Gestión de Información en la Web
Master en Ingeniería Informática

Práctica 2: Análisis y Evaluación de Redes en Twitter

Luis Alberto Segura Delgado
DNI: 45922174-Y
segura2010@correo.ugr.es

Martes 5 de Abril de 2016

Índice

1 Introducción

El objetivo de esta segunda práctica es formalizar todos los conocimientos adquiridos en el curso aplicándolos a un caso real de análisis de una red social online generada a partir de un medio social. Para ello, se ha seleccionado un medio social concreto (Twitter) y una pregunta de investigación. A partir del medio social elegido, se obtendrá el conjunto de datos y se construirá una red social, que será analizada con objetivo de responder a la pregunta de investigación planteada.

2 Trabajo Realizado

En esta sección se detalla el trabajo realizado en la práctica, indicando en primer lugar el problema concreto que se ha planteado y el conjunto de datos y la forma de obtenerlos para resolver dicho problema. A continuación se explicará el análisis realizado sobre los datos y la red social obtenida y finalmente las conclusiones obtenidas del estudio.

2.1 Descripción del Problema

El problema a estudiar es detectar cuales son los usuarios más relevantes en la discusión de Twitter sobre la emisión en **Periscope**[1] que tuvo lugar el día 25 de marzo, organizada por Gerard Piqué[2].
Para abordar el problema, se han recopilado tweets publicados durante la emisión en los que se mencionaba a Piqué (@3gerardpique) y se incluía la palabra "Periscope". Y como la obtención de los datos se realizó unos días después, se han limitado la búsqueda a los tweets que se publicaron el día 25 de Marzo, día de la emisión[3]. Para obtener los tweets, se ha utilizado la herramienta NodeXL.

De cara a evaluar la red correctamente, se ha decidido eliminar el nodo de Piqué de la red, pues todos los tweets lo mencionan, por tanto se conecta con todos los usuarios, y esto dificulta el análisis de la red y su visualización al mismo tiempo que no resulta interesante.

2.2 Cálculo de los valores de las medidas de análisis

Para el análisis de la red se ha utilizado la herramienta **Gephi**.
Nuestra red social tiene los siguiente valores para las medidas de análisis:

- **Número de Nodos (N)**: 1763
- **Número de Enlaces (L)**: 1464
- **Densidad (D)**: 0.001
- **Grado Medio** ($< k >$): 1.661
- **Diámetro** (d_{max}): 2
- **Distancia Media** ($< d >$): 1.013
- **Distancia Media para la red aleatoria equivalente** ($< d_{aleatoria} >= \frac{log(N)}{log(<k>)}$): 14.73
- **Coeficiente de Clustering Medio** ($< C >$): 0.05

[1] https://www.periscope.tv
[2] http://as.com/videos/2016/03/25/portada/1458916408_738738.html
[3] Búsqueda avanzada de Twitter: @3gerardpique periscope since:2016-03-25 until:2016-03-26 (https://twitter.com/search?vertical=default&q=%403gerardpique%20periscope%20since%3A2016-03-25%20until%3A2016-03-26&src=typd)

- **Coeficiente de Clustering Medio para la red aleatoria equivalente** ($< C_{aleatoria} >= \frac{<k>}{N}$): 0.0009

El número de componentes conexas es de 929, mientras que 883 de los nodos no están conectados con ningún otro, ya que los usuario mencionan principalmente a Piqué (eliminado de la red) y a Iker Casillas. En general los usuarios no se mencionan entre si, salvo excepciones. La componente gigante de nuestra red es Iker Casillas (@casillasworld), ya que recibe la mayor parte de menciones de los usuarios. Tiene un grado de 434 (grado de entrada=434 ; grado de salida=0), por tanto 434 aristas de las 1464 totales son dirigidas a Casillas (un 29.64%). Como vemos, Casillas es, principalmente, el centro de la red. Cosa que tiene sentido, pues es conocido y es el protagonista de la emisión junto a Piqué. Sin embargo, como veremos más adelante, hay otros usuarios importantes en nuestra red, que son mencionados por los espectadores a la hora de comentar el evento.

