



Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación
Máster Profesional en Ingeniería Informática

Curso 2020/2021

PRÁCTICA 3

Gestión de Información en la Web

Breve descripción

Análisis Preliminar y Visualización Básica de una Red Social con Gephi

Autor

Álvaro de la Flor Bonilla (alvdebon@correo.ugr.es) 15408846-L

Propiedad Intelectual

Universidad de Granada

RESUMEN

El objetivo de esta primera práctica es doble. Por un lado, familiarizarse con los procedimientos de análisis de redes y con las medidas habitualmente consideradas para esta tarea. Por otro, aprender el manejo de una herramienta estándar de análisis y visualización de redes como Gephi 1, disponible para su descarga en <https://gephi.org/users/download/>.

Para ello, se requerirá que el estudiante realice un análisis de una red social cargándola en la herramienta, visualizándola y calculando los valores de una serie de medidas estándar de análisis de redes para estudiar las características principales de la misma.



ÍNDICE DEL PROYECTO

Resumen 1

1 Introducción 4

2 Análisis 5

 2.1 Métricas globales..... 5

 2.2 Estudio 5

ÍNDICE DE ILUSTRACIONES

Ilustración 1 – Histograma de distribución de grados de entrada	6
Ilustración 2 – Histograma de distribución de grados de salida.....	6
Ilustración 3 – Histograma de excentricidad.....	7
Ilustración 4 – Red coloreada en función de grados de entrada.....	7

1 INTRODUCCIÓN

En esta primera práctica se analizará una red procedente del [este repositorio](#), en concreto este repositorio forma parte del entorno oficial de “Gephi” en “Github”. En concreto en nuestro caso hemos elegido la red “CPAN authors” que puede descargar a través de [este enlace](#).

CPAN es un proyecto de visualización cuyo objetivo es analizar las relaciones entre los desarrolladores y los paquetes del lenguaje Perl, conocidos como la comunidad CPAN. Esta instantánea fue creada por Linkfluence en julio de 2009. Este archivo contiene la red de desarrolladores, vinculados cuando utilizan el mismo módulo “Perl”.

Por otro lado, me ha parecido interesante utilizar esta red ya que uno de los profesores de nuestra escuela, [Juan Julián Merelo Guervós](#) forma parte de la red de desarrolladores de “Perl”.

El fichero descargado se encuentra en la extensión “.gexf” por lo que no será necesario realizar ningún pre-procesamiento para comenzar a trabajar con él.

2 ANÁLISIS

2.1 Métricas globales

La interfaz de “Gephi” nos proporciona el siguiente cálculo de alguna de las medidas de la red, las cuales son:

MEDIDA	VALOR
Número de nodos N	840
Número de enlaces L	2248
Número máximo de enlaces L	704 760
Densidad del grafo L / L_{MAX}	0.003
Grado medio $\langle K \rangle$	2.676
Diámetro d_{MAX}	9
Distancia media $\langle d \rangle$	4.357882096069869
Coeficiente medio de <i>clustering</i> $\langle C \rangle$	0.233
Número de componentes conexas	2
Número de nodos de la componente gigante (y %)	839 (99.88 %)
Número de enlaces de la componente gigante (y %)	2248 (100 %)

2.2 Estudio

Nuestra red esta compuesta por un total de 840 nodos, que corresponden a los 840 desarrolladores que participaron en los módulos de “Perl”. Por otro lado, cuenta con 2248 enlaces, que corresponden con las relaciones entre los paquetes y los desarrolladores.

Para realizar el cálculo del número máximo de enlaces hemos utilizado la fórmula $(N * (N - 1))$ donde N es el número de nodos de la red.

En cuanto a la densidad, como podemos ver, es bastante baja (0.003). Como consecuencia de este dato podemos afirmar que la conectividad de nuestra red es bastante baja, por lo que habrá muchos nodos aislados.

