



**UNIVERSIDAD
DE GRANADA**

GESTIÓN DE INFORMACIÓN EN LA WEB
**Análisis preliminar y
visualización básica de una red
social con *Gephi***

PRÁCTICA 1

Víctor Vázquez Rodríguez
victorvazrod@correo.ugr.es
76664636R

Máster Universitario en Ingeniería Informática
Curso 2019/20

Índice

1. Introducción	2
2. Análisis	4
Referencias	10

1. Introducción

La red que se ha elegido para realizar esta práctica es la de *Political blogs* [1] de la [página web de Mark Newman](#). Como su propio nombre indica, se trata de una red de blogs sobre política de los Estados Unidos de América, donde los nodos son los blogs y las aristas, los enlaces de una web a otra, tratándose entonces de un grafo dirigido.

En la figura 1 se puede ver la representación gráfica de la red. Como se puede apreciar, la red se ve muy pequeña, esto se debe a que hay muchos nodos que no tienen relación con ningún otro, de manera que están aislados. En la figura, estos nodos forman un círculo casi imperceptible alrededor de la componente principal.

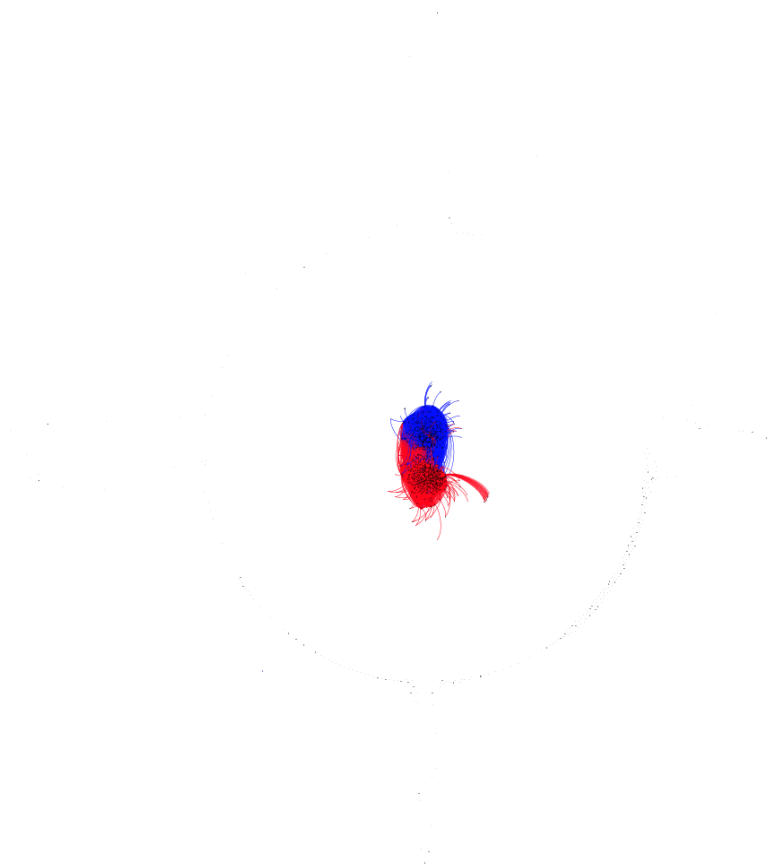


Figura 1: Representación gráfica de la red

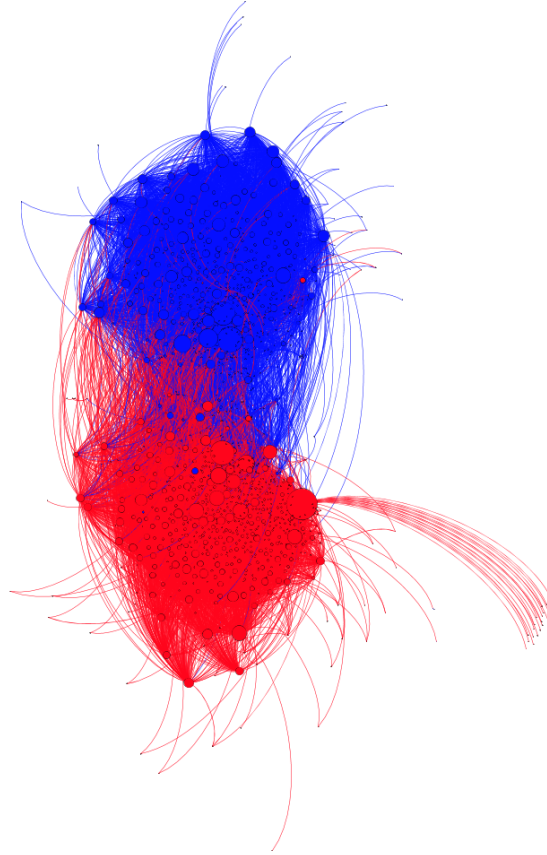


Figura 2: Representación gráfica de la componente gigante

Si filtramos la red para quedarnos solo con esta componente gigante, obtenemos el grafo de la figura 2, que es mucho más apreciable. Se han utilizado los colores clásicos para representar las ideologías políticas en los EEUU: el azul para la izquierda o los liberales y el rojo para la derecha o los conservadores. El tamaño de un nodo refleja el grado total del mismo, de forma que los nodos más grandes representan blogs que realizan/reciben más enlaces. Como era de esperar, los blogs de cada ideología están fuertemente agrupados entre ellos, habiendo también bastantes referencias de algunos blogs de cada ideología hacia la otra.

