

# Minería de Medios Sociales

## Análisis y Visualización básica de una Red Social con Gephi

Red “Airlines”

Curso 2018/2019

Antonio Manuel Milán Jiménez  
77449165T

[antoniomj@correo.ugr.es](mailto:antoniomj@correo.ugr.es)

La red “Airlines” representa la conexión entre los diferentes aeropuertos de Estados Unidos. Así, se representan un total de 235 aeropuertos existiendo un enlace entre dos aeropuertos si existe alguna ruta aérea entre ellos..

Señalar que en la red original existen algunas aristas paralelas. Sin embargo, no se ha encontrado información acerca de estas aristas (podrían ser rutas aéreas alternativas) y dado que consiste en un estudio sobre la conexión existente entre los diferentes aeropuertos de Estados Unidos, se ha decidido tratar estas aristas como únicas, teniendo en cuenta un enlace entre los nodos si de alguna forma están conectados los aeropuertos.

## Análisis básico de la red

En esta sección se analizarán los atributos básicos de la red:

Medida	Valor
Número de nodos $N$	235
Número de enlaces $L$	1297
Número máximo de enlaces $(N*(N-1)/2) L_{max}$	27495
Densidad del grafo $L/L_{max}$	0,047
Grado medio $\langle k \rangle$	11,038
Diámetro $d_{max}$	4
Distancia media $d$	2,318
Coefficiente medio de clustering $\langle C \rangle$	0,652
Número de componentes conexas	1
Número de nodos componente gigante (y %)	235 (1)
Número de aristas componente gigante (y %)	1297 (1)

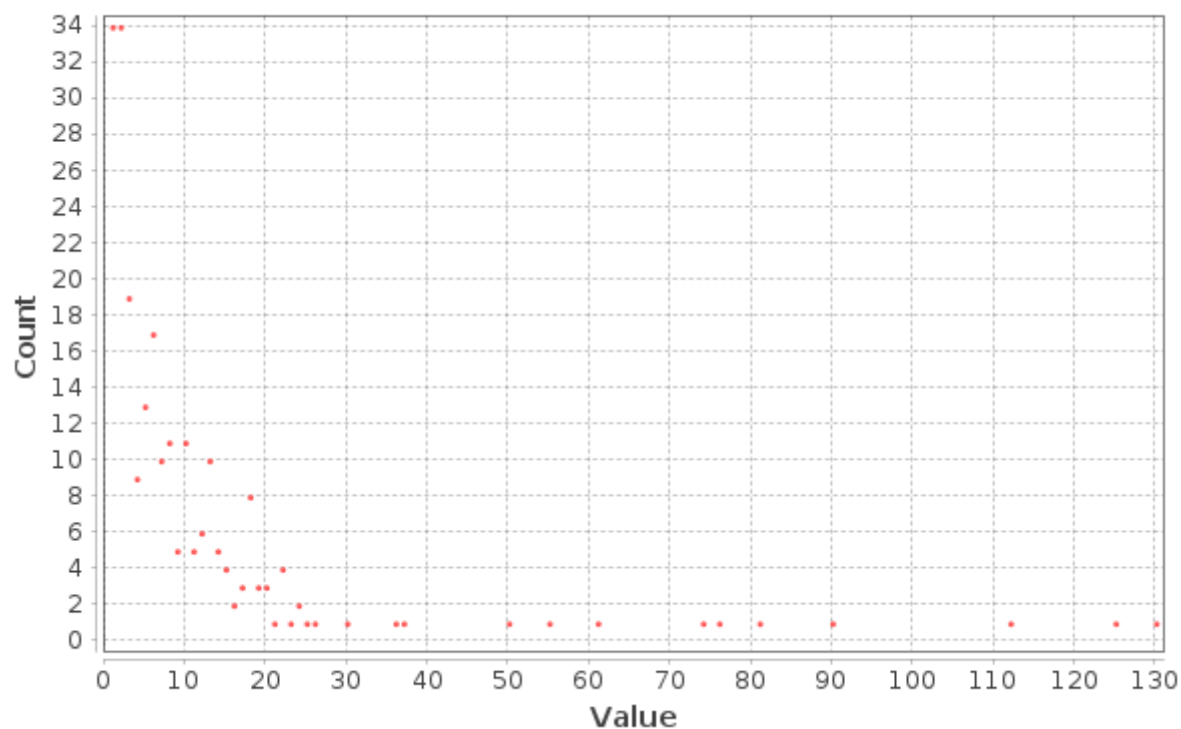
La red consiste en un total de 235 nodos con 1297 enlaces entre ellos, lo cual implica que haya una densidad en la red de 0,047.