Por otro lado, el valor del grado medio también es bastante bajo, en nuestro caso de 2.676 lo que indica que de media cada nodo tiene un máximo de 4 a 5 nodos. Es decir, de media cada desarrollador esta conectado con de 4 a 5 desarrolladores más.

Continuando con el histograma de las distribuciones de los grados de entrada se observa que la mayoría de los desarrolladores son pocos referenciados, ya que la mayoría de ellos se concentran en la zona del lateral izquierdo que podrán ver en la siguiente imagen. Sin embargo, existen algunas excepciones, como los casos aislados que vemos sobre la zona de la derecha. Estos casos son los denominados “hubs” que son aquellos desarrolladores a los que se les ha referenciado mucho.

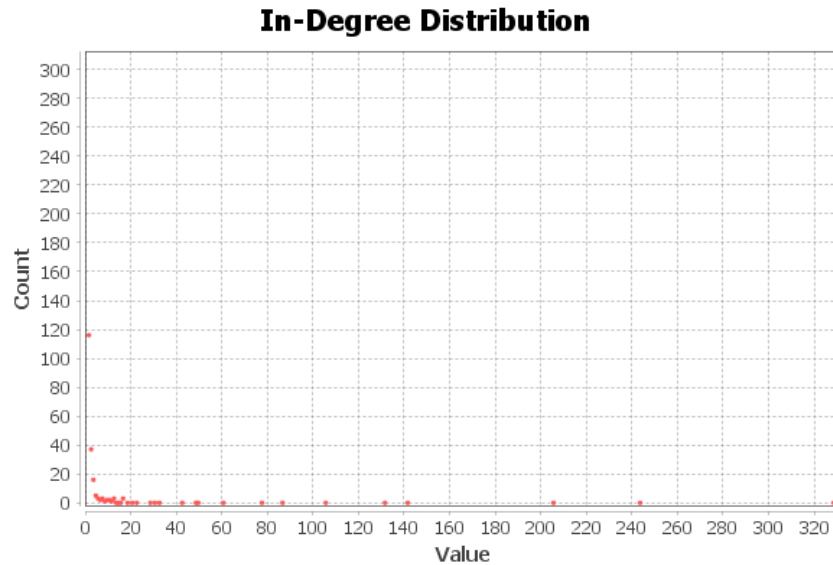


Ilustración 1 – Histograma de distribución de grados de entrada

Continuando con la distribución de los grados de salida se continua con la dinámica anterior ya que la mayoría de los desarrolladores no referencian mucho a otros colegas. Sin embargo, en esta ocasión hay una menor variedad de desarrolladores que referencian a otros, incluso hay menos cantidad de los llamados “hubs” aislados que mencionamos anteriormente.

¿Cómo podemos explicar esto? Pues se me ocurre que existen un grupo sólido de desarrolladores aislados que se han encargado de la mayoría del código, mientras que el resto se ha ocupado de pequeñas dependencias.