Los nodos que hemos visto, cuyo grado es 0, no están conectados a ningún otro nodo de la red, pues mencionaban únicamente a Piqué (que ha sido eliminado de la red por ser el centro de la misma). Estos usuarios, no son de gran interés y en general son usuarios aislados que en algún momento han comentado la emisión. El resto de usuario que si están conectados a otros usuarios nos ayudarán a detectar comunidades y actores relevantes de nuestra red. En cuanto a la detección de comunidades, veremos más adelante exactamente que nos cuentan, pero a priori las comunidades de nuestra red deberían representar conversaciones entre diferentes usuarios sobre el evento.

2.3 Propiedades de la red

2.3.1 Distribuciones de Grados

En las figuras 1 y 2 podemos ver las distribuciones de grados de entrada y salida respectivamente, mientras que en la figura 3 podemos ver la distribución global de grados.

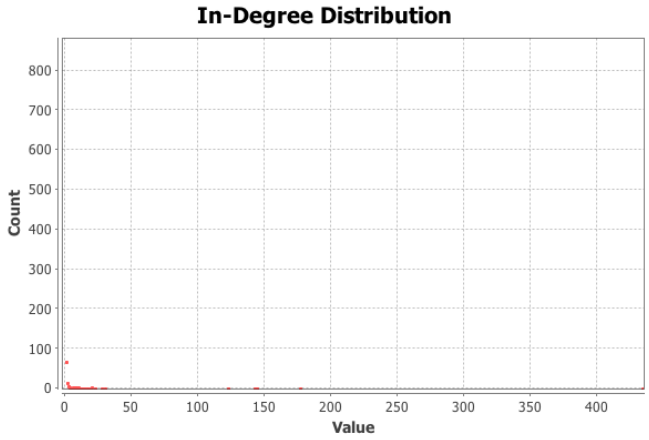

Figura 1: Distribución de grados de entrada

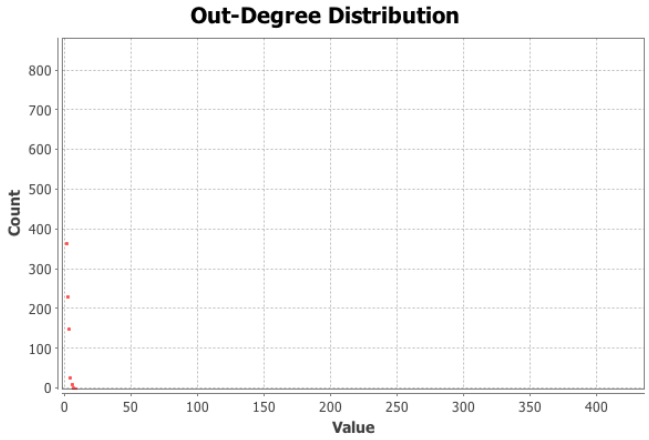

Figura 2: Distribución de grados de salida

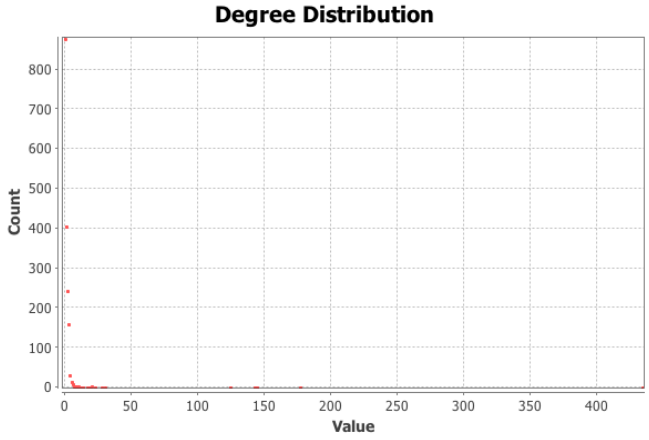

Figura 3: Distribución de grados

A partir de estas distribuciones de probabilidad de los grados de entrada y salida de los nodos de nuestra red social, podemos determinar que nuestra red es libre de escala, y por tanto, que sigue la **Ley de la Potencia**. Las gráficas nos muestran una distribución de larga estela, como ya se ha mencionado en teoría, por ello, **podemos deducir que la red es libre de escala**.