2. Análisis

Usando *Gephi*, se han obtenido algunas medidas básicas de la red, las cuales se pueden ver en el cuadro 1. Fijándonos en estas medidas, lo primero que destacamos es que la densidad del grafo es muy baja, lo cual ya intuíamos por el anillo de nodos aislados del que hablábamos en la figura 1. Esto también se refleja en la medida del número de nodos y de aristas que contiene la componente gigante, ya que se "pierden" el 17.99 % de los nodos pero se mantienen casi el 100 % de las aristas.

Cuadro 1: Medidas obtenidas de la red

Medida	Valor
Número de nodos N	1490
Número de enlaces L	19025
Número máximo de enlaces L_{max}	2113889
Densidad del grafo L/L_{max}	0.009
Grado medio $\langle k \rangle$	12.768
Diámetro d_{max}	9
Distancia media d	3.3902
Coefficiente medio de <i>clustering</i> $\langle C \rangle$	0.172
Número de componentes conexas	268
Número de nodos componente gigante (y %)	1222 (82.01 %)
Número de aristas componente gigante (y %)	19024 (99.99 %)

En la figura 3 vemos la distribución del grado de la red, es decir, del número de enlaces que conectan cada nodo con otros. Tenemos 4 *hubs* que tienen más de 350 enlaces, pero el 56.31 % de los nodos no tienen más de 12 enlaces (el grado medio es de 12.768).

Al tratarse de un grafo dirigido, vamos a fijarnos también en las distribuciones del prestigio de entrada y de salida, representadas en las figuras 4 y 5, respectivamente. Mientras que solo hay un nodo que realiza referencias a más de 250 blogs, tenemos hasta 9 nodos a los que llegan enlaces de más de 200. Entendemos entonces que hay unos pocos blogs más importantes o conocidos a los que enlazan muchos otros blogs más pequeños.

