

Tarea 5

1985269

29 de abril de 2019

1. Descripción del experimento

Fueron generados cinco grafos de órdenes distintos, luego fueron escogidos dos nodos entre los cuales se calculó el máximo flujo que puede transportarse de uno a otro. Este procedimiento fue repetido con varios pares de nodos y en cada una de las repeticiones se determinaron los coeficientes de agrupamiento (clustering), centralidad (closeness centrality, load, centrality), excentricidad (eccentricity) etc. Los pesos de las aristas fueron asignados siguiendo una distribución normal. El objetivo de este experimento es determinar si existe alguna relación entre dichos coeficientes y el valor de la función objetivo o con el tiempo de ejecución. Así como determinar cuales son los mejores nodos fuentes y sumideros.

2. Generador de grafos

Fue utilizado el generador Watts–Strogatz small-world, este genera conexiones con una probabilidad predefinida. Se encuentra cierta similitud entre la manera en la que se genera este grafo con asentamientos poblacionales donde alrededor de las ciudades con una probabilidad mayor se fundan nuevos asentamientos y el flujo podría verse como el comercio.

A continuación se comparte el código de Python con el que se recopiló la información:

```

1 for i in range(5):
2     rango=random.randint(ordenes[i], ordenes[i]*2)
3     #G=nx.dense_gnm_random_graph(ordenes[i], rango)
4     G=nx.watts_strogatz_graph(ordenes[i], int(ordenes[i]/2) , 0.33 , seed=
None)
5     lista=[]
6     lista[:]=G.edges
7     width=np.arange(len(lista)*1, dtype=float).reshape(len(lista),1)
8     for r in range(len(lista)):
9         R=np.random.normal(loc=20, scale=5.0, size=None)
10        width[r]=R
11        G.add_edge(lista[r][0], lista[r][1], capacity=R)
12    for w in range(ordenes[i]):
13        initial=final=0
14        while initial==final:
15            initial=random.randint(0, round(len(G.nodes)/2))
16            final=random.randint(initial, len(G.nodes)-2)
17
18
19        tiempo_inicial=time()
20        T=nx.maximum_flow(G, initial, final)
21        tiempo_final =time()
22        tiempo_ejecucion=tiempo_final- tiempo_inicial
23
24        data[contador,2]=nx.clustering(G, nodes=initial)
25        data[contador,3]=nx.load_centrality(G, v=initial)
26        data[contador,4]=nx.closeness_centrality(G, u=initial)
27        data[contador,5]=nx.eccentricity(G, v=initial)
28        data[contador,6]=nx.pagerank(G, alpha=0.9)[initial]
29    #
30
31        data[contador,7]=nx.clustering(G, nodes=final)
32        data[contador,8]=nx.load_centrality(G, v=final)
33        data[contador,9]=nx.closeness_centrality(G, u=final)
34        data[contador,10]=nx.eccentricity(G, v=final)
35        data[contador,11]=nx.pagerank(G, alpha=0.9)[final]
36    #
37
38        data[contador,0]=T[0]
39        data[contador,1]=tiempo_ejecucion
40        data[contador,12]=ordenes[i]
41        contador+=1

```

3. Visualizació de los grafos Grafo

A continuación se presentan los grafos generados, los nodos marcados con formas y colores representan la fuente y el sumidero con verde y rojo respectivamente. Las aristas azules representan el flujo y su anchura es proporcional a su capacidad. Es posible observar como los grafos tienen ordenes distintos.