2.3.2 Distribuciones de Distancias

En nuestra red pequeña, nos encontramos ante un caso de *mundo ultra-pequeño*, ya que como vimos anteriormente la distancia media ($< d >$) tenía un valor de 1.01, mientras que la distancia media para una red aleatoria equivalente era de 14.73. Para un mundo ultra-pequeño, la distancia media debe ser aún menor que $\frac{log(N)}{log(log(N))}$, que en este caso obtiene un valor de 3.715, por lo que la distancia media de nuestra red tiene un valor más bajo incluso que la distancia media para un mundo ultra-pequeño. Por tanto, podemos concluir que **nuestra red social es un mundo ultra-pequeño**.

2.3.3 Coeficiente de Clustering

El coeficiente de Clustering nos permite conocer la densidad local de la red, o dicho en otras palabras, la proporción de los vecinos de cada nodo que están conectados. En nuestra red, el coeficiente de clustering

medio es de 0.05. Tenemos un coeficiente de clustering muy bajo, cosa que es lógica si nos fijamos en la gran cantidad de nodos que no están conectados con ningún otro nodo de la red. Como ya vimos antes, al eliminar a Piqué de la red, gran cantidad de nodos quedan totalmente desconectado y no tienen ninguna arista que los una a otros.

2.4 Calculo de los valores de las medidas de análisis de redes sociales

Ya hemos visto el valor medio de los grados, ahora vamos a ver para cada usuario el grado concreto y a tratar de analizar, a partir de esta y otras de las medidas de centralidad, cuales son los usuarios más importantes de nuestra red social. Para empezar, vamos a analizar los actores más interesantes de nuestra red en base al grado. Al trabajar con una red dirigida, trabajaremos con dos grados, el de entrada (que nos indica el prestigio de un usuario a la hora de ser citado/mencionado) y el de salida (que nos indica el alcance de la influencia de un usuario). En nuestro problema, nos interesa más el grado de entrada, pues nos indica los usuarios que han sido más mencionados, y por tanto, los más conocidos y/o interesantes para el resto de usuarios. Como no podría ser de otra manera, el usuario más mencionado (con mayor grado de entrada) es Iker Casillas. Esto ya lo habíamos visto antes, y como decíamos, es lógico al ser el protagonista de la emisión. Veamos que otros usuarios son también mencionados de forma frecuente.

Usuario	Grado Entrada	Grado Salida
casillasworld	434	0
as_tomasroncero	177	0
elchirincirco	144	0
jpedrerol	143	0
elsimiolopez	123	1
juanmacastano	30	0
hoyendeportes4	28	0
barcastuff	23	0
chuycorona25	20	0
miguel_layun	20	0
mundodeportivo	18	1
abc_deportes	17	0
eukarolyi	15	3
miseleccionmx	14	0
txikiforero	13	0
sefutbol	12	0
sergioramos	10	0
sientelaroja	10	0
jordialba	9	0
marcbartra	9	0

Tabla 1: Usuarios ordenados por grado de entrada

En la tabla 1 podemos ver la lista de usuarios más mencionados. Como podemos ver, a parte de Casillas, el usuario más importante de nuestra red es Tomás Roncero (@as_tomasroncero), seguido por @elchirincirco y @jpedrerol. Estos resultados me parecen interesantes y curiosos, pues muestran algunos detalles interesantes en nuestro problema. Es curioso que los usuarios más mencionados, y por tanto más importantes de nuestra red desde el punto de vista del grado de entrada, no sean jugadores de fútbol que se encontrasen convocados con la selección española, ya que en el momento de la emisión los jugadores estaban convocados y concentrados

en el hotel para afrontar un serie de partidos amistosos. De hecho, me resulta curioso que haya usuarios mas mencionados que Cesc Fabregas, que aparecía junto a Piqué y Casillas en la emisión (aunque con mucho menos protagonismo). Pero de estas curiosidades que no muestra la red, hablaremos más tarde en la sección dedicada a conclusiones.

2.5 Descubrimiento de comunidades

2.6 Visualización de la red social

2.7 Discusión de los resultados y Conclusiones