Aunque pueda no parecer una densidad elevada, estudiando más a fondo la conectividad dentro de la red encontramos que la distancia media es de tan sólo 2,31, indicando que sí existe una gran conectividad entre los nodos. De hecho, los dos nodos más alejados están a una distancia máxima de 4. También se ha calculado que el grado medio de los nodos es de 11,038 y que el coeficiente medio de clustering es de 0,652, indicando un grado bastante alto de clustering local.

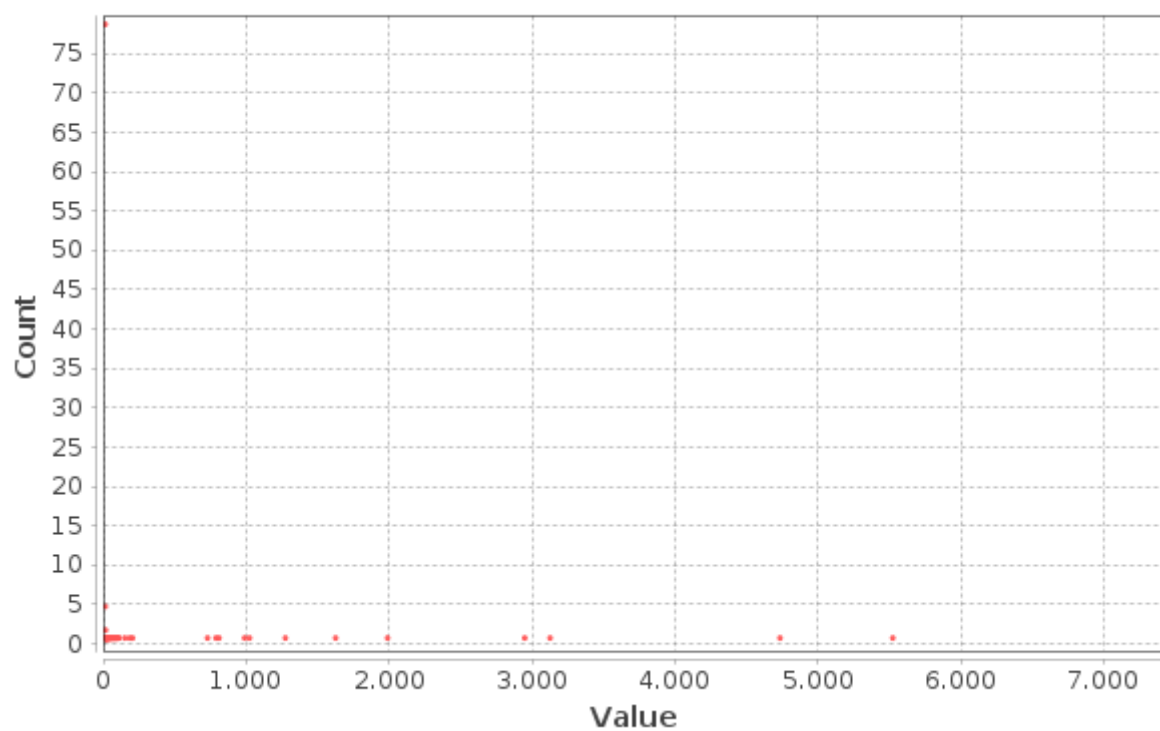
Además, sólo existe una componente conexas, estando conectados todos los nodos de alguna forma al resto de nodos.

A continuación se muestran las diferentes gráficas sobre las distribuciones en la red:

