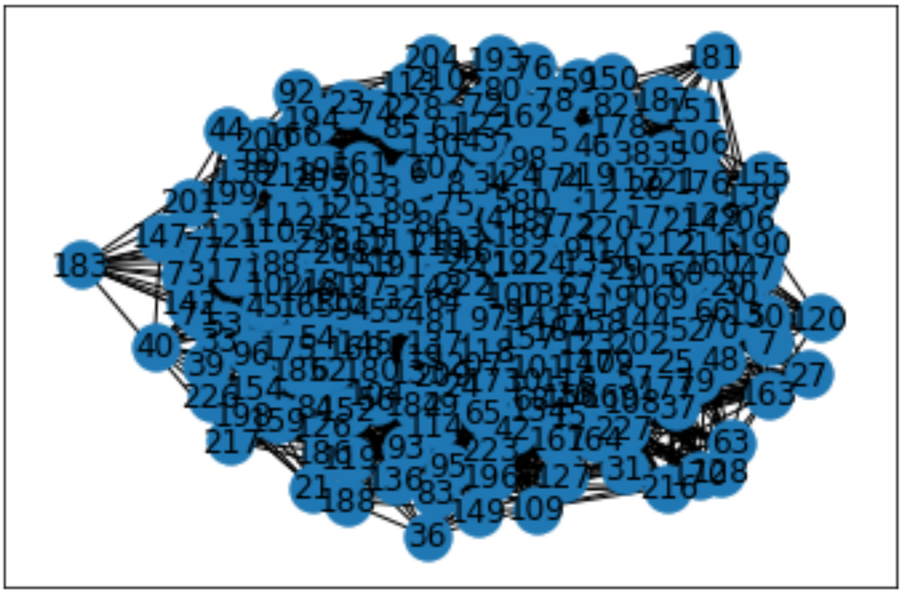
$$p = 1 - \frac{1}{\alpha - 1}$$

p= 1- 1/2,45-1 = **0.31 ——> Probabilidad de que dos nodos se unan al azar**

- Como a = 2,46 la distancia promedio es igual a :

$$\frac{\log \log n}{\log(\alpha - 1)}$$

d_promedio = log log 229 / log (2,45 - 1) = 2,27



- ***Representacion de Anonymous Walks:***
  - Random walks muy cortos donde no se distinguen los vértices entre sí . Proporcionan una forma flexible de reconstruir una red .
  - Para la red de vuelos original calculamos el embedding de annonymous walks de largo 7:
  - Esto nos da un vector de probabilidades ——> p_red_original
  - Lo mismo hacemos para la red de Erdos Renyi —> p_red_erdos_renyi
  - Y por ultimo repetimos el mismo procedimiento para la red de preferential attachment ——>p_red_pref
  - Luego teniendo los 3 embeddings de probabilidades calculamos la distancia coseno entre ellos:

Alumna: Paula Brück

Padron: 107533

- Distancia coseno entre:
  - p_red_original <——> p_erdos_renyi =0.0004793
  - p_red_original <———> p_pref = 0.0004634
  - Por lo tanto podemos decir que el modelo a elegir seria el de prefferential attachment pero la realidad es que la diferencia de distancia entre ambos modelos es muy pequeña , lo cual nos lleva a decir que ambos podrían ser bueno modelos