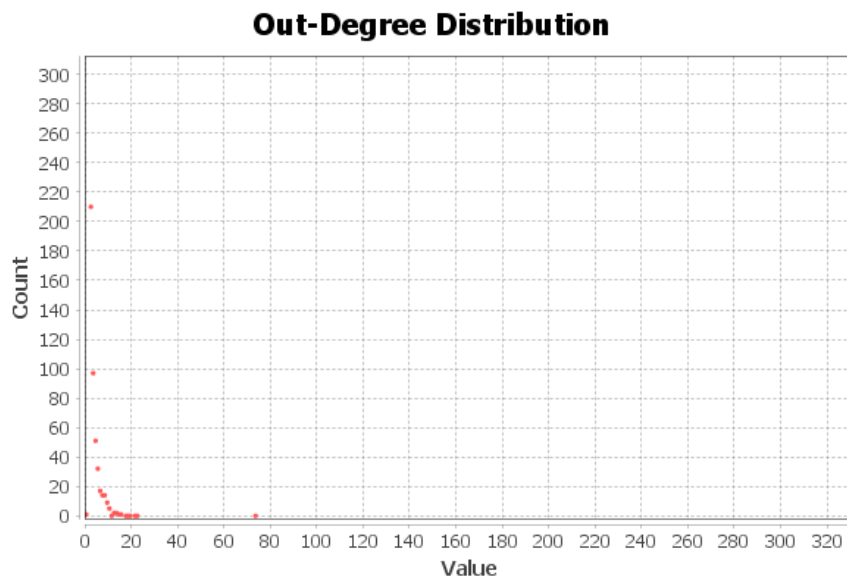


Ilustración 2 – Histograma de distribución de grados de salida

Respecto a las medidas al diámetro, que representa el número de nodos que se deben usar para alcanzar al nodo más lejano, en nuestro caso es 9. Es una distancia bastante corta pero también tenemos que señalar que tampoco tenemos una red muy grande (839 nodos).

La distancia media (que en nuestro caso es 4.36) nos indica el camino más corto para llegar de un desarrollador a otro.

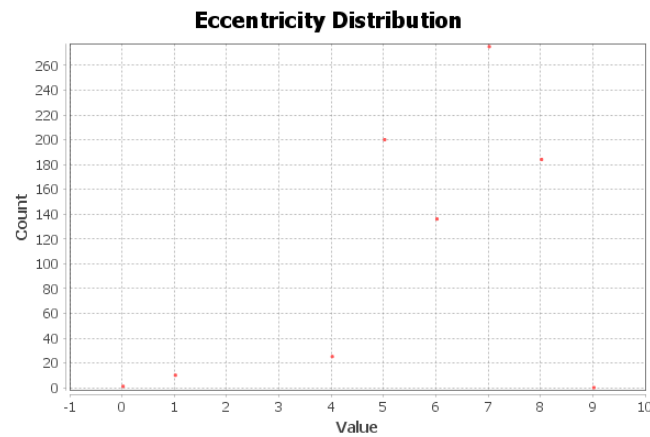


Ilustración 3 – Histograma de excentricidad

En cuanto a la conectividad, en nuestro caso tenemos solo dos componentes conexas, lo cual indica que esta red se puede dividir en hasta 2 subgrafos. La mayor de ellas, conocida como la componente gigante alberga el 99.88% de los nodos totales y el 100% de aristas, es decir, lo que no forma parte de la componente gigantes son nodos aislados.

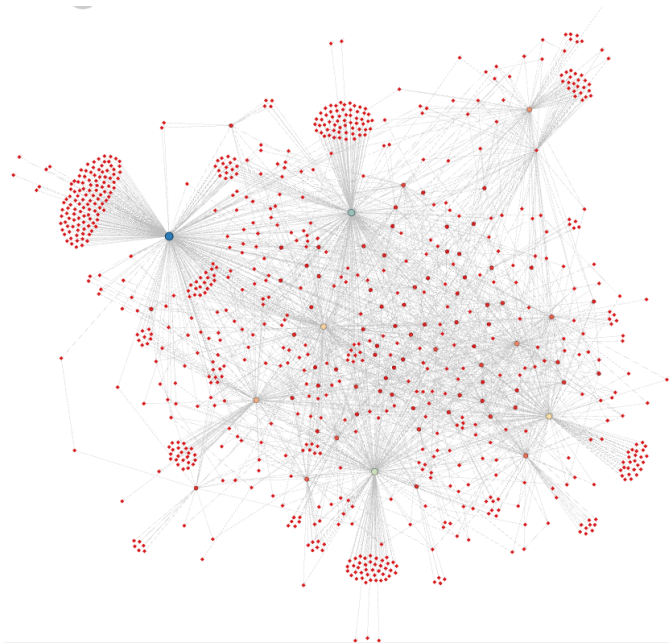


Ilustración 4 – Red coloreada en función de grados de entrada

Continuando con la explicación que dimos antes, es en esta imagen donde podemos apreciar claramente como, a partir de un grupo reducido de desarrolladores (colores diferentes al rojo), podemos ver como se distribuyen el resto de relaciones.