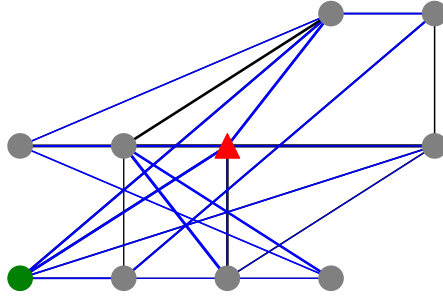


Figura 1: Grafo no dirigido cíclico de orden 10

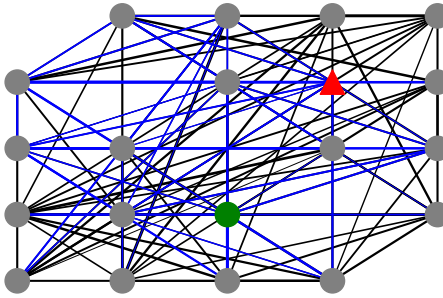


Figura 2: Grafo no dirigido cíclico de orden 20

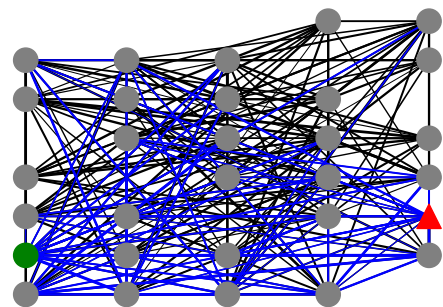


Figura 3: Grafo no dirigido cíclico de orden 30

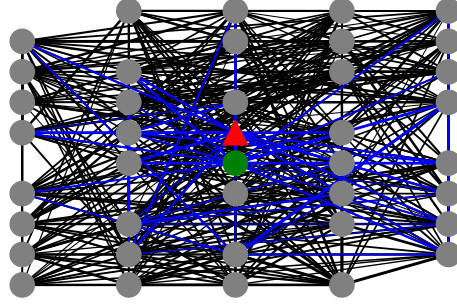


Figura 4: Grafo no dirigido cíclico de orden 40

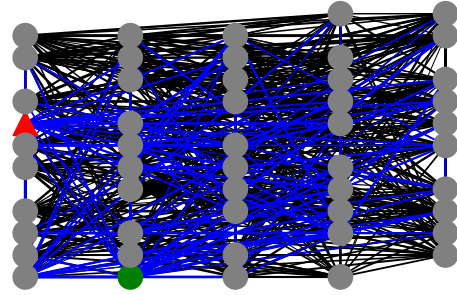
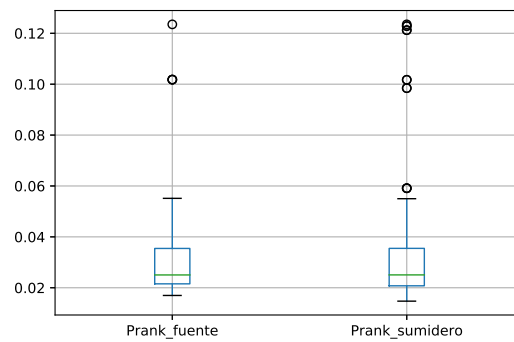
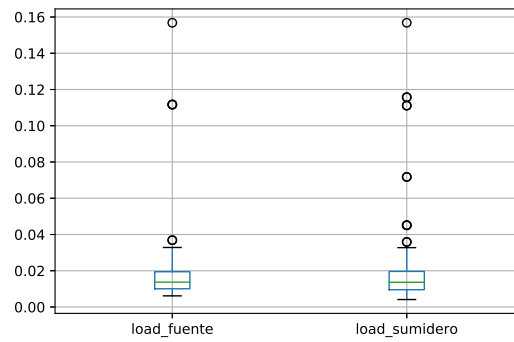
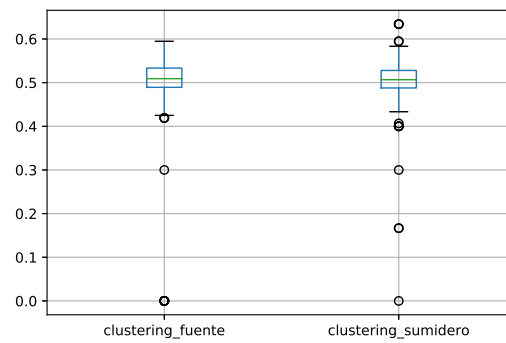
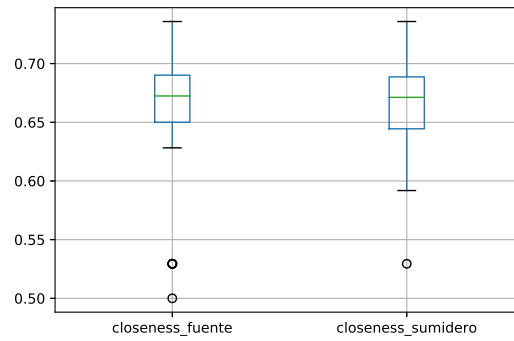
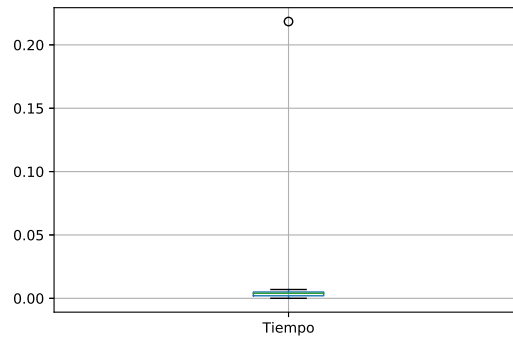


