
MÁSTER EN BIG DATA & BUSINESS ANALYTICS

MÓDULO IX: APLICACIONES ANALÍTICAS

Documentación del Caso de Estudio
Analítica en Redes Sociales

17 de Diciembre del 2018

Autor

José María Sánchez Salas
josemaria.sanchezsalas@gmail.com

Índice

Módulo IX: Aplicaciones Analíticas		
1	Introducción	5
2	Enunciado	5
3	Solución	6
3.1	Ejercicio básico	6
3.2	Ejercicio intermedio	9
3.3	Ejercicio avanzado	10

1 Introducción

Este documento contiene la resolución del caso de estudio Analítica en Redes Sociales del Módulo IX: Aplicaciones Analíticas, del M. Big Data & Business Analytics impartido por IMF Business School. El documento contiene dos secciones: el enunciado y la solución de cada uno de los ejercicios propuestos.

2 Enunciado

La actividad se estructura en tres partes crecientes en esfuerzo requerido. La premisa es similar: el alumno deberá elegir un hashtag y ejecutar un ejercicio parecido.

Programas como Volando voy suelen tener una respuesta generalmente positiva (es decir, el interesado en el programa o la región comenta en positivo si quiere, pero si no le gusta no suele comentar). Esto hace que las comunidades se dividan en áreas de influencia o interés. Pero hay otro caso interesante: el debate. Para seguir con la televisión, programas como Al Rojo Vivo o La Sexta noche suelen girar en torno a temas que generan división de opiniones y reacciones. Además, suelen incluir un hashtag concreto para cada uno de esos temas, así que es relativamente sencillo monitorizarlos. De cara a este ejercicio, por tanto, hay varios casos posibles, como los debates ya mencionados, los eventos deportivos (que se dividen por simpatías), concursos... Cualquiera de estas opciones os dará facilidades a la hora de entender los resultados obtenidos en el ejercicio intermedio, pero son solo recomendaciones; el tema es libre.

- **Ejercicio básico.** Para el ejercicio básico vamos a utilizar únicamente Gephi. Se debe cargar la red proporcionada (1_uclfinal.gephi), que corresponde, por ejemplo, a una extracción de unos minutos antes de la final de la Champions League, y ejecutar el mismo guion/idea base que en el apartado 6.5. Como entregable se espera:
 - Capturas de pantalla de Gephi (al menos, la inicial tras la carga del fichero, la final y una intermedia del proceso).
 - ¿Cuáles son las principales comunidades obtenidas? ¿Y los usuarios de mayor importancia? ¿Los podrías dividir en bandos y en neutrales?
- **Ejercicio intermedio.** Para el ejercicio intermedio vamos a utilizar las posibilidades de la API de Twitter de forma distinta: a partir de un plugin de Gephi llamado Twitter Streaming Importer que facilita la tarea de almacenar tuuits de forma automática. Como entregable se espera el grafo obtenido de la extracción, utilizando File > Save As (o guardar como) y guardado en el formato por defecto, (.gephi).
- **Ejercicio avanzado.** Con los datos extraídos en el ejercicio intermedio (el grafo obtenido), se solicita buscar las comunidades y realizar una interpretación de estas, es decir:
 - Que ejecutéis la misma dinámica que en el ejercicio básico, pero sobre vuestra propia extracción.
 - Y que extraigáis unas conclusiones mucho más detalladas, al estilo de las del caso práctico visto en la unidad (en el ejercicio base se pide menos detalle).

3 Solución

3.1 Ejercicio básico

A continuación, se muestran las capturas de pantalla del proceso para obtener las comunidades:

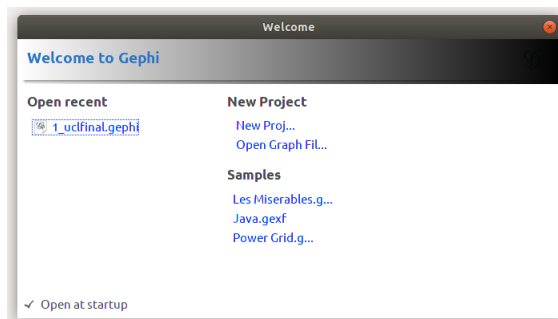


Figura 1: Carga de los datos - Abrir fichero.

