



ugr

Universidad
de Granada

ANÁLISIS Y VISUALIZACIÓN BÁSICA DE UNA RED SOCIAL CON GEPHI

MÁSTER DATCOM

Minería de medios sociales

Autor

Alberto Armijo Ruiz

26256219V

armijoalb@correo.ugr.es



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

13 de abril de 2019

Índice general

0.1. Información sobre la red	5
0.1.1. Medidas de la red	5
0.1.2. Red de vuelos	5
0.1.3. Distribuciones de medidas de la red	7
0.2. Análisis de la red	9
0.3. Análisis de la centralidad de los actores	9
0.4. Estudio de comunidades en la red	10
0.5. Gráficos adicionales	14

Índice de figuras

1.	Distribución de grados de la red	7
2.	Coeficiente medio de la red	7

Índice de cuadros

1.	Métricas de la red estudiada	5
2.	Aeropuertos más importantes según diferentes medidas	9
3.	resultados algoritmo Lovaina	11

0.1. Información sobre la red

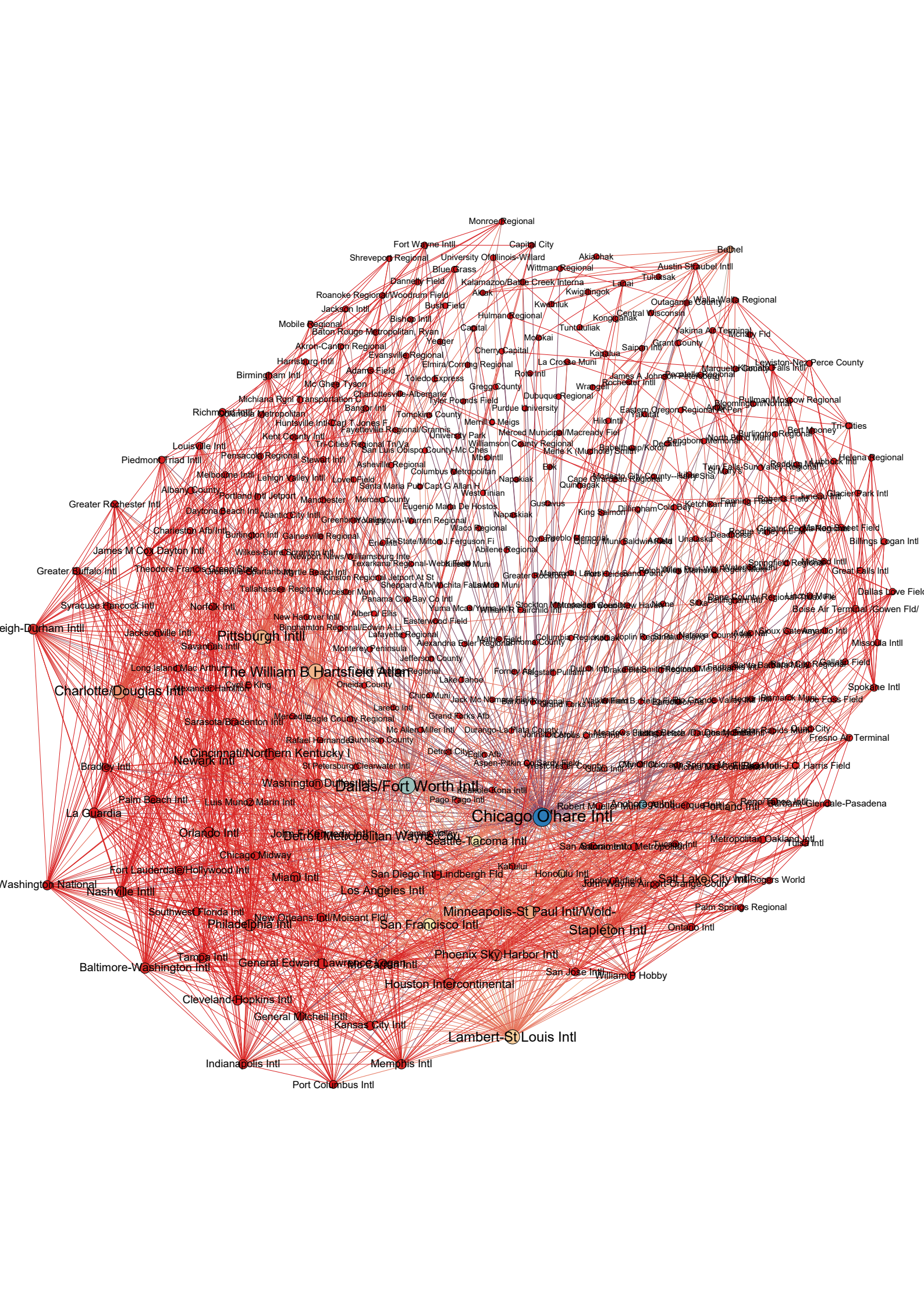
La red que se va a estudiar se trata de una red de vuelos en Norte América en el año 1997. Esta red se trata de una red dirigida en la cuál cada nodo representa un aeropuerto en Norte América y las aristas muestran los vuelos que entran y salen de cada aeropuerto.