Figura 5: Grafo no dirigido cíclico de orden 50

4. Análisis de los datos

Los gráficos de caja y bigotes muestran como los datos obtenidos para cada uno de los nodos presentan ciertas similitudes, esto producto a que dependen del generador utilizado, es decir si fuera realizado un análisis entre distintos generadores resultaría una relación entre ellos y las características medidas para los nodos. El tiempo de ejecución registrado para el algoritmo de flujo máximo es pequeño para el tamaño de las instancias estudiadas. En la matriz de correlaciones el tiempo fila, columna 1 no se encuentra correlacionado fuertemente con ninguna otra variable, esto será ratificado más adelante. Aun así sí existen correlaciones fuertes entre las otras variables.





5. Análisis de varianza

Se realizó una prueba ANOVA para demostrar desigualdad entre las medias. Como indican los resultados de la prueba los valores p son pequeños las medias no son iguales, por lo cual se procedió a utilizar una prueba de mínimos cuadrados ordinarios a ver si con las variables registradas se pueden explicar los valores de la función objetivo.

ANOVA.txt

	sum_sq	...	PR>F
clustering_fuente	14340	...	0
load_fuente	60054	...	0
closeness_fuente	101311	...	0
eccentricity_fuente	140950	...	0
Prank_fuente	132493	...	0
clustering_sumidero	10	...	0
load_sumidero	15165	...	0
closeness_sumidero	505	...	0
eccentricity_sumidero	151097	...	0
Prank_sumidero	146	...	0.07
Residual	272297	...	NaN
c@FancyVerbLinee	272297	...	NaN

6. Mínimos cuadrados ordinarios

Los resultados de la regresión indican que con las variables medidas es posible explicar el comportamiento de los valores del flujo máximo con un error aceptable, el modelo hace una buena predicción. Las variables clustering no son significativas en el modelo al igual que load y page rank sumidero. Las características más influyentes de los nodos fuentes son load, closeness y eccentricity mientras que los sumideros closeness y load en ese orden.

En el segundo experimento se realizó la misma prueba para el tiempo de ejecución y el resultado fue como se esperaba que el modelo no explica el comportamiento del tiempo ya que este depende de otras variables. Como se puede observar en la matriz de correlación el tiempo se correlaciona con el valor objetivo y esto tiene sentido ya que si existen varias aristas por las cuales pasa flujo la exploración es más grande y el tiempo de ejecución aumenta.