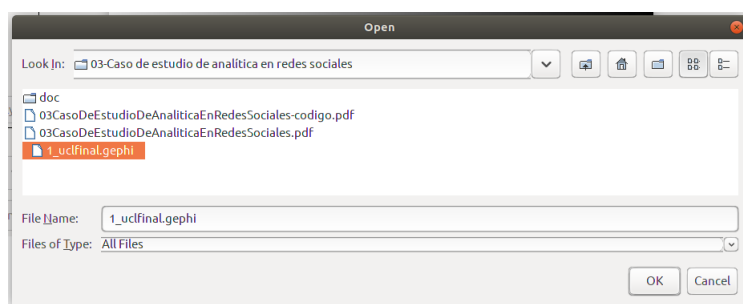


Figura 2: Carga de los datos - Seleccionar fichero 1_uclfinal.gephi.

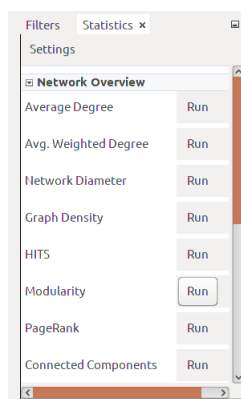


Figura 3: Ejecución de la opción Modularity.

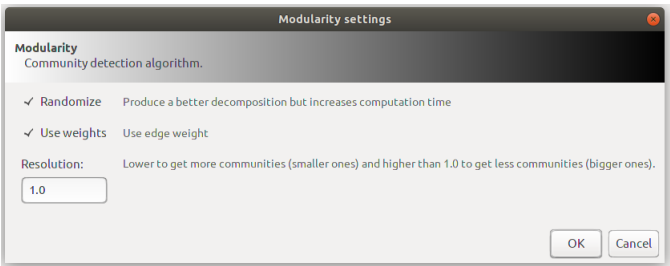


Figura 4: Configuración de la opción Modularity (dejarlo por defecto).

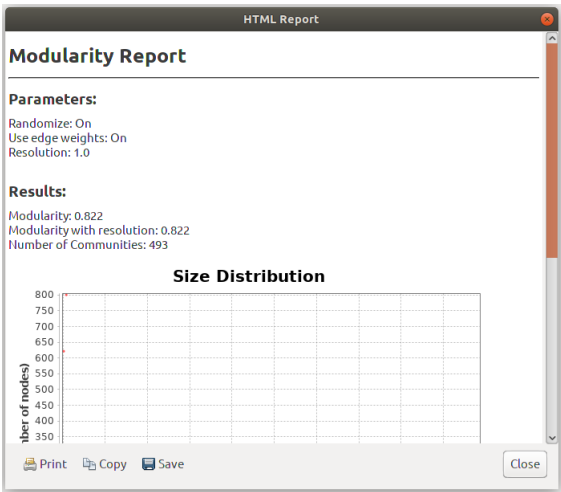


Figura 5: Reporte de la ejecución de Modularity.

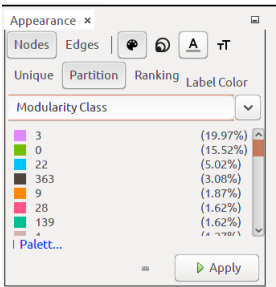


Figura 6: Selección Modularity Class para diferenciar los nodos por colores.

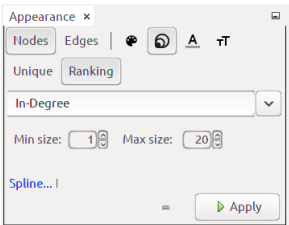


Figura 7: Selección In-Degree para diferenciar el tamaño de los nodos por grado de entrada.

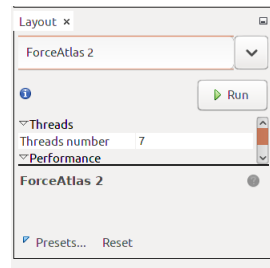


Figura 8: Selección Force Atlas 2 como opción de layout.