0.1.1. Medidas de la red

Medida	Valor
<i>Número de nodos N</i>	332
<i>Número de enlaces L</i>	2126
<i>Número máximo de enlaces L_{max}</i>	4517750
<i>Densidad del grafo L/L_{max}</i>	0,019
<i>Grado medio $\langle k \rangle$</i>	6,404
<i>Diámetro d_{max}</i>	6
<i>Distancia media d</i>	2,564
<i>Coefficiente medio de clustering $\langle C \rangle$</i>	0,313
<i>Número de componentes conexas</i>	1
<i>Número de nodos componente gigante (y %)</i>	332 (1.0)
<i>Número de aristas componente gigante (y %)</i>	2126 (1.0)

Cuadro 1: Métricas de la red estudiada

0.1.2. Red de vuelos



0.1.3. Distribuciones de medidas de la red

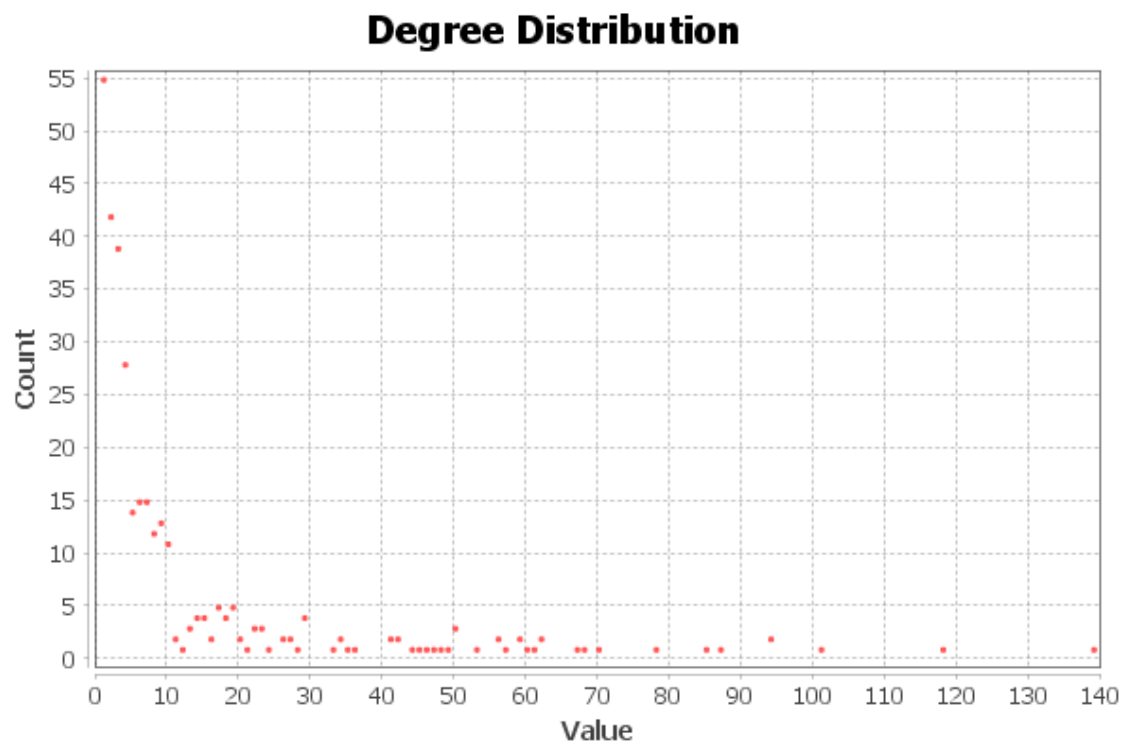


Figura 1: Distribución de grados de la red

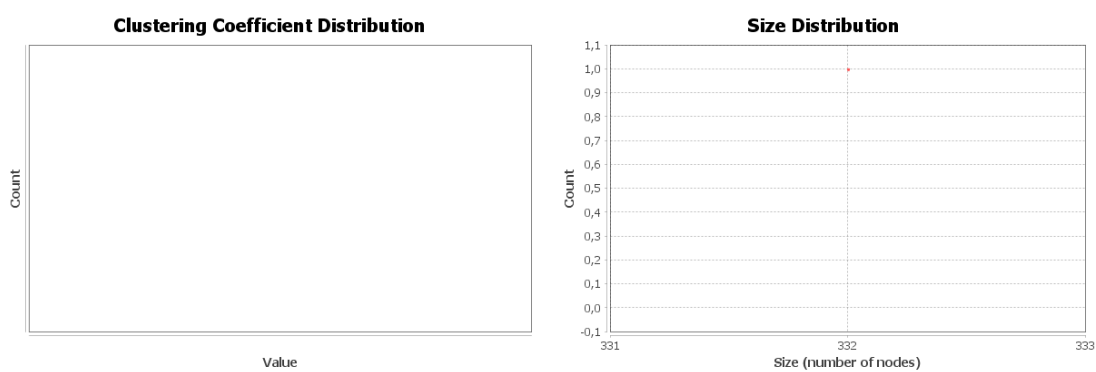
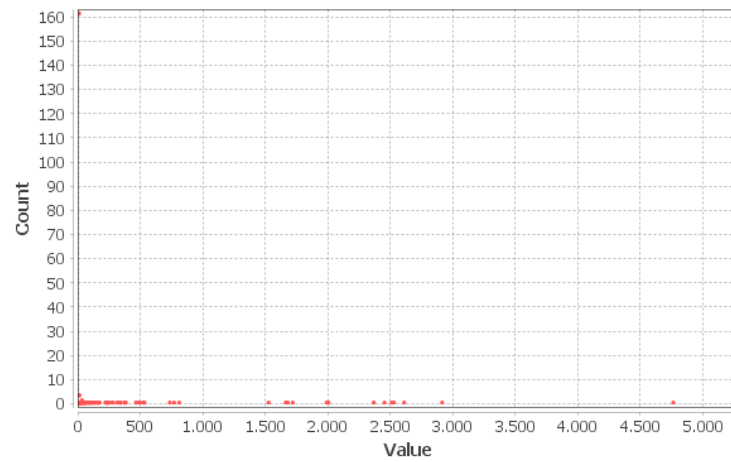
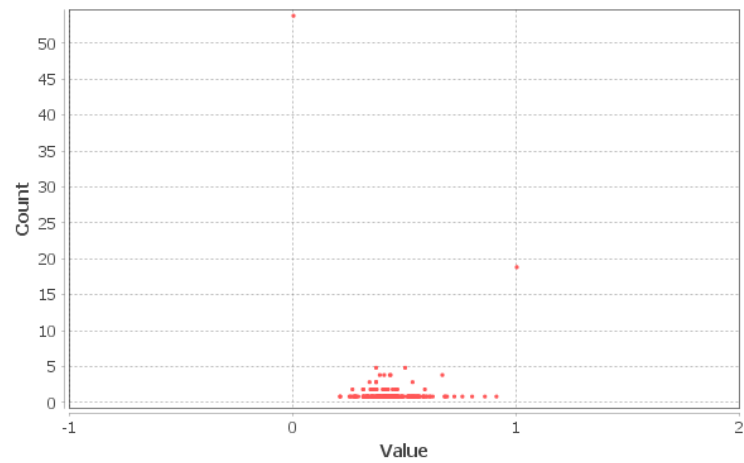
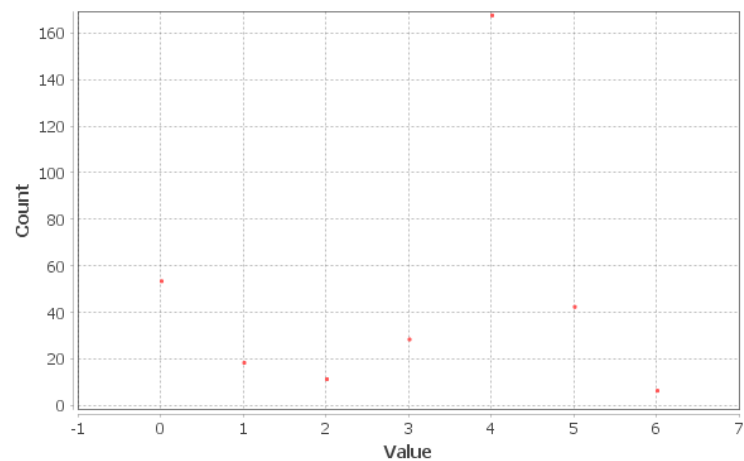