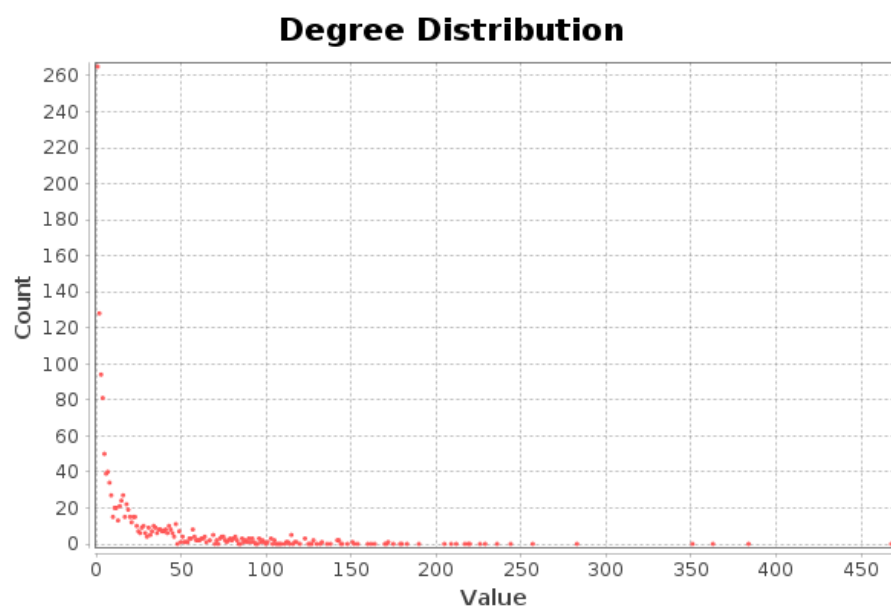


Figura 3: Distribución del grado

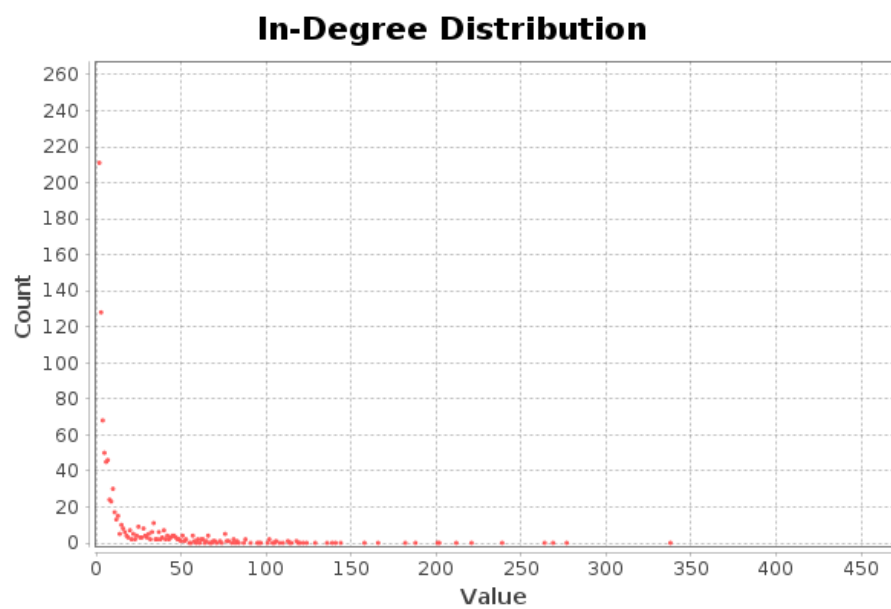


Figura 4: Distribución del prestigio de entrada

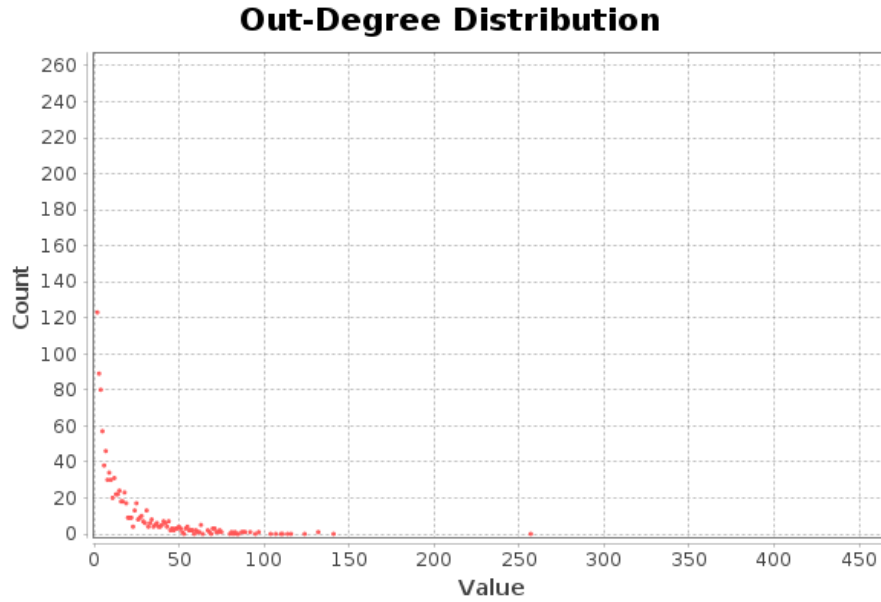


Figura 5: Distribución del prestigio de salida

Volviendo a fijarnos en el cuadro 1, vemos como la red tiene diámetro 9, mientras que la distancia media de los caminos es de 3.3902. Calculando esta medida, también hemos obtenido distribuciones relativas a las distintas medidas de cercanía, que se pueden ver en las figuras 6 a 9. Todas estas medidas han sido normalizadas.

En la distribución de la intermediación (figura 8), nos encontramos con valores muy bajos, por lo que no tenemos ningún nodo concreto que actúe como puente o portero entre grupos de nodos.

En cuanto a la cercanía (figura 6), sí que tenemos algunos nodos con un valor muy alto, es decir, que se encuentran en una posición muy céntrica dentro de la red.