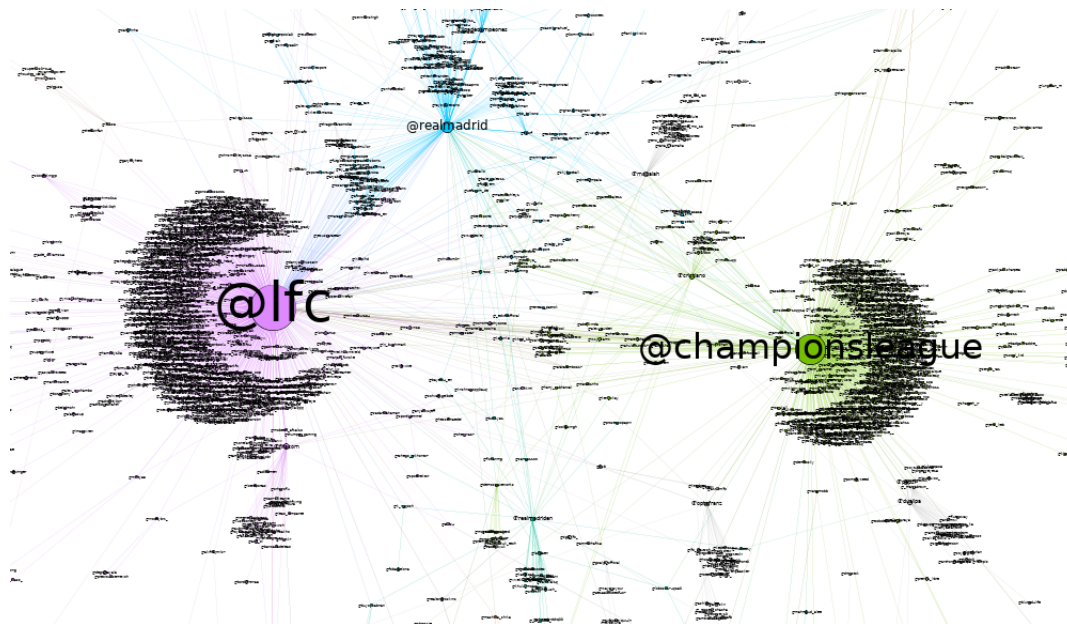


Figura 9: Gráfico generado.

- **¿Cuáles son las principales comunidades obtenidas?** Las principales comunidades obtenidas son:
 - En color lila claro (izquierda) se encuentra la comunidad del equipo Liverpool Football Club (LFC) y todos sus seguidores.
 - En color verde (derecha) se encuentra la comunidad de la UEFA Champion League y todos sus seguidores.
 - Y por último, en color azul claro (arriba) se encuentra la comunidad del equipo Real Madrid y todos sus seguidores.
- **¿Y los usuarios de mayor importancia?** Los usuarios con mayor importancia en orden decreciente son: el Liverpool Football Club (@lfc), la UEFA Champion League (@championsleague) y el Real Madrid (@realmadrid).
- **¿Los podrías dividir en bandos y en neutrales?** Claramente hay dos bandos: el del Liverpool Football Club y el del Real Madrid. Como neutrales, están los que siguen a la UEFA Champion League.

3.2 Ejercicio intermedio

El domingo 13 de enero fue la final de Master Chef Junior y se quería analizar durante todo el programa la actividad global en Twitter, por lo que desde el inicio hasta el final del programa se estuvo recogiendo datos con Gephi. Así pues, se muestran las capturas de pantalla del proceso para obtener los datos de Twitter:

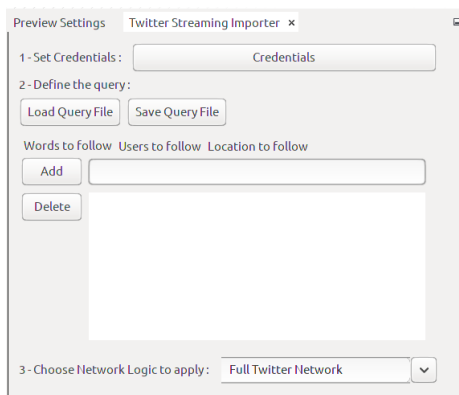


Figura 10: Configuración del plugin - Inicio.



Figura 11: Configuración del plugin - Introducción de credenciales.

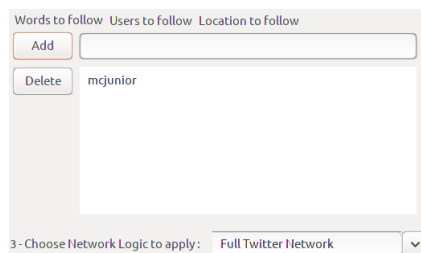


Figura 12: Configuración del plugin - Introducción del hashtag MCJunior.

Tras la emisión del programa, se obtuvo un gráfico dirigido de 6347 nodos y 11414 aristas, y que se encuentra almacenado en el fichero `master-cher-junior.gephi`. La Figura 13 muestra el grafo obtenido sin tratamiento.