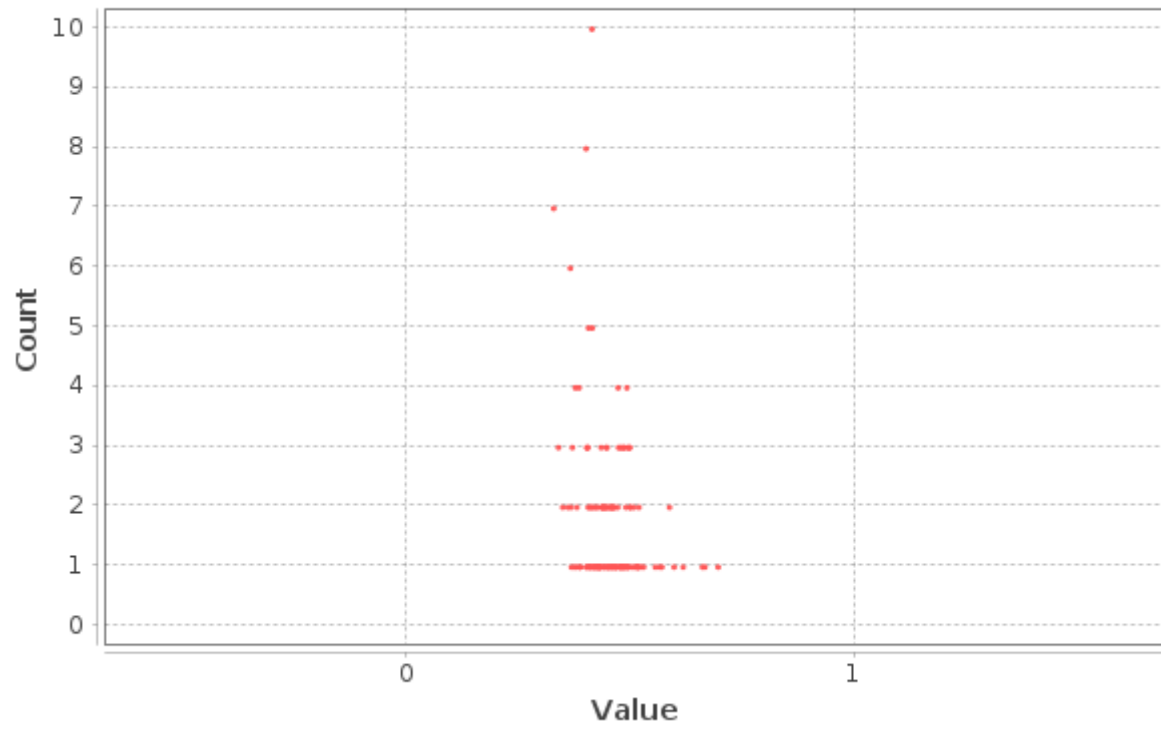
**Degree Distribution**



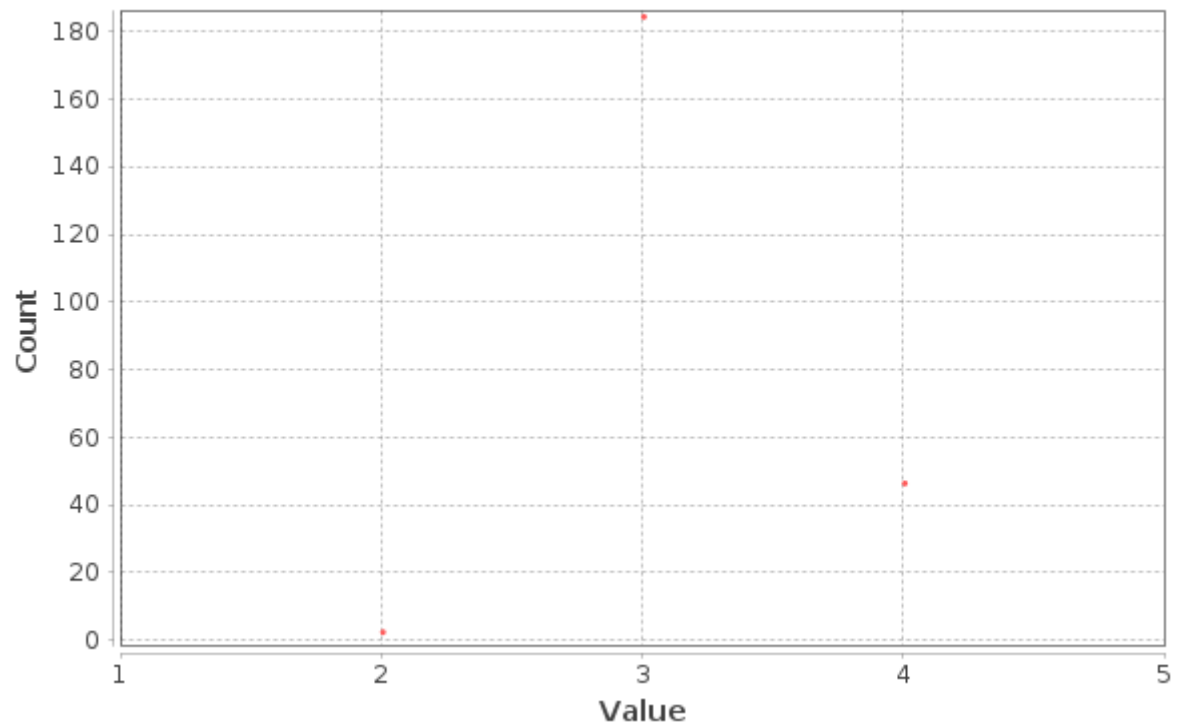
**Betweenness Centrality Distribution**



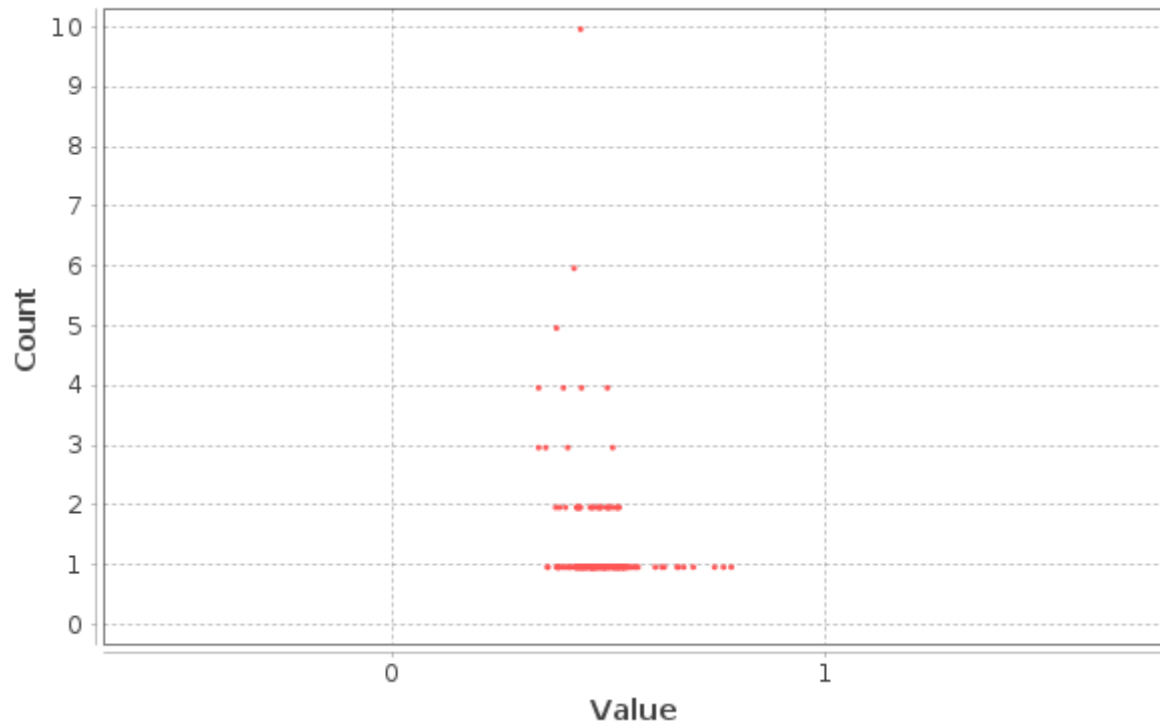
### Closeness Centrality Distribution



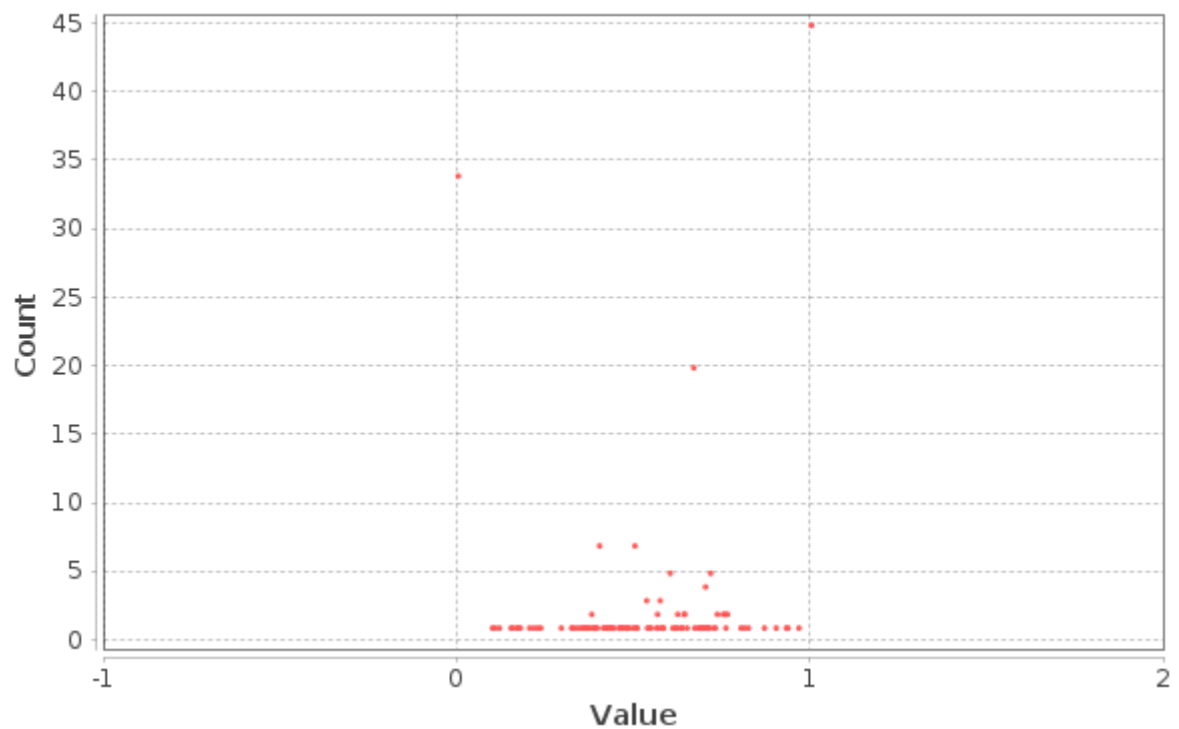
