

Trabajo Práctico N°2:

[75.30] Teoría De Algoritmos II- FIUBA

Autor

Buceta, M. Belen	102121	bbuceta@fi.uba.ar
------------------	--------	-------------------

Índice

1. Introducción	2
2. Comunidades	2
2.1. Consigna 1.a	2
2.1.1. Visualización	2
2.2. Consigna 1.b	2
2.2.1. Sobre la homofilia	2
2.3. Consigna 1.c	3
2.3.1. Subcomunidades	3
2.3.2. Visualización	4
3. Motifs	4
3.1. Consigna 2.a	4
3.1.1. Resultados	4
3.2. Consigna 2.b	4
3.2.1. Resultados	4
3.2.2. Visualización	5
3.3. Consigna 2.c	5
3.3.1. Conclusiones	5
4. Roles	6
4.1. Consigna 3	6
4.1.1. Visualización	6
4.2. Conclusiones	6

1. Introducción

El presente trabajo reúne la resolución del segundo trabajo práctico de la materia. El desarrollo consiste en analizar una red de vuelos dada.

La red a considerar puede obtenerse en:

<https://drive.google.com/file/d/1bZ5xauNego4i3oV-qeQQrk-1DbQmF9a8/view>.

2. Comunidades

2.1. Consigna 1.a

Obtener una visualización de las comunidades presentes en dicha red (indicando el algoritmo utilizado).

2.1.1. Visualización

Para obtener la visualización se elige el algoritmo de Louvain. Dado que se trata de un algoritmo Greedy que ejecuta en $O(E \log V)$. Además cuenta con buenos resultados.

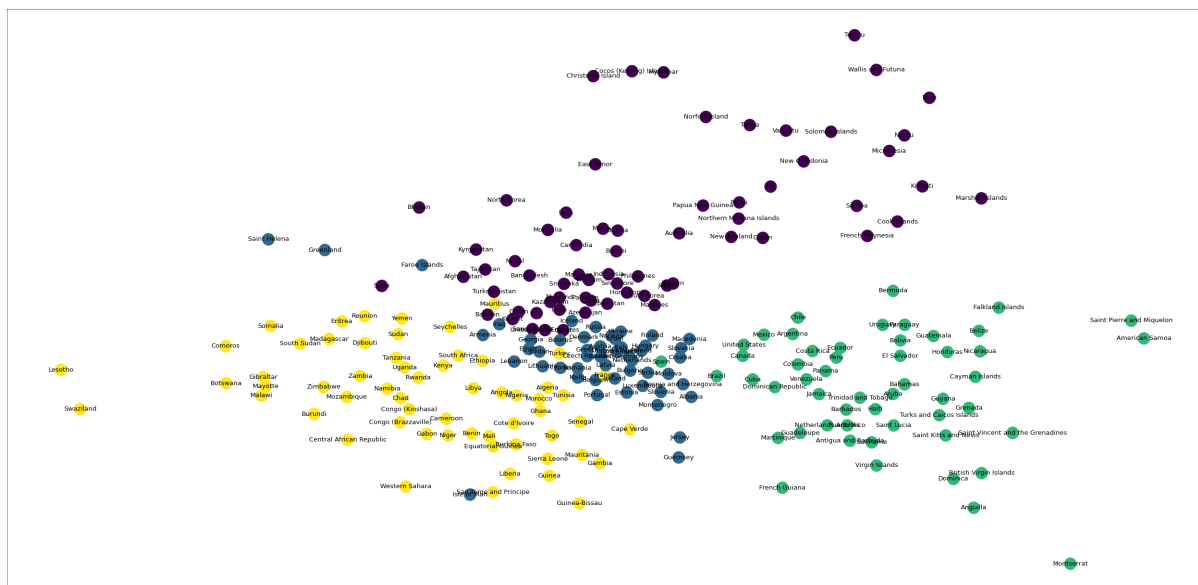


Figura 1: Comunidades presentes en el grafo

2.2. Consigna 1.b

Considerando lo que respondiste en el parcialito 1 (ejercicio 2):

i. Si mencionaste que había homofilia, ¿corresponde por el mismo tipo que mencionaste anteriormente? ¿por qué?