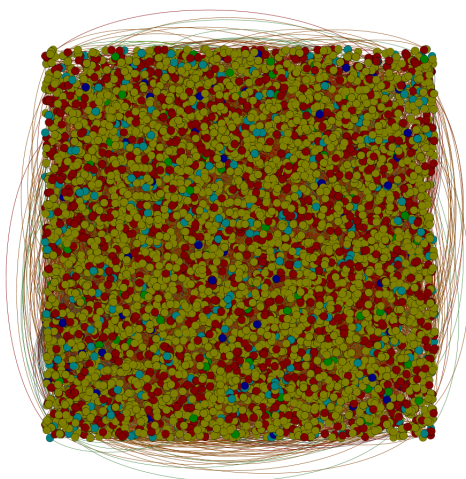


Figura 13: Gráfico obtenido - Fichero `master-cher-junior.gephi`.

3.3 Ejercicio avanzado

Para la realización de este ejercicio se han seguido los mismos pasos que en el ejercicio básico, pero cambiando la configuración del algoritmo ForceAtlas 2 por la que se muestra en la Figura 14.

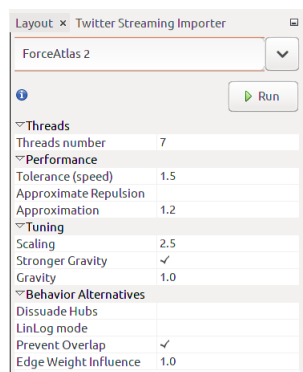


Figura 14: Configuración del algoritmo ForceAtlas 2.

La Figura 15 muestra el resultado del tratamiento de todos los pasos anteriores y que se encuentra almacenado en el fichero `master-chef-junior-dev.gephi`.

Es importante destacar aquí que se han recogido toda la información de Twitter, no se ha buscado usuarios ni hashtags asociados en concreto, si no toda la información que se generó. Por lo tanto, el análisis que se quiere hacer es encontrar qué usuarios, tweets, hashtags, páginas asociadas, etc, han sido más influyentes o importantes durante la emisión del programa. En ella, podemos observar lo siguiente:

– Comunidades

- En color rosa, tenemos la comunidad representada por todas las personas que siguen a la cuenta de MasterChef @masterchef_es, así como los tweets e interacciones que hizo la cuenta durante la emisión.
- En color gris oscuro, tenemos la comunidad representada por todas las interacciones que se tuvieron con la página web del programa. Estas interacciones vienen de todos los tweets de la cuenta de MasterChef, pues en todos mencionaba a la web.
- En color azul, se encuentra la comunidad de todas las interacciones, tweets del usuario @subnorbook y usuarios que le siguen.
- En color verde claro, la comunidad de todas las interacciones, tweets que tuvo la cuenta @abelarana y usuarios que le siguen.
- En color naranja, la comunidad de todas las interacciones, tweets que tuvo la cuenta @crisrupa y usuarios que le siguen.
- En verde más oscuro y en pequeño (situado a la derecha de la imagen), se encuentra la comunidad de las interacciones, tweets que hizo la cuenta @lolicastro3 y usuarios que le siguen.
- En rojo y en pequeño (situado debajo del anterior), está la comunidad de las interacciones que hizo la cuenta @alfonso y usuarios que le siguen.
- En gris claro, el resto de tweets, interacciones y usuarios que no son fueron tan influyentes como el resto.

– Análisis de resultados

- En cuanto a la influencia de la cuenta de MasterChef, era de esperar. Sin embargo, hay que darse cuenta de que hubieron dos tweets que más interacciones tuvieron. El primero (muy cerca de la bola de la página del programa) es un tweet (enlace) en la que se muestra la parte del programa muy emotiva, en la que se le dice que no pasa a la final a Candela, una niña de 8 años y Pepe (uno de los chefs) la consuela. El segundo, es el tweet (enlace) en el que se muestra la parte del programa en la que se dice el ganador.
- Como se ha comentado antes, la influencia de la web del programa viene dada por todos los tweets de MasterChef.
- En cuanto a la influencia de @subnorbook, se trata de una cuenta que tiene más de doce mil seguidores. La bola más grande es un tweet que puso MasterChef y que él citó (enlace).
- La influencia de la cuenta @abelarana es notoria debido a que se trata de la cuenta de Abel Arana, presentador online de MasterChef y que cuenta con casi 38 mil seguidores.
- Lo más destacable es la influencia de la cuenta @crisrupa, que se trata de la cuenta de una chica con 18 seguidores. La gran influencia de esta chica reside en la cantidad de retweets y likes que le dio a los tweets generados durante el programa.