Figura 2: Coeficiente medio de la red

Betweenness Centrality Distribution**Closeness Centrality Distribution****Eccentricity Distribution**

0.2. Análisis de la red

Según la información que hay en la tabla de métricas, la red cuenta con 332 nodos y 2126 aristas (332 aeropuertos y 2126 vuelos). Si nos fijamos en el grado medio, podemos ver que cada aeropuerto tiene entre 6 y 7 vuelos de media, es decir, en total llegan y salen 6/7 vuelos. Si nos fijamos en la distribución de grados de la red; se puede ver que hay unos pocos vuelos que tienen un gran número de vuelos (6 aeropuertos con más de 80 vuelos); el mayor de todos con 140 vuelos. Si miramos la densidad de la red, vemos que es bastante baja, esto nos dice que hay muy pocos nodos que estén altamente conectados. Mirando la distancia máxima y la distancia media se puede ver que no hay grandes variaciones, siendo la media 2.6 nodos y 6 el diámetro de la red. Si nos fijamos en las métricas vemos que la red tiene una sola componente conexas, por ello sabemos que no hay zonas que estén desconectadas de la red y que solamente existe un único clúster que contiene todos los nodos de la red, lo cual parece bastante normal para una red de vuelos.

0.3. Análisis de la centralidad de los actores

Centralidad de Grado

Chicago O'hare Intl
Dallas/Fort Worth Intl
The William B Hartsfield Atlan
Lambert-St Louis Intl
Pittsburgh Intl

Intermediación

Chicago O'hare Intl
Dallas/Fort Worth Intl
Lambert-St Louis Intl
Seattle-Tacoma Intl
San Francisco Intl

Cercanía

San Antonio Intl
Luis Munoz Marin Intl
Cyril E King
Keahole-Kona Intl
Kahului

Vector propio

Miami Intl
Houston Intercontinent
Dallas/Fort Worth Intl
Orlando Intl
Fort Lauderdale

Cuadro 2: Aeropuertos más importantes según diferentes medidas

Primero nos centraremos en la medida de centralidad del grado, dado que esta red es dirigida, debemos usar también las medidas de influencia y soporte. El primer nodo que aparece es el aeropuerto de Chicago, este es el aeropuerto con más conexiones de la red (139), si nos fijamos en los valores de Soporte e Influencia, se puede ver que tiene un Soporte de 41 y una Influencia de 99, por lo que podemos decir que es un aeropuerto desde el cual salen muchos vuelos al resto del país y por ello es el hub principal de esta red. Los aeropuertos de Dallas y The William B Hartsfield Atlan tienen un grado de 118 y 101; lo que tienen en común estos dos que su Soporte es mayor que su Influencia, por ello podemos saber que estos aeropuertos son sitios a los cuales la gente viaja más (al ser ciudades grandes de EEUU) pero no son tan influyentes como el aeropuerto de Chicago; el aeropuerto de St Louis tiene un Soporte e Influencia parecidos, con una influencia mayor que los dos anteriores, por lo que podríamos decir que este aeropuerto sirve también de enlace a otros lugares que simplemente como destino; por último el aeropuerto de Pittsburg que también tiene un grado de 94 y una distribución entre Soporte e Influencia parecida a la del aeropuerto de Chicago pero en menor medida.

La siguiente medida que se va a analizar es la Intermediación, esta medida mide la frecuencia con la que un nodo aparece en el camino más corto entre dos nodos de la red; esto significa que cuanto mayor sea el valor de este, más nodos necesitan pasar por él para llegar a otro nodo. Para el caso de la red significa que cuanto mayor sea el valor en un aeropuerto más vuelos necesitan pasar por él para llegar a algún lugar. En los primeros puestos se encuentran los aeropuertos de Chicago y Dallas, con unos valores muy superiores al resto, esto es normal por el gran grado de Influencia en el caso de Chicago y de Soporte en el caso de Dallas. En los siguientes puestos se encuentran los aeropuertos de St Louis, Seattle y San Francisco; lo que significa que estos tres aeropuertos son muy influyentes a la hora de llegar de un sitio a otro, posiblemente porque al ser aeropuertos de ciudades muy grandes sea necesario pasar por ellos para llegar a algún lugar cercano.

Otra medida que se va a estudiar es la Cercanía, esta medida premia estar en la parte central de la red. En este caso, todos los aeropuertos que aparecen tiene el mismo valor para esta componente (1.0), posiblemente sean estos porque sean los más centrales dentro de los diferentes estados del país. Por ejemplo, el aeropuerto de San Antonio se encuentra en el centro de Texas, y es más central que el aeropuerto de Dallas aunque este sea más influyente.

Por último se va a analizar la medida de centralidad del vector propio. La medida de centralidad del vector propio tiene en cuenta la centralidad de los nodos vecinos al nodo que se estudia, es decir, un nodo es tan importante como los nodos que tiene alrededor. Los aeropuertos que más valor tienen dentro de esta medida son aquellos que tienen un Soporte relativamente alto y que las entradas que tienen son de otros aeropuertos importantes; por ello podríamos decir que estos aeropuertos son los que más conexiones con aeropuertos grandes tienen.

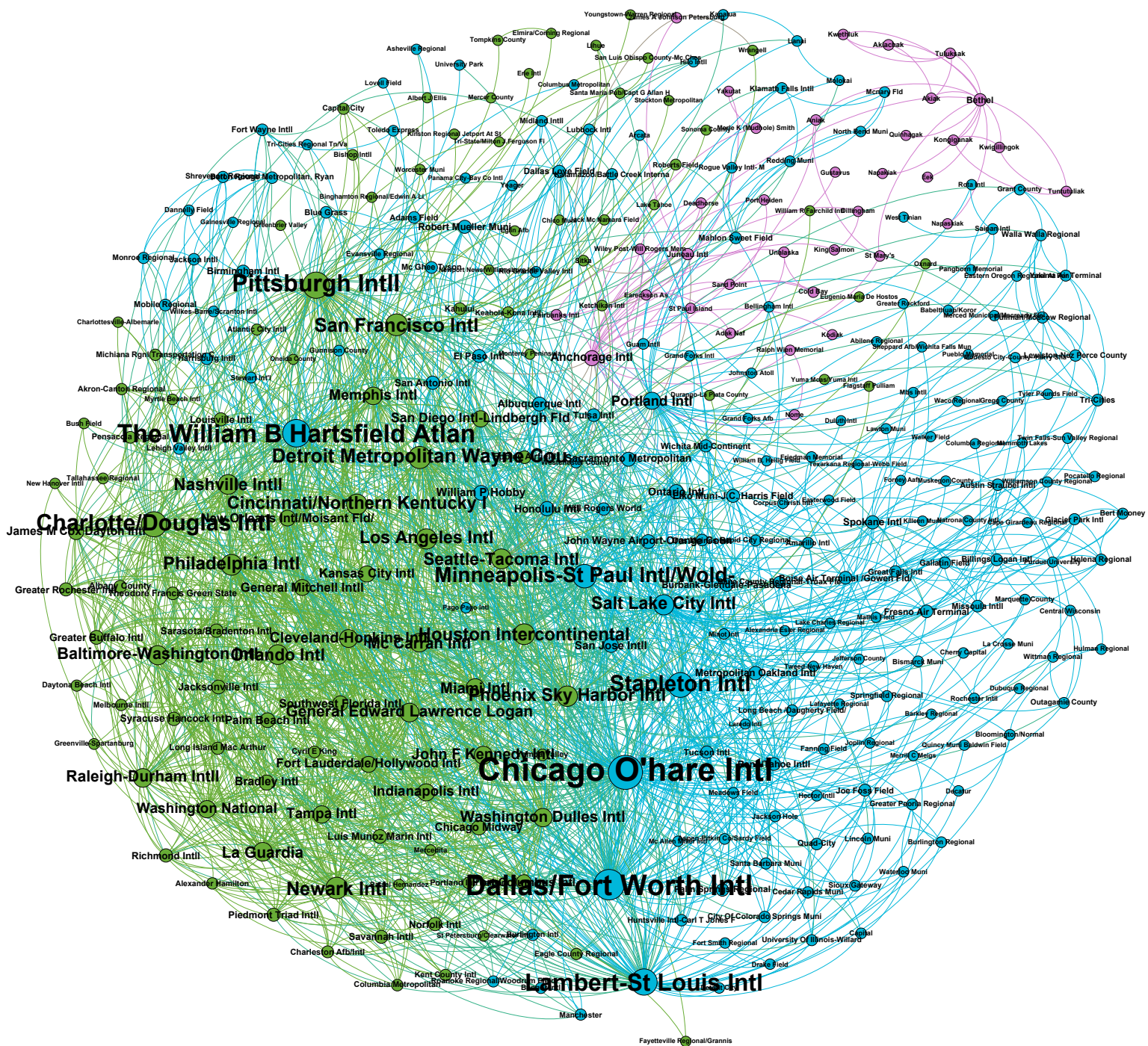
0.4. Estudio de comunidades en la red

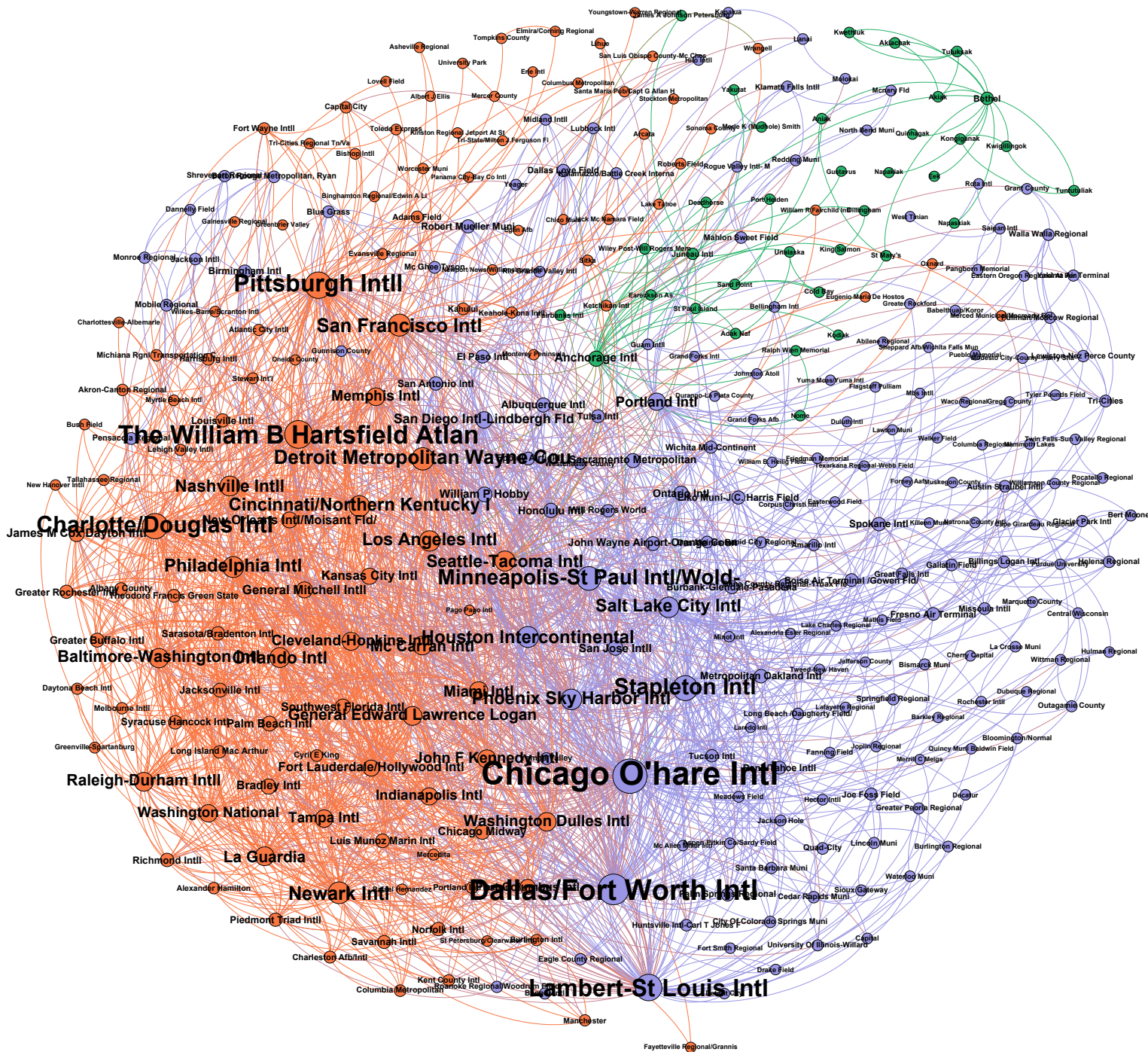
Para hacer un estudio sobre las comunidades de la red, se utilizará el algoritmo Lovaina cambiando la resolución y viendo el número de comunidades que obtiene. Se muestran las dos mejores soluciones obtenidas, no solo por el valor de modularidad obtenido, sino por el sentido que tenga en la red. Si nos fijamos en la tabla, se puede ver que el valor de la modularidad es bastante baja; esto seguramente ocurra porque aunque un nodo seleccionado para una comunidad tenga una mayoría de nodos de esa misma comunidad, tenga también otras aristas a nodos de comunidades diferentes. Además, el peso de las aristas también influye en el cálculo, por lo que si un nodo está unido a un nodo importante en la red, es fácil que caiga en la misma comunidad que este. Los resultados obtenidos son los siguientes.

Resolución	Valor	Número de comunidades
1	0,202	4
1,1	0,187	4
1,2	0,178	3
1,3	0,190	3
1,4	0,176	2
1,5	0,16	2
1,6	0,037	4
1,7	0,034	5
1,8	0,034	4
1,9	0,032	3
2	0,031	3

Cuadro 3: resultados algoritmo Lovaina

Las dos imágenes de la comunidad que se muestran en la práctica son las que más sentido parecen tener y con una buena modularidad. Por lo que se puede ver en ambas imágenes, existen dos comunidades grandes dentro de esta red; estas dos comunidades pueden representar vuelos en ciertas zonas de Norte América, por ejemplo vuelos por el este de Norte América (Detroit, Los Ángeles, ...). Otro detalle que se puede ver en la red, es que aparece una pequeña comunidad; el nodo más importante de esta comunidad es el aeropuerto de Anchorage, este aeropuerto se encuentra en Alaska, y comunica con aeropuertos más pequeños de esa zona; por ello, es normal que aparezca esta comunidad dentro de la red, ya que desde este nodo se existen vuelos a los aeropuertos más importantes de la red, y desde ese aeropuerto se puede llegar al resto de nodos de la comunidad.





0.5. Gráficos adicionales