2.2.1. Sobre la homofilia

Utilizando este algoritmo para detectar comunidades podemos ver que no hay homofilia por el mismo tipo que se eligió anteriormente: Por hemisferios: Occidental y Oriental. En el gráfico podemos ver como se detectan cuatro comunidades por continente o más bien cercana a ser detectada por continentes: América, Europa, África y Asia. Cabe destacar que América está tomada como

uno solo (Sur/Norte) así también como Asia y Oceanía. Se concluye entonces que la homofilia que se visualiza se da por la pertenencia de los países a un determinado continente.

2.3. Consigna 1.c

Obtener los nodos correspondientes a una de las subredes (con al menos 20 % de los nodos), y realizar una visualización de las sub-comunidades presentes.

2.3.1. Subcomunidades

Para la elección de la subred a elegir, primero se revisa la cantidad de nodos con la que cuenta cada comunidad, tal que se elija aquella que cumpla con el mínimo requerido en el enunciado. Como todas las comunidades contaban con el mínimo necesario, se eligió la comunidad que contenía a los países de América (tanto del Sur como del Norte). Luego, se volvió a utilizar el algoritmo de Louvain para detectar las subcomunidades presentes en el subgrafo elegido. Aclaración de la subred elegida: es cierto que existen algunos outliers, como por ejemplo, España. Pero en su mayoría está subcomunidad contiene a los países de América.