### Eccentricity Distribution



## Harmonic Closeness Centrality Distribution



## Clustering Coefficient Distribution



Observando la gráfica de la distribución de grados de los nodos, encontramos que presenta una larga cola a la derecha, lo cual indica que esta red es una red libre de escala.



## Estudio de la Centralidad de los Actores

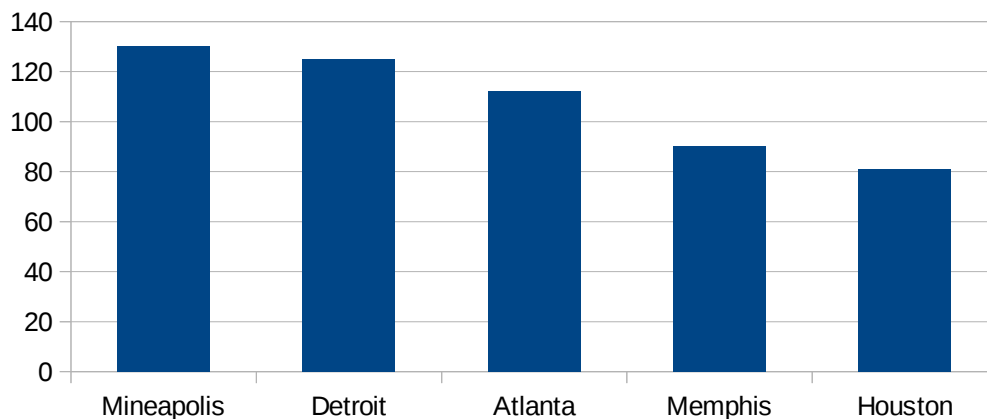
Como se acaba de comentar, dado el problema que se presenta, deben existir actores (aeropuertos) que funciones como puntos de conexión entre diferentes regiones de Estados Unidos y que por lo tanto presenten una alta centralidad en la red.