OLS.txt

```

=====
                        OLS Regression Results
=====
Dep. Variable:                F0      R-squared:                0.920
Model:                        OLS      Adj. R-squared:            0.915
Method:                       Least Squares      F-statistic:            160.5
Date:                         Sat, 20 Apr 2019      Prob F-statistic:        3.61e-71
Time:                         12:27:00      Log-Likelihood:          -739.54
No. Observations:              150      AIC:                     1501.
Df Residuals:                  139      BIC:                     1534.
Df Model:                      10
Covariance Type:               nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept                -1268.0017      223.711      -5.668      0.000     -1710.318     -825.686
clustering_fuente           -7.3702       76.549     -0.096      0.923     -158.722     143.982
load_fuente                6235.1056      565.337     11.029      0.000     5117.334     7352.877
closeness_fuente           718.4359      163.935      4.382      0.000      394.308     1042.564
eccentricity_fuente        222.4259       57.859      3.844      0.000      108.029     336.823
Prank_fuente              -1.034e+04     1317.746     -7.848      0.000    -1.29e+04    -7736.633
clustering_sumidero         39.8599       79.430      0.502      0.617     -117.187     196.907
load_sumidero              200.1120      467.998      0.428      0.670     -725.203     1125.427
closeness_sumidero        1156.9662      176.495      6.555      0.000      808.005     1505.928
eccentricity_sumidero       78.2676       27.680      2.828      0.005       23.540     132.996
Prank_sumidero            -1989.1513     1327.265     -1.499      0.136     -4613.390     635.087
=====
Omnibus:                   1.709      Durbin-Watson:           1.384
ProbOmnibus:                0.425      Jarque-Bera JB:          1.486
Skew:                      -0.243      ProbJB:                  0.476
Kurtosis:                   3.038      Cond. No.                2.18e+03
=====
c@FancyVerbLinee=====

```

OLS Regression Results

```

=====
Dep. Variable:      Tiempo      R-squared:      0.125
Model:              OLS        Adj. R-squared:  0.062
Method:             Least Squares      F-statistic:    1.978
Date:               Sat, 20 Apr 2019    Prob F-statistic: 0.0400
Time:                12:27:00          Log-Likelihood:  485.89
No. Observations:   150              AIC:             -949.8
Df Residuals:       139              BIC:             -916.7
Df Model:           10
Covariance Type:    nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0893	0.063	1.409	0.161	-0.036	0.215
clustering_fuente	-0.0162	0.022	-0.747	0.456	-0.059	0.027
load_fuente	0.0033	0.160	0.021	0.984	-0.313	0.320
closeness_fuente	0.0226	0.046	0.486	0.628	-0.069	0.114
eccentricity_fuente	-0.0082	0.016	-0.499	0.618	-0.041	0.024
Prank_fuente	-0.1936	0.373	-0.519	0.605	-0.931	0.544
clustering_sumidero	0.0083	0.022	0.367	0.714	-0.036	0.053
load_sumidero	0.0069	0.133	0.052	0.959	-0.255	0.269
closeness_sumidero	-0.0917	0.050	-1.835	0.069	-0.190	0.007
eccentricity_sumidero	-0.0064	0.008	-0.820	0.413	-0.022	0.009
Prank_sumidero	0.0528	0.376	0.141	0.888	-0.690	0.796

```

=====
Omnibus:      143.935    Durbin-Watson:      2.152
ProbOmnibus:  0.000    Jarque-Bera JB:      1529.659
Skew:         3.756    ProbJB:              0.00
Kurtosis:     16.722    Cond. No.            2.18e+03
=====
c@FancyVerbLinee=====

```

7. Conclusiones

Fue realizada la experimentación propuesta y los resultados fueron analizados, dicho análisis arrojó que con las variables medidas es posible encontrar un modelo que explique el valor objetivo del algoritmo del flujo máximo utilizado. Los mejores nodos fuentes son aquellos que tienen un mayor valor de load, closeness y eccentricity. Mientras que los mejores sumidero son aquellos que presentan mayor valor de closeness. Además se comprobó que no existe una correlación fuerte entre el tiempo de ejecución y estas variables aunque existe cierta relación positiva entre el tiempo y el valor del flujo máximo.

Referencias

- [1] Janet M Six and Ioannis G Tollis. A framework for circular drawings of networks. In *International Symposium on Graph Drawing*, pages 107–116. Springer, 1999.
- [2] Yehuda Koren. On spectral graph drawing. In *International Computing and Combinatorics Conference*, pages 496–508. Springer, 2003.

- [3] EW Mayr. Praktikum algorithmen-entwurf (teil 6), nov. 2002, 6–11. *Technische Universität München* <http://www.mayr.in.tum.de/lehre/2002WS/algoprak/part6.ps.gz>.
- [4] E. Shaeffer. <https://elisa.dyndns-web.com/>.