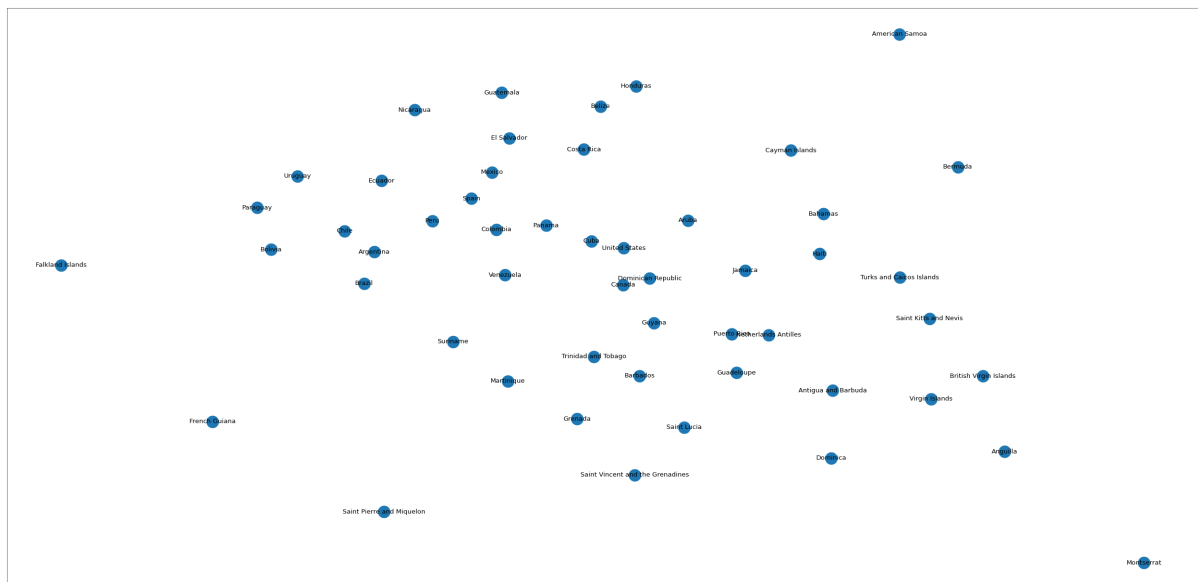


Figura 2: Subgrafo de la comunidad elegida

2.3.2. Visualización

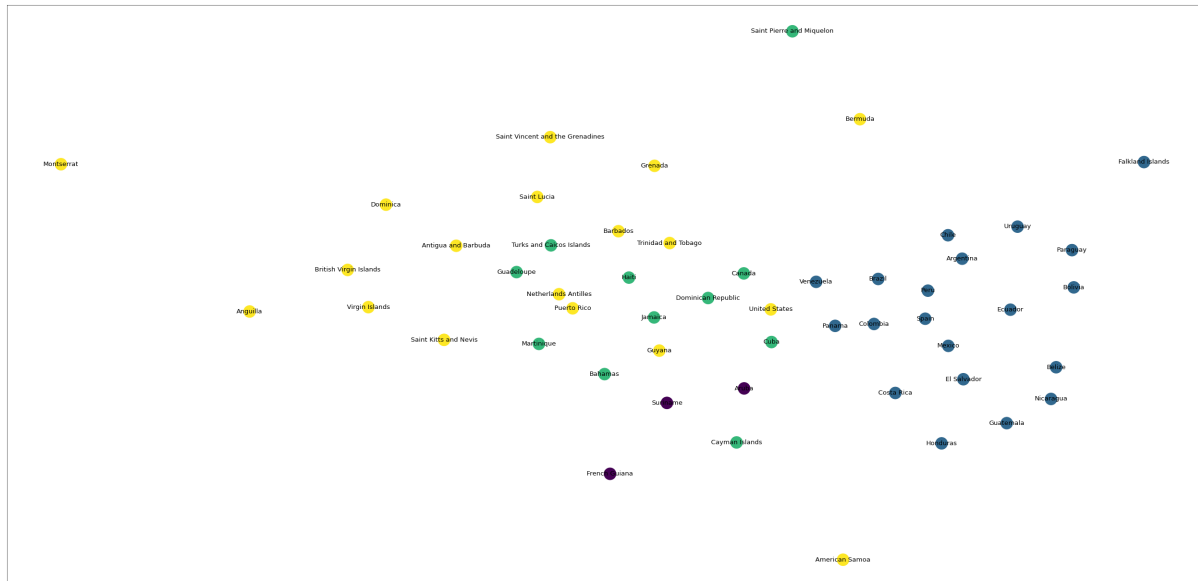


Figura 3: Subcomunidades del grafo elegido

3. Motifs

3.1. Consigna 2.a

Calcular los motifs de hasta 5 nodos de la subred definida en el punto 1.c.

3.1.1. Resultados

Utilizamos la función `calcular_smotifs` para calcular los motifs de la subred (`sub_G`) y los resultados obtenidos fueron un array con los siguientes valores:

[3925, 949, 14476, 23068, 1180, 24525, 7397, 1674, 35714, 108701, 107732, 64950, 53814, 207513, 1983, 22990, 161393, 25871, 53569, 1627, 14622, 25325, 68560, 54892, 5852, 50265, 3156, 12129, 2024]

3.2. Consigna 2.b

Calcular el promedio y desvío estándar de los motifs de una red de baseline. Calcular el significant profile de la red, y hacer un gráfico.

3.2.1. Resultados

Para utilizar la función que calcula el promedio y el desvío estándar debo primero calcular una función que devuelva un modelo aleatorizado, como puede ser cualquiera, utilicé una de las funciones de `calculos.py` para calcular el modelo (`configuration_smodel`). Una vez obtenido el modelo que sirve como baseline se utiliza la función `motifs_grafo_seleatorios` y obtienen los siguientes resultados:

- Promedio:
[3.717800e+03, 3.220000e+02, 2.369320e+04, 1.662920e+04, 4.416800e+03, 1.133060e+04, 2.341800e+03, 1.292000e+02, 1.013934e+05, 1.883704e+05, 4.600860e+04, 7.126660e+04,

3.796940e+04, 6.692280e+04, 1.659680e+04, 1.056632e+05, 6.341140e+04, 5.002600e+03,
 3.317100e+04, 1.335920e+04, 4.620760e+04, 5.594800e+03, 8.113800e+03, 2.123060e+04,
 1.430000e+04, 5.421400e+03, 2.715800e+03, 7.300000e+02, 1.800000e+01]

■ Desvío Estándar:

[1.52398688e+02, 6.06630036e+00, 1.09338893e+03, 1.27018052e+03, 6.18687126e+02, 5.44315570e+02,
 1.24684241e+02, 1.73827501e+01, 2.65830499e+03, 1.16854002e+04, 6.06529836e+03, 4.54562834e+03,
 1.00453225e+03, 6.86312009e+03, 1.02446384e+03, 1.38652506e+04, 6.09479928e+03, 4.59324548e+02,
 2.04319818e+03, 3.49941152e+03, 5.93787819e+03, 4.38230259e+02, 9.73881184e+02, 2.17110678e+03,
 3.09912639e+03, 4.26042064e+02, 4.16287112e+02, 1.49917311e+02, 1.21655251e+01]

■ Significance Profile: Una vez calcula lo anterior, se puede calcular el significance profile.