De esta forma, este es el estudio realizado de los 5 actores con mayor centralidad en la red para diferentes medidas tales como el número de aeropuertos a los que está conectado o cómo de importante es para conectar diferentes regiones del país:

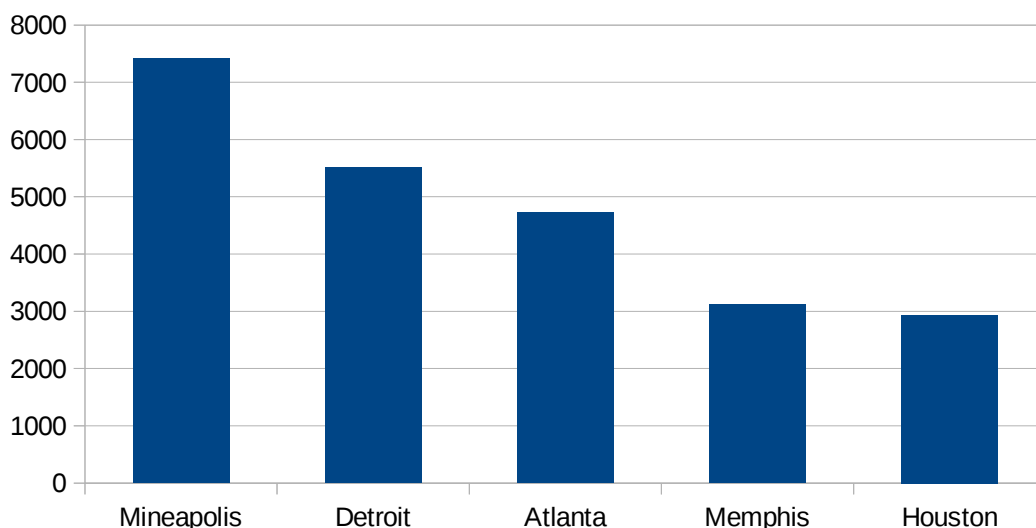
Centralidad de Grado	Intermediación	Cercanía	Vector propio
MSP: 130	MSP: 7413,48	MSP: 0,6923	MSP: 1
DTW: 125	DTW: 5510,82	DTW: 0,6628	DTW: 0,9982
ATL: 112	ATL: 4722,11	ATL: 0,6573	ATL: 0,9441
MEM: 90	SEA: 3112,84	MEM: 0,6141	MEM: 0,8197
IAH: 81	MEM: 2935,39	IAH: 0,5939	IAH: 0,8001

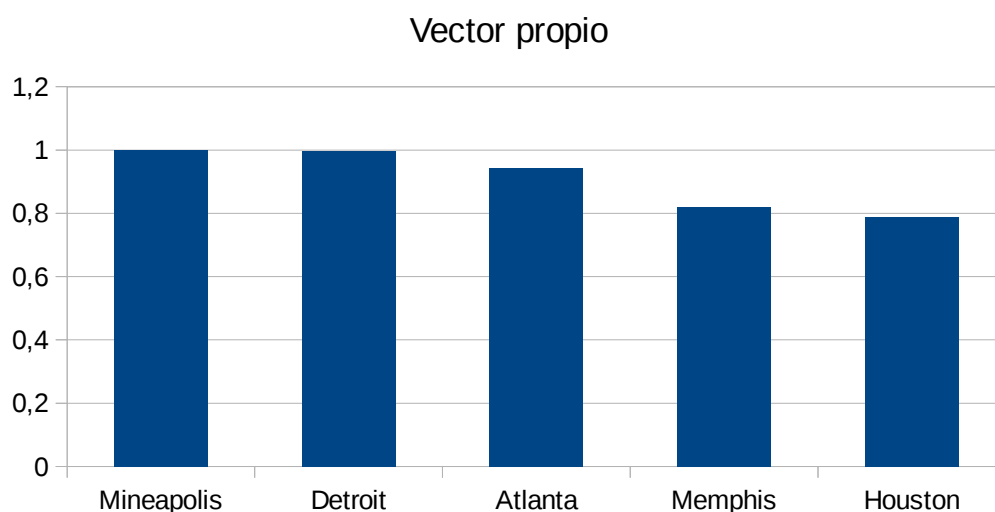
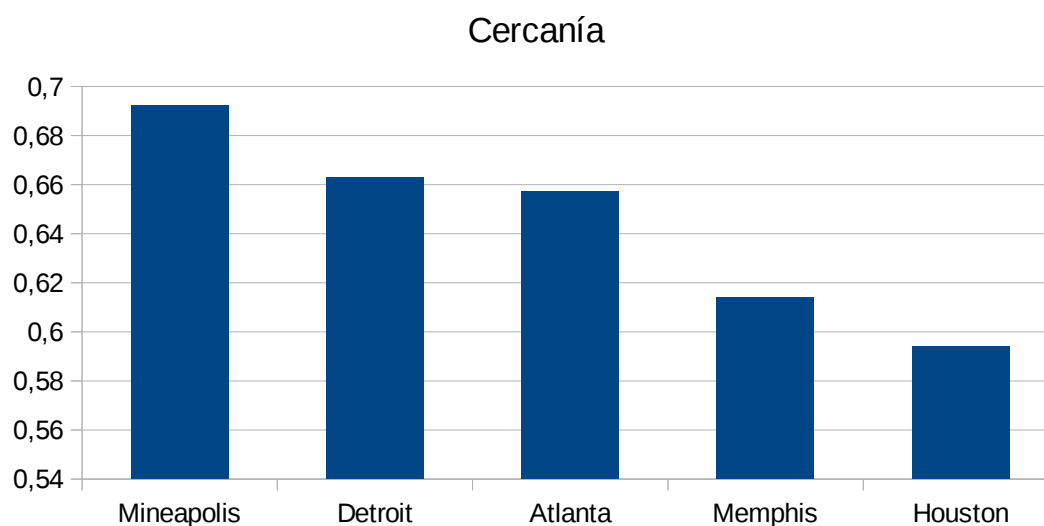
Los tres actores con mayor centralidad son claramente, en orden: MSP (Mineápolis), DTW (Detroit) y ATL (Atlanta). Los siguientes nodos más relevantes serían ciudades como Memphis, Seattle o Houston.