– Conclusiones

- Como cabía esperar, la cuenta de MasterChef fue la más interactiva durante el programa. Como consecuencia de que en todos sus tweets, ponía la web del programa, esta también fue de la más influyente. Lo mismo pasa con la cuenta de Abel Arana, que retransmitió online el programa.
- Se podría decir, que hubo dos momentos muy importantes en el programa: uno fue el emotivo momento en el que Pepe consuela a la pequeña Candela y el otro fue la proclamación del ganador.
- Además, es muy importante que cuentas como @subnorbook interaccionen con tweets del programa, pues se da más visibilidad y puede hacer que alguien que no estuviera viendo o no le interesara MasterChef, se interesara o viera el programa.
- Por último, también es importante llegar a conocer aquellas cuentas que interaccionan con todos los tweets que puso MasterChef, como es el caso de @crisrupa. Aunque no tenga muchos seguidores.

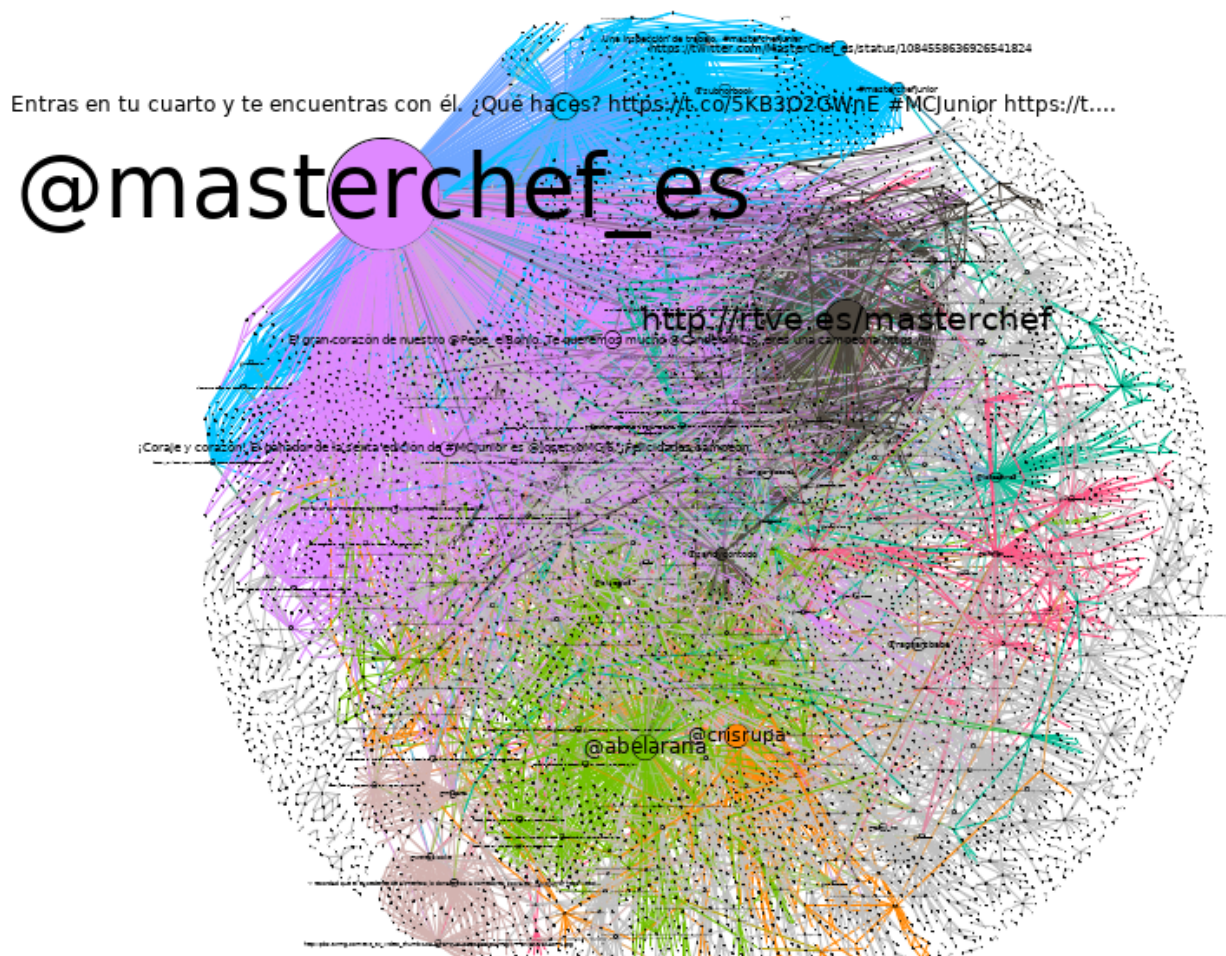


Figura 15: Gráfico obtenido - Fichero master-chef-junior-dev.gephi.