La cual dio como resultado:

[0.00967864, 0.37369564, -0.02895943, 0.02912764, -0.01885849, 0.09628482, 0.14695522,
 0.32032562, -0.08749473, -0.02196906, 0.05900598, -0.00092907, 0.05998912, 0.09396983, -
 0.05141963, -0.02139149, 0.06831881, 0.16866586, 0.03790775, -0.01208502, -0.01917435, 0.1627997
 , 0.2424297 , 0.05674906, -0.00982601, 0.38153459, 0.0038117 , 0.27407849, 0.59432946]

3.2.2. Visualización

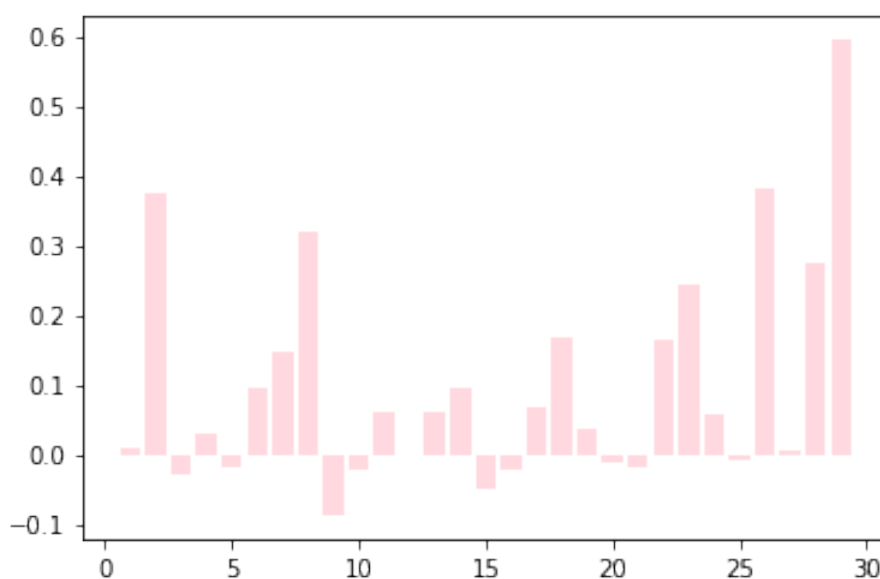


Figura 4: Significance Profile

3.3. Consigna 2.c

Dar con una explicación del resultado obtenido en el punto anterior.

3.3.1. Conclusiones

Lo que el significance profile nos aporta es un vector de importancia de los motifs normalizados. Observando en el gráfico podemos ver que el motifs 30 es el que más probable es. Esto indica que este motif referencia a los que más conectados están en la red.

4. Roles

4.1. Consigna 3

Detectar los roles en dicha red utilizando el algoritmo RolX, explicando el resultado obtenido.

4.1.1. Visualización

Para detectar los roles en la red utilizamos el algoritmo de RolX.