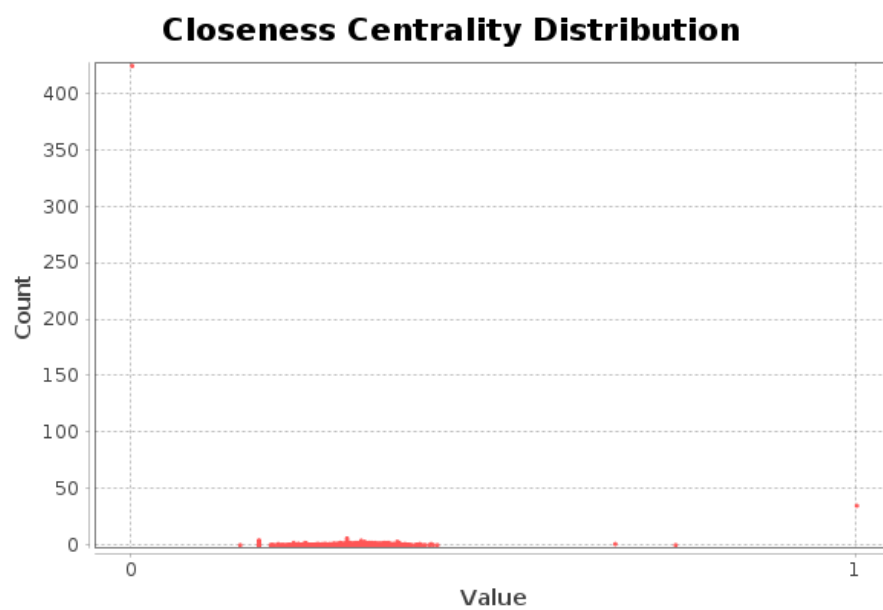


Figura 6: Distribución de la centralidad de cercanía

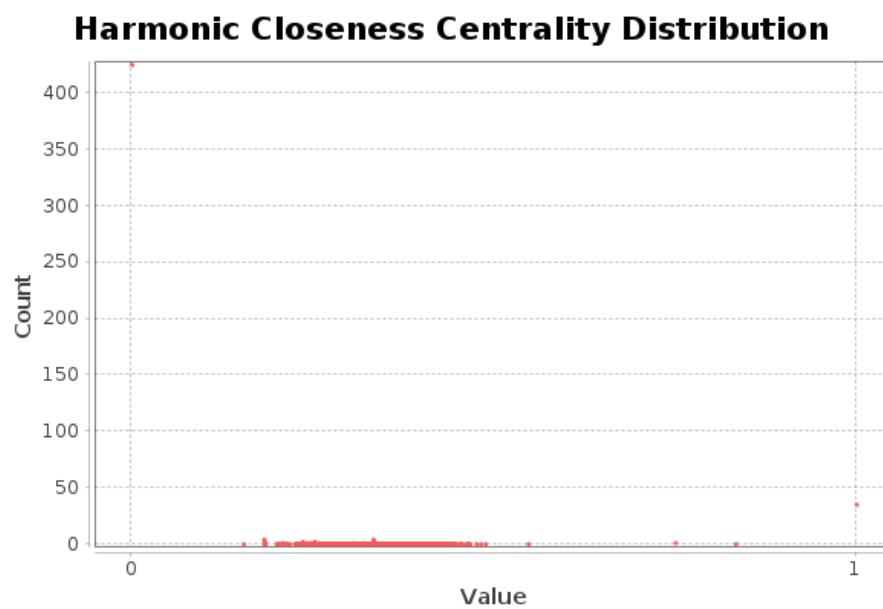


Figura 7: Distribución de la centralidad armónica de cercanía

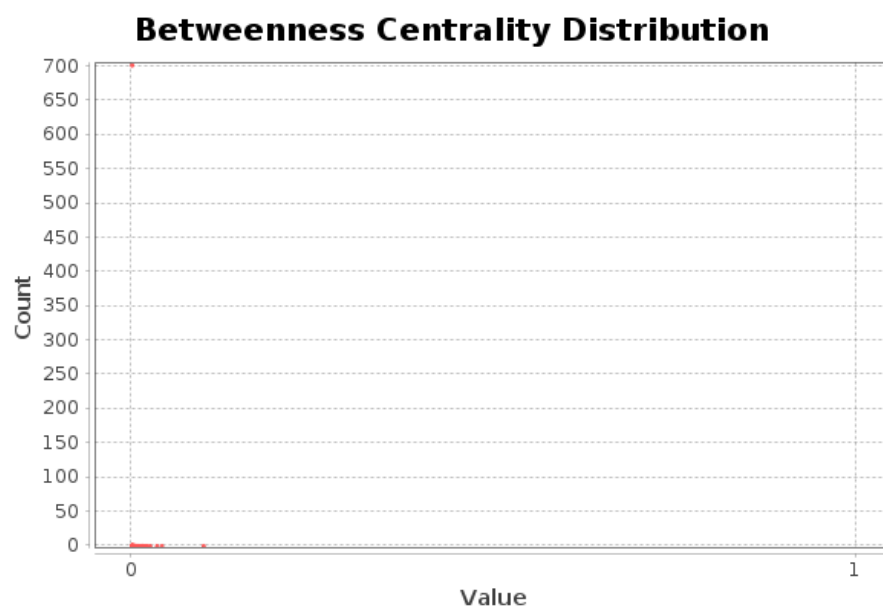


Figura 8: Distribución de la centralidad de intermediación

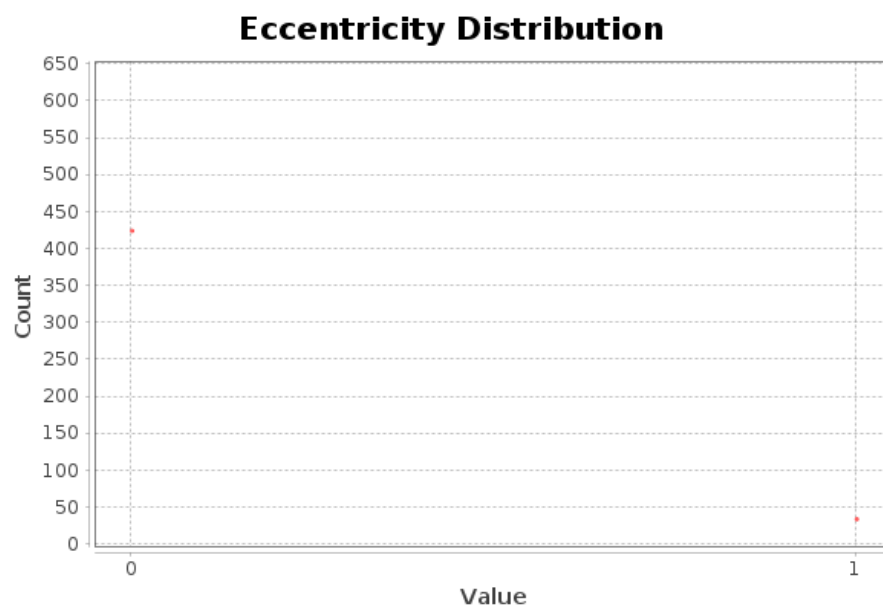


Figura 9: Distribución de la excentricidad

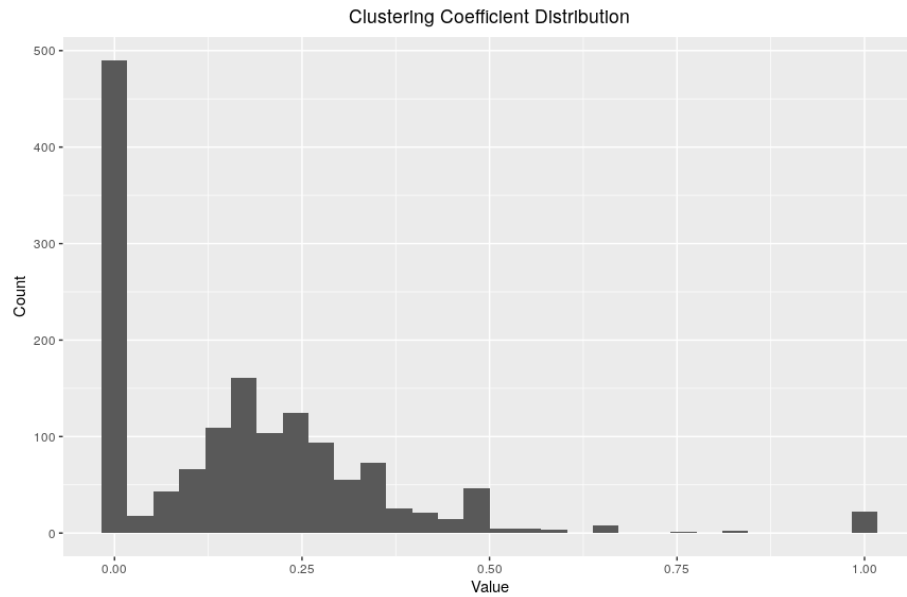


Figura 10: Distribución del coeficiente de *clustering*

En cuanto a la conectividad, tenemos 268 componentes conexas, muchas de ellas presentes en el “cinturón” de nodos aislados del que hemos hablado en la figura 1.

El coeficiente de clustering, cuya distribución tenemos en la figura 10, tiene un valor medio de 0.172, lo que es bastante bajo. En la distribución vemos que la mayoría de nodos tienen un valor de este coeficiente inferior a 0.5. Esto puede deberse a que, como ya hemos visto, la red es muy poco densa.

Referencias

- [1] L. A. Adamic and N. Glance, “The political blogosphere and the 2004 us election,” in *Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem*, 2005.