Grado



Intermediación





Es curioso que, aunque son todas ellas grandes ciudades de Estados Unidos, no son precisamente las más grandes y conocidas como Nueva York o Los Angeles tal y como se podría pensar en un principio. Esto se debe, en este caso, a que una ciudad es más importante respecto a la centralidad en función de su distribución geográfica en el mapa, que se encuentre en una mejor zona para conectar regiones del país o incluso conexión con el resto de países.

Gracias a este estudio realizado, podemos ya conseguir una mejor representación gráfica de la red en la que se tengan mayores nodos para aquellos actores más importantes respecto a su centralidad en la red calculada (El tamaño del nodo es proporcional a la intermediación de éste y la tonalidad de azul proporcional al grado):





Detectando ya las comunidades en Gephi:

## Modularity Report

### Parameters:

Randomize: On  
Use edge weights: Off  
Resolution: 1.0

### Results:

Modularity: 0.231  
Modularity with resolution: 0.231  
Number of Communities: 6



Con un “Resolution” de 1,0 se obtienen 6 comunidades, un número aceptable por lo comentado anteriormente. Sin embargo, la modularidad es de 0,239, bastante baja por lo que es recomendable probar con otros parámetros:

## Modularity Report

### Parameters:

Randomize: On  
Use edge weights: Off  
Resolution: 0.7

### Results:

Modularity: 0.241  
Modularity with resolution: 0.129  
Number of Communities: 10



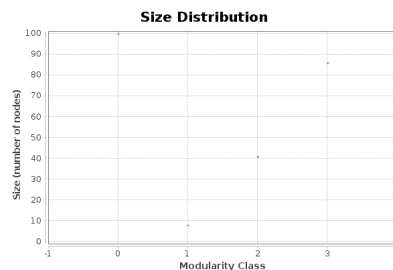
## Modularity Report

### Parameters:

Randomize: On  
Use edge weights: Off  
Resolution: 1.2

### Results:

Modularity: 0.229  
Modularity with resolution: 0.351  
Number of Communities: 4



## Modularity Report

### Parameters:

Randomize: On  
Use edge weights: Off  
Resolution: 2.0

### Results:

Modularity: 0.011  
Modularity with resolution: 1.006  
Number of Communities: 2



Descubrimos que aún probando con otros parámetros en “Resolution” (0,7 1,2 2,0), se sigue obteniendo una baja modularidad (0,241 0,229 0,011) respectivamente. Además, el número de comunidades en todas ellas (10 4 2) no se encuentra dentro de lo esperado. Teniendo en cuenta todo esto, se decide escoger la primera detección de comunidades, con “Resolution” 1,0 y obteniendo 6 comunidades diferentes. Visualizándolas en la red:

Otro ejemplo claro de las comunidades existentes sucede en la comunidad más pequeña encontrada:

Se trata de una comunidad muy clara pues se tienen 7 aeropuertos muy pequeños de Montana conectados únicamente, escala obligatoria, al aeropuerto de Billings Logan (BIL), el aeropuerto mayor de Montana que sirve de conexión con el resto de estados y regiones del país.

