

# **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

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>Trabajo Realizado</b>	<b>3</b>
2.1	Descripción del Problema . . . . .	3
2.2	Cálculo de los valores de las medidas de análisis . . . . .	3
2.3	Propiedades de la red . . . . .	4
2.4	Calculo de los valores de las medidas de análisis de redes sociales . . . . .	4
2.5	Descubrimiento de comunidades . . . . .	4
2.6	Visualización de la red social . . . . .	4
2.7	Discusión de los resultados y Conclusiones . . . . .	4

# 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**<sup>1</sup> que tuvo lugar el día 25 de marzo, organizada por Gerard Piqué<sup>2</sup>.

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<sup>3</sup>. 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 ( $\langle k \rangle$ ):** 1.661
- **Diámetro ( $d_{max}$ ):** 2
- **Distancia Media ( $\langle d \rangle$ ):** 1.013
- **Distancia Media para la red aleatoria equivalente ( $\langle d_{aleatoria} \rangle = \frac{\log(N)}{\log(\langle k \rangle)}$ ):** 5.885
- **Coefficiente de Clustering Medio ( $\langle C \rangle$ ):** 0.05

---

<sup>1</sup><https://www.periscope.tv>

<sup>2</sup>[http://as.com/videos/2016/03/25/portada/1458916408\\_738738.html](http://as.com/videos/2016/03/25/portada/1458916408_738738.html)

<sup>3</sup>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>)

- **Coefficiente de Clustering Medio para la red aleatoria equivalente** ( $\langle C_{aleatoria} \rangle = \frac{\langle k \rangle}{N}$ ): 0.002

El número de componentes conexas es de 929, mientras que 883 de los nodos no están conectados con ningún otro, ya que los usuarios mencionan principalmente a Piqué (eliminado de la red) y a Iker Casillas, en general los usuarios no se mencionan entre sí, 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%).

## 2.3 Propiedades de la red

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

## 2.5 Descubrimiento de comunidades

## 2.6 Visualización de la red social

## 2.7 Discusión de los resultados y Conclusiones