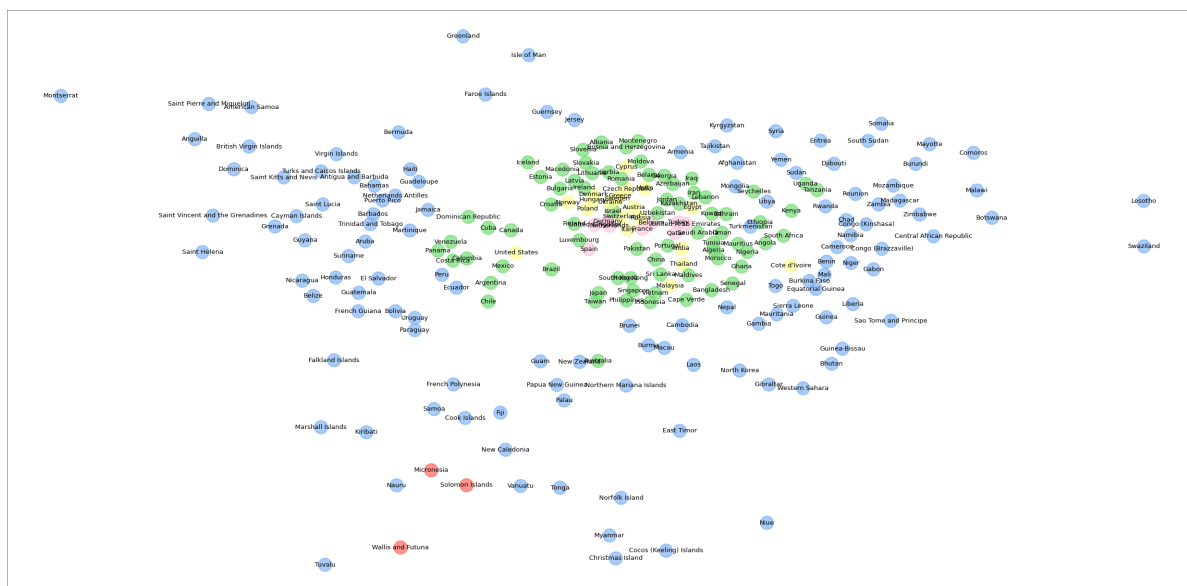


Figura 5: Roles detectados en la red

4.2. Conclusiones

Para este análisis primero vemos que los distintos roles se diferencian por color y que además, hay roles mucho más densos de nodos que otros. Si tomamos en cuenta que estamos considerando una red de vuelos parece lógico pensar que el rol mayoritario (leáse aquellos con mayor cantidad de nodos del mismo color), en este caso los de color azul, son aquellos aeropuertos que no son los más famosos y que se conectan solo con algunos aeropuertos. Pensemos que este razonamiento tiene sentido si consideramos que esto ocurre en la realidad: La mayoría de los países no tienen grandes conexiones de vuelos, ni mayoría de vuelos al lugar en cuestión, excepto que sean lugares más centrales, famosos, atractivos, etc, como por ejemplo Francia. Por otro lado, tenemos otros nodos con color verde que son el segundo rol con color dominante del gráfico, estos son nodos que tienen más conexiones con otros aeropuertos, ya sea porque son países que se usan de paso o bien como destino, en comparación con el rol anterior descripto (los de color azul). Luego, con mayor escala podemos ver que los colores amarillo y rosa corresponden a otros dos roles aún más reducidos que los anteriores, simbolizan los nodos más centrales de la red es decir aquellos que corresponden a países que poseen más conexiones y más vuelos al país, pero a diferencia de los nodos de color verde, estos nodos lo hacen a una escala mayor. La rosa aún mayor que la amarilla. Por último, también podemos detectar un rol de nodos de color rojo (tan solo 3 nodos), dichos nodos pueden interpretarse como aquellos lugares que son diametralmente opuestos a los nodos que veníamos describiendo, es decir, son nodos muy poco conocidos, o por lo menos con muy pocas conexiones de vuelos. Podemos entonces discernir que existen roles que tienen un carácter menos general que otras: O bien son lugares centrales que constituyen más conexiones de vuelos

ya sea porque son destinos atractivos o vuelos de tránsito. O bien porque son lugares muy poco transitados y apenas cuentan con conexiones aéreas o vuelos en sus aeropuertos. En ambos casos, hablamos de roles que no son tan frecuentes como otros. Con todo esto pueden concluirse que los roles siguen una relación de orden: De más general a menos general. Esto puede comprenderse bien teniendo en mente la distribución Normal, el comportamiento de los roles es similar: En los extremos encontraremos a los menos generales (que son menos que los más generales por alguna razón) mientras que en la parte central acampanada ubicamos